Multispectral Crop Yield Prediction Using 3D-Convolutional Neural Networks and Attention Convolutional LSTM Approaches

Seyed Mahdi Mirhoseini Nejad, Dariush Abbasi-Moghadam, Alireza Sharifi, Nizom Farmonov, Khilola Amankulova, and Mucsi László

***Abstract*—In recent years, national economies are highly affected by crop yield predictions. By early prediction, the market price can be predicted, importing, and exporting plan can be provided, social, and economic effects of waste products can be minimized, and a program can be presented for humanitarian food aid. In addition, agricultural fields are constantly growing to generate products required. The use of machine learning (ML) methods in this sector can lead to the efficient production and high-quality agricultural products. Traditional predictive machine models were unable to find nonlinear relationships between data. Recently, there has been a revolution in prediction systems via the advancement of ML, which can be used to achieve highly accurate decision-making networks. Thus far, many strategies have been used to evaluate agricultural products, such as DeepYield, CNN-LSTM, and ConvLSTM. However, preferable prediction accuracy is required. In this study, two architectures have been proposed. The first model includes 2D-CNN, skip connections, and LSTM-Attentions. The second model comprises 3D-CNN, skip connections, and ConvLSTM Attention. The Input data given from MODIS products such as Land-Cover, Surface-Temperature, and MODIS-Land-surface from 2003 to 2018 on the county level over 1800 counties, where soybean is mainly cultivated in the USA. The proposed methods have been compared with the most recent models. Then, the results showed that the second proposed method notably outperformed the other techniques. In case of MAE, the second proposed method, DeepYield, ConvLSTM, 3DCNN, and CNN-LSTM obtained 4.3, 6.003, 6.05, 6.3, and 7.002, respectively.**

***Index Terms*—3D-CNN, ConvLSTM, forecasting, LSTM attention, skip connection.**

I. Introduction

ALONG time ago, manual surveys were one of the most widely used sources in crop forecasting [1]. Afterward, mathematical models have been introduced with several parameters, and data are required in the original place to be collected [2]. Meanwhile, collecting data also are so expensive, and difficult to measure. Recently, machine learning (ML) techniques were used for accurate prediction models [3]. Indeed, it results in food availability in the future, and also the product resources demand can be used optimally [4]. Moreover, digital and intelligence farming with the usage of remote sensing data leads the farmers to get closer to new advanced innovation methods. To have incredibly optimum lands, it requires increasingly to extract data from satellites. Therefore, it demands several datasets such as soil, weather, water, climate, use of fertilizers, etc. [5]. This illustrates that crop yield forecasting is not only a straightforward assignment, but also comprises different complex stages. After collecting the data, it is necessary to provide an optimal decision-making system that leads to ensuring the sustainability of human food resources [6].

Although crop yield forecasting systems can reasonably predict accurate crop yields, a higher yield forecasting quality is always desirable [7]. ML methods, which have been utilized considerably from supermarket to customer treatment evaluation, can be used to make an accurate crop yield forecasting [8]. ML can discover the patterns and relationships between extracted features, and also can determine valuable information. Some traditional ML techniques, including support vector machines, decision trees, and artificial neural networks, have been used recently in prediction applications [9]. In additional, deep learning, which is a subset of ML, can be used to achieve high accurate rates during the cultivation and harvesting period, because of having deeper networks. Therefore, when a system becomes deeper, the network's efficiency will increasingly improve [10].

In recent years, a great number of researchers have evaluated different ML methods in crop yield prediction field. Convolution neural network, which is a subset of the ML, is the most usable ML method. In [11], MLP and CNN architectures have been used to feature extraction by authors. They have employed two methods Watershed Segmentation and Circular Hough Transform, for counting fruits from extracted features. One of the algorithms named WS gained as best output with $R2 = 0.826$. Habaragamuwa et al. [12] used region convolutional neural network methods to detect ripe, and unripe strawberries in the field, and 82.61% was the best accuracy. For detecting apples on trees, Kang and Chen [13] used an RC-CNN, which is a deeper

Manuscript received 2 September 2022; revised 24 October 2022; accepted 15 November 2022. Date of publication 21 November 2022; date of current version 7 December 2022. This work was supported by the University of Szeged Open Access Fund under Grant 5902. *(Corresponding author: Nizom Farmonov.)*

Seyed Mahdi Mirhoseini Nejad and Dariush Abbasi-Moghadam are with the Department of Electrical Engineering, Shahid Bahonar University of Kerman, Kerman 76169-14111, Iran (e-mail: mirhoseini@eng.uk.ac.ir; abbasi-moghadam@uk.ac.ir).

Alireza Sharifi is with the Surveying Engineering, Faculty of Civil Engineering, Shahid Rajaee Teacher Training University, Tehran 16788-15811, Iran (e-mail: a_sharifi@sru.ac.ir).

Nizom Farmonov, Khilola Amankulova, and Mucsi László are with the Department of Geoinformatics, Physical and Environmental Geography, University of Szeged, H-6722 Szeged, Hungary (e-mail: farmonov.nizom@stud.u-szeged.hu; amankulova.khilola@stud.u-szeged.hu; mucsi@geo.u-szeged.hu).

Digital Object Identifier 10.1109/JSTARS.2022.3223423

model. They have reached 86% accuracy. Nosratabadi et al. evaluated the crop yield prediction performance by comparing artificial neural networks-imperialist competitive algorithm and artificial neural networks-gray wolf optimizer models [14]. They found that ANN-GWO predicted better with R of 0.48, RMSE of 3.19, and MEA of 26.65 than the ANN-ICA model. Li et al. [15] have used a prediction model, which indicated an evolutionary method according to LSTM equipped with the attention technique. The results showed that the defined strategy reached an excellent prediction performance compared to the other models.

LSTM, GRU, BiLSTM, and BiGRU were mostly used for forecasting systems. Alibabaei et al. [16] used bidirectional for both LSTM and GRU to crop yield forecasting. They employed time series information like temperature data, irrigation plans, and soil data in their prediction model. The results showed that BiLSTM performs better than the LSTM, GRU, and BiGRU by means of Tomato and Potato prediction. YiledNet was revealed by Khaki et al. [17]. They used this model for transfer learning between soybean and corn yield forecasting. Gong et al. [18] combined RNN and temporal convolutional networks for tomato yield prediction. Results proposed that RSME of their method outperformed both classical and traditional models. Ju et al. [19] compared seven favorite ML methods, on three products: paddy rice in South Korea and soybean, and corn in two states of the US, Illinois, and Iowa, 14 years were trained for prediction for each crop. They have used a series of data indexed consisting of vegetation indices from MODIS data, climate information, product field measurements, and land cover data of city-level spatial resolution. Gholizadeh et al.[20] revealed that the ANN model was a more accurate tool than MLR for predicting fruit yield in coriander. Alwis proposed depth smart LSTM, which consists of DNN with environmental agents, where the data are evaluated with such information [21]. The chemical structures of the vegetables can be assessed with an accuracy of 89%.

In addition, few number of articles has been addressed skip connection for the purpose of predicting. However, in the filed of classification, lots of researchers have revealed that deeper neural networks can extract more features. ResNet is known as one of the most prominent deep networks [22], which is able to remove the vanishing gradient issues. DenseNet has been offered to tackle the ResNet problems. It has higher performance and fewer parameters in comparison to ResNet [23]. The training challenges of deeper systems is enormously decreased. Zhong et al. [24] used spatial-spectral residual system and fast dense connection spatial-spectral. Wang et al. [25] proposed to hyperspectral classification.

Attention mechanism has been first used for the translation duties [26]. The mechanism of attention systems is such that focus the main features and reduce effect of the incoherent data. The convolutional block attention module is applied to HSI input for classification tasks in [27] and [28]. The attention mechanism is applied after each CNN in spatial-spectral attention method [29]. Two structure named double-branch multiattention [29] and double-branch dual-attention (DBDA) mechanism systems [30] used attention mechanism.

According to our investigations, no paper has researched the combination of Attention ConvLSTM and 3DCNN for crop yield prediction. Also, a network with higher accuracy and faster speed is demanded. Therefore, in this article, more attentive spatio-temporal feature extractions have been presented by using the 3D-CNN, convolutional LSTM, and attention mechanism. The proposed models show that the architectures offer more precise crop yield prediction.

Scientists indicate that deep learning methods are the most popular method for crop yield prediction. Therefore, the main purpose behind this research to use satellite images and deep learning methods, to generate yield predictions for any input types, either full dimension images or histograms. The first proposed model actually consists of skip connection and CNN to extract relevant features. Then, a sequence of the 2D convolutional neural networks paralleled with an LSTM equipped with an attention mechanism. The series of CNNs can extract spatial features from the previous stage and combine them efficiently with a time series extractor. Attention LSTM also focuses on those weights, which are intensely interesting. But in the second architecture, 2D-CNNs and LSTM have been replaced by 3D-CNN and ConvLSTM, respectively. Spectral-spatial features are effectively execrated by 3D-CNN and ConvLSTM can extract the Spatio-temporal features impressively. In this article, It is presented that the proposed model will effectively predict crop yield compared to the other competent models.

