EARTH OBSERVATION AND GEOMATICS ENGINEERING
website: https://eoge.ut.ac.ir

Comparison of Some Deep Neural Networks for Corn and Soybean Mapping in Iowa State using Landsat imagery

Mahdiyeh Fathi[1], Reza Shah-Hosseini[1*], Armin Moghimi[2]

[1] School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran
[2] Ludwig-Franzius-Institute for Hydraulic, Estuarine and Coastal Engineering, Leibniz University Hannover, Hanover, Germany

Article history:
Received: 2023-03-17, Revised: 2023-05-17, Accepted: 2023-06-11

ABSTRACT

Corn and soybeans are crucial crops for feeding the world's population, and it is essential to monitor and map these fields for effective agricultural planning. With recent advancements in remote sensing technology and deep learning algorithms, more intelligent management of these crops has become possible. In this paper, we compare modern deep learning (DL) architectures for mapping corn and soybean crops in the state of Iowa, in United States of America (USA), using temporal Landsat 8 (OLI) images. Our analysis focuses on the performance of six different DL networks, namely 1-D CNN, 1-D CNN-LSTM, 2-D UNet, 2D UNet 3+, 2D Attention UNet, and 2D Recurrent Residual UNet. For each network, we use time-series Normalized Difference Vegetation Index (NDVI) data derived from Landsat 8 images taken between April and November 2020 as input, while ground truth labels are taken from the United States Department of Agriculture (USDA). Our experimental results show that the 2-D Recurrent Residual U-net model achieved the highest accuracy for identifying corn and soybean classes, with an F-score of 92.85. This indicates the model's ability to distinguish complex features and patterns with similar spectral characteristics from multi-temporal remote sensing data. Conversely, the CNN and CNN-LSTM models had the worst performance among the considered models, with an F-score of about 89.50. Nevertheless, all the DL methods examined in this study achieved acceptable classification kappa coefficients (above 82%), indicating their significant potential for accurately mapping and monitoring Corn and Soybean crops.

KEYWORDS

Corn
Remote sensing
Soybeans
Landsat-8
UNet
UNet 3+
Attention UNet
Recurrent Residual UNet
Deep Learning

1. Introduction

Corn and soybean production in the United States of America (USA) holds tremendous significance as these crops serve as vital sources of energy, protein, and oil, accounting for a third of the global production, and they are primary types of summer crops in the central States (Zhong et al, 2016). Understanding the spatial distribution of corn and soybean planting areas is critical in different aspects (e.g., food security and environmental monitoring) since corn has high water use, while soybeans have the ability to fix nitrogen (Zhong et al, 2014). Therefore, accurate monitoring of corn and soybean crops are necessary not only for USA, but also for all of the countries on the worlds (McNairn et al, 2014).

In recent years, due to technological advancements and the increased availability of open-access Remote Sensing (RS) data acquired from space/airborne sensors with varying characteristics (e.g., spectral, spatial, temporal, and radiometric resolutions) (Moghimi et al, 2021a, Amani et al., 2020), RS technology has been viewed as an efficient solution for overcoming the challenges of costly traditional crop field mapping. Moreover, variety of space-borne RS variety of images including optical (Landsat, Sentinel-2, etc.) and radar images (e.g., Sentinel-1), can be used to deal with economic and social challenges caused by supply, demand, and a food crisis by providing information such as agricultural area, crop types, crop rotation dynamics, and so on (Zhong et al, 2014).

Multitemporal maps of Vegetation Indices (VIs) (e.g., the Normalized Difference Vegetation Index (NDVI), Enhanced

Vegetation Index (EVI), Soil Adjusted Vegetation Index (SAVI), Transformed Vegetation Index (TVI) and so on) derived from Optical RS images widely used for crop mapping using researchers. Based on these indices crops can be mapped through phenology analysis and/or Machine Learning (ML)-based methods. For example, Land Surface Phenology (LSP) extracted from multitemporal vegetation indicator maps is widely used to track biological events in response to climatic conditions. This involves analyzing varying values of vegetation indices during different stages of crop growth, such as the beginning and end of the growing season, the maximum value of the vegetation index, and the length of the growing season (Fathi, & Shah-Hosseini, 2021).

Random Forest (RF) and Decision Tree (DT) classifier as the traditional ML classifiers have been frequently used for crop mapping and monitoring. For example, the RF algorithm has been applied on the MODerate-resolution Imaging Spectroradiometer (MODIS) reflectance product to separate corn and soybeans based on the best features by Liu et al, 2019. Zhong et al, 2016 used the DT classifier to map corn and soybeans using extracted phenology from the surface reflectance of the shortwave infrared band of the Time series MODIS reflectance product (MCD43A4). McNairn et al, 2014 also proposed using TerraSAR-X and RADARSAT-2, along with a DT classifier, for early season monitoring of corn and soybeans.

Some studies have used various input features, such as spectral features at imaging dates, phenological metrics derived from EVI time series, accumulated temperature during phenological stages, and features derived from Landsat, Sentinel-2, and SRTM+ data (Paludo et al, 2020; Zhong et al, 2014, Zhong et al, 2016).

In recent years, deep learning algorithms such as hybrid CNN-RF, LSTM-RNN and DeepCropMapping have been developed for crop field monitoring and mapping (Yang et al, 2020, Sun et al, 2019; Xu et al, 2020). These methods take advantage of the temporal patterns of crops across image time series to improve accuracy and reduce complexity. For instance, Xu et al, 2020 used six optical bands of Landsat as input features for DeepCropMapping, which is based on a long short-term memory (LSTM) structure with attention mechanisms.

