Expert Systems With Applications 184 (2021) 115511

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

DeepYield: A combined convolutional neural network with long short-term memory for crop yield forecasting

Keyhan Gavahi [*,1], Peyman Abbaszadeh [2], Hamid Moradkhani [3]

Center for Complex Hydrosystems Research, Department of Civil, Construction and Environmental Engineering, University of Alabama, Tuscaloosa, AL, USA

ARTICLE INFO

Keywords:
Crop yield forecasting
Deep learning
Remote sensing
Convolutional neural networks (CNN)
Convolutional long short-term memory (ConvLSTM)

ABSTRACT

Crop yield forecasting is of great importance to crop market planning, crop insurance, harvest management, and optimal nutrient management. Commonly used approaches for crop prediction include but are not limited to conducting extensive manual surveys or using data from remote sensing. Considering the increasing amount of data provided by remote sensing imagery, this approach is becoming increasingly important for the task of crop yield forecasting and there is a need for more sophisticated approaches to extract the inherent spatiotemporal patterns of these data. Although considerable progress has been made in this field by using Deep Learning (DL) methods such as Convolutional Neural Networks (CNN), no study before has investigated the use of Convolutional Long Short-Term Memory (ConvLSTM) for crop yield forecasting. Here, we propose DeepYield, a combined structure, that integrates the ConvLSTM layers with the 3-Dimensional CNN (3DCNN) for more accurate and reliable spatiotemporal feature extraction. The models are trained by using county-based historical yield data and MODIS Land Surface Temperature (LST), Surface Reflectance (SR), and Land Cover (LC) data over 1836 primary soybean growing counites in the Contiguous United States (CONUS). The forecasting performance of the developed models is compared against the competing approaches including Decision Trees, CNN + GP, and CNN-LSTM and results indicate that DeepYield significantly outperforms these techniques and also performs better than both ConvLSTM and 3DCNN.

1. Introduction

Accurate and timely crop yield forecasting is of great importance for a variety of reasons. It allows societies to understand the future available food supply and helps the demand side to optimize the utilization of crop resources. From a management point of view, future yield estimation helps farmer plan better for the end-of-season by establishing risk management policies, insurance premiums, and evaluating the value of input costs (Johnson, 2014). It can also help better understanding the impacts of severe weather or changing the climatic conditions such as drought and hurricanes on crops (Ceglar et al., 2018; Gavahi, Abbaszadeh, Moradkhani, Zhan, & Hain, 2020; Liakos, Busato, Moshou, Pearson, & Bochtis, 2018).

The commonly used crop yield forecasting methods use manual surveys (United States Department of Agriculture. (2012), 2012), crop simulation models (Hoogenboom, White, & Messina, 2004), or remote sensing data (Gallego, Carfagna, & Baruth, 2010). Among these approaches, remote sensing can provide more affordable yield forecasting tools as several free and open-source remote sensing databases are available online (Bolton & Friedl, 2013; Mendes, Araújo, Dutta, & Heeren, 2019). A variety of pertinent information can be extracted through remote sensing data for yield forecasting. In particular, vegetation indices such as the Normalized Difference Vegetation Index (NDVI) (Lofton et al., 2012; Shrestha et al., 2016; Shrestha, Di, Eugene, Kang, & Bai, 2017), Green Leaf Area Index (GLAI) (Duchemin, Maisongrande, Boulet, & Benhadj, 2008), Enhanced Vegetation Index (EVI) (Xue & Su, 2017), Normalized Difference Water Index (NDWI) (Bolton & Friedl, 2013) have been widely utilized for crop yield forecasting.

The existence of disturbances, modeling errors, and various uncertainties in the real systems, makes the task of modeling a highly nonlinear phenomenon with spatiotemporal variability a daunting challenge (Stojanovic, He, & Zhang, 2020; Wei, Li, & Stojanovic, 2021;

* Corresponding author.
E-mail addresses: kgavahi@crimson.ua.edu (K. Gavahi), pabbaszadeh@ua.edu (P. Abbaszadeh), hmoradkhani@ua.edu (H. Moradkhani).
[1] https://orcid.org/0000-0002-1313-2286.
[2] https://orcid.org/0000-0002-4079-5149.
[3] https://orcid.org/0000-0002-2889-999X.

https://doi.org/10.1016/j.eswa.2021.115511
Received 8 February 2021; Received in revised form 20 May 2021; Accepted 28 June 2021
Available online 3 July 2021

Zhang, He, Stojanovic, Luan, & Liu, 2021). This is particularly true where missing data exists in the input data as well (Chen, Zhang, Stojanovic, Zhang, & Zhang, 2020). With the advances in Machine Learning (ML) techniques, considerable attention has been paid to their application to multispectral satellite images for crop yield forecasting. These include Decision Trees (DT) (Johnson, 2014; Kim & Lee, 2016), Support Vector Machine (SVM) (Kim & Lee, 2016; Kuwata & Shibasaki, 2015), Artificial Neural Network (ANN) (Kim & Lee, 2016), and Restricted Boltzmann Machine (RBM) (Kuwata & Shibasaki, 2015). Despite the widespread use of ML techniques, Deep Learning (DL) has recently been considered a breakthrough data mining platform in agricultural remote sensing studies and other applications (Sun, Di, Sun, Shen, & Lai, 2019). This includes Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), Stacked Sparse Autoencoder (SSAE), and Recurrent Neural Networks (RNN), which have been applied and outperformed conventional ML algorithms in many studies (Kim et al., 2019).

For the first time, You, Li, Low, Lobell, and Ermon (2017) used CNN and LSTM for crop yield prediction using Moderate-Resolution Imaging Spectroradiometer (MODIS) satellite images. They incorporated a Gaussian Processor (GP) into the last layer of the CNN to account for the spatiotemporal variability of the inputs. Their proposed method outperformed other competing techniques with a 30% Root Mean Squared Error (RMSE) improvement. Wang, Tran, Desai, Lobell, and Ermon (2018) continued the work and used deep transfer learning to extend the method to other regions in Argentina and Brazil. Their findings demonstrated that this approach can effectively learn features from raw data and improve performance compared to other ML techniques. Russello (2018) proposed a 3DCNN architecture for crop yield prediction and showed that it outperforms other traditional ML methods. In addition, the RNN and LSTM structures have been widely used to predict crop yield due to their ability to account for the temporal characteristics of the plant growth phenology. Jiang et al. (2018) applied a deep LSTM for county-based corn yield prediction using soil and weather data. The results in the state of Iowa showed the predictive power of the LSTM for yield estimation. Sun et al. (2019) developed a CNN-LSTM model for county-based soybean yield prediction and showed that their combined approach outperforms the single CNN or LSTM. Khaki, Wang, and Archontoulis (2020) proposed a CNN-RNN model for the similar study and showed that the combined method significantly outperforms traditional ML techniques.

To the best of our knowledge, no study has investigated the use of Convolutional LSTM (ConvLSTM) for crop yield forecasting. Additionally, in all those studies, the spatial dimension of the remote sensing images were discarded by turning them into histograms of pixel intensities or having their pixels averaged per county. Whereas including the spatial dimension can provide crucial information about crops' pertinent variables such as soil properties and elevation and thus increase the models' forecasting skills. Moreover, this study provides a more rigorous spatiotemporal feature extraction by combining the 3DCNN and ConvLSTM layers. The comparison with the individual architectures shows that the proposed combined approach (DeepYield) provides more accurate crop yield forecasts.

Previous studies show that DL algorithms are becoming the mainstream of forecasting and crop yield prediction (Khaki et al., 2020; Sun et al., 2019). Thus, the main motivation of this study is to introduce an integrated model that uses satellite imagery and produces yield prediction without a need to reduce the spatial dimensionality of the images or use the handcraft features. In this study, we propose a procedure to effectively combine the ConvLSTM and 3DCNN structures for county-based crop yield forecasting in the contiguous United States. The main contributions of this study are as follows: (1) instead of taking the average of pixel values or using histograms of pixel intensities, this study preserves the spatial characteristics of the input images by using the full image as input. As a result, the spatial correlation of adjacent pixels is preserved which enhances the performance of convolutional filters, (2) ConvLSTM is used for the first time for the crop yield forecasting accounting for the inherent spatiotemporal patterns of the input images, (3) a combined architecture, namely DeepYield, based on ConvLSTM and 3DCNN is introduced for more accurate and robust crop yield forecasting. The proposed approach uses an end-to-end learning scheme, to automatically process the input and provide a more accurate and reliable yield forecast.