The structure of this article is organized as follows. Section II indicates the material of some deep learning; Section II-A describes the architecture of the proposed model; Section II-B shows details about the datasets, which is used in this article, and their preprocessing techniques, training details, and evaluation metrics as well; Results and Discussion of the prediction models are shown in Sections III and finally Section IV concludes this article.

II. Material and Methods

Deep learning models have the ability to learn the extracted information over time. These deep learning models can be as listed below: recurrent neural networks, 2D/3D-convolutional neural networks, attention mechanism, and skip connection. In general, the abovementioned networks are composed of several stacked layers, where the input of one layer is the output of the previous layers [31]. This article has tried to use these models efficiently.

A. Proposed Neural Network Model

1) Prediction Framework Settings: In this section, two novel models have been proposed, which have indicated in Figs. 1 and 2, respectively. The first model contains 2D-CNN with the help of skip connections, then it is followed by several 2D-CNNs, then attention LSTM mechanism paralleled with multilayered 2D-CNN for final forecasting. The second method is the same as the first model, with the difference that instead of 2D-CNN layers and an LSTM layer, 3D-CNN and ConvLSTM are used sequentially. The combination of the 2D-CNN and attention LSTM was used in the first model, and 3D-CNN and attention ConvLSTM were applied for the second proposed model, which makes a significant improvement in prediction accuracy.

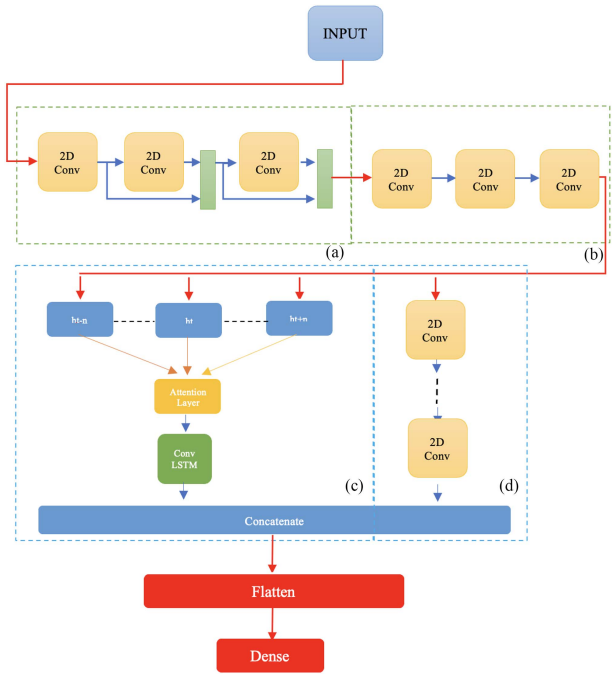


Fig. 1. Deep learning architecture of the first proposed prediction model.

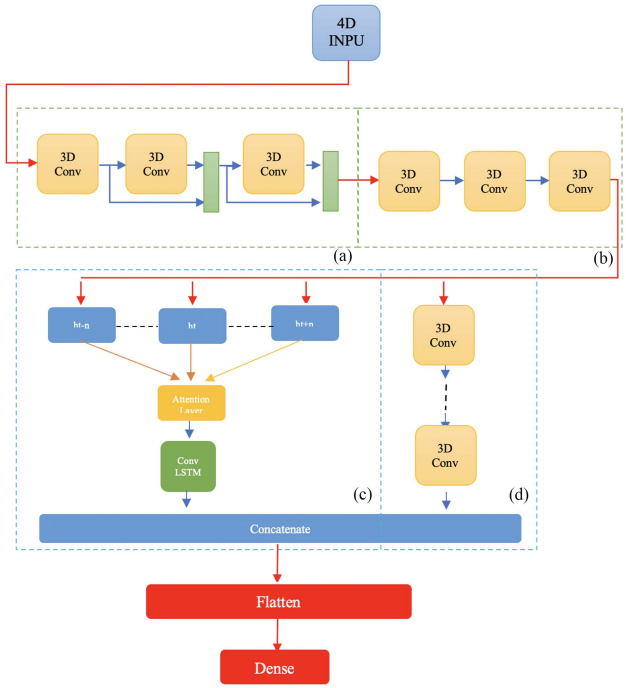


Fig. 2. Deep learning architecture of the second proposed prediction model.

The purpose of this article is to estimate soybean yield in the United States of America. It is worth noting that various types of validation tests have been performed to confirm the best number of layers, which will be fully explained in the following sections. Briefly, 1–5 CNN layers were investigated in the testing phase. The best training rate was selected.

2) Proposed Models Crop Yield Prediction Architecture: Overview of the first proposed model has been shown in Fig. 1.

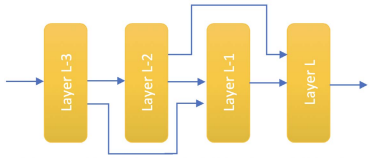


Fig. 3. Skip connections network mechanism.

The proposed prediction model includes four phases: (a) spatial feature extractor, (b) spatial encoder, and a parallel learning deep temporal and spatial feature extractor using attention LSTM, and 2D-CNNs which are indicated in the part (c) and (d). The process is composed of the five-step as follows:

Step 1: In the first part, a 2-D feature extractor, which consists of a three-layer convolution, was used, and shown in Fig. 1(a), where each layer is made of a 3×3 2-D convolution layer with 128 filters, then followed by a batch normalization layer. The nonlinear batch normalization is performed to speed up and sharpen the network [32].

One of the most popular deep learning modules is Conventional Neural Networks which can be used as a feature extractor. CNNs are usually combined with several stacked layers including convolutional, activation function, pooling, and batch normalization [33]. A convolutional layer applies filters to create a convolution operation. Various features are extracted and learned from previous layers. Therefore, differences in each layer can be distinguished at the depths of the network. The results of convolutional layers with a window size of N on input data I can be shown as follows:

$$Z(x,y) = f\left(\sum_{i=0}^{I}\sum_{j=0}^{J} I(x+i, y+j) * N(i,j) + b\right). \quad (1)$$

After convolution layers are done, a nonlinear function is applied to the system named activation function. RLUs are the most preferred activation function [34]. It showed a smooth behavior compared with the other activation functions and also can converge quickly [35].

This stage is also equipped with two skip connections that make the network behavior reliable. Because the tuned weights are adjusted to quiet the upstream layer. In fact, the use of skip connection provides several advantages to the system. First, it avoids saturation of the network, where higher error occurs when the number of layers increases to deepen the models. Second, it leads to a huge reduction in the problems of vanishing gradients. Third, it helps to move the data to the lower layers, which makes it easier to reach the optimal point. The scheme of skip connections has been shown in Fig. 3. The formulation below shows how the skip connections works

$$a_l = f(W_{l-1,l}a_l - 1 + B_l + W_{l-2,l}a_l - 2) \quad (2)$$

where $W_{l-2,l}$ and $W_{l-1,l}$ are tuned weights from layer $l-2$ and $l-1$ to layer l for connection weights which is used for forward propagation. a_l indicates as activation in layer l. g represents as activation function. Meanwhile, back-propagation is formulated

as follows:

$$\Delta W_{l-1,l} = -\eta \frac{\sigma E_l}{\sigma w_{l-1,l}} \tag{3}$$

$$\Delta W_{l-2,l} = -\eta \frac{\sigma E_l}{\sigma w_{l-2,l}} \tag{4}$$

where (3) is used for the standard route and (4) is used for the skipped route. η, here, is the learning rate.

Then, the results of the first and the second CNN block are concatenated to create the input of the next CNN. This scenario is repeated for the next blocks. Since padding has been adjusted to the same, the spatial dimension of each block will be 128×32×9.

Step 2: In Fig. 1(b) represents the second part of the proposed architecture. This part takes a 3-D input and gradually compresses it into an encoded compacted feature tensor representation. This is done by using 3×3 convolutions layers with 128, 256, and 512 filters, then followed by a batch normalization layer. After that, a dropout layer with probability 0.5 is placed at the end of each CNN block, which regularizes the network in each epoch. Dropout is used to prevent networks from overfitting. The spatial dimensions are reduced where the stride is adjusted to 2. Then, the output of convolutional layers is reshaped to a 2-D tensor, which makes it readable for LSTM in step 3.

Step 3: The last two steps of the proposed method are constituted for the temporal (LSTM) and spatial feature extraction. In the LSTM networks, which are mainly used for time series variables [21], [36], [37], [38], [39], [40], [41]. The previous output values will be retained for a short time. Hence, it acts like a memory unit that simplifies the feedback analysis in such a network.