Although the DL models have shown promising results for corn and soybean crop mapping, they still face challenges due to inter-annual variations. These variations include changes in imaging dates, climate changes, cloud cover, phenological variation, and other factors that can make trained classification models for a specific year unusable for other years. Creating a spatially generalizable model due to the spatial heterogeneity of weather conditions, topographic features, and phenological diversity can be also another challenge.

To overcome the challenges mentioned earlier, it is recommended to train classification models for multiple years or use phenological metrics. Furthermore, to determine the methods that perform well despite these challenges, a comparison of the performance of various DL models for crop and soybean mapping using time series vegetation indices is required.

To address these issues, this study aims to compare the performance of six different DL networks, including 1-D CNN, 1-D CNN-LSTM, 2-D UNet, 2D UNet 3+, 2D Attention UNet, and 2D Recurrent Residual UNet, for mapping soybean and corn fields. The study will extract the NDVI index from time series Landsat 8 images and assess the performance of the models under the influence of inter-annual variations, spatial heterogeneity of weather conditions, topographic features, and phenological diversity. By comparing the results of different DL models, this study hopes to provide insights into the best approaches for accurately mapping crops, which could have important implications for agriculture and food security.

2. The studied area and material

2.1. Study area

The study covers a total of seven counties in Iowa State, including Hamilton, Hardin, Boone, Story, Dallas, Polk, and Jasper. The research regions that were considered for the investigation are shown in Figure 1.

2.2. Dataset

Landsat-8 OLI images were used in this study to map crops. Since corn and soybeans are typically planted in April and May, and harvested in September and November, cloud-free images from March to November were utilized. It is essential to perform radiometric normalization prior to processing and analyzing multitemporal remote sensing images (Moghimi et al, 2021b). There is two type of radiometric correction methods: absolute and relative approaches (Moghimi et al, 2022a, Moghimi et al, 2022b). In this study, as we want to preserve the characteristics of vegitataion per each image, an absolute radiometric correction was used to radiometricaly calibrate each of multitemporal Landsat-8 images. In the next step, derived spectral bands and VIs from pre-processed Landsat-8 images (i.e., listed in Table 1) were used as inputs for considered DL models to map corn and soybean fields.

2.2.1. Feature selection

The Landsat-8 multi-temporal images were used to derive the NDVI Index (i.e., NDVI=(NIR-RED)/(NIR+RED)), which was the feature used for the research. Using NDVI multi-temporal indices, it enables us to reconstruct a chronology of crop production (Ramadhani et al, 2020). The extracted phenology curves for corn and soybean fields during the planting season are shown in Figure 2. We next

assessed the mapping of corn and soybean fields using six alternative networks.

Table 1. The date of considered Landsat-8 images from study area

Number	Acquisition 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

2.2.2. Ground truth

The ground truth map (Cropland Data Layer) with spatial resolution 30 m was retrieved from USDA. Prior to 2006, crop fields were categorized using Landsat TM/ETM satellites and the Maximum Likelihood classifier algorithm. However starting in 2006, the USDA employed the Decision Tree approach to categorize agricultural areas using the Landsat 8 sensor, the Disaster Monitoring Constellation DEIMOS-1, UK2, ISRO ResourceSat-2 LISS-3, and the ESA Sentinel-2 A/B sensors. Since satellite imagery and polygon reference data were not georeferenced with the same level of precision in the past (i.e., everything "stacked" less perfectly), the training and validation data used to calculate and assess classification accuracy had to be spatially conservative and eliminate spectrally-mixed pixels at land cover boundaries. The accuracy evaluations as they have been provided are slightly exaggerated since buffered data is utilized for accuracy assessment, which implies that the edge pixels are not evaluated entirely with the rest of the classification. The kappa coefficient for corn and soybean fields in 2020 (respectively) is 97% and 88.9%, according to the error matrix supplied for the state of Iowa (Fathi & Shah-Hosseini, 2023). The ground truth map is shown in Figure 1.

