AUTOMATIC CORN AND SOYBEAN MAPPING BASED ON DEEP LEARNING METHODS (CASE STUDY: HAMILTON, HARDIN, BOONE, STORY, DALLAS, POLK, AND JUSPER COUNTIES IN LOWA STATE)

M. Fathi¹, R. Shah-Hosseini^{1*}

¹ School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran - (mahdiyeh.fathi, rshahosseini) @ut.ac.ir

Commission IV, WG IV/3

KEY WORDS: Corn, Soybean, Landsat-8, NDVI, multi-temporal, Deep Learning

ABSTRACT:

Corn and Soybean are important crops for the world's people. Agricultural planning relies heavily on monitoring and mapping corn and soybean fields. With the development of remote sensing technology and deep learning algorithms, corn and soybean fields are being managed more intelligently nowadays. By using Landsat-8 images with multi-temporal maps of NDVI index, we intend to compare deep learning models such as 1-D CNN, 1-D CNN-LSTM, and 2-D U-net for separating corn and soybean fields from other crops (because soybean and corn fields have similar NDVI curves) in the United States in 2020. It was found that the 2-D U-net model performed the most accurately for corn and soybean classes with Kappa coefficients of 88.48 and 88.89, respectively. This can be explained by the identification of complex features using NDVI multi-temporal indexes of March to November in the United States.

1. INTRODUCTION

Corn and soybeans are important agricultural commodities in the world. The United States is the world's largest producer of corn and soybeans. As an example, in 2020, corn and soybeans were produced in the United States at a rate of 14,2 and 4,14 billion bushels, respectively. In light of these factors, monitoring corn and soybean fields with remote sensing technology is crucial for food security, monitoring climate change, and detecting crop diseases, etc. A variety of satellite images can be used to map crop fields, including optical images (e.g., Landsat, Sentinel-2, etc.) and radar (e.g., Sentinel-1). Due to the ability of optical images to observe the surface of the earth within a spectral range of 0.4 to 2.5 meters, multitemporal maps are obtained of vegetation indexes like Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Soil Adjusted Vegetation Index (SAVI), and Transformed Vegetation Index (TVI). Low spatial resolution prevents MODIS images from detecting small crop fields, while the Sentinel 2 and Landsat serise are more compatible in this way. Radar images are also ideal for mapping crop fields due to their independence from weather conditions such as cloud cover, rain, and snow (Mosleh, Hassan, Chowdhury, 2015). During the growing season, radar backscatter varies over time and is therefore critical when mapping crop fields using radar images. Since optical images are more interpretable than SAR images, they have been used more frequently in crop monitoring. Thus, this study focuses on monitoring crop land using optical images.

The multi-temporal maps which are derived from vegetation indices are used to map crop fields (Mosleh, Hassan, Chowdhury, 2015). In monitoring and mapping crop fields, Land Surface Phenology (LSP) can be extracted from the multitemporal maps of vegetation indicators. Biological events occur periodically according to climatic conditions, known as phenology. It is calculated by varying values of vegetation indexes during the planting, holding, and harvesting of crop fields for a period of time, such as the beginning of the growing season, the end of the growing season, the maximum value of the vegetation index, and the length of the growing season (Wang, Jin, Zhou, Guo, Song, 2015).

In order to map crop fields based on their phenology, satellite images have been used in a variety of ways. For example, crop fields have been mapped with phenology-based algorithms using Landsat-7/8, MODIS, Pi-SAR-L2 images, etc. (Ding, Guan, Li, Zhang, Liu, Zhang, 2020; Liu, Huang, Xiong, Zhang, Song, Huang, Wang, 2020; Lobell, Asner, 2004; Yonezawa, Watanabe, 2020). Machine learning algorithms (e.g., Random Forest (RF), Support Vector Machine (SVM), etc.) have also been suggested for mapping crop fields using polarimetric and phenology features, as well as optical image indices (e.g., NDVI, EVI, RVI, MNDWI, and LSWI) (Chen, Zhang, Shen, Zeng, Hu, Niyogi, 2020; Indolia, Goswami, Mishra, Asopa, 2020; Mansaray, Huang, Zhang, Huang, Li, 2017; Talema, Hailu, 2020; Wang, Zang, Tian, 2020; Yang, Shao, Li, Liu, Liu, Brisco, 2017). Despite the fact that machine learning algorithms mentioned above are not able to fully extract the spectral and spatial features of interest features, deep learning algorithms (e.g., Convolutional Neural Networks, LeNet-5, LSTM, Autoencoder, Fusion in-Decoder) have been proposed as a means of improving mapping accuracy by extracting high-level features from low-level features in crop fields. (Zhao, Liu, Ding, Liu, Wu, Wu, 2020; Zhang, Lin, Wang, Sun, Fu, 2018; Guo, Jia, Paull, 2018; Jo, Lee, Park, Lim, Song, Lee, Lee, 2020; Zhang, Liu, Wu, Zhan, Wei, 2020; Zhao, Liu, Ding, Liu, Wu, Wu, 2020; Jiang, Liu, Wu, 2018; Rawat, Kumar, Upadhyay, Kumar, 2021, Fathi, Shah-Hosseini, 2021). Researchers have recently applied deep learning to map corn and soybean fields by analyzing spectral features extracted from Landsat-8 images using LSTM (Deep Crop Mapping) (Xu, Zhu, Zhong, Lin, Xu, Jiang, Lin, 2020). Our study compares CNN, LSTM-CNN, and U-net networks for automatic mapping of corn and soybean