In addition, the LSTM network is associated with an attention mechanism, which works on the basis that prediction of outcomes can be made by applying a conditional probability distribution to the input and past sample outcomes [42]. It is given in the equation as follows:

$$p(y_i|x_1, ..., x_{i-1}, y_{i-1}). \tag{5}$$

A nonlinear approximation function is adopted due to the infeasibility of computation conditional probability distribution. So : $f = (y_i, h_i, C_i,)$, where f is a function for LSTM, h_i is the LSTM inner state, and C_i is the present context, which means, a vector maintaining data of which entrance data are essential at the present stage. Context is taken from the current state, h_i, and the input string x. After the LSTM steps through the entire input sequence, the network of attention system determines what attention should be paid to the annotations provided at each step. By the calculation of e_t, which is stated in equations, the mechanism of attention will start.

$$e_t = v^T.\tanh(W_e.h_t + U_e.d_{t-1} + b) \tag{6}$$

where d describes the input and score of the attention are computed by the soft-max function as follows:

$$a^{t,t\cdot} = \frac{\exp(e_t)}{\sum_{j=1}^{T} \exp(e_t)} \tag{7}$$

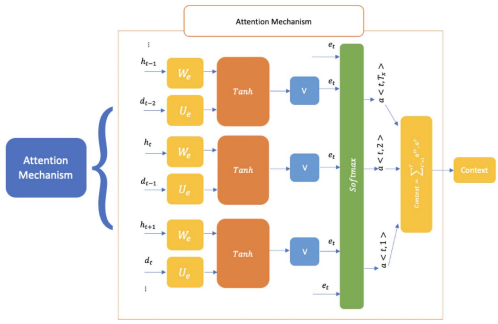


Fig. 4. Attention mechanism.

where $t\cdot$ denoted each t time.

By the weighted sum of the C_i the context vector is computed. Fig. 4 shows the mechanism of attention which is described as follows.

$$C_t = \sum_{t\cdot=1}^{T} a^{t,t\cdot}.h_{t\cdot} \tag{8}$$

As shown in Fig. 1(c), the attention systems are used for the output of each LSTM module to create the corresponding long-term dependencies, which was first implemented by Bahdanau [43]. The main idea behind the attention mechanism is to gives the network the flexibility to use the most relevant parts of the input sequence through a weighted combination of all encoded input vectors. The most relevant vector is assigned the highest weight. In the proposed method, the attention mechanism has been used for LSTM layers to predict crop yield by making a context vector as a weighted sum of all provided information. All hyperparameters were calculated, such as related context, and learned states, then weights of the attention related to states were applied to perform the attention system.

Step 4: Multi 2D-convolutional layers have been paralleled with attention LSTM, which consists of 3×3 convolution layers with 128 filters, then followed by a batch normalization layer. The main reason for doing this technique is that the prediction model responded to superior behavior.

Step 5: The results of the attention LSTMs and 2D-CNNs are concatenated, then flatten to a 1×12800 output. Finally, a single layer of dense is responsible for the latest forecasting of the predicted value.

In Fig. 2, the architecture of the second model has been shown. Same as the previously proposed model, this method also contains four parts: (a) spectral-spatial feature extractor, (b) spectral-spatial encoder, a parallel learning deep model including spectral feature extractor were applied using 3D-CNN which is shown in (c), and (d) spectral-spatial-temporal feature extractors have been used by attention convolutional LSTM.

The input is transformed to a 4-D matrix before entering the network and its dimensions will be in the form 32×32 × 9×1. Unlike the first proposed method, this model used a 3-D convolutional neural network. 3D-convolutional neural networks are another model of CNNs that includes an additional step. In the year 2013, 3D-CNN model has been first released by

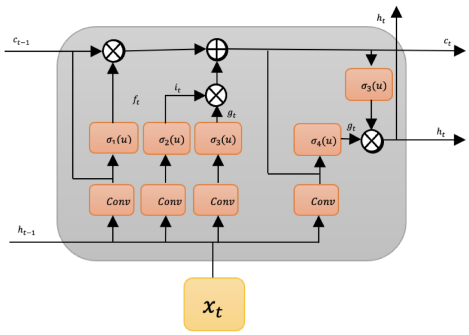


Fig. 5. Architecture of a ConvLSTM.

Shuiwan [44] and it can be used efficiently in remote sensing applications as these data include spectral and temporal features. When 3D-CNN is applied to the hyperspectral input, the results will be as follows:

$$Z(x,y,d) = f\Bigg(\sum_{i=0}^{I}\sum_{j=0}^{J}\sum_{d=0}^{K} I \times (x+i, y+j, d+k) * N(i,j,k) + b\Bigg) \quad (9)$$

where K is demonstrated as the spectral size of the 3-D kernel.

A 3D-CNN can extract both spatial and spectral features at the same time, which can provide more specific extracted features in comparison with 2D-CNN [4]. The used 3D-CNN includes a 3×3×1 kernel size, with 128 filter size, followed by batch normalization. Fig. 2(a) shows the architecture of the first part. The other settings are as the same the first proposed model, the output dimension size of each 3D-CNN will be in form of 128×32 × 9×1, and the padding is adjusted to the same.

In part (b), the dimension of the spectral features will be 256×16 × 5×1, due to applying a 3D-CNN with 3×3×1 kernel size of 128, 512, and 256 filter size, respectively. Since the input data have both spectral, and spatial features, 3D-CNN can be used efficiently. After that, it is followed by batch normalization, and dropout is adjusted to 0.5, which leads to preventing overfitting issues.

The results from part (b) flowed into two sections including the convolution LSTM network, which is coupled with attention mechanism shown in (c), and three layers of the 3D-CNN indicted in part (d), which leads to better overall network performance.

In part (c), ConvLSTM has been used. The combination of LSTM and CNN was presented first time by Shi [45]. In comparison to LSTM, the ConvLSTM network significantly reduces the network's parameters. Therefore, the model can be deep enough during the training phase. Also, ConvLSTM is well known to extract the inherent Spatio-temporal features of wide-ranging input [46]. In Fig. 5, simple architecture of the ConvLSTM has been indicated, where $\sigma_1(u)$, $\sigma_2(u)$, $\sigma_3(u)$, $\sigma_4(u)$ are forget, input, $\tanh$, and output gate, respectively. Here, (10) to 14 represent the functionality of ConvLSTM architecture

TABLE I
SUMMARY OF THE SECOND PROPOSED METHOD

Layer	Output Shape	# of the Parameters
InputLayer	(32,32,9,1)	0
conv3d	(128,32,9,1)	110720
BatchNormalization	(128,32,9,1)	512
conv2d_1	(128,32,9,1)	442496
batch_normalization_1	(128,32,9,1)	512
concatenate	(256,32,9,1)	0
conv3d_2	(128,32,9,1)	884864
batch_normalization_2	(128,32,9,1)	512
concatenate_1	(384,32,9,1)	0
conv3d_3	(128,32,9,1)	442496
batch_normalization_3	(128,32,9,1)	512
dropout	(128,32,9,1)	0
conv3d_4	(512,32,9,1)	590336
batch_normalization_4	(512,16,5,1)	2048
dropout_1	(512,16,5,1)	0
conv3d_5	(256,16,5,1)	1179904
batch_normalization_5	(256,16,5,1)	1024
dropout_2	(256,16,5,1)	0
conv_lstm2d	(128,16,5)	1769984
batch_normalization_6	(128,16,5)	512
reshape_1	(2048,5)	0
last_hidden_state(Lambda)	(2048)	0
attention_score_vec (Dense)	(2048,5)	4194304
attention_score (Dot)	5	0
reshape_3 (Reshape)	(128,16,5,1)	0
attention_weight (Activation)	(5)	0
conv3d_6	(128,16,5,1)	147584
context_vector (Dot)	(2048)	0
batch_normalization_8	(128,16,5,1)	512
attention_output	(4096)	0
conv3d_7	(128,16,5,1)	147584
attention_vector (Dense)	(512)	2097152
batch_normalization_9	(128,16,5,1)	512
dropout_3	(512)	0
conv3d_8	(128,16,5,1)	147584
batch_normalization_7	(512)	2048
batch_normalization_10	(128,16,5,1)	512
reshape_2	(1,4,128,1)	0
reshape_4	(20,4,128,1)	0
concatenate_2	(21,2,4,128,1)	0
flatten	(10752)	0
MAE	1	10753
MAPE	1	10753
MSLE	1	10753
RSME	1	10753
Total params:	12,207,236	
Trainable params:	12,202,628	

and * represents convolution operation

$$i_t = \sigma(W^*_{xi}x_t + W^*_{ai}a_{t-1} + W_{ci}c_{t-1} + b_i) \quad (10)$$

$$f_t = \sigma(W^*_{xf}x_t + W^*_{af}a_{t-1} + W_{cf}c_{t-1} + b_f) \quad (11)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W^*_{xc}x_t + W^*_{ac}a_{t-1} + b_c) \quad (12)$$

$$g_t = \sigma(W^*_{xo}x_t + W^*_{ao}a_{t-1} + W_{co}c_{t-1} + b_0) \quad (13)$$

$$h_t = o_t \tanh(c_t). \quad (14)$$

Also, the attention mechanism is used to creating the corresponding long-term dependencies. Then, three 3-D CNN layers have been paralleled with attention ConvLSTM, which consists of a 3×3×1 as kernel size layers of 128 filters, and each CNN is followed by a batch normalizations layer. Last, a single layer of the proposed network is responsible for the latest forecasting of

the forecasted value. Table I, shows the summary of the second proposed method.