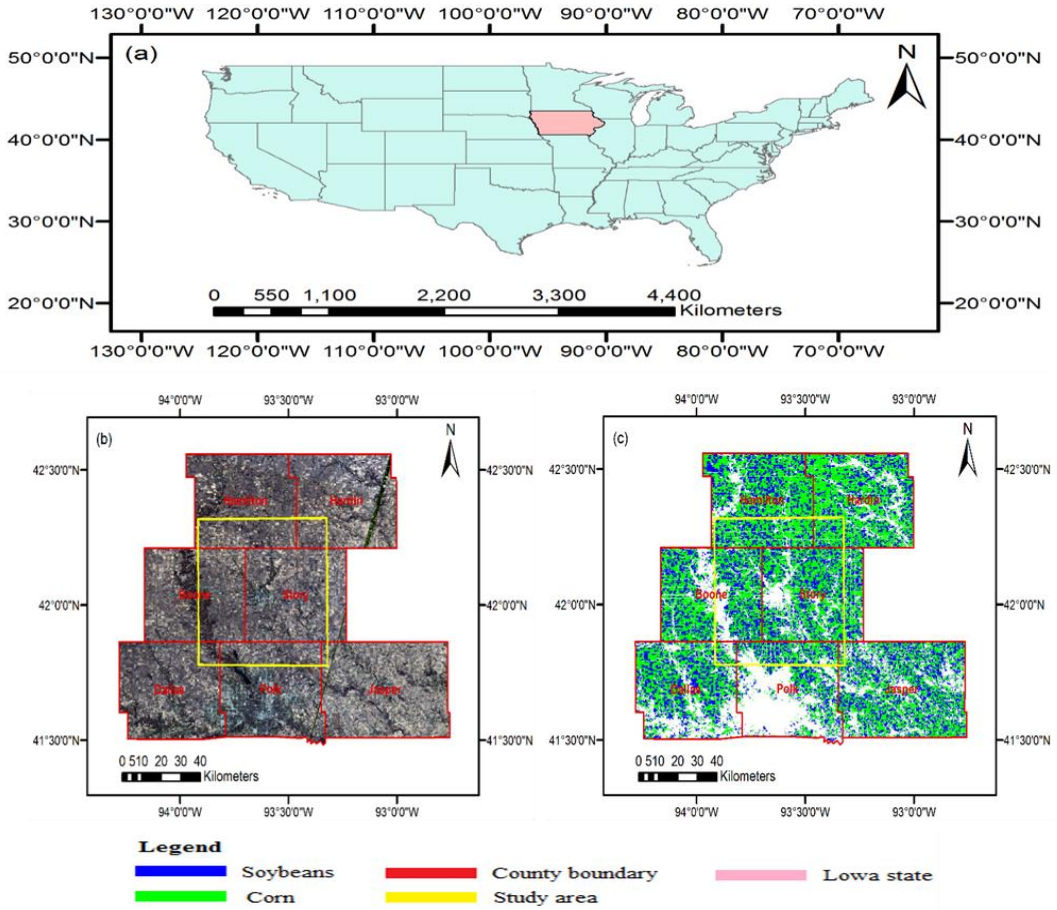


Figure 1. Study Area Overview. a) United States of America. b) Study area. c) Cropland Data Layer (CDL).

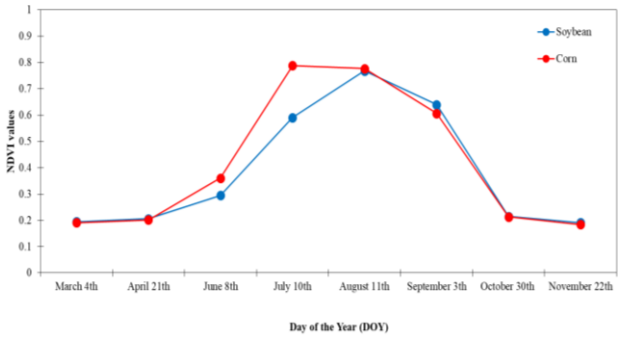


Figure 2. Extracted phenology curve of NDVI multi-temporal index for corn and soybean fields.

3. Method

For the automated mapping of corn and soybean fields in the state of Iowa, we used optical images from Landsat-8 in this research to compare some DL architectures. As discussed before, multitemporal NDVI indices retrieved from Landsat-8 images was utilized as input to map corn and soybean fields using U-net, Attention U-net, Recurrent Residual UNet, and UNet 3+, 1D CNN and 1D CNN-LSTM networks. The ideal classification technique was chosen for mapping corn and soybean fields after the analysis of the confusion matrix. Figures 3–4 depict the method's flowchart and the network design.

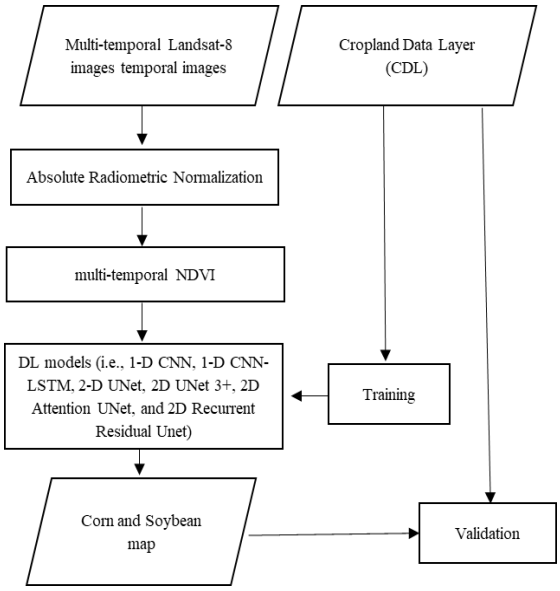


Figure 3. Flowchart of the proposed method.

3.1. Convolution Neural Network

*The convolution neural network has four parts: a convolution layer, a pooling layer, an activation function, and a fully connected layer. The input for the convolution layer is an image with r feature band of size m*n. The input feature layer and output feature layer are concatenated by a K filter with a size of l*l*q in the convolution layer. The layer convolution output of the feature layer (X) with weight W and bias b is shown by the formula Z=W*X + b. The activation function (a=f(Z)) is a nonlinear function that applies to Z. Considering that the feature dimensions are large, a pooling layer (like max-pooling) is employed to lower the feature dimensions after the convolution layer in order to avoid over fitting. The dense layer, which is the last layer of the CNN network and is completely linked, connects every neuron to every output node from the layers before it (Indolia et al, 2018).*

3.2. CNN-LSTM Network

The CNN-LSTM network combines the CNN and LSTM networks. An input gate, a forget gate, and an output gate make up an LSTM. LSTM permits memorization of the context information in sequence data over a lengthy time period. The input vector, the output from the previous cell, the current cell's output, the memory value from the values cell, the forget gate (a neural network with sigmoid activation function), the candidate date (a neural network with tanh activation function), the input gate (a neural network with sigmoid activation function), and the output gate (a neural network with sigmoid activation function) make up each LSTM cell (Rawat et al, 2021).

3.3. UNet Network

UNet design comprises of an encoder and a decoder elements and effectively utilized in numerous image processing applications. The input image is downsampled by a succession of convolutional and pooling layers that make up the encoder, while an upsampling and convolutional layer sequence creates a segmentation map in the decoder. The UNet architecture's ability to enable the smooth transfer of high-level information learnt during the encoding step to the decoding stage is one of its benefits (Ronneberger et al, 2015).