^{*} Corresponding author

fields from Landsat-8 optical images based on NDVI index extracted from these images.

2. THE STUDIED AREAS AND MATERIAL

2.1 The studied areas

A total of seven counties in Lowa State are being studied, including Hamilton, Hardin, Boone, Story, Dallas, Polk, and Jusper. Figure 1 shows the study areas that were included in the study.



Figure 1. The google earth image (a) and the studied areas (b).

2.2 Data set

This study used Landsat-8 OLI images to map crops. As corn and soybeans are planted in April and May (respectively) and harvested in September and November (respectively), cloudless images were used from March to November. It is important to perform radiometric normalization before processing and analysing multitemporal remote sensing images (Moghimi, Celik, Mohammadzadeh, Kusetogullari, 2021, Moghimi, Mohammadzadeh, Celik, Brisco, Amani, 2022a, Moghimi, Celik, Mohammadzadeh, 2022b). Therefore, radiometric calibration was used to pre-process Landsat-8 images. The mapping of corn and soybean fields was therefore accomplished using extracted spectral and vegetation features from multitemporal Landsat-8 images (i.e., listed in Table 1).

Number	Date
1	2020/03/04
2	2020/04/21
3	2020/06/08
4	2020/07/10

5	2020/08/11	
6	2020/09/03	
7	2020/10/30	
8	2020/11/22	
	0 1.1	

 Table 1. Used Landsat-8 multi-temporal images.

2.3 Feature selection

The selected feature in the study included NDVI Index (NDVI=(NIR-RED)/(NIR+RED)) was extracted from the Landsat-8 multi-temporal images. It allows us to recreate a timeline of crop production by using NDVI multi-temporal indexes (Ramadhani, Pullanagari, Kereszturi, Procter, 2020). Figure 2 illustrates the extracted phenology curves for corn and soybean fields during the planting period. We then evaluated the mapping of corn and soybean fields using three different networks.



Figure 2. Extracted phenology curve of NDVI multi-temporal index for corn and soybean fields.

2.4 Ground truth

This ground truth map was downloaded from USDA with a spatial resolution of 30 m. The field classification was based on the Maximum Likelihood classifier, using Landsat TM/ETM satellites, prior to 2006. However, USDA from 2006 has used DEIMOS-1, UK2, LISS-3, and ESA Sentinel-2 A/B sensors to classify crop fields by using Landsat 8 sensor, DEIMOS-1, and UK2. The training and validation data used to calculate and accuracy assess classification for reasons like 1- satellite imagery (as well as the polygon reference data) in the past was not georeferenced to the same precision as now (i.e. everything "stacked" less perfectly), 2- to eliminate from training spectrally-mixed pixels at land cover boundaries, and 3- to be spatially conservative during the era when coarser 56 meter AWiFS satellite imagery was incorporated, based on ground truth data that is buffered inward 30 meters. Since buffered data is used for accuracy assessment, edge pixels are not fully evaluated with the rest of the classification, leading to somewhat inflated accuracy assessments. Using the error matrix provided for Lowa, the kappa coefficient for corn and soybean fields in 2020 is 97% and 88.9%, respectively. The ground truth map and NDVI index are shown in Figure 3.



Figure 3. NDVI index (a) and ground truth map (b).

3. METHODOLOGY

NDVI index extracted from Landsat-8 optical images was used to map corn and soybean fields using U-net, CNN, and CNN-LSTM networks. After analyzing the confusion matrix, the best classification algorithm was selected for mapping corn and soybean fields. Figure 4 illustrates the proposed method's flowchart.



Figure 4. Flowchart of the proposed method.