B. Investigational Study

1) Datasets: This section describes the datasets, which have been used in the experiment. The datasets consist of MODIS satellite data (moderate resolution imaging spectroradiometer) as the input data and the soybean products are used as observed data. All US counties where soybeans have been grown were selected with no product restrictions. The land products MODIS of the NASA, surface reflectance, surface temperature, and land cover kind are prepared by the Google Earth Engine (GEE) [47]. MODIS satellite images are collected 32 times per year, every eight days, into about 1834 counties from 2003 to 2018.

2) Ground Truth Data: The United State Division of Agriculture provides open-source information for agribusiness. The ground truth data are achieved from the USDA Quick Stats Database [48] for years of our interest from 2003 to 2018 on the county level. The sum of 1848 U.S. provinces cultivated soybean products.

3) Dataset Split: Neural network training data usually split the input data into three different phases including the training, validating, and test phases, which are used to assess the performance of the final model.

The training datasets were accidentally chosen. 80% of the counties in each state were selected in training sets by random to make sure that the cities are geomorphology equally divided. The rest of the counties were chosen as the validation data. Twelve training sets have been trained for the baseline repetition.

4) Evaluation Metrics: The efficiency of the forecasting method has been measured using mean absolute error, root mean square error, mean absolute percent error, and mean square logarithmic error. The MAE, RMSE, MAPE, and MSLE for these measurements are shown as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\frac{A_t - F_t}{A_t}\right)^2} \tag{15}$$

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|A_t - F_t| \tag{16}$$

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \tag{17}$$

$$\text{MSLE} = \frac{1}{n}\sum_{t=1}^{n}(\log(A_t + 1) - \log(F_t + 1))^2 \tag{18}$$

where observed data as ground truth data shown by A_t and forecasted values depicted by F_t.

5) Feature Normalization: Minimum–Maximum normalization has been applied over all features, which have been rescaled the values between 0 and 1

$$S_i = \frac{g_i - \min(g)}{\max(g) - \min(g)} \tag{19}$$

where S_i and g_i are denoted normalized features and primary data in the ith feature, and maximum and minimum values of the features are shown as $min(g)$ and $max(g)$.

6) Data Preprocess: Training data directly from raw images is impossible due to lake of labeled training data. Also, using the other famous benchmarks for pretraining such as ImageNet, is infeasible. Therefore, after extracting data from MODIS, In the first step, land cover is applied to set the values to zero, where croplands are not labeled by creating a mask. Then, a 32-bins histogram is considered for each band, in which a 32×9 histogram matrix was generated per image. Also, the images are captured 32 times per annual and stacked to create a $32 \times 32 \times 9$ 3D-histogram per county per year. Several individual pixels provide helpful information. Since the position of the cropland has been illustrated, crop yield will not be changed if the position of the image pixels changes.

III. Results and Discussion

In this section, the outcome of the proposed work was presented. First, it is illustrated how the number of CNN layers impacts the performance of the prediction systems. Second, the results from proposed method compared with DeepYield [4], ConvLSTM [49], 3DCNN [50], and CNN-LSTM [38].

1) Execution Details: In the experiment, models were implemented by python using Tensor-flow and Keras libraries 2.8.0. A high-performance computing platform has been used at the Shahid Bahonar University of Kerman to accelerate computational, simulation, and modeling processes in this research. This platform has 11 computing nodes and uses three GPUs. This configuration provides 16TFLOPS computing power for the university's researchers. Our data consists of multispectral images from 2003 to 2018. In the training phase, the data from 2003 to 2014 are selected to tuning the hyperparameters. This means that 12 years $\times$ 1834 counties, equivalent to 22 008 data, must be trained. Also, the next four years from 2015 to 2018 were selected for the tasting phase. During the training process, several optimizers such as Adam and SGD optimizer were tested in this experiment. Finally, the Adam Optimizer is selected with a learning rate of 10^{-5}. To prevent overfitting of the trained model, early stopping has been used in the validation. In addition, different batch sizes such as 32, 64, 128, and 256 were used in the training phase. The training phase lasts about 4 h and 15 min.

2) Effect of the Number of Layers on the Crop Forecasting Efficiency: One of the essential issues that should be considered in deep learning is the optimal number of layers in the deep neural systems. Therefore, the results will be directly affected by changing the number of neural networks. The effects of changing the number of layers have been measured in the proposed method. As shown in Fig. 1(a)–(c), which includes 2D-CNN Skip Connection and CNNs layers were paralleled with attention LSTM, the different number of CNN layers was evaluated to check prediction performance. Fig. 6 shows the forecasting error by changing the layers from 1 to 5.

The results of Fig. 6 show that with the increase in the number of layers, the error trend gradually decreases. In all tested cases

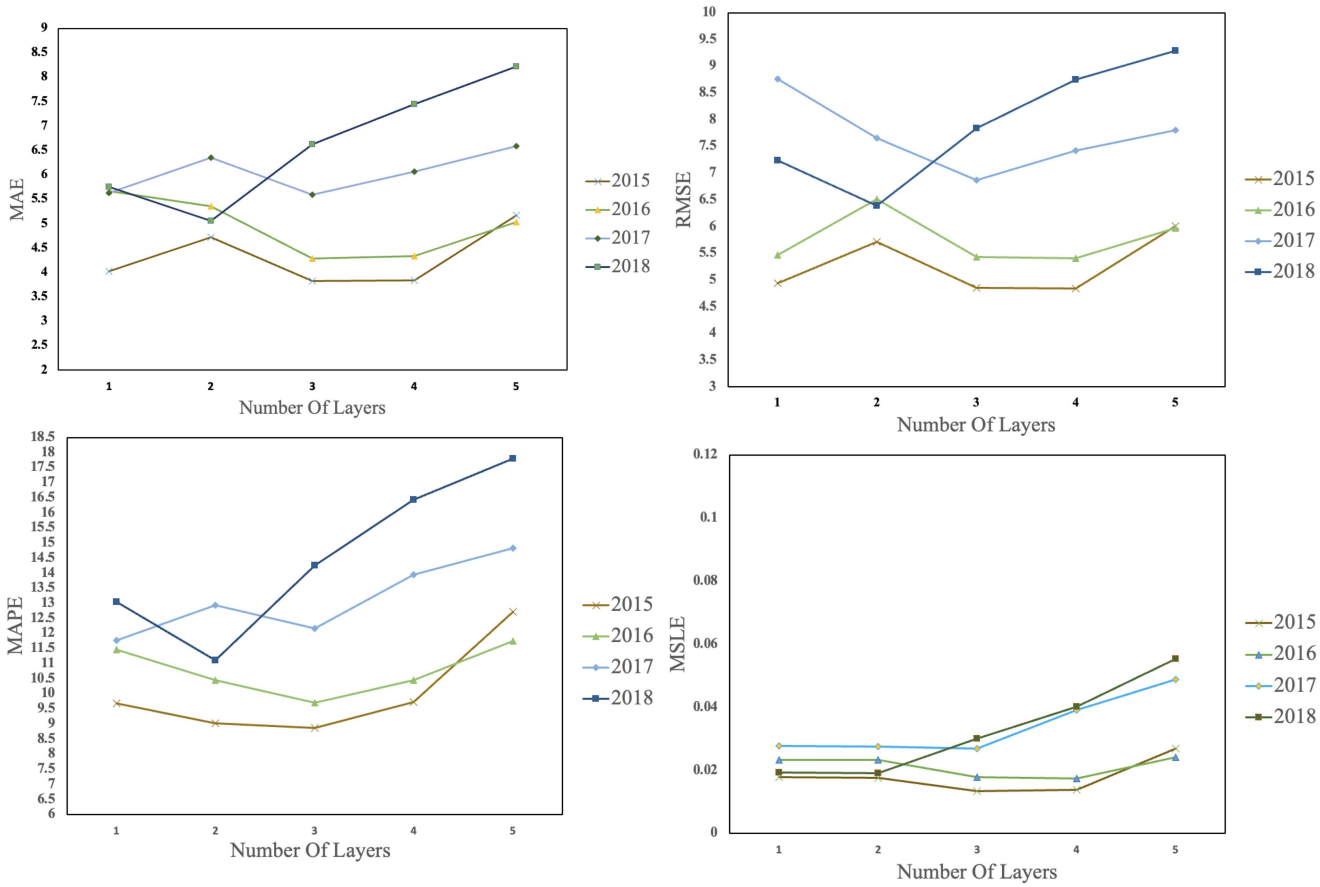


Fig. 6. Error orientation at different number of layers on the test set.

TABLE II
ERRORS WITH VARYING LAYERS ON TEST DATA

Layers	MAE	MAPE	RMSE	MSLE
1	5.6747	11.4653	5.4744	0.0234
2	5.3603	10.4593	6.5084	0.0234
3	**4.2938**	**10.2732**	5.4306	**0.0199**
4	4.3366	10.4589	**5.414**	0.0219
5	5.0311	11.7542	5.9788	0.0242

Bold font indicates best result obtained for analysis.