3.4. RR-UNet Network

Recurrent Residual U-Net combines the power of both residual connections and recurrent neural networks (RNNs). The RRU-Net architecture has a similar structure to the U-Net, with encoder and decoder parts that are connected by skip connections. Recurrent connections between the layers are included into RRU-Net, which enables the model to maintain temporal information. Moreover, RRU-Net employs residual connections rather than recurrent ones, which helps to solve the vanishing gradient issue and enables training to complete more quickly (Alom et al, 2018).

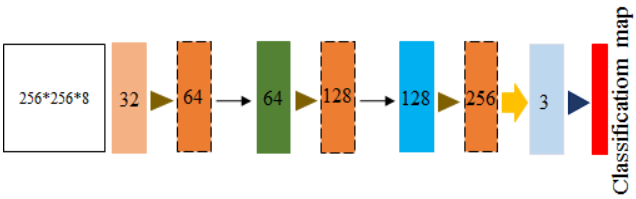
3.5. Attention-UNet Network

Layers of Attnention-UNet are commonly referred to be soft attention gates. This layer extracts spatial data from detailed feature maps. The information is then included in the decoding stage. The inputs to the attention gate consist of two vectors: vector x, which undergoes a strided

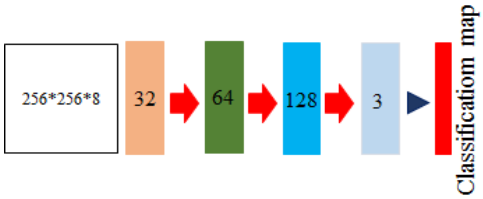
convolution, and vector g, which is obtained from the network's next-lowest layer and undergoes a 1x1 convolution. Aligned weights increase in size as a consequence of the sum of the x and g vectors. Result vector's dimensions are reduced after it passes through a ReLU activation layer and a 1x1 convolution. The attention coefficients (weights) are created by scaling this vector between the [0, 1] range, with coefficients closer to 1 denoting more relevant information (ReLU was used in our work). Trilinear interpolation is used to upsample attention coefficients to the x vector's original dimensions. The initial x vector is multiplied by the attention coefficients, and the result vector is handed along in the skip connection (Oktay et al, 2018).

3.6. UNet 3+ Network

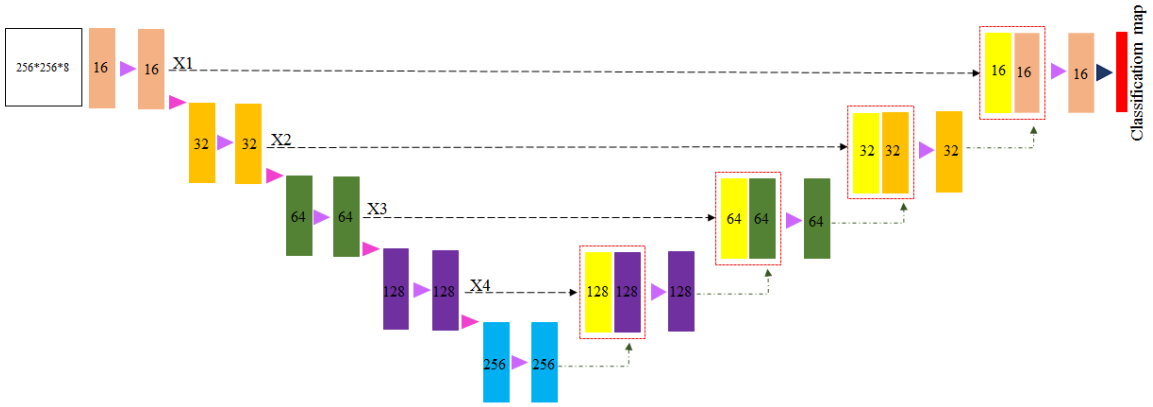
The UNet 3+ architecture has several parts such as the encoder, the decoder, and the skip connections allow the decoder to access information from the corresponding layers in the encoder. This helps to preserve fine-grained details during upsampling and allows the network to better handle objects of different sizes and shapes. Thus, UNet 3+ combines the multi-scale features by re-designing skip connections as well as utilizing a full scale deep supervision, which provides fewer parameters (Huang et al, 2020).