The remainder of the paper is organized as follows: Section 2 describes the datasets and methods. Section 3 explains the proposed method, its implementation, and capabilities. The experimental results and analysis are presented in Section 4. Finally, a summary and concluding remarks are given in Section 5.

2. Materials and methods

2.1. Datasets

2.1.1. Yield data

County-based soybean statistics were collected from the USDA National Agricultural Statistical Services (NASS) Quick Stat tool available at https://www.nass.usda.gov/Quick_Stats/index.php. The yield data from 2003 to 2019 were used as ground truth labels for model training.

2.1.2. MODIS Surface reflectance

The MODIS/Terra Surface Reflectance (SR) product provides 7 bands of surface spectral reflectance at 500 m spatial resolution every 8 days (Vermote, 2015). Each pixel contains the best possible SR observation value selected from all the acquisitions within the 8-day window. The product is publicly available at https://lpdaac.usgs.gov/products/mod09a1v006/. Here, we used all 7 bands of version 6 of this product for soybean yield forecasting.

2.1.3. MODIS Land cover

The Terra and Aqua combined MODIS Land Cover type (LC) product provides yearly land cover types derived from six classification schemes (Sulla-Menashe & Friedl, 2019) at 500 m spatial resolution. The annual University of Maryland (UMD) classification (land cover type 2) scheme was used in this study to mask cropland areas. The dataset is publicly available at https://lpdaac.usgs.gov/products/mcd12q1v006/.

2.1.4. MODIS Land Surface temperature

The MODIS Version 6 Land Surface Temperature (LST) provides an average 8-day per-pixel daytime and nighttime surface temperature at 1 km spatial resolution (Wan, 2015). The temperature is collected by using 7 thermal infrared bands using the LST algorithm (Wan, 2006). This dataset has been widely used in multitude of studies (Abbaszadeh et al., 2021; Benali et al., 2012; Wang et al., 2021). In this study, both daytime and nighttime LSTs were used. The product is publicly available at https://lpdaac.usgs.gov/products/myd11a2v006/.

2.2. 3D convolutional networks

For the first time, Ji, Xu, Yang, and Yu (2013) proposed a 3DCNN structure for human action recognition applying 3D convolutions along both temporal and spatial dimensions. As opposed to directly inferring the temporal information from raw data, 3DCNNs have shown to be more suitable for spatiotemporal presentations (Elboushaki, Hannane, Afdel, & Koutti, 2020). This method has been successfully used in many applications such as gesture recognition (Elboushaki et al., 2020; Ji, Zhang, Xu, Shi, & Duan, 2018; Lin et al., 2016; Liu, Zhang, & Tian, 2016; Tran, Bourdev, Fergus, Torresani, & Paluri, 2015), learning 3D structures from LiDAR (Maturana & Scherer, 2015), and learning spatio-spectral patterns from hyperspectral images (Li, Zhang, & Shen, 2017). In general, the 3DCNNs are not as widely used as the 2DCNN since the temporal dimension is usually ignored in computer vision studies (Ji et al., 2018). However, remote sensing images often contain temporal information (feature) which can be more efficiently exploited

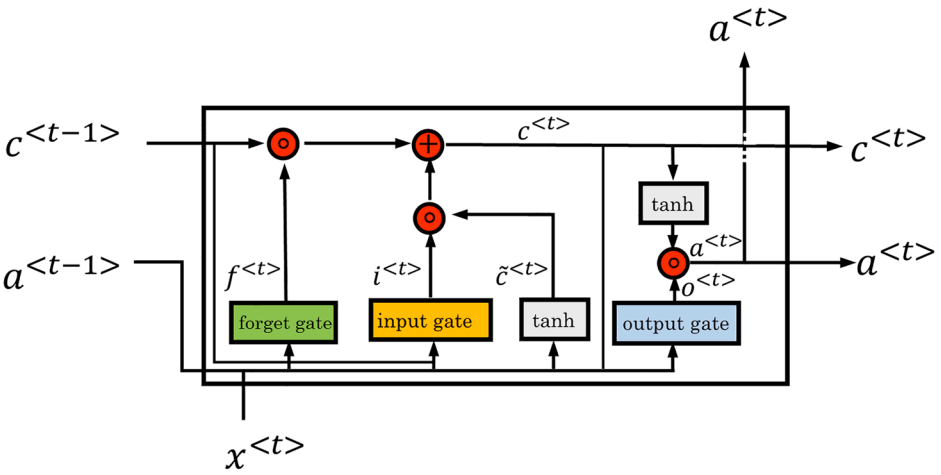

Fig. 1. The inner structure of an LSTM cell.

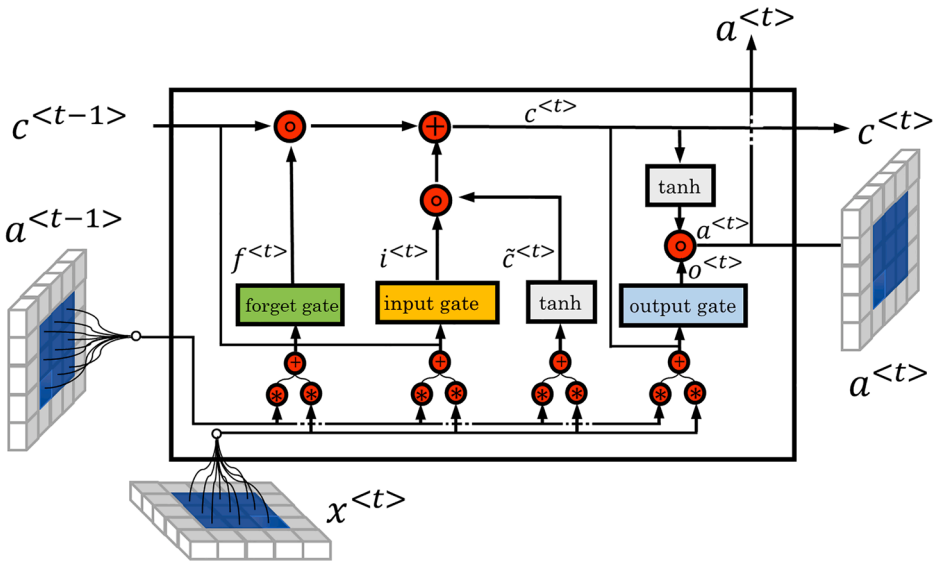

Fig. 2. The inner structure of a ConvLSTM cell.

using the 3DCNN. For instance, surface reflectance images captured during the growth season of soybean at different stages of the crop phenology contain crop growth information that is necessary for a more accurate end-of-season yield forecast. A 3DCNN structure can simultaneously extract both temporal and spatial information and potentially provide more accurate and robust feature extraction.

In 2DCNNs, extracting features from a local neighborhood on a specific feature map is performed by 2D convolutional filters. Then the bias is added and the result is passed through a sigmoid function. Equation (1) represents the value of a unit at position $y(x, y)$ in a specific layer and for a specific feature.

$$y(x,\ y) = \sigma\left(b + \sum_{n}\sum_{i=0}^{I}\sum_{j=0}^{J} w_{ijn}X(x+i)(y+j)_n\right) \tag{1}$$

where σ is the sigmoid function. X denotes the input 2D image. b represents bias. w_{ijn} is the kernel weight for the nth feature at position (i, j) of the filter, and I and J represent the kernel width and height, respectively.

While in the 2DCNNs, convolutions are applied on the 2D feature maps to extract features from spatial dimension only, in the 3DCNNs, the convolutional filters exploit features from both temporal and spatial dimensions. Formally, the value at position $y(x, y, t)$ in a specific layer and for the nth feature is given by Eq. (2):

$$y(x,\ y,\ t) = \sigma\left(b + \sum_{n}\sum_{r=0}^{R}\sum_{i=0}^{I}\sum_{j=0}^{J} w_{tijn}X(t+r)(x+i)(y+j)_n\right) \tag{2}$$

where R is the kernel size along the temporal dimension, w_{tijn} is the weight at position (r, i, j) of the 3D kernel with size (R, I, J) and for the nth feature. In practice, temporal images also consist of multispectral channels i.e. spatial, temporal, and spectral dimensions creating 4D tensors. Like the 2DCNNs where the relations among spectral bands are treated independently, in 3DCNNs each spectral band is treated separately (e.g., RGB bands in a 2DCNN). This allows for more rigorous information exploitation from various MODIS bands comparing to applying 2DCNN Fig. 1.