TABLE III
COMPARISON BETWEEN TWO PROPOSED MODELS

MAE	2015	2016	2017	2018	Average
P.Model 1	**3.8335**	4.2938	5.5907	6.6229	5.0852
P.Model 2	3.8348	**4.2883**	**5.3811**	**4.0715**	**4.393925**
RMSE	2015	2016	2017	2018	Average
P.Model 1	**4.8501**	5.4306	6.8681	7.8382	6.2467
P.Model 2	4.9422	**5.3624**	**6.6855**	**6.972**	**5.934**
MAPE	2015	2016	2017	2018	Average
P.Model 1	**4.9.2413**	10.2732	13.0338	14.7547	11.8257
P.Model 2	9.2597	**10.1715**	**11.9645**	**8**	**9.848925**
MSLE	2015	2016	2017	2018	Average
P.Model 1	**0.0173**	0.0199	0.03011	0.0328	0.0250
P.Model 2	0.0178	**0.0197**	**0.0298**	**0.0119**	**0.0198**

Bold font indicates best result obtained for analysis.

of the error trend, the decrement process stops until the layer number reaches three. After that, the error trend remains constant or increases when the number of layers increases to more than three. As shown in Fig. 6, the trends of MAE, MAPE, and MSLE fluctuate in the first and second layers. In layer three, the errors were significantly reduced. Although 2015 and 2018 showed slight changes, 2016 and 2017 saw a sharp drop in MAE. In MAPE, when the layers reached three, the error reached the lowest value in 2015, 2016, and 2017. Also, when the network was tested with three CNN layers, the changes stopped. Therefore, it can be concluded that when the model is simulated with three layers of CNN, the errors are at their lowest value.

Also, Table II shows the comparative information equivalent to Fig. 6 of the test data. As the results are shown in Table II the MAE, MAPE, RMSE, and MSLE of the proposed method on the test data with three layers were 4.2938, 10.2732, 5.4306, and 0.0199, respectively. Although RMSE reached the lowest error when the number of layers is four, the other three evaluating methods confirm that three layers show efficient performance. Meanwhile, after the number of layers, increased to 4 and 5, MAE, MAPE, and MSLE increased by approximately 17%, 14.43%, and 21.5%, respectively. Also, the complexity of the network and the number of parameters increase, which leads to the aggravation of network errors.

3) Comparing the Both Proposed Models: In Table III, two models are compared based on evaluation metrics. The first

TABLE IV
RMSE OF THE PROPOSED MODELS COMPARED WITH THE OTHER COMPETING MODELS

RMSE	CNN-LSTM [38]	3D-CNN [50]	ConvLSTM [49]	DeepYield [4]	Proposed Model 1	**Proposed Model 2**
2015	6.9164	6.8968	6.4058	6.1203	**4.8501**	4.9422
2016	8.2889	8.6095	7.8252	7.4404	5.4306	**5.2444**
2017	7.6199	7.3485	7.1507	7.2770	6.8681	**6.5804**
2018	8.5763	8.3434	8.0523	7.8509	7.8382	**6.972**
Average	7.8503	7.7994	7.3585	7.1721	6.24675	**5.934**

Bold font indicates best result obtained for analysis.

TABLE V
MAE OF THE PROPOSED MODELS COMPARED WITH THE OTHER COMPETING MODELS

MAE	CNN-LSTM [38]	3D-CNN [50]	ConvLSTM [49]	DeepYield [4]	Proposed Model 1	**Proposed Model 2**
2015	5.9812	5.546	5.1310	4.9291	**3.8335**	3.8348
2016	7.8632	7.2546	6.5836	6.3126	4.2938	**4.2883**
2017	6.9824	5.925	5.8663	6.0812	5.5907	**5.3811**
2018	7.9750	6.698	6.6357	6.6903	6.6229	**4.0715**
Average	7.2004	6.3559	6.0541	6.0033	5.0852	**4.3939**

Bold font indicates best result obtained for analysis.

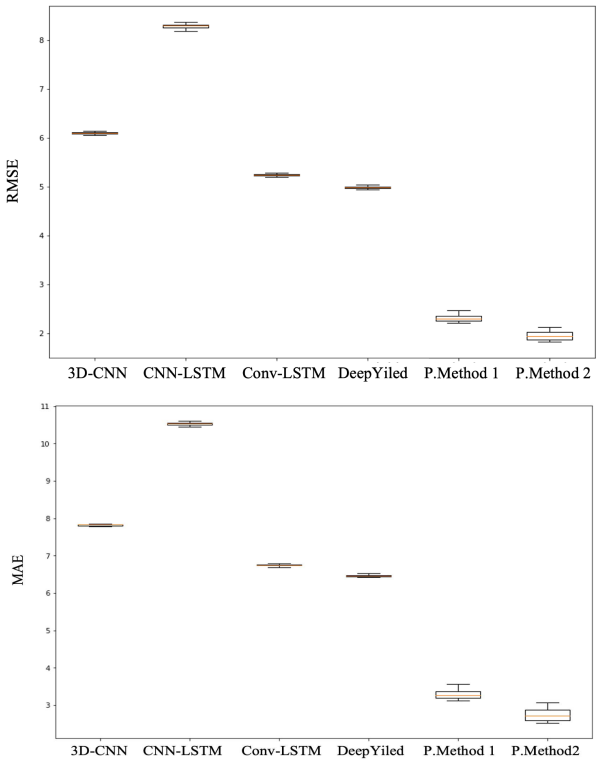


Fig. 7. Box-plots of MAE and RMSE values on the training dataset.

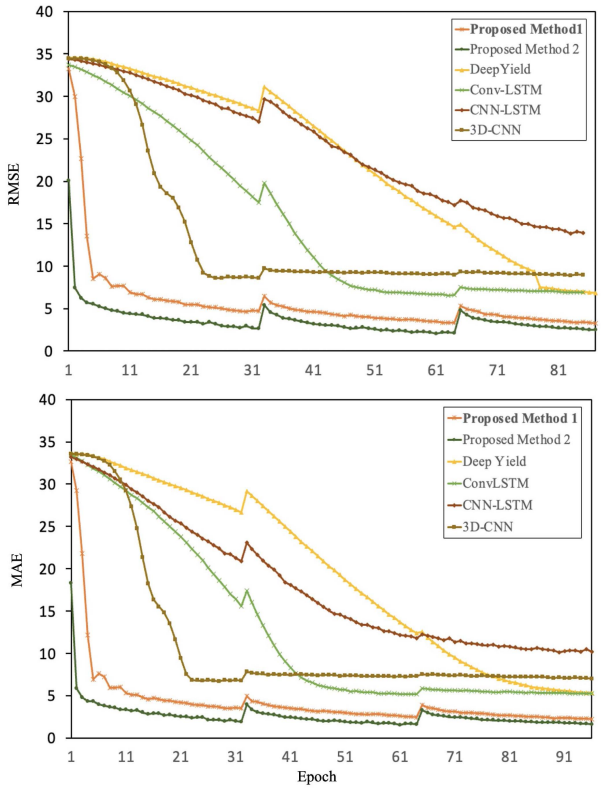


Fig. 8. Loss versus the number of training epochs for training sets.

proposed model, which used 2D-CNN can merely extract spatial features, while the 3D-CNN-based model not only can extract spatial features, but also spectral data can be extracted. Actually, those models using 3D-CNN are superior to 2-D based. Therefore, spectral-spatial-feature-extractors perform much better than spatial extractor-based models. As it is shown, Although in 2015, the first model showed better performance in all evaluation metrics, in the years 2016, 2017, and 2018 the method used 3D-CNN, and ConvLSTM had the lowest error in compared to 2-D based one. Overall, the 3D-based model performed better MAE, RMASE, MAPE, and MSLE, and the error reduced by approximately 2%.

4) Crop Yield Perdition Comparing Approach: The results from proposed methodologies compared with DeepYield [4], ConvLSTM [49], 3DCNN [50], and CNN-LSTM [38]. Simulations have been repeated in the same condition for all competitive methods to have a fair comparison. The simulations have been repeated under the same conditions for all competing methods to make a fair comparison. The simulations were performed with 1834 counties as input data. Data from the years 2003 to 2104 (12 years) were considered as training sets and also unseen data packages from 2015 to 2018 were considered as test sets. Results are done in 5 runs.