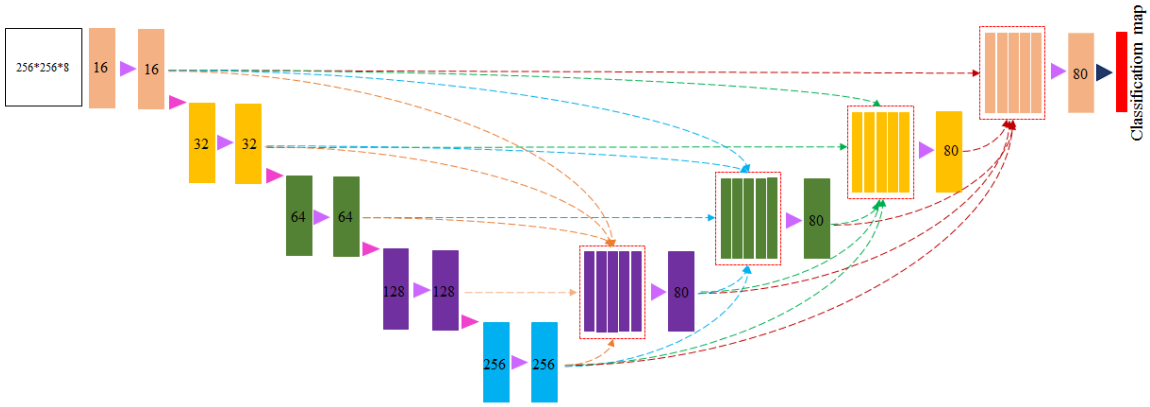
(a) 1D CNN-LSTM



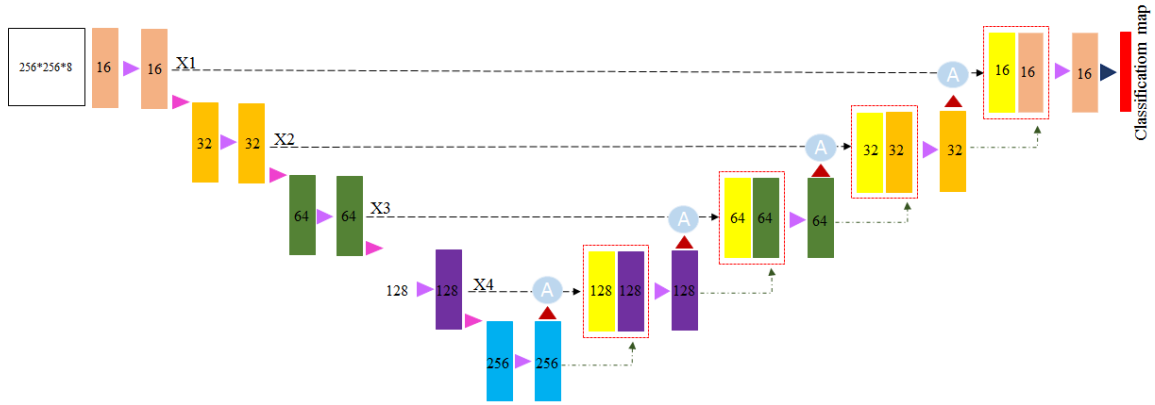
(b) 1D CNN



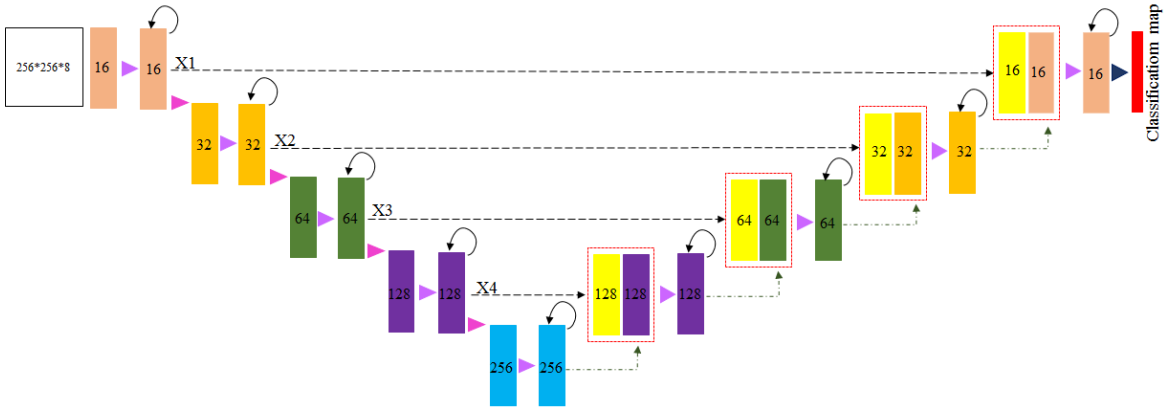
(c) 2D UNet



(d) 2D UNet 3+



(e) 2D Attention UNet



(f) 2D Recurrent Residual UNet

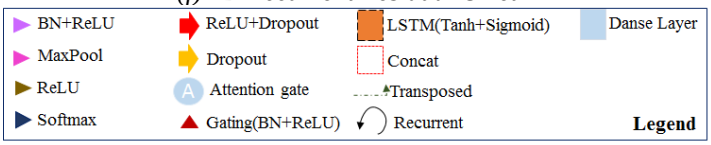


Figure 4. Architecture of proposed networks.

4. Results

*According to Table 1, NDVI multi temporal index is calculated by using Landsat-8 multi temporal images as input networks. We use from image patches with the size of 256*256*8 as input for each of the networks. In this study, 1773936 and 368890 number pixels were selected to train and test the networks, respectively (Table 2).*

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

The number of patches	*The number of pixels for Soybean class*	*The number of pixels for corn class*
35	*686077*	*1087859*
7	*130485*	*238405*

*According to Fjgure 4, the common part of network architecture is divided into two phases: encoding and decoding. The encoding step included a convolution block, a ReLU activation function, max pooling, and downsampling (the number of feature channels is doubled). The decoding step comprises of an up-sampling of the feature map, convolution for half the number of feature map (up-convolution), a concatenation, convolution block, and ReLU activation function. Finally, a 1*1 layer convolution is utilized with the required number of classes (Ronneberger et al., 2015). Dropout is used to minimize overfitting and build various topologies by randomly eliminating neurons (Garbin et al., 2020). Batch Normalization is used to maintain the distribution of each layer's input values and to accelerate learning (Garbin et al.,2020). The decoding phase had five layers: a 2D Convolution block, a ReLU activation function, a Batch Normalization layer, and a max-pooling layer. The number of filters in each layer of the decoding step was 16, 32, 64, 128, and 256, respectively. The decoding phase also included four layers: a 2D Convolution block, a ReLU activation function, a Batch Normalization layer, and an up-sampling layer. The number of filters in each stage of the decoding step was 128, 64, 32, and 16, respectively.*

Table 3. Results by UNet family networks, CNN, and CNN-LSTM. (green-best results, red-worst results)