3.1 CNN network

Convolution Neural Networks are a deep learning algorithm that including four components convolution layer, pooling layer, activation function, and fully connected layer. Convolution layer input is the image with r feature band with size m*n. The convolution layer contains K filter in size l*l*q that concatenates the input and output feature layer. Z=W*X + b shows layer convolution output (Z) of feature layer (X) with weight W and bias b. Activation function is nonlinear function that apply to Z (a=f(Z)). Given that feature dimensions are high layer to prevent Over Fitting the Pooling layer (such as maxpooling) is used to reduce the feature dimensions after the convolution layer. Finally, the dense layer (last layer) is a fully connected that each neuron is connected to each node of the output available from the prior layers (Indolia, Goswami, Mishra, Asopa, 2018). Figure 5 shows the proposed architecture for the CNN network. The proposed one-dimensional CNN model is included three one-dimensional convolution layers including 32, 64, and 128 neurons, ReLU activation function, and a Dropout layer(rate=0.5). In our method, the categorical Entropy loss function is applied to calculate weight and bias parameters, and ADAM algorithm is used for network training with patch size 5 for 120 epochs (Wahlberg, Boyd, Annergren, Wang, 2012; Zhang, Sabuncu, 2018). Finally, a Dense layer is used with a softmax activation function (Dunne, Campbell, 1997).

3.2 LSTM network

LSTM network has been used for text recognition. LSTM consists of three components an input, forget, and an output gate. LSTM memorizes of the context information in sequence data for a large time period. Figure 6 shows a LSTM cell. Where X(t), M(t-1), M(t), N(t-1), N(t), f, c, I, O, U and W are the input vector, the output from the prior cell, the output from the current cell, memory value from the values cell, forget gate including sigmoid activation function, the candidate date including tanh activation function, the input gate including sigmoid activation function, the output gate including sigmoid activation function, and the weight vectors for forget gate, candidate gate, input gate, and output gate, respectively (Rawat, Kumar, Upadhyay, Kumar, 2021). Figure 7 shows the proposed architecture for the CNN-LSTM network. The proposed onedimensional CNN-LSTM model has consisted of three onedimensional convolution layers including 32, 64, and 128 neurons along with the LSTM layers with 64, 128, 256 neurons, ReLU activation function, and a Dropout layer (rate=0.5). Finally, a Dense layer with a softmax activation function is applied, and the categorical Entropy loss function is used to calculate weight and bias parameters, and ADAM algorithm is used for network training with patch size 5 for 120 epochs.

3.3 U-net network

U-net is a deep learning algorithm that is used to predict masks in tasks of image segmentation. U-net is consisted of a contraction path and an expansive path. The contraction path includes two 3*3 convolutions, two ReLU activation functions, max pooling, and down-sampling. The expansive path includes an up-sampling of the feature map, 2*2 convolution for halving the number of feature maps (up-convolution), a concatenation, two 3*3 convolutions, and two ReLU activation functions. Finally, a 1*1 layer convolution is used with the desired number of classes (Ronneberger, Fischer, Brox, 2015). Dropout is used to reduce overfitting and create different architectures by using removing neurons randomly in the last layer of the expansive path (Garbin, Zhu, Marques, 2020, Fathi, Shah-Hosseini, 2021). Also, distribution of the input values of each layer was keep with Batch Normalization (Garbin, Zhu, Marques, 2020). Figure 8 shows the proposed architecture for the two-dimensional Unet network. For the proposed two-dimensional U-NET model, the contraction path included 5 layers containing a twodimensional Convolution block (with 2 convolution layers), ReLU activation function, Batch Normalization layer, and maxpooling layer. The number of filters in each layer of the contraction path was 64, 128, 256, 512, and 1024, respectively. Also, the expansive path included 4 layers containing a twodimensional Convolution block (with 2 convolution layers), ReLU activation function, Batch Normalization layer, and upsampling layer. The number of filters in each layer of the expansive path was 512, 256, 128, and 34, respectively. In addition, a Dropout layer (rate=0.5) was applied in the last expansive path layer. Following the expansive path, the last layer contained two-dimensional Convolution and Softmax activation function. Finally, the categorical Entropy loss function is applied to calculate weight and bias parameters, and ADAM algorithm is used for network training with patch size 5 for 120 epochs.



Figure 6. Representation of cell LSTM.



Figure 7. Proposed CNN-LSTM network.



Figure 8. Proposed U-net network.

4. RESULT

According to Table 1, NDVI multi temporal index is calculated by using Landsat-8 multi temporal images as input networks. According to Figure 3. a, on a False Colour Composite image for the date 2020/04/03 (R), 2020/07/10 (G), and 2020/11/22 (B), the soybean fields appear to be dark green. We are used from image patches with the size of 256*256*8 as input for each of the networks. In this study, 2293760 and 458752 number pixels were selected to train and validation the networks, respectively (Table 2). Table 3 shows the classification results for three classes of corn, soybean, and other classes. The obtained values of these measures are quite satisfactory for the two-dimensional U-net models (Kappa coefficient for corn class=88.48, Kappa coefficient for soybean class=88.89, accuracy for corn class =94.31, and accuracy for soybean class =95.64).