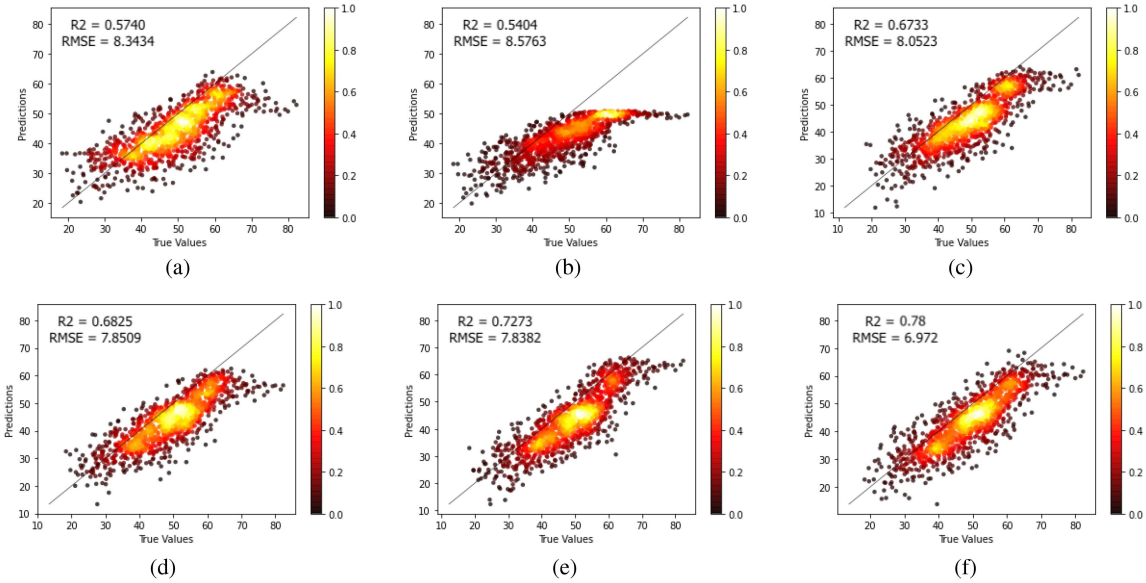


Fig. 9. Scatter plots of the predicted versus true yield values for different methods. (a) 3D-CNN [50]. (b) CNN-LSTM [38]. (c) ConvLSTM [49]. (d) DeepYield [4]. (e) Proposed Method 1. (f) Proposed Method 2.

RMSE and MAE prediction performances of crop yield forecasting at the county level are illustrated in Tables IV and V, respectively. These tables show the comparison between the RMSE's and MAE's models. For better understanding, the average values of those models have been shown in the last row. By introducing both proposed models, the errors decreased rapidly. On average, the second proposed method improved approximately 14.40% and 16.77% of prediction in terms of RMSE and MAE compared to DeepYield [4], respectively. In Table IV, it is observed that the proposed model is improved from 16.5% to about 25% in compared to ConvLSTM, 3D-CNN, and CNN-LSTM models. The same behavior is repeated in Table V, and it outperformed better than ConvLSTM, 3D-CNN, and CNN-LSTM, respectively by 17.47%, 26.44%, and 30.6%. In the literature, the second proposed method notably performs better than the stated four deep models.

In Fig. 7, the difference between the proposed methods and the other methods is shown in a Box-Plot that shows the loss error in the training set in terms of RMSE and MAE. It is done over 32 itineration in every training year, which is learned on the five runs for each network. According to RMSE and MAE, it is clearly shown that the proposed methods have he least error loss. MAE results show that the worst variance belongs to 3D-CNN, ConvLSTM, and DeepYiled. However, the second proposed model shows the lowest error. Also, from the RMSE plot, the second proposed model showed amazing performance compared to DeepYield, ConvLSTM, 3D-CNN, and CNN-LSTM.

The convening step of the six deep neural networks is shown in Fig. 8 during the training phase. The RMSE and MAE figure losses show a sharp decreasing in the early stages of the starting training for proposed methods, and will relatively stay steady on the optimal values for the next iterations. DeepYield represented a prolonged reduction and decreased the loss in the next three step of year iteration (Each year has 32 iterations). Although the other models like 3D-CNN, and ConvLSTM converged faster than DeepYield for the first year of the training, DeepYield reaches below them gradually after passing a long time. Last but not least, Fig. 8 also represents that the second proposed model had a faster convergence in comparison with the first proposed model.

Scatter plots are another ways to show which model performs better in an identical condition, showed In Fig. 9. In fact, it shows forecasted against observed crop yield for the prediction models. It demonstrated that proposed architectures have achieved the highest correlation coefficient score, and got the lowest root mean squared error (RMSE). By looking through plots, the other models have relatively close results. While, DeepYield had the highest value, which reached 0.68, comparing the rest models including CovLSTM, 3DCNN, and CNN-LSTM. The simplest deep learning models have relatively the lowest efficiency. By comparing the RMSE and R2 of the models, both approach methods had the best performance against the other models. CNN-LSTM had the worst performance, which dropped dramatically to 0.54. While ConvLSTM, DeepYield, and 3D-CNN achieved better performance compared to CNN-LSTM. As it is expected, the correlation coefficient improved significantly by using 3-D feature extractors, and ConvLSTM, compared with the first proposed model, which considerably increased from 0.71 to 0.78.

In Fig. 10, prediction sharing error map of different models has been shown from 2016 to 2018. The second proposed model overcame remarkably the other models. The second proposed method has the most minor prediction error, which means the counties are in light yellow and blue. In contrast, the deviation in the CNN-LSTM model is too high which predicts an error of more than 15 bushels per acre. DeepYiled, 3D-CNN, and ConvLSTM are predicting safer values in comparison with CNN-LSTM. However, the most reliable model belongs to the proposed models. Most counties have predicted the error between -5 and $+5$ bushels per acre.

5) Discussion: Also, it should be noted that the proposed methods have several advantages over the other models. The first advantage is both proposed methods try to have stable outputs

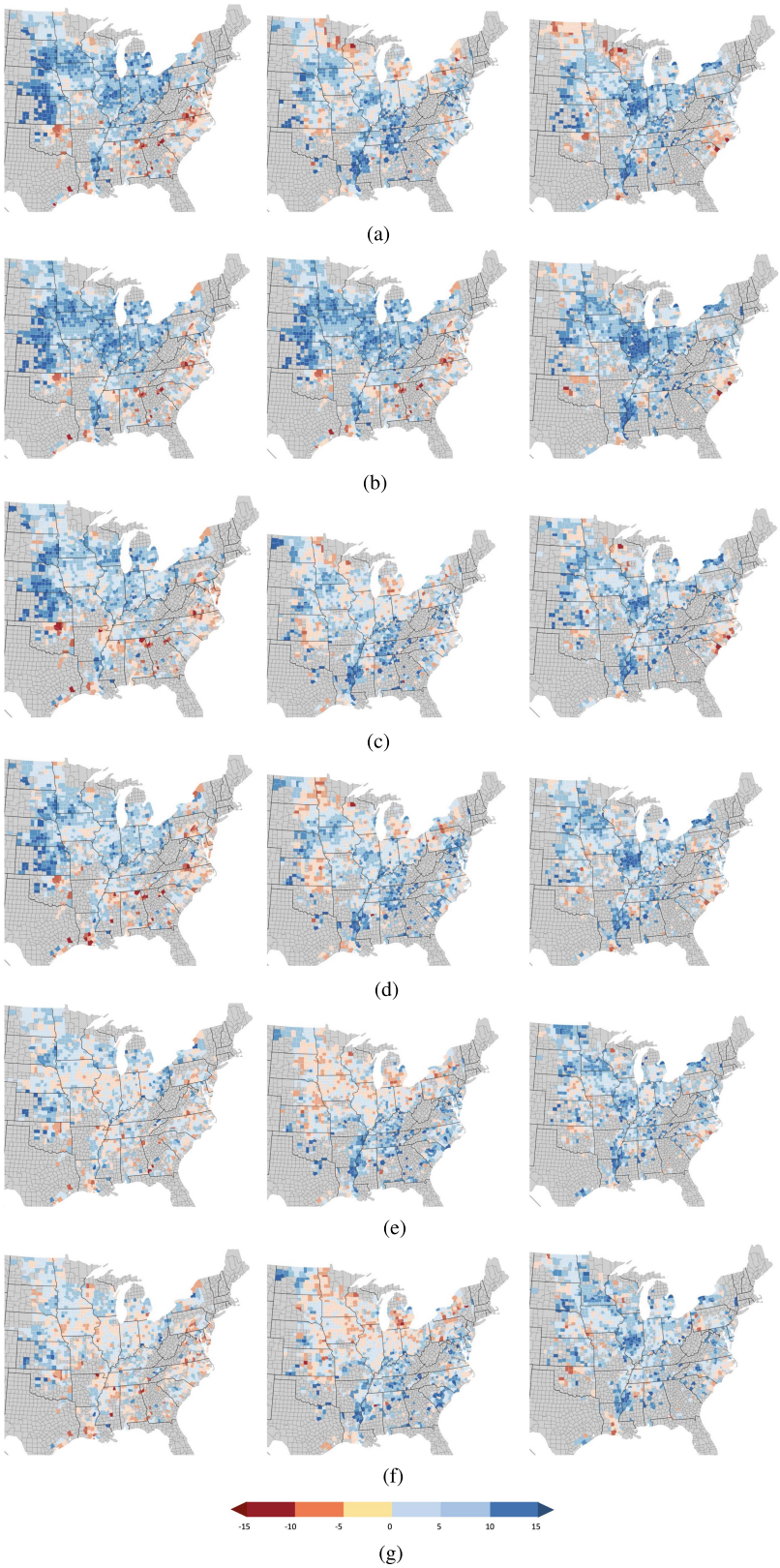


Fig. 10. Maps of forecasting error distribution of different approaches compared with proposed method. (a)–(c) 3D-CNN [50] Method predicting error distribution from 2016 to 2018, from left to right. (d)–(f) CNN-LSTM [38] Method predicting error distribution from 2016 to 2018, from left to right. (g)–(i) ConvLSTM [49] Method predicting error distribution from 2016 to 2018, from left to right. (j)–(l) DeepYield [4] Method predicting error distribution from 2016 to 2018, from left to right. (m)–(o) Proposed Method predicting error distribution from 2016 to 2018, from left to right. (p)–(r) Proposed Method 2 predicting error distribution from 2016 to 2018, from left to right.

of each convolution layer in part (a) of the architecture. They are typically clarified by the truth that other models diminish the size of the input tensor utilizing convolution piece with a stride of 3×3 from the beginning of the model, then that employments a max pooling layer. Those operations lessen the size outcomes of the middle layers of the model significantly. As a result, the number of multiply and-mass operations; in any case, this exceptional diminish causes a loss of valuable spatial data that influences the performance of the model.