2.3. ConvLSTM networks

LSTM is a special form of RNN, which has been proven to be stable for capturing long-term patterns (Hochreiter & Schmidhuber, 1997). One important aspect of an LSTM network is its ability to maintain a cell state from the previous sequence of observations while eliminating irrelevant information. In the LSTM network this is performed by maintaining the information through three gates: input gate, forget gate,

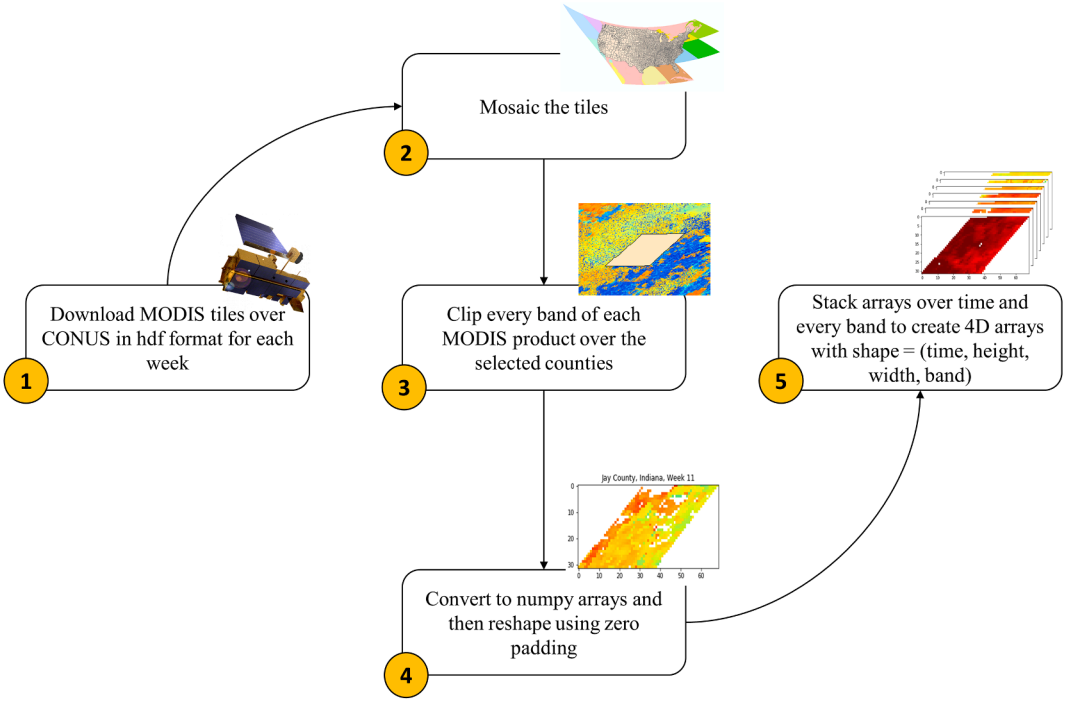

Fig. 3. MODIS data preprocessing workflow.

and output gate. Each of these gates returns a state variable, $i^{<t>}$, $f^{<t>}$, and $o^{<t>}$, respectively, along with the cell output, $a^{<t>}$. See Eqs. (1)–(6) in which (°) represents the elementwise product. Fig. 2, shows the inner structure of an LSTM cell (Gers & Schmidhuber, 2000).

$$i^{\langle t\rangle} = \sigma\left(W_{xi}x^{\langle t\rangle} + W_{ai}a^{\langle t-1\rangle} + W_{ci}\circ c^{\langle t-1\rangle} + b_i\right) \tag{3}$$

$$f^{\langle t\rangle} = \sigma\left(W_{xf}x^{\langle t\rangle} + W_{af}a^{\langle t-1\rangle} + W_{cf}\circ c^{\langle t-1\rangle} + b_f\right) \tag{4}$$

$$c^{\langle t\rangle} =\circ f^{\langle t\rangle}\circ c^{\langle t-1\rangle} + i^{\langle t\rangle}\circ tanh\left(W_{xc}x^{\langle t\rangle} + W_{ac}a^{\langle t-1\rangle} + b_c\right) \tag{5}$$

$$o^{\langle t\rangle} = \sigma\left(W_{xo}x^{\langle t\rangle} + W_{ao}a^{\langle t-1\rangle} + W_{co}\circ c^{\langle t\rangle} + b_o\right) \tag{6}$$

$$a^{\langle t\rangle} = o^{\langle t\rangle}\circ tanh\left(c^{\langle t\rangle}\right) \tag{7}$$

Shi et al. (2015) proposed a novel combination of convolutional filters and LSTM layers, called ConvLSTM, for precipitation nowcasting. In this method, the convolutional filters are applied to the input-to-state and state-to-state transitions of the LSTM. Equations (7)–(12) describe the architecture of the method. In these equations (*) denotes the convolution operator and σ represents the sigmoid activation function.

$$i^{\langle t\rangle} = \sigma\left(W_{xi}*x^{\langle t\rangle} + W_{ai}*a^{\langle t-1\rangle} + W_{ci}\circ c^{\langle t-1\rangle} + b_i\right) \tag{8}$$

$$f^{\langle t\rangle} = \sigma\left(W_{xf}*x^{\langle t\rangle} + W_{af}*a^{\langle t-1\rangle} + W_{cf}\circ c^{\langle t-1\rangle} + b_f\right) \tag{9}$$

$$c^{\langle t\rangle} = f^{\langle t\rangle}\circ c^{\langle t-1\rangle} + i^{\langle t\rangle}\circ tanh\left(W_{xc}*x^{\langle t\rangle} + W_{ac}*a^{\langle t-1\rangle} + b_c\right) \tag{10}$$

$$o^{\langle t\rangle} = \sigma\left(W_{xo}*x^{\langle t\rangle} + W_{ao}*a^{\langle t-1\rangle} + W_{co}\circ c^{\langle t\rangle} + b_o\right) \tag{11}$$

$$a^{\langle t\rangle} = o^{\langle t\rangle}\circ tanh\left(c^{\langle t\rangle}\right) \tag{12}$$

ConvLSTM is known to be well suited for capturing the inherent spatiotemporal patterns of large-scale datasets (Lee & Kim, 2020). Like the traditional CNN networks, the output dimension of a ConvLSTM layer is specified by the number of filters used in the network. However, in the ConvLSTM structure, eight filters are required for each desired output. It is important to note that applying the convolutional filters to the LSTM significantly reduces the number of model parameters compared to a single LSTM structure and thus allows for training even deeper models (Elboushaki et al., 2020; Petersen, Rodrigues, & Pereira, 2019).

3. Proposed approach

3.1. Data preprocessing

The datasets used in this study include MODIS LST (2 bands), SR (7 bands), and Land Use Land Cover (1 band). The latter was used to mask the cropland areas over each county. 14 tiles of the MODIS satellite cover the CONUS, which were downloaded using an Application Programming Interface (API) developed by the authors. The tiles were then mosaiced into a singular raster image that covers the extent of the CONUS. The mosaiced raster was later clipped over each county and images for the selected periods were concatenated creating 3D tensors. All the datasets used in this study have a temporal resolution of 8 days (per image) from January 2003 to December 2019. However, the spatial resolution of MODIS SR and LC is 500 m which is different from the MODIS LST dataset that has a 1 km spatial resolution. Hence the 500 m images were upscaled to 1 km resolution using a linear interpolation method. Finally, bands of each product were concatenated to the 3D tensors created before to produce 4D tensors with the dimension of Time × Height × Width × band. Zero paddings, a process for expanding the size of the input images by adding rows and columns of zero values, were used to make image sizes identical before feeding to the CNN. The detailed preprocessing pipeline of the MODIS images is described in Fig. 3. Owing to the high spatial resolution of the MODIS data and the number of counties, the preprocessing workflow operation creates a tremendous computational burden. To overcome this challenge, the preprocessing workflow was parallelized over the University of Alabama High Performance Computing (UAHPC) server to speed up the process.