Method	Metric	Soybean	Corn	Total	Time
CNN	OA	92.90	91.18	92.04	22m 29s
	Kappa	82.49	82.20	82.35	
	Recall	86.4	93.30	89.67	
	F-score	87.47	91.19	89.33	
CNN-LSTM	OA	93.20	91.29	92.25	2h 30m
	Kappa	83.14	82.42	82.78	
	Recall	87.31	92.54	89.93	
	F-score	87.89	91.40	89.64	
Attention-UNet	OA	95.19	93.17	94.48	23m 51s
	Kappa	87.99	87.41	87.70	
	Recall	91.37	94.99	93.18	
	F-score	91.34	93.88	92.61	
UNet 3plus	OA	94.22	92.71	93.46	1h 52m
	Kappa	85.59	85.28	85.37	
	Recall	90.67	93.33	92.00	
	F-score	89.46	92.85	91.15	
RR-UNet	OA	95.37	93.90	**94.63**	1h 11m 19s
	Kappa	88.46	87.66	**88.06**	
	Recall	91.73	95.00	**93.33**	
	F-score	91.69	94.00	**92.85**	
UNet	OA	94.94	93.45	94.19	31m 23s
	Kappa	87.33	86.77	87.05	
	Recall	91.66	94.67	93.16	
	F-score	90.85	93.56	92.21	

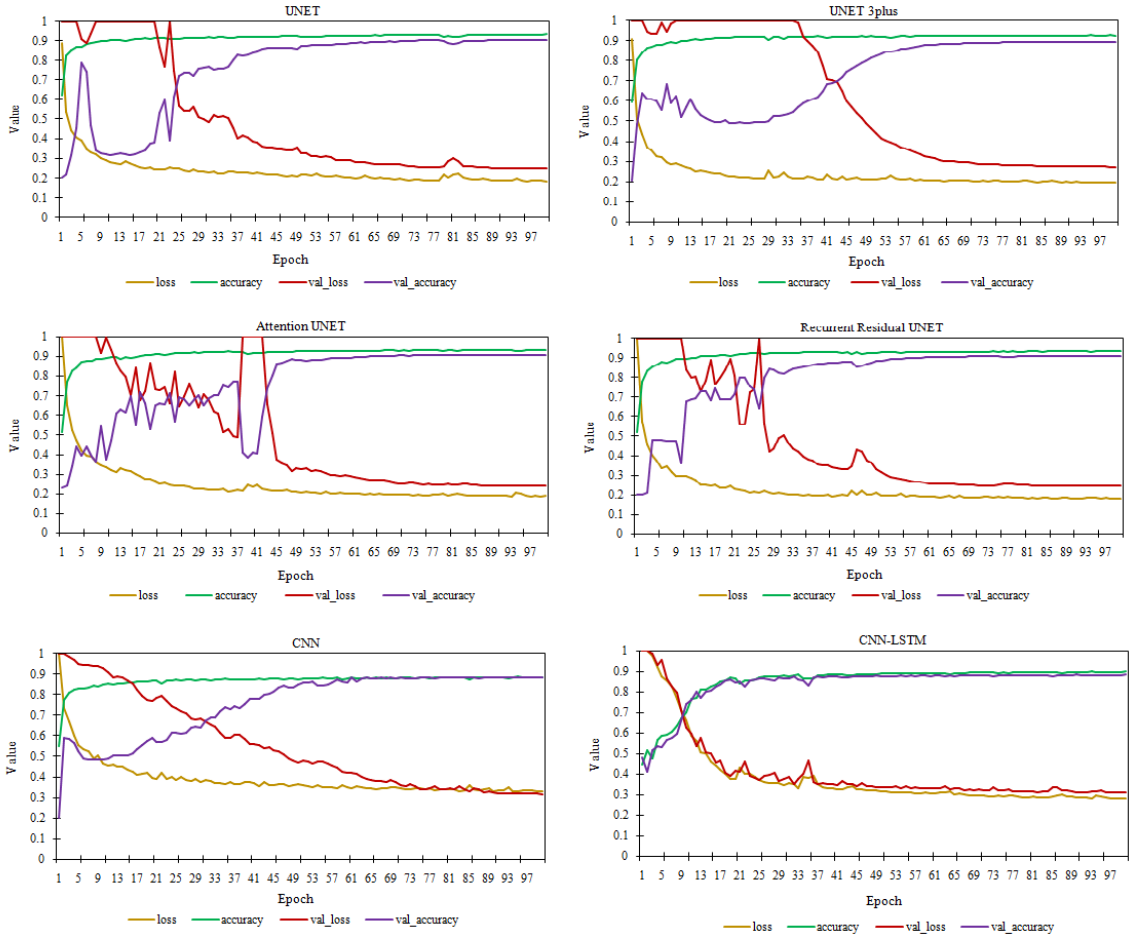


Figure 5. Accuracy and Loss curves for the six architectures used in the training phase.

After the decoding phase, the last layer included a 2D Convolution and Softmax activation function. Ultimately, the category Entropy loss function is utilized to determine weight and bias parameters, and the network is trained using the ADAM method. In the final model, the batch size was 5 and the number of epochs was 100. In this work, the callback function was employed to regulate the loss function with a learning rate of 0.002.

The UNet family was compared to the CNN and CNN-LSTM models. The suggested 1D-CNN model consists of three 1-D convolution layers with 32, 64, and 128 neurons, a ReLU activation function, a Dropout layer with a rate of 0.5, and a fully connected (Dense layer) with a softmax activation function. The suggested 1-D CNN-LSTM model has three 1-D convolution layers with 32, 64, and 128 neurons, as well as LSTM layers with 64, 128, 256 neurons, a ReLU activation function, a Dropout layer with a rate of 0.5, and a fully connected (Dense layer) with a softmax activation function. Table 3 displays the categorization findings for two types of corn and soybean. Figure 5 depicts the accuracy and loss curves for the six architectures with kernel size 3×3 using training datasets. Figure 6 shows the maps generated by six distinct architectures for test images.