The second is the use of the attention mechanism, which is equipped with both LSTM and skip connection in CNN architecture to increase the accuracy. In detail, CNN with the help of skip connections has been defined, which made a stable model. Then, the approach method used LSTM which were enhanced with an attention mechanism to increase the ability of the model. The attention mechanism is the concept of freeing the deep neural networks from a certain length of internal representation. This is often done by keeping the intermediate outcomes from each step of the input sequence and training the network to learn and pay selective attention to these inputs and relate them to performing predation.

The third which is specifically used explicitly for the second proposed method is the usage of 3D-CNN for extracting spectral, spatial, and temporal features. As stated before, some remote sensing data are captured in the rising time in the different phases which 3D-CNN can be efficiently applied to extract spectral, spatial, and temporal features. Therefore, the results can be more reliable.

Last but not least, the combination of LSTM and convolution highly decreases the network's complexity. As a result, the network can be deep enough to receive desirable results. Also, since remote sensing data are high-dimensional information, ConvLSTM has been used to handle those high-dimensional data.

However, both proposed models suffered from a major limitation in the input data. Due to hardware and software limitations, we have not managed to generate full-size images as input data. That is true the model has been trained by a high-performance computing platform (HPC). Since, the university HPC resource only works on local storage, we were having trouble generating the input data. Because, input data can only be generated by connecting to the cloud storage. Therefore, due to lake of Cloud Storage and RAM, we were able only to create Histogram input data rather than generating full-size images. By doing so, the change in input data will lead to differences in the results from reference papers as the spatial specifications of input data are protected via utilizing the full image size as input. Accordingly, the spatial correlation pixels are guarded, which raises the efficiency of convolutional filters.

IV. Conclusion

Two novel methods have been proposed by this article, which are the combination of the 2D-CNN and LSTM attention as the first model, and the usage of 3D-CNN and ConvLSTM instated of 2D-CNN and single LSTM as the second model for county-level crop yield prediction. As the first step, multi-2D-CNNs are used with help of the skip connection to extract features. After that, the outputs of the previous step are used for attention LSTM and multicascaded CNN parallelly. Attention mechanism was used to focus on main features and disqualify the unimportant ones. Finally, a single dense layer has been applied to make predictions. Although the second model has the same architecture as the first model, 3D-CNN and ConvLSTM have been used instated of the 2D-CNN and LSTM. 3D-CNNs can extract both spectral and spatial simultaneously, and ConvLSTM is bale to temporal and spatial-spectral together. By using these features, it distinguishes the second model from the first model. Significant improvements have been seen compared to the most recent models, which are used for crop yield prediction. Remote sensed data were used such as MODIS Land Cover, Temperature, and MODIS Surface Reflectance. The data have been extracted over 1840 counties from 2003 to 2018. The proposed models have been evaluated with different hyperparameters to achieve the best performance and after that different layers of the CNN have been used. It is found that a model with three layers of CNN is the most effective forecasting system. Both proposed models have been evaluated with relevant works, including DeepYeild, ConvLSTM, 3D-CNN, and CNN-LSTM, which are tested from 2016 to 2018 in the identical conditions to have a fair comparison. It is finally discovered that the second proposed model achieved the highest score in comparison with the other methods. The results of this study can be used in different sectors of the agencies in the US and agriculturalists. There are some approaches for future works, which are listed below. Scientists and researchers can add climatic data, which are important for growing the products as input to have better accuracy for forecastings, such as watering, genotype, rainfall information, and environmental data. Moreover, optimal use of satellite data with higher resolution will be strongly recommended. Last but not least, due to the sparsity of input data the use of Spiking neural networks is also recommended.

References

[1] N. A. S. S. the Statistical Methods Branch, Statistics Division, *The Yield Forecasting Program of NASS. Department of Agriculture*, Washington, D.C., USA, Apr. 2012.

[2] G. Hoogenboom, J. W. White, and C. D. Messina, "From genome to crop: Integration through simulation modeling," *Field Crops Res.*, vol. 90, no. 1, pp. 145–163, 2004.

[3] D. K. Bolton and M. A. Friedl, "Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics," *Agricultural Forest Meteorol.*, vol. 173, pp. 74–84, 2013.

[4] K. Gavahi, P. Abbaszadeh, and H. Moradkhani, "DeepYield: A combined convolutional neural network with long short-term memory for crop yield forecasting," *Expert Syst. Appl.*, vol. 184, 2021, Art. no. 115511.

[5] X. Xu et al., "Design of an integrated climatic assessment indicator (ICAI) for wheat production: A case study in Jiangsu Province, China," *Ecological Indicators*, vol. 101, pp. 943–953, 2019.

[6] H. Sundmaeker, C. Verdouw, J. Wolfert, and L. Perez Freire, *Internet of Food and Farm 2020, Ser. River Publishers Series in Communications*. Denmark: River Publishers, 2016, pp. 129–150.

[7] P. Filippi et al., "An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning," *Precis. Agriculture*, vol. 20, no. 5, pp. 1015–1029, 2019.

[8] A. Sharifi and J. Amini, "Forest biomass estimation using synthetic aperture radar polarimetric features," *J. Appl. Remote Sens.*, vol. 9, no. 1, 2015, Art. no. 097695, doi: 10.1117/1.jrs.9.097695.