3.2. Networks topology

Fig. 3 shows the 3DCNN and ConvLSTM architectures used in this

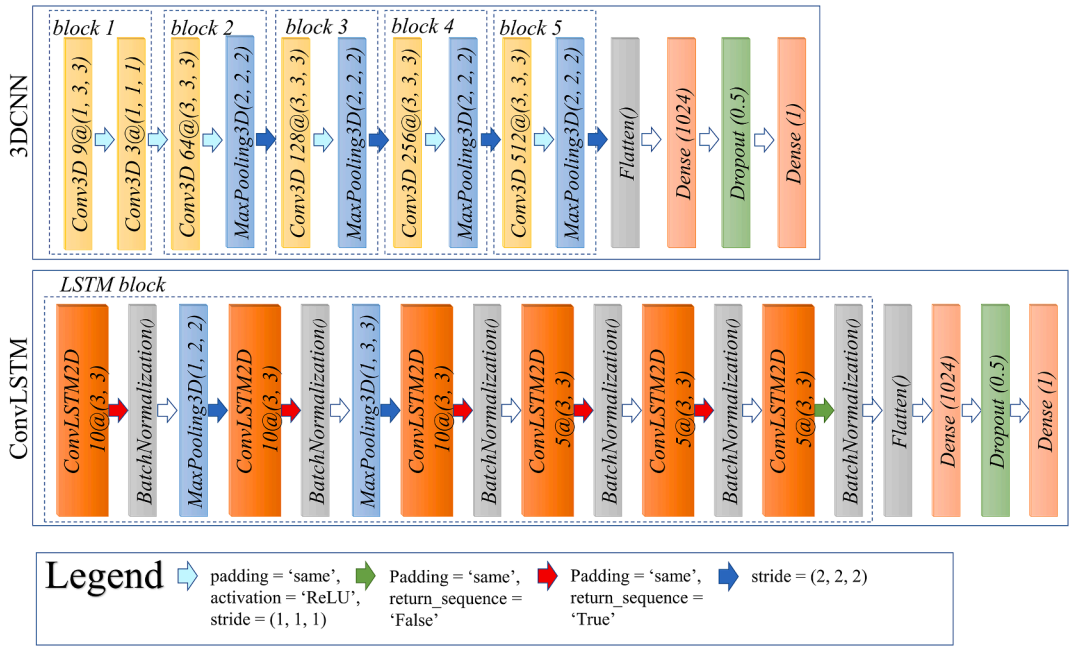

Fig. 4. Model architectures of the developed ConvLSTM and 3DCNN networks.

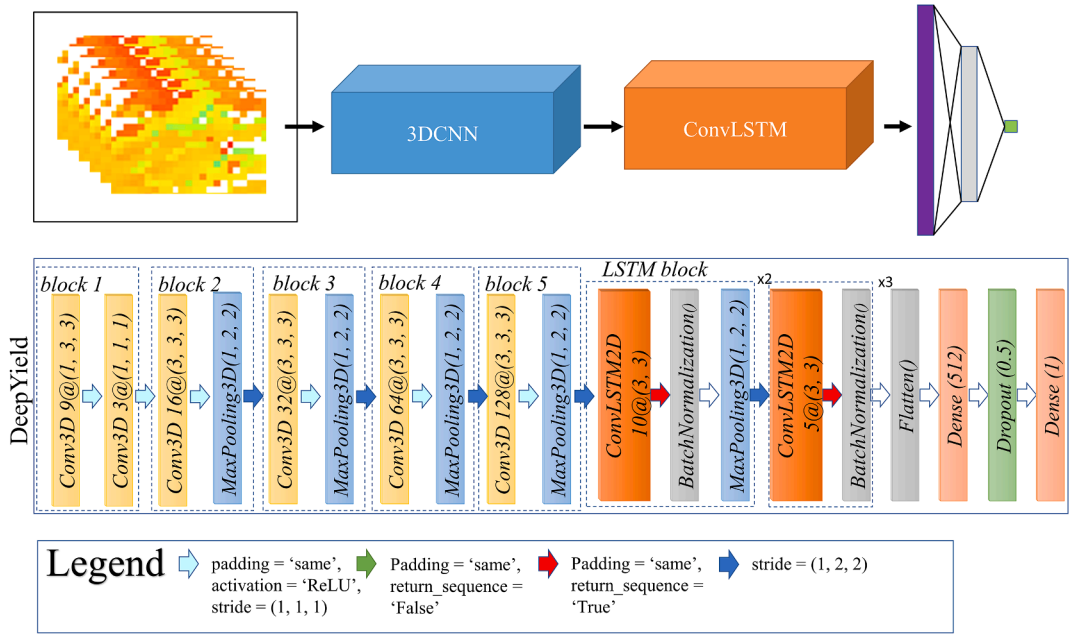

Fig. 5. The model architecture of DeepYield.

study for county-based soybean forecasting over the CONUS. MODIS images were converted into 4D tensors during the data preprocessing step and used as inputs to the networks. As depicted in this figure, the tensors passed through multiple convolutions and max-pooling layers. Multiple experiments were performed to find the optimum number of blocks and their respective parameters. In total 4 to 7 blocks were tested. For each block, 32, 64, 128, 512, or 1024 3D filters were considered, conditioned that the number of filters must increase by moving forward to the next block. For example, considering a total of 4 blocks, this will create 20 combinations. The number of neurons in the final dense layer was selected to be either 512 or 1024. Thus, in total, 42*2 = 84 combinations were tested and the best combination with 5 blocks of convolution and max-pooling layers followed by a flatten layer and a final dense layer with 1024 neurons was selected. The first block is a dimension reduction block in which 9 spectral bands of MODIS LST and SR are converted to 3 feature maps. The next four blocks perform the spatiotemporal feature extraction. ReLU activation function was used in these blocks for faster convergence of the network (Nair & Hinton, 2010). Extracted features are then connected to a flatten layer followed by two dense layers and a dropout layer with 0.5 probability in between. Dropout is a regularization method that at each epoch, randomly ignores some neurons during the training process. The method has proven to be effective in reducing the chance of overfitting (Srivastava, Hinton, Krizhevsky, & Salakhutdinov, 2014).

Fig. 3 also depicts the deep ConvLSTM architecture. In deep architecture, the output from one ConvLSTM layer is the input for the next

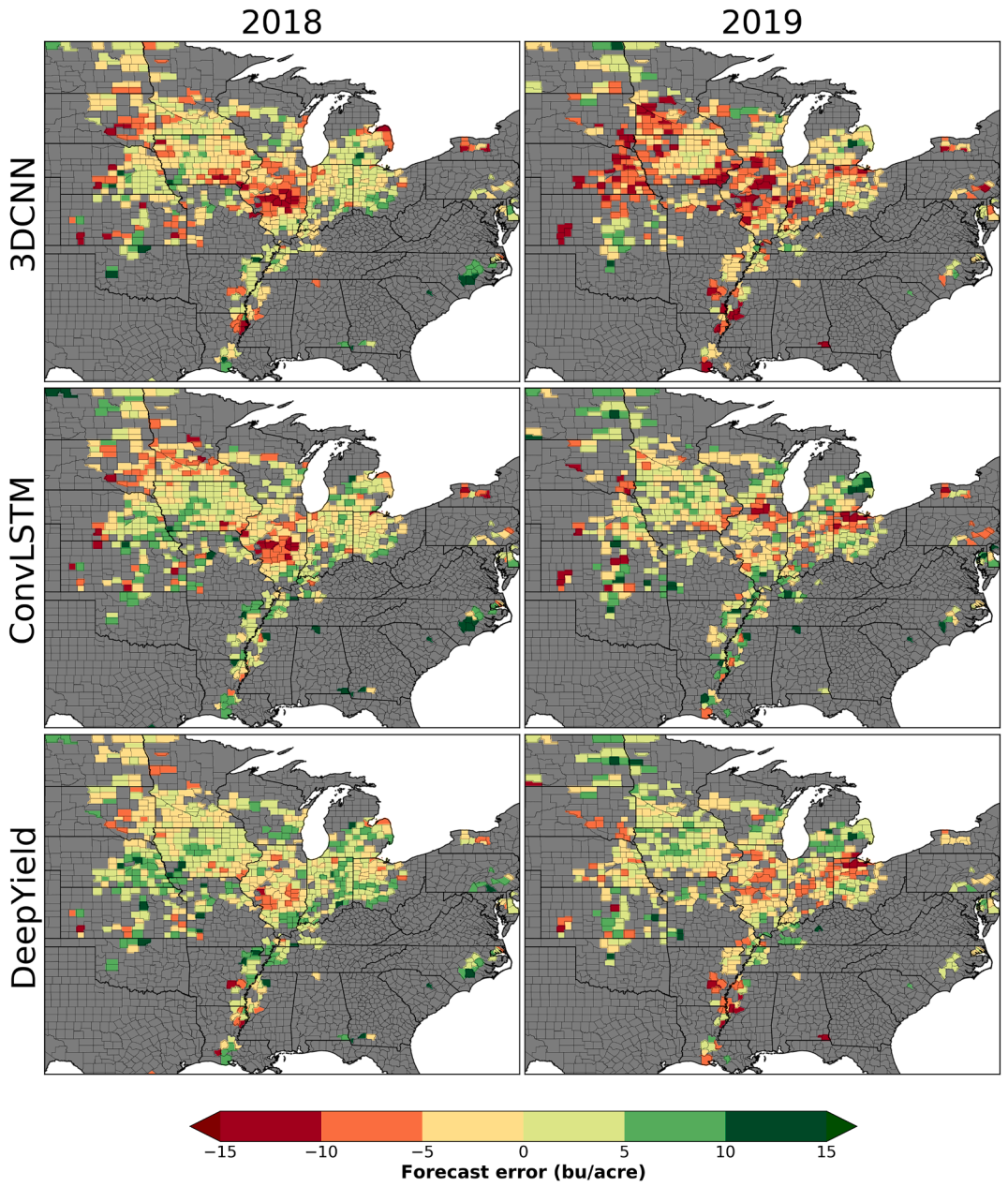

Fig. 6. A county-based map of forecasting error results for the test dataset.