5. Discussion

To map crop fields at the state and county level, remote sensing methods and deep learning algorithms have been suggested. Optical and radar sensors are employed to map agricultural lands. The primary obstacles to mapping crop fields are (1) the cloud in optical views, (2) the inability to recognize tiny crop fields, (3) the poor accuracy of mapping crop fields using radar images relative to optical images, and (4) the spectral similarity of crop classes (Fathi & Shah-Hosseini, 2021). To map agricultural fields, machine learning and deep learning algorithms are used. Since high-level characteristics are derived from spectral bands and indices, deep learning techniques are more accurate than machine learning approaches. (Fathi & Shah-Hosseini, 2021). Machine learning and deep learning algorithms are used to map crop fields. Deep learning techniques are more accurate than machine learning methods due to the extraction of high-level characteristics from spectral bands and indices by deep learning algorithms. Using deep learning algorithms with phenological factors, the accuracy of mapping agricultural fields is enhanced.

According to Figure 6, the primary issue with mapping crop fields using 1D CNN and 1D CNN-LSTM is the inability to recognize crop field edge pixels and the lack of class separation in some portions. Unet family networks might be used as a means of resolving this difficulty. In this research, we mapped corn and soybean fields using UNet, UNet 3+, Attention UNet, and Recurrent Residual UNet networks. UNet 3+ incorporates skip-connections across several resolutions and densely linked convolutional blocks, in compared to UNet. UNet 3+ is more sophisticated than UNet, resulting in a longer training period but fewer parameters than UNet. In our research, full-scale aggregated feature maps have eighty channels. In this investigation, UNet 3+ had the lowest accuracy, kappa, recall, and f-score. Compared to UNet, Attention UNet has an extra attention module that prioritizes more relevant characteristics during encoding and decoding. Since the architecture of the Internet is increasingly complicated, it requires more memory. In terms of precision, kappa, recall, and f-score, Attention UNet performed better than UNet and UNet 3+. Attention UNet was quicker than three other UNet in terms of speed. In this study, the ReLU function was used to gate attention. Recurrent Residual UNet employs a mix of recurrent and residual neural network blocks (RRNN) to enhance the performance of picture categorization in comparison to UNet. The recurrent blocks aid in capturing long-range relationships in a picture, while the residual blocks address the issue of gradients that disappear. Recurrent Residual UNet had the greatest accuracy in this investigation. Recurrent Residual UNet is slower than Attention UNet and UNet, but quicker than UNet 3+. Recurrent Residual UNet has the most stringent settings compared to the other three UNet networks. Table 4 lists the number of network parameter values. Thus, according to Tables 4, the Recurrent Residual UNet network had the highest accuracy (94.63), kappa (88.06), recall (93.33), and f-score (92.85). Among the UNet family, the Recurrent Residual UNet network has the most parameters. Attention UNet was the fastest network in the UNet family (23m 51s). The UNet 3+ network showed the lowest accuracies (93.46), kappa (85.37), recall (92.00), and f-score (91.15). UNet 3+ was the slowest of the UNet series of networks (1h 52m). According to Table 4, UNet 3+ had the fewest parameters (1.6 M), whereas Recurrent Residual UNet had the most (6.3 M).

Table 4. The number of network parameters

Method	Parameter Number
CNN	32,995
CNN-LSTM	581,091
UNet	2,164,595
UNet 3+	1,638,515
Attention UNet	2,000,799
Recurrent Residual UNet	6,391,699

6. Conclusions

This research was conducted to evaluate the performance of different deep learning models in distinguishing between soybean and corn fields. The models that were tested included UNet, UNet 3+, Attention UNet, Recurrent Residual UNet, 1D CNN, and 1D CNN-LSTM. The experimental results showed that the Recurrent Residual UNet model performed the best in accurately identifying soybean and corn fields, with Kappa values of 88.46% and 87.66%, respectively. This means that the model had a very high level of agreement with human raters in identifying the correct crop type.