[9] D. M. Johnson, "An assessment of pre and within season remotely sensed variables for forecasting corn and soybean yields in the United States," *Remote Sens. Environ.*, vol. 141, pp. 116–128, 2014.
[10] G. Nguyen et al., "Machine learning and deep learning frameworks and libraries for large-scale data mining: A survey," *Artif. Intell. Rev.*, vol. 52, no. 1, pp. 77–124, 2019.
[11] S. Bargoti and J. P. Underwood, "Image segmentation for fruit detection and yield estimation in apple orchards," *J. F. Robot.*, vol. 34, no. 6, pp. 1039–1060, 2017, doi: 10.1002/rob.21699.
[12] H. Habaragamuwa, Y. Ogawa, T. Suzuki, T. Shiigi, M. Ono, and N. Kondo, "Detecting greenhouse strawberries (mature and immature), using deep convolutional neural network," *Eng. Agriculture Environ. Food*, vol. 11, no. 3, pp. 127–138, 2018.
[13] H. Kang and C. Chen, "Fast implementation of real-time fruit detection in apple orchards using deep learning," *Comput. Electron. Agriculture*, vol. 168, 2020, Art. no. 105108.
[14] S. Nosratabadi, K. Szell, B. Beszedes, I. Felde, S. Ardabili, and A. Mosavi, "Comparative analysis of ANN-ICA and ANN-GWO for crop yield prediction," EasyChair Preprint no. 2759, EasyChair, 2020.
[15] Y. Li, Z. Zhu, D. Kong, H. Han, and Y. Zhao, "EA-LSTM: Evolutionary attention-based LSTM for time series prediction," *Knowledge-Based Syst.*, vol. 181, 2019, doi: 10.1016/j.knosys.2019.05.028.
[16] K. Alibabaei, P. D. Gaspar, and T. M. Lima, "Crop yield estimation using deep learning based on climate Big Data and irrigation scheduling," *Energies*, vol. 14, no. 11, 2021, Art. no. 3004.
[17] S. Khaki, H. Pham, and L. Wang, "Simultaneous corn and soybean yield prediction from remote sensing data using deep transfer learning," *Sci. Rep.*, vol. 11, no. 1, 2021, Art. no. 11132.
[18] L. Gong, M. Yu, S. Jiang, V. Cutsuridis, and S. Pearson, "Deep learning based prediction on greenhouse crop yield combined TCN and RNN," *Sensors*, vol. 21, no. 13, 2021, Art. no. 4537.
[19] S. Ju et al., "Optimal county-level crop yield prediction using modis-based variables and weather data: A comparative study on machine learning models," *Agricultural Forest Meteorol.*, vol. 307, 2021, Art. no. 108530.
[20] A. Gholizadeh, M. Khodadadi, and A. Sharifi-Zagheh, "Modeling the final fruit yield of coriander (Coriandrum sativum l.) using multiple linear regression and artificial neural network models," *Arch. Agronomy Soil Sci.*, vol. 68, pp. 1398–1412, 2022.
[21] S. D. Alwis, Y. Zhang, M. H. Na, and G. Li, "Duo attention with deep learning on tomato yield prediction and factor interpretation," in *Proc. Pacific Rim Int. Conf. Artif. Intell.*, 2019, pp. 704–715.
[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016, Dec. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
[23] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2261–2269.
[24] Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral–spatial residual network for hyperspectral image classification: A 3-D deep learning framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 847–858, Feb. 2018.
[25] W. Wang, S. Dou, Z. Jiang, and L. Sun, "A fast dense spectral–spatial convolution network framework for hyperspectral images classification," *Remote Sens.*, vol. 10, no. 7, 2018, Art. no. 1068.
[26] A. Vaswani et al., "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 2017, Dec. 2017.
[27] M. Zhu, L. Jiao, F. Liu, S. Yang, and J. Wang, "Residual spectral–spatial attention network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 449–462, Jan. 2021.
[28] W. Ma, Q. Yang, Y. Wu, W. Zhao, and X. Zhang, "Double-branch multi-attention mechanism network for hyperspectral image classification," *Remote Sens.*, vol. 11, no. 11, 2019, Art. no. 1307.
[29] H. Sun, X. Zheng, X. Lu, and S. Wu, "Spectral–spatial attention network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3232–3245, May 2020.
[30] R. Li, S. Zheng, C. Duan, Y. Yang, and X. Wang, "Classification of hyperspectral image based on double-branch dual-attention mechanism network," *Remote Sens.*, vol. 12, no. 3, 2020, Art. no. 582.
[31] J. You, X. Li, M. Low, D. Lobell, and S. Ermon, "Deep gaussian process for crop yield prediction based on remote sensing data," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 4559–4565.
[32] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn.*, vol. 1, 2015, pp. 448–456.
[33] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
[34] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, 807–814.
[35] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Comput.*, vol. 14, no. 8, pp. 1771–1800, Aug. 2002.
[36] B. Alhnaity, S. Pearson, G. Leontidis, and S. Kollias, "Using deep learning to predict plant growth and yield in greenhouse environments," *Acta Horticulturae*, vol. 1296, pp. 425–431, 2020, doi: 10.17660/ActaHortic.2020.1296.55.
[37] L. Nguyen et al., "Spatial-temporal multi-task learning for within-field cotton yield prediction," in *Proc. Adv. Knowl. Discov. Data Mining 23rd Pacific-Asia Conf.*, 2019, pp. 343–354.
[38] J. Sun, L. Di, Z. Sun, Y. Shen, and Z. Lai, "County-level soybean yield prediction using deep CNN-LSTM model," *Sensors*, vol. 19, no. 20, 2019, Art. no. 4363.
[39] J. Shook, T. Gangopadhyay, L. Wu, B. Ganapathysubramanian, S. Sarkar, and A. K. Singh, "Crop yield prediction integrating genotype and weather variables using deep learning," *PLOS ONE*, vol. 16, no. 6, 2021, Art. no. e0252402.
[40] A. Sharifi, J. Amini, J. T. Sri Sumantyo, and R. Tateishi, "Speckle reduction of PolSAR images in forest regions using fast ICA algorithm," *J. Indian Soc. Remote Sens.*, vol. 43, no. 2, pp. 339–346, 2015, doi: 10.1007/s12524-014-0423-3.
[41] R. A. Schwalbert, T. Amado, G. Corassa, L. P. Pott, P. Prasad, and I. A. Ciampitti, "Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil," *Agricultural Forest Meteorol.*, vol. 284, 2020, Art. no. 107886.
[42] B. Alhnaity, S. Kollias, G. Leontidis, S. Jiang, B. Schamp, and S. Pearson, "An autoencoder wavelet based deep neural network with attention mechanism for multi-step prediction of plant growth," *Inf. Sci.*, vol. 560, pp. 35–50, 2021.
[43] D. Bahdanau, K. H. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2015.
[44] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2013.
[45] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W. C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. 28th Int. Conf. Neural Inf. Process. Syst.*, 2015, pp. 802–810.
[46] S. W. Lee and H. Y. Kim, "Stock market forecasting with super-high dimensional time-series data using convlstm, trend sampling, and specialized data augmentation," *Expert Syst. Appl.*, vol. 161, 2020, Art. no. 113704.
[47] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google earth engine: Planetary-scale geospatial analysis for everyone," *Remote Sens. Environ.*, vol. 202, pp. 18–27, 2017.
[48] "USDA national agricultural statistics service," *Choice Rev. Online*, vol. 51, no. 7, pp. 51-3802–51-3802, 2014, doi: 10.5860/choice.51-3802.
[49] F. Guo, J. Yang, H. Li, G. Li, and Z. Zhang, "A convLSTM conjunction model for groundwater level forecasting in a karst aquifer considering connectivity characteristics," *Water*, vol. 13, no. 19, 2021, Art. no. 2759.
[50] P. Nevavuori, N. Narra, P. Linna, and T. Lipping, "Crop yield prediction using multitemporal UAV data and spatio-temporal deep learning models," *Remote Sens.*, vol. 12, no. 23, 2020, Art. no. 4000.

Seyed Mahdi Mirhoseini Nejad received the B.Sc. degree in electrical & electronic engineering from the Islamic Azad University of Kerman, Kerman, Iran, in 2010, and the M.Sc. degree in electrical and communication engineering from Tehran University of Technology, Tehran, Iran, in 2013. He is currently working toward the Ph.D. degree in electrical and communication engineering from Shahid Bahonar University of Kerman, Kerman, Iran.

He is currently researching deep learning methods for crop yield predictions and classification approaches with hyper and multispectral images. His research interests include image processing, machine learning, deep learning, and its applications on hyper multispectral image classification and predictions.

Dariush Abbasi-Moghadam received the B.S. degree in electrical engineering from Shahid Bahonar University, Kerman, Iran, in 1998, and the M.S. and Ph.D. degrees in Iran University of Science and Technology, Tehran, Iran, in 2001 and 2011, respectively, both in electrical engineering.

He was primary with the Advanced Electronic Research Center, Iran from 2001 to 2003 and worked on the design and analysis of satellite communication systems. In September 2004, he joined Iranian Telecommunications Company, Tehran, as a Research Engineer. He is currently with Department of Electrical Engineering, Shahid Bahonar University of Kerman (SBUK), Kerman, Iran, as an Associate Professor. His research interests are in the area of wireless communications, satellite communication systems, remote sensing, and signal processing.

Alireza Sharifi was born in Tehran, Iran, in 1981. He received the M.Sc. and Ph.D. degrees in remote sensing engineering from the University of Tehran, Tehran, Iran, in 2008 and 2015, respectively.

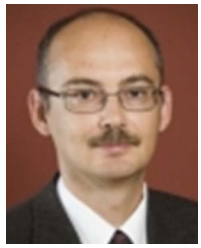
He is currently an Assistant Professor of remote sensing with the Faculty of Civil Engineering from the Shahid Rajaee Teacher Training University, Tehran, Iran. In particular, he is involved in GEOAI program for Food Security and Environmental Monitoring.

Nizom Farmonov received the B.S. degree in land management and land cadastre from the Department of Land use and Land Cadastre, Tashkent Institute of Irrigation and Agricultural Mechanization Engineers, Tashkent, Uzbekistan, in 2018, and the M.S. degree in geodesy and geoinformations from the Department of Geodesy and Geoinformatics, Tashkent Institute of Irrigation and Agricultural Mechanization Engineers, Tashkent, Uzbekistan, in 2015. He is currently working toward the Ph.D. degree in crop type classification and yield prediction using hyperspectral images with the Department of Geoinformatics, Physical and Environmental Geography, University of Szeged, Szeged, Hungary.

His research interests include remote sensing, GIS and crop yield forecasting with machine learning using high spatial-temporal multisource satellite data.

Khilola Amankulova received the B.S. degree in land management and land cadastre from the Department of Land use and Land Cadastre, in 2018, and the M.S. degree in geodesy and geoinformations from the Department of Geodesy and Geoinformatics, Tashkent Institute of Irrigation and Agricultural Mechanization Engineers, Tashkent, Uzbekistan. She is currently working toward the Ph.D. degree with the Department of Geoinformatics, Physical and Environmental Geography, University of Szeged, Szeged, Hungary.

Her research focuses on remote sensing, precision agriculture, and geoinformatics.



Mucsi László received the Ph.D. degree in earth sciences from University of Szeged, Szeged, Hungary, in 1997.

He is currently a Professor (Associate) with the Department of Geoinformatics, Physical and Environmental Geography, University of Szeged. His current research activities are remote sensing, time series analysis, image classification, artificial intelligence.