layer. Several experiments, similar to the method explained for 3DCNN, were performed with up to 10 ConvLSTM layers and the best architecture was selected. The proposed ConvLSTM network consists of 8 layers such that each layer has 10 filters of size $(3, 3)$ and a stride size of 1. The same padding is used in each ConvLSTM layer to preserve the spatial dimension of the extracted spatiotemporal features during the convolutional process. Batch normalization, an operation that normalizes layers of inputs, is also performed before each ConvLSTM layer to speed up the learning process and ensure reasonable inputs for activations (Ioffe & Szegedy, 2015). Each ConvLSTM layer returns a sequence of last k time steps which is fed into the next layer. Except for the last layer which returns the final long short-term spatiotemporal features with the temporal dimension of one. The features are then flattened and connected to a dense layer with 512 neurons. Similar to the 3DCNN architecture, the linear activation function is used for the dense layers, and dropout with 0.5 probability is used to avoid overfitting.

3.3. DeepYield arcitecture

Fig. 4 presents the DeepYield architecture. In DeepYield, 3DCNN and ConvLSTM networks are combined for improved crop yield forecasting. Several experiments, similar to what was explained for 3DCNN and ConvLSTM, were performed to find the best number of 3DCNN and ConvLSTM blocks, and their respective parameters, to achieve the best performance. The LSTM blocks are connected to 5 blocks of 3DCNN for more rigorous spatiotemporal feature extraction. Since the ConvLSTM layers require fewer learning parameters DeepYield architecture allows for a deeper network to be trained. After processing the complete sequence of 3D inputs, the 3DCNN network performs several convolutions and max-pooling operations on the input images and prepares the final feature maps to be fed to the LSTM blocks. Then the LSTM blocks receive the extracted feature maps provided by the 3DCNN to perform more rigorous spatiotemporal feature extraction. Finally, the final feature maps will be fed to a dense layer with 1024 neurons. A dropout layer with 0.5 probability is used between two dense layers to avoid

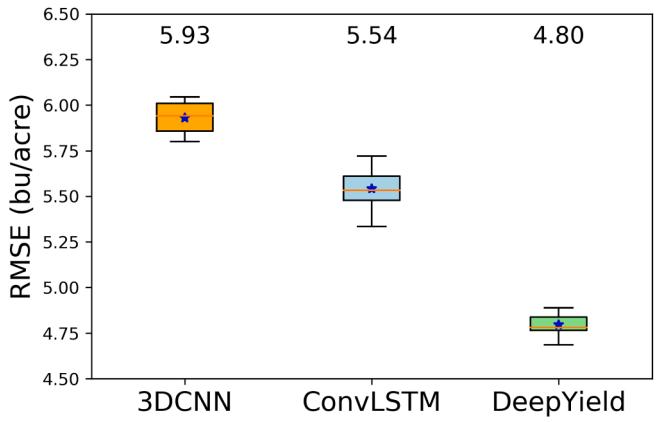

Fig. 7. Boxplots of loss values on the test dataset. The results are over 10 experiments with the same parameters for each model. The blue star represents the mean value which its value is also showed on top of each boxplot. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
The RMSE of the developed models comparing to other competing methods.

Year	DT	CNN + GP	CNN-LSTM	3DCNN	ConvLSTM	DeepYield
2018	7.20	5.89	5.97	5.97	5.62	**4.85**
2019	7.68	5.67	5.78	5.82	5.55	**4.73**
Average	7.44	5.78	5.88	5.90	5.59	**4.79**

overfitting.

4. Experimental results and analysis

4.1. Evaluation metrics

In order to be consistent with the literature, we used Root Mean Squared Error (RMSE) and Pearson Correlation Coefficient (PCC) measures to evaluate and compare the performance of the methods used in this study. Furthermore, errors are squared in RMSE before they are averaged giving a relatively high weight to large errors. This is most useful in the case of crop yield forecasting where large errors are particularly undesirable (Abbaszadeh, Gavahi, & Moradkhani, 2020; Gavahi, Mousavi, & Ponnambalam, 2019; Ravichandran, Gavahi, Ponnambalam, Burtea, & Mousavi, 2021).

$$RMSE = \sqrt{\frac{\sum_{i=0}^{N}(M_i - O_i)^2}{n}} \tag{13}$$

$$PCC = \frac{\sum_{i=0}^{N}\left(M_i - \overline{M}\right)\left(O_i - \overline{O}\right)}{\sqrt{\sum_{i=0}^{N}\left(M_i - \overline{M}\right)^2}\sqrt{\sum_{i=0}^{N}\left(O_i - \overline{O}\right)^2}} \tag{14}$$

where M_i and O_i denote the model forecast and observed yield value, respectively, and $\overline{M}$ and $\overline{O}$ are their respective mean values.

4.2. Implementation details

In this study, Tensorflow 1.14 (Abadi et al., 2016) and Keras library (Gulli & Pal, 2017) in python were used to implement the proposed models on the University of Alabama High-Performance Computing (UAHPC) server with two Tesla V100-PCIE GPU.

For the training process, different batch sizes of 16, 32, and 64 were tested and the best performance was achieved by the batch size of 32. The number of epochs was set to 50. We experimented with Stochastic Gradient Decent (SGD) (Fuh & Hu, 2006) and Adam optimizer (Kingma & Ba, 2015) with different learning rates and selected Adam optimizer with a learning rate of 0.001. Early stopping was used on the validation set to prevent models from overfitting.

The input images are 4D tensors with the dimension of Time × Height × Width × band. This will create an enormous computational burden for the training process. Thus it is recommended to train such deep architectures with big datasets as inputs over GPU configurations (Wang, Wei, & Brooks, 2019). Considering the above implementation details, each training process for DeepYield architecture takes about 7 h and 42 min. Also, this high computational intensity makes it difficult to test different configurations which can be considered as one of the setbacks of such deep architectures.

4.3. Forecasting performance for years 2018 and 2019

We used 15 years of data (from 2003 to 2017) for training and 2 years (from 2018 to 2019) for testing the trained models. A 20% validation set was also used for tuning model hyperparameters and checking on the stopping criteria. Fig. 5 shows the forecast error in bu/acre for the years 2018 and 2019 over the CONUS. As it is shown in this figure, DeepYield results in the best performance with errors of less than 5 bu/acre for the majority of the counties. The performance is specifically improved in major soybean-producing states including Ohio, Indiana, Illinois, and Iowa. States with lower rates of soybean production such as

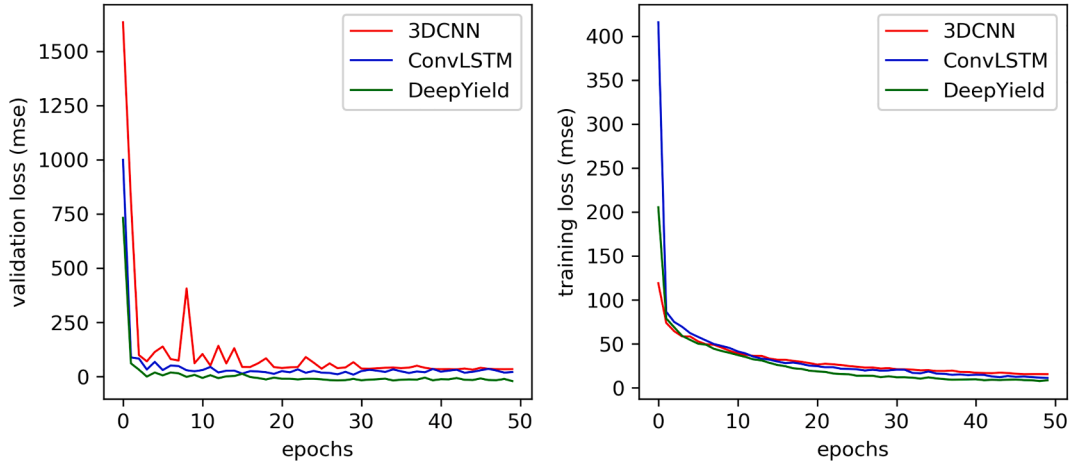

Fig. 8. Loss versus the number of training epochs for training and validation sets.