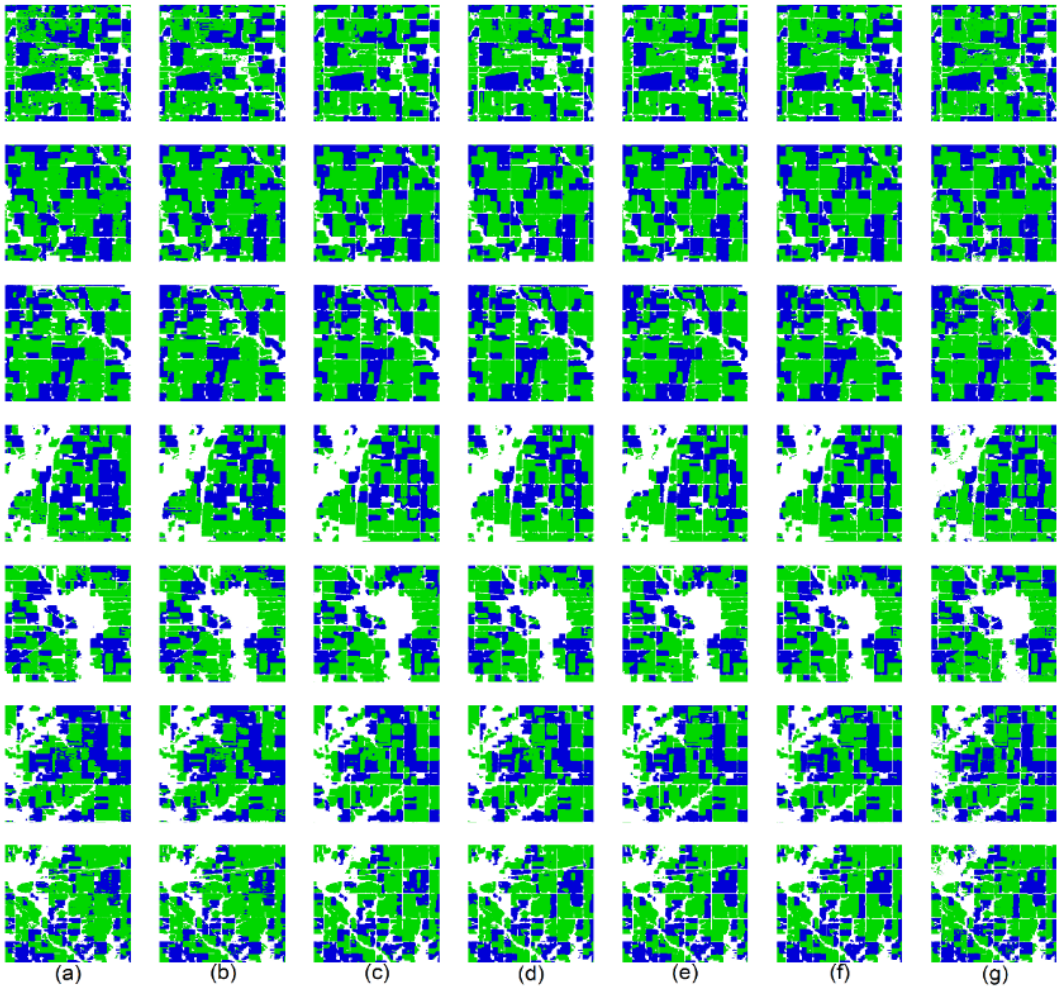


Figure 6. Generated maps by six different architectures (Green=corn and Blue=soybean). (a) CNN, (b) CNN-LSTM, (c) UNet, (d) UNet 3plus, (e) Attention UNet, (f) Recurrent Residual UNet, and (g) ground truth.

The study also found that the 2D Recurrent Residual UNet network was particularly effective in mapping multiple crop varieties. This is significant because it suggests that deep learning models could potentially be used to distinguish between different types of crops and map them accurately, which could be beneficial for agricultural research and management. Overall, this research highlights the potential of using advanced deep learning models for crop classification and mapping, which could have important implications for improving agricultural efficiency and sustainability.

References

Alom, M. Z., Hasan, M., Yakopcic, C., Taha, T. M., & Asari, V. K. (2018). Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. arXiv preprint arXiv:1802.06955.

Amani, M., Ghorbanian, A., Ahmadi, S. A., Kakooei, M., Moghimi, A., Mirmazloumi, S. M., ... & Brisco, B. (2020). Google earth engine cloud computing platform for remote sensing big data applications: A comprehensive review. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, 5326-5350.

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.

Fathi, M., & Shah-Hosseini, R. (2023). Automatic Corn and Soybean Mapping Based on Deep Learning Methods (case Study: Hamilton, Hardin, Boone, Story, Dallas, Polk, and Jusper Counties in Iowa State). ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 10, 187-193.

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.

Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., ... & Wu, J. (2020, May). UNet 3+: A full-scale connected UNet for medical image segmentation. In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1055-1059). IEEE.

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.

Liu, X., Yu, L., Zhong, L., Hao, P., Wu, B., Wang, H., ... & Gong, P. (2019). Spatial-temporal patterns of features selected using random forests: A case study of corn and soybeans mapping in the US. International Journal of Remote Sensing, 40(1), 269-283.

McNairn, H., Kross, A., Lapen, D., Caves, R., & Shang, J. (2014). Early season monitoring of corn and soybeans with TerraSAR-X and RADARSAT-2. International Journal of Applied Earth Observation and Geoinformation, 28, 252-259.

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., Sarmadian, A., Mohammadzadeh, A., Celik, T., Amani, M., & Kusetogullari, H. (2021). Distortion robust relative radiometric normalization of multitemporal and multisensor remote sensing images using image features. IEEE Transactions on Geoscience and Remote Sensing, 60, 1-20.

Moghimi, A., Mohammadzadeh, A., Celik, T., Brisco, B., & Amani, M. (2022). Automatic Relative Radiometric Normalization of Bi-Temporal Satellite Images Using a Coarse-to-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.

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., ... & Rueckert, D. (2018). Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999.

Paludo, A., Becker, W. R., Richetti, J., Silva, L. C. D. A., & Johann, J. A. (2020). Mapping summer soybean and corn with remote sensing on Google Earth Engine cloud computing in Parana state–Brazil. International Journal of Digital Earth, 13(12), 1624-1636.

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). U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18 (pp. 234-241). Springer International Publishing

Sun, Z., Di, L., & Fang, H. (2019). Using long short-term memory recurrent neural network in land cover classification on Landsat and Cropland data layer time series. International journal of remote sensing, 40(2), 593-614.

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, S., Gu, L., Li, X., Jiang, T., & Ren, R. (2020). Crop classification method based on optimal feature selection and hybrid CNN-RF networks for multi-temporal remote sensing imagery. Remote sensing, 12(19), 3119.

Zhong, L., Gong, P., & Biging, G. S. (2014). Efficient corn and soybean mapping with temporal extendability: A multi-year experiment using Landsat imagery. Remote Sensing of Environment, 140, 1-13.

Zhong, L., Yu, L., Li, X., Hu, L., & Gong, P. (2016). Rapid corn and soybean mapping in US Corn Belt and neighboring areas. Scientific reports, 6(1), 1-14.