	The	The	The	The
	number	number	number	number
	of	of pixels	of pixels	of pixels
	patches	for	for corn	for
		Soybean	class	others
		class		classes
training	35	686077	1087859	519824
validation	7	130485	238405	89862

 Table 2. The number of used patches/pixels for training and validation of each class.

Method		Other classes	Soybean	Corn
1-D CNN	Accuracy	93.74	92.26	90.55
	Kappa	77.95	80.96	80.98
1-D CNN- LSTM	Accuracy	93.53	92.49	90.77
	Kappa	76.65	81.64	81.37
2-D U-NET	Accuracy	95.61	95.64	94.31
	Kappa	84.24	88.89	88.48

Table 3. The generated results by proposed Methods.

According to Table 3 and Kappa coefficient values, the onedimensional CNN-LSTM model has performed better than the one-dimensional CNN model for classes of corn and soybean. According to Figure 5, Figure 7, and Figure 8 was noticed that the one-dimensional CNN and one-dimensional CNN-LSTM models had a relatively simpler structure relative twodimensional U-NET model. One of the reasons for the twodimensional U-NET model performing better than models of one-dimensional CNN and one-dimensional CNN-LSTM is the identification of complex features in the NDVI multi-temporal index. One of the advantages of deep learning algorithms is the extraction of higher-level features from higher-level features by convolution layers. Also, the reason one-dimensional CNN-LSTM is better-performing in comparison to one-dimensional CNN for corn and soybean classes the extraction of sequential relationships by LSTM layers. A visual interpretation of the results shown in Figure 9 suggests that the two-dimensional Unet model generated better classification outputs than models of one-dimensional CNN and one-dimensional CNN-LSTM for NDVI multi-temporal index for classifying multi class classification.





Figure 9. Generated corn and soybean maps by proposed methods.

5. CONCLUSIONS

To separate soybean and corn fields, three deep learning models were used and analysed in this study, including one-dimensional CNN, one-dimensional CNN-LSTM, and a two-dimensional U-net. For soybean and corn fields (with Kappa coefficients of 88.89% and 88.48%), the two-dimensional U-NET model performed better than the one-dimensional CNN and one-dimensional CNN-LSTM models. A two-dimensional U-net network, on the other hand, appears to be more effective when it comes to mapping multiple classes of interest, specifically crop types.

REFERENCES

Chen, N., Yu, L., Zhang, X., Shen, Y., Zeng, L., Hu, Q., &vNiyogi, 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.

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.

Fathi, M., & Shah-Hosseini, R. 2021: Automatic rice fields mapping by fusion of in-decoder CNN and data augmentation techniques on Landsat-8 multi-temporal images. Earth Observation and Geomatics Engineering, 5(2), 74-95.

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.

Moghimi, A., Celik, T., Mohammadzadeh, A., & Kusetogullari, H. 2021: Comparison of keypoint detectors and descriptors for relative radiometric normalization of bitemporal remote sensing images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 4063-4073. Moghimi, A., Mohammadzadeh, A., Celik, T., Brisco, B., & Amani, M. 2022: Automatic Relative Radiometric Normalization of Bi-Temporal Satellite Images Using a Coarseto-Fine Pseudo-Invariant Features Selection and Fuzzy Integral Fusion Strategies. Remote Sensing, 14(8), 1777.

Moghimi, A., Celik, T., & Mohammadzadeh, A. 2022: Tensorbased keypoint detection and switching regression model for relative radiometric normalization of bitemporal multispectral images. International Journal of Remote Sensing, 43(11), 3927-3956.

Indolia, S., Goswami, A. K., Mishra, S. P., & Asopa, P. 2018: Conceptual understanding of convolutional neural network-a deep learning approach. Procedia computer science, 132, 679-688.

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.

Jiang, T., Liu, X., & Wu, L. 2018: Method for mapping rice fields in complex landscape areas based on pretrained convolutional neural network from HJ-1 A/B data. ISPRS International Journal of GeoInformation, 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 labelled data: The case of detecting rice paddy in South Korea. IEEE Transactions on Geoscience and Remote Sensing, 58(11), 7589 7601.

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.

Ramadhani, F., Pullanagari, R., Kereszturi, G., & Procter, J. 2020: Automatic mapping of rice growth stages using the integration of SENTINEL-2, MOD13Q1, and SENTINEL-1. Remote Sensing, 12(21), 3613.

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.

Ronneberger, O., Fischer, P., & Brox, T. 2015, October: U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

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.

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.

Xu, J., Zhu, Y., Zhong, R., Lin, Z., Xu, J., Jiang, H., ... & Lin, T. 2020: DeepCropMapping: A multi-temporal deep learning approach with improved spatial generalizability for dynamic corn and soybean mapping. Remote Sensing of Environment, 247, 111946.

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.

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.