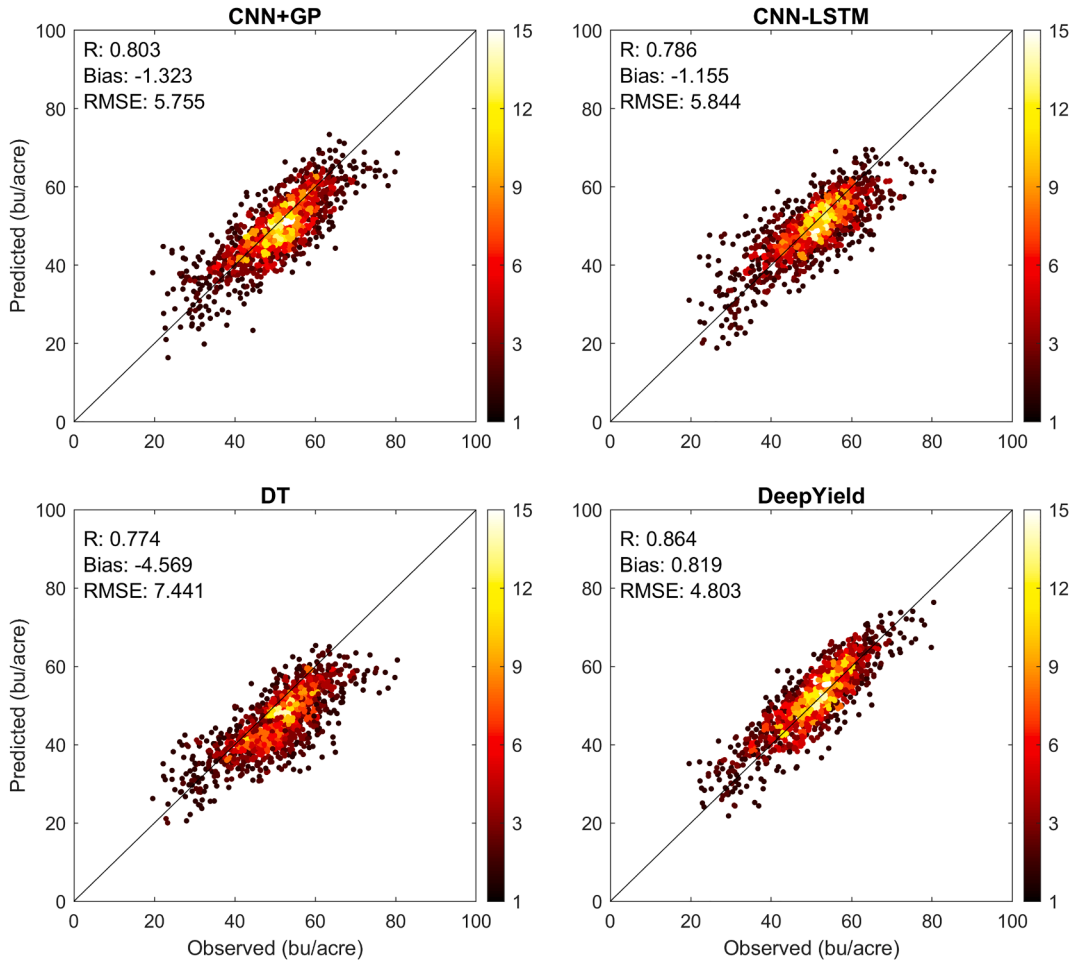

Fig. 9. Heat scatter plots of the predicted versus observed yield values for different models.

Mississippi, Arkansas, and Louisiana are showing the worst performance. This is because most of the soybean-producing counties in these states have less than 10 to 30 percent cropland area which makes it difficult for models to accurately forecast their yield.

The main comparison between the developed structures is illustrated in Fig. 6. It shows the boxplots of loss values on the test dataset by the network structure. Results are based on 10 runs of experiments for each network structure. The results based on RMSE clearly show that on average the DeepYield outperforms the ConvLSTM and 3DCNN networks by 16 and 25 percent respectively. Also, it has the lowest variance of 0.0035 comparing to 0.0121 and 0.0073 for ConvLSTM and 3DCNN respectively. This shows that the DeepYield structure is not only performing better but also has higher stability.

Fig. 7 demonstrates the convergence speed of the three deep networks used in this study during the training phase for one of the sample runs. All the methods experience a sharp reduction in loss value in the first few epochs. However, despite some fluctuations, DeepYield converges faster to the optimal values of parameters. In addition, comparing the networks from the validation loss rate angle, it achieves the lowest RMSE value.

4.4. Comparing with competing approaches

To provide a comprehensive comparison with other methodologies presented in the literature, we have replicated the works by Johnson (2014) (Decision Tree), You et al. (2017) (histogram-based CNN + GP), and Sun et al. (2019) (histogram-based CNN-LSTM) to assess their performance in the same testing period (years 2018 and 2019). The county-level RMSE forecasts performances are presented in Table 1. The results are the average over 10 runs. Each row corresponds to the forecast for that year based on a model trained from 2003 to 2017. Model hyperparameters and stopping criteria are tuned on a 20% hold-out validation set. The results show that DeepYield significantly outperforms other deep networks introduced in the literature.

Fig. 8 shows the heat-scatter plots of the predicted versus observed soybean yield for each model. This figure indicates, our proposed approach has had the lowest bias and RMSE and the highest correlation coefficient compared to other networks presented in the literature. As it is expected the DT method is the simplest approach and has the lowest performance. The CNN + GP and CNN-LSTM results are relatively close while the CNN + GP slightly outperforms the CNN-LSTM. Comparing the performance metrics of DeepYield with CNN + GP shows that this method improves the RMSE and correlation coefficient by 16.5 and 7.6 percent, respectively. The error distribution maps of these approaches are also depicted in Fig. 9, which indicates that DeepYield significantly outperforms the other methods. The DT model is mostly underestimating the yield values (counties with dark red colors) which is consistent with the heat-scatter plot provided in Fig. 8 in which the points are mostly below the $y = x$ line. The CNN + GP and CNN-LSTM models are performing much better by showing few sporadic counties with error values higher than 10 bu/acre. However, the best performance is achieved by the DeepYield model in which the majority of the counties have forecasting errors below 5 bu/acre.

To further bolster the effectiveness of the proposed DeepYield, the method was also tested for forecasting corn yield in major corn-producing counties in the CONUS. To illustrate more on differences

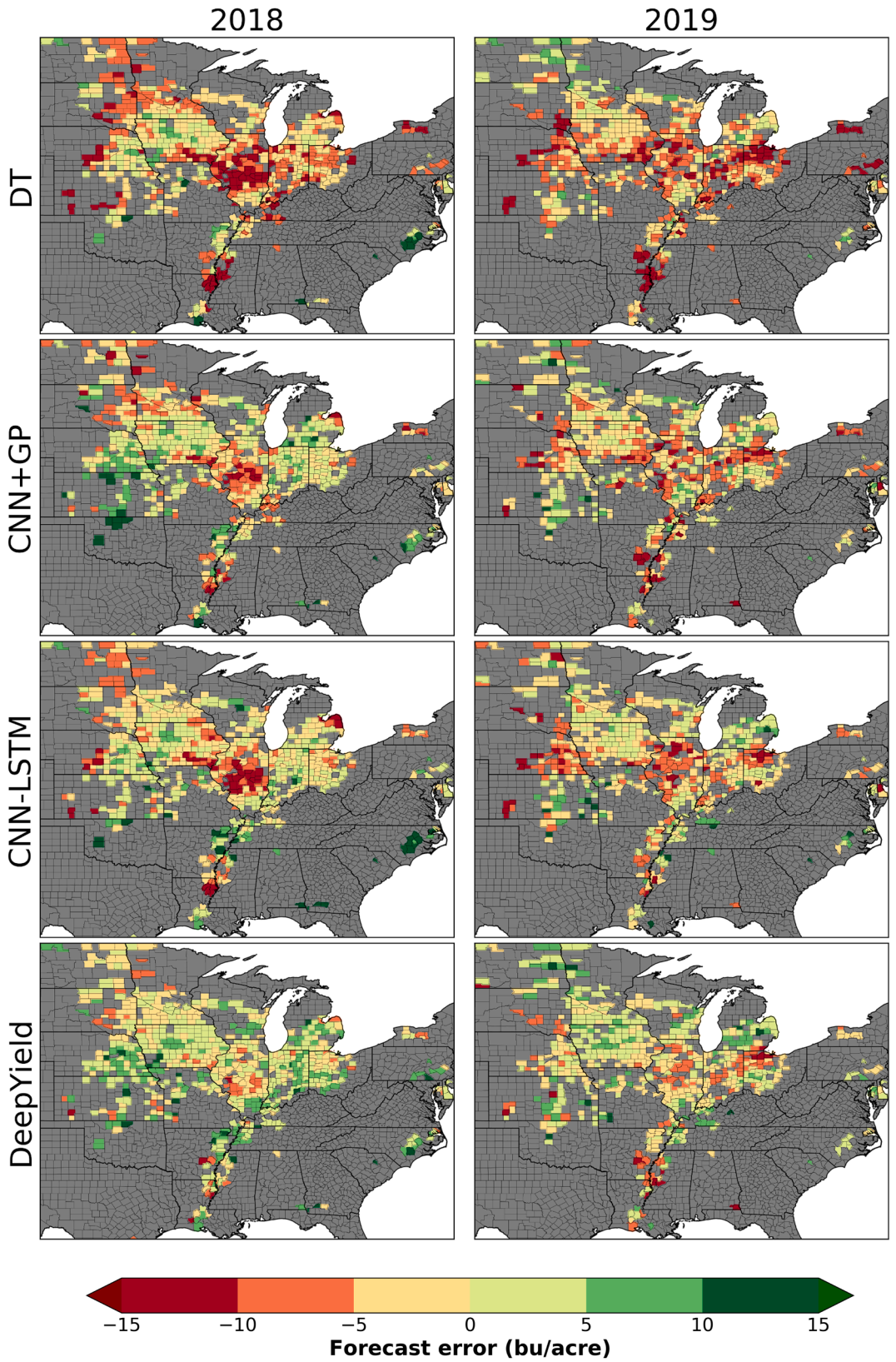

Fig. 10. Maps of forecasting error distribution of different approaches compared with DeepYield.

between corn and soybean, Fig. S1 (in supplementary file) shows a comparison between corn and soybean distributions for all the counties that produce these crops for the years 2003 to 2019. As it has been shown in Figs. S2 and S3, DeepYield has been successful in forecasting county-based corn yields for the test years as well. Similar to soybean, DeepYield still outperforms individual models and also shows higher performance in terms of RMSE and R, when compared with competing approaches in Fig. S3. This shows that the proposed DeepYield

architecture can effectively capture the spatiotemporal patterns and successfully be used for the task of crop yield forecasting.

5. Conclusion and future work

In this study, a combined deep convolutional network based on ConvLSTM and 3DCNN, called DeepYield, was proposed for county-based crop yield forecasting across the CONUS. For the first time, a ConvLSTM network was used for crop yield forecasting. We further improved the effectiveness and usefulness of the deep network by combining the ConvLSTM with 3DCNN for better spatiotemporal feature extraction. Furthermore, instead of taking the average of pixel values or using histograms of pixel intensities, this study preserves the spatial dimension of the input images by using the full image as input. Remote sensing images from the MODIS satellite have been used as predictors to forecast the end-of-season soybean yield. The findings of this study indicated that the most efficient deep features were determined by the proposed approach, outperforming other individual methods i.e., ConvLSTM, and 3DCNN. The models were tested for forecasting soybean yields for two years of 2018 and 2019 and the results showed that they provide reasonably accurate forecast yields for these years. Finally, we compared the results of our proposed approach with those obtained by other methods such as Decision Tree, CNN + GP, and CNN-LSTM and concluded that our developed approach significantly outperforms the other competing methods. The outcome of this study can be beneficial for farmers, agricultural planners, and other agencies such as the United States Department of Agriculture (USDA), responsible for the national and regional crop yield forecasting.

In this study, only SR and LST have been considered as the most attributing factors. Future works include adding other important inputs that affect plant growth and final yields such as hydrological variables, weather and environmental data, and plant genotypes. Moreover, the results of this study can only provide yield forecasts on a county level. More research is required to find solutions that could provide information at a finer scale. Also, as it is apparent in Figs. 6 and 10, there exists a high spatial correlation between the counties in proximity. This will provide room for further investigation by taking the spatial correlations into account. Furthermore, one important aspect to be investigated is the impacts of extreme events such as drought and hurricanes on the predictability of machine learning models and how to improve their accuracy during these events.

CRediT authorship contribution statement

Keyhan Gavahi: Conceptualization, Methodology, Software, Visualization, Data curation, Validation, Investigation, Writing - original draft. **Peyman Abbaszadeh:** Conceptualization, Visualization, Writing - review & editing. **Hamid Moradkhani:** Conceptualization, Methodology, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work appeared in this paper.

Acknowledgment

Partial financial support for this research was provided by NSF-INFEWS grant # 1856054.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eswa.2021.115511.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., … Kudlur, M. (2016). In *In 12th USENIX symposium on operating systems design and implementation ({OSDI}* (pp. 265–283).

Abbaszadeh, P., Gavahi, K., & Moradkhani, H. (2020). Multivariate remotely sensed and in-situ data assimilation for enhancing community WRF-Hydro model forecasting. *Advances in Water Resources, 145*, 103721.

Abbaszadeh, P., Moradkhani, H., Gavahi, K., Kumar, S., Hain, C., Zhan, X., … Karimiziarani, S. (2021). High-resolution SMAP satellite soil moisture product: Exploring the opportunities. *Bulletin of the American Meteorological Society, 102*(4), 309–315.

Benali, A., Carvalho, A. C., Nunes, J. P., Carvalhais, N., and Santos, A. (2012). "Estimating air surface temperature in Portugal using MODIS LST data." Remote Sensing of Environment, Elsevier Inc., 124, 108–121.

Bolton, D. K., & Friedl, M. A. (2013). Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agricultural and Forest Meteorology, Elsevier B.V., 173*, 74–84.

Ceglar, A., Toreti, A., Prodhomme, C., Zampieri, M., Turco, M., & Doblas-Reyes, F. J. (2018). Land-surface initialisation improves seasonal climate prediction skill for maize yield forecast. *Scientific Reports, Springer, US, 8*(1), 1–9.

Chen, Z., Zhang, B., Stojanovic, V., Zhang, Y., & Zhang, Z. (2020). Event-based fuzzy control for T-S fuzzy networked systems with various data missing. *Neurocomputing, Elsevier B.V., 417*, 322–332.

Duchemin, B., Maisongrande, P., Boulet, G., & Benhadj, I. (2008). A simple algorithm for yield estimates: Evaluation for semi-arid irrigated winter wheat monitored with green leaf area index. *Environmental Modelling and Software, 23*(7), 876–892.

Elboushaki, A., Hannane, R., Afdel, K., & Koutti, L. (2020). MultiD-CNN: A multi-dimensional feature learning approach based on deep convolutional networks for gesture recognition in RGB-D image sequences. *Expert Systems with Applications, Elsevier Ltd, 139*, Article 112829.

Fuh, C. Der, and Hu, I. (2006). "Bayesian stochastic estimation of the maximum of a regression function." Random Walk, Sequential Analysis and Related Topics: A Festschrift in Honor of Yuan-Shih Chow, 269–280.

Gallego, J., Carfagna, E., & Baruth, B. (2010). Accuracy, objectivity and efficiency of remote sensing for agricultural statistics. *Agricultural Survey Methods, John Wiley & Sons Ltd, Chichester, UK*, 193–211.

Gavahi, K., Abbaszadeh, P., Moradkhani, H., Zhan, X., & Hain, C. (2020). Multivariate assimilation of remotely sensed soil moisture and evapotranspiration for drought monitoring. *Journal of Hydrometeorology, 21*(10), 2293–2308.

Gavahi, K., Mousavi, S. J., & Ponnambalam, K. (2019). Adaptive forecast-based real-time optimal reservoir operations: Application to Lake Urmia. *Journal of Hydroinformatics, IWA Publishing, 21*(5), 908–924.

Gers, F. A., & Schmidhuber, J. (2000). Recurrent nets that time and count. In *Proceedings of the International Joint Conference on Neural Networks* (pp. 189–194).

Gulli, A., & Pal, S. (2017). Deep Learning with Keras. *Packt Publishing Ltd, <https://books.google.com/books?hl=en&lr=&id=20EwDwAAQBAJ&oi=fnd&pg=PP1&ots=lHhC5ngSS1&sig=pwB45vaYe_XABCsbtxnYDjIoSO0#v=onepage&q&f=false> (Aug. 11*, 2020).

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation, MIT Press Journals, 9*(8), 1735–1780.

Hoogenboom, G., White, J. W., & Messina, C. D. (2004). From genome to crop: Integration through simulation modeling. *Field Crops Research, 90*(1), 145–163.

Ioffe, S., & Szegedy, C. (2015). In *Batch normalization: Accelerating deep network training by reducing internal covariate shift* (pp. 448–456). International Machine Learning Society (IMLS).

Ji, S., Xu, W., Yang, M., & Yu, K. (2013). 3D Convolutional neural networks for human action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE, 35*(1), 221–231.

Ji, S., Zhang, C., Xu, A., Shi, Y., & Duan, Y. (2018). 3D convolutional neural networks for crop classification with multi-temporal remote sensing images. *Remote Sensing, 10*(1).

Jiang, Z., Liu, C., Hendricks, N. P., Ganapathysubramanian, B., Hayes, D. J., and Sarkar, S. (2018). "Predicting County Level Corn Yields Using Deep Long Short Term Memory Models." 1–26.

Johnson, D. M. (2014). An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. *Remote Sensing of Environment, Elsevier, 141*, 116–128.

Khaki, S., Wang, L., & Archontoulis, S. V. (2020). A CNN-RNN framework for crop yield prediction. *Frontiers in Plant Science, 10*(January), 1–14.

Kim, N., Ha, K. J., Park, N. W., Cho, J., Hong, S., & Lee, Y. W. (2019). A comparison between major artificial intelligence models for crop yield prediction: Case study of the midwestern United States, 2006–2015. *ISPRS International Journal of Geo-Information, 8*(5).

Kim, N., & Lee, Y. W. (2016). Machine learning approaches to corn yield estimation using satellite images and climate data: A case of Iowa State. *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography, Korean Society of Surveying, 34*(4), 383–390.

Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* (pp. 1–15).

Kuwata, K., & Shibasaki, R. (2015). Estimating crop yields with deep learning and remotely sensed data. In *International Geoscience and Remote Sensing Symposium (IGARSS), Institute of Electrical and Electronics Engineers Inc* (pp. 858–861).

Lee, S. W., & Kim, H. Y. (2020). Stock market forecasting with super-high dimensional time-series data using ConvLSTM, trend sampling, and specialized data augmentation. *Expert Systems with Applications, Elsevier Ltd, 161*, Article 113704.

Li, Y., Zhang, H., & Shen, Q. (2017). Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network. *Remote Sensing, 9*(1).

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors (Switzerland), 18*(8), 1–29.

Lin, L., Wang, K., Zuo, W., Wang, M., Luo, J., & Zhang, L. (2016). A Deep Structured Model with Radius–Margin Bound for 3D Human Activity Recognition. *International Journal of Computer Vision, Springer, US, 118*(2), 256–273.

Liu, Z., Zhang, C., & Tian, Y. (2016). 3D-based Deep Convolutional Neural Network for action recognition with depth sequences. *Image and Vision Computing, Elsevier B.V., 55*, 93–100.

Lofton, J., Tubana, B. S., Kanke, Y., Teboh, J., Viator, H., & Dalen, M. (2012). Estimating sugarcane yield potential using an in-season determination of normalized difference vegetative index. *Sensors (Switzerland), 12*(6), 7529–7547.

Maturana, D., & Scherer, S. (2015). 3D Convolutional Neural Networks for landing zone detection from LiDAR. In *Proceedings - IEEE International Conference on Robotics and Automation, Institute of Electrical and Electronics Engineers Inc* (pp. 3471–3478).

Mendes, W. R., Araújo, F. M. U., Dutta, R., & Heeren, D. M. (2019). Fuzzy control system for variable rate irrigation using remote sensing. *Expert Systems with Applications, Elsevier Ltd, 124*, 13–24.

Nair, V., and Hinton, G. E. (2010). Rectified Linear Units Improve Restricted Boltzmann Machines.

Petersen, N. C., Rodrigues, F., & Pereira, F. C. (2019). Multi-output bus travel time prediction with convolutional LSTM neural network. *Expert Systems with Applications, Elsevier Ltd, 120*, 426–435.

Ravichandran, T., Gavahi, K., Ponnambalam, K., Burtea, V., & Mousavi, S. J. (2021). Ensemble-based machine learning approach for improved leak detection in water mains. *Journal of Hydroinformatics, 23*(2), 307–323.

Russello, H. (2018). *Convolutional Neural Networks for Crop Yield Prediction using Satellite Images*. University of Amsterdam.

Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., and Woo, W. C. (2015). "Convolutional LSTM network: A machine learning approach for precipitation nowcasting." Advances in Neural Information Processing Systems, 2015-Janua (June), 802–810.

Shrestha, R., Di, L., Eugene, G. Y., Kang, L. Shao, & Bai, Y. Q. (2017). Regression model to estimate flood impact on corn yield using MODIS NDVI and USDA cropland data layer. *Journal of Integrative Agriculture, 16*(2), 398–407.

Shrestha, R., Di, L., Eugene, G. Y., Kang, L., Li, L., Rahman, M. S., & Yang, Z. (2016, July). *Regression based corn yield assessment using MODIS based daily NDVI in Iowa state* (pp. 1–5). IEEE.

Srivastava, N., Hinton, G., Krizhevsky, A., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research.*

Stojanovic, V., He, S., & Zhang, B. (2020). State and parameter joint estimation of linear stochastic systems in presence of faults and non-Gaussian noises. *International Journal of Robust and Nonlinear Control, 30*(16), 6683–6700.

Sulla-Menashe, D., and Friedl, M. (2019). "MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006." distributed by NASA EOSDIS Land Processes DAAC, https://doi.org/10.5067/MODIS/MCD12Q1.006.

Sun, J., Di, L., Sun, Z., Shen, Y., & Lai, Z. (2019). County-level soybean yield prediction using deep CNN-LSTM model. *Sensors (Switzerland), 19*(20), 1–21.

Tran, D., Bourdev, L., Fergus, R., Torresani, L., & Paluri, M. (2015). Learning spatiotemporal features with 3D convolutional networks. *Proceedings of the IEEE International Conference on Computer Vision*.

United States Department of Agriculture. (2012). "The Yield Forecasting Program of NASS." SMB staff report number SMB 12-01, (April), NASS Staff Report No. SMB 12-01.

Vermote, E. (2015). "MOD09A1 MODIS/Terra Surface Reflectance 8-Day L3 Global 500m SIN Grid V006." distributed by NASA EOSDIS Land Processes DAAC, https://doi.org/10.5067/MODIS/MOD09A1.006.

Wan, Z. (2006). "Modis land surface temperature products users guide." Institute for Computational Earth System Science, University of California: Santa Barbara, CA, USA, (March), 1–33.

Wan, Z. (2015). "MYD11A2 MODIS/Aqua Land Surface Temperature/Emissivity 8-Day L3 Global 1km SIN Grid V006." distributed by NASA EOSDIS Land Processes DAAC, https://doi.org/10.5067/MODIS/MYD11A2.006.

Wang, A. X., Tran, C., Desai, N., Lobell, D., and Ermon, S. (2018). "Deep transfer learning for crop yield prediction with remote sensing data." Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies, COMPASS 2018.

Wang, D., Chen, Y., Hu, L., Voogt, J. A., Gastellu-Etchegorry, J. P., and Krayenhoff, E. S. (2021). "Modeling the angular effect of MODIS LST in urban areas: A case study of Toulouse, France." Remote Sensing of Environment, Elsevier Inc., 257(19), 112361.

Wang, Y., Wei, G. Y., and Brooks, D. (2019). "Benchmarking TPU, GPU, and CPU platforms for deep learning." arXiv.

Wei, T., Li, X., & Stojanovic, V. (2021). Input-to-state stability of impulsive reaction–diffusion neural networks with infinite distributed delays. *Nonlinear Dynamics, Springer, Netherlands, 103*(2), 1733–1755.

Xue, J., and Su, B. (2017). "Significant remote sensing vegetation indices: A review of developments and applications." Journal of Sensors, 2017.

You, J., Li, X., Low, M., Lobell, D., & Ermon, S. (2017). Deep Gaussian process for crop yield prediction based on remote sensing data. *31st AAAI Conference on Artificial Intelligence*.

Zhang, X., He, S., Stojanovic, V., Luan, X., & Liu, F. (2021). Finite-time asynchronous dissipative filtering of conic-type nonlinear Markov jump systems. *Science China Information Sciences, 64*(5), 1–12.