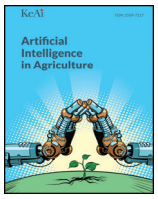
Artificial Intelligence in Agriculture 5 (2021) 196–207

Contents lists available at ScienceDirect

Artificial Intelligence in Agriculture

journal homepage: http://www.keaipublishing.com/en/journals/artificial-intelligence-in-agriculture/



Reducing deep learning network structure through variable reduction methods in crop modeling

Babak Saravi [a], A. Pouyan Nejadhashemi [a,b,*], Prakash Jha [b,c], Bo Tang [d]

[a] Department of Biosystems and Agricultural Engineering, Michigan State University, East Lansing, MI 48824, USA
[b] Department of Plant, Soil and Microbial Sciences Michigan State University, East Lansing, MI 48824, USA
[c] Feed the Future Innovation Lab for Collaborative Research on Sustainable Intensification, Kansas State University, Manhattan, KS 66506, USA
[d] Department of Electrical and Computer Engineering, Mississippi State University, MS 39762, USA

ARTICLE INFO

Article history:
Received 23 March 2021
Received in revised form 30 September 2021
Accepted 30 September 2021
Available online 2 October 2021

Keywords:
Deep learning
Artificial intelligent
Variable reduction
Crop modeling
Yield prediction
Irrigation

ABSTRACT

Crop models are widely used to predict plant growth, water input requirements, and yield. However, existing models are very complex and require hundreds of variables to perform accurately. Due to these shortcomings, large-scale applications of crop models are limited. In order to address these limitations, reliable crop models were developed using a deep neural network (DNN) – a new approach for predicting crop yields. In addition, the number of required input variables was reduced using three common variable selection techniques: namely Bayesian variable selection, Spearman's rank correlation, and Principal Component Analysis Feature Extraction. The reduced-variable DNN models were capable of estimating future crop yields for 10,000,000 different weather and irrigation scenarios while maintaining comparable accuracy levels to the original model that used all input variables. To establish clear superiority of the methodology, the results were also compared with a very recent feature selection algorithm called min-redundancy max-relevance (mRMR). The results of this study showed that the Bayesian variable selection was the best method for achieving the aforementioned goals. Specifically, the final Bayesian-based DNN model with a structure of 10 neurons in 5 layers performed very similarly (78.6% accuracy) to the original DNN crop model with 400 neurons in 10 layers, even though the size of the neural network was reduced by 80-fold. This effort can help promote sustainable agricultural intensifications through the large-scale application of crop models.

1. Introduction

A major challenge in the twenty-first century is meeting the needs of the fast-growing human population that burgeons demand on food, water, and energy (triple nexuses) (Slavin, 2016). The world's population has grown exponentially in the past 100 years and is expected to reach 10 billion by 2055 (Kitzes et al., 2008). The Food and Agriculture Organization of the United Nations predicts that this growth will demand 50% more food, equating to a 70% increase in food production (FAO, 2017). Given that most of the world's agricultural land is already in production (Bruinsma, 2003), management efficiency must improve to match the demand. As a result, new techniques are emerging, which take both resource and climate limitations into account (Gebbers and Adamchuk, 2010).

Increases in crop productivity have been largely attributed (50–60%) to breeding and the development of hybrid cultivars, followed by improved management practices (Connor et al., 2011; Sacks and Kucharik, 2011). Technological advancements have helped fine-tune management with increased adaptation of yield simulation (Ali and Deo, 2020; Ali et al., 2018a), field monitoring (Rao and Sridhar, 2018), and other data-driven practices (Pathak et al., 2018). Most recently, precision in resource management has been continuously fine-tuned through crop models and use of satellite navigation systems (Abbasi et al., 2014; Basso et al., 2001; Lobell and Burke, 2010).

Digitized agriculture, or smart farming, has made a significant contribution to improving productivity, ensuring food security, and protecting the environment (Tyagi, 2016). Wide-scale employment of smart agriculture is necessary for meeting the coming challenges of food security and water deficiency; applying inputs in the right amount, time, and place are the basis of efficient crop management systems (Gebbers and Adamchuk, 2010). External influences can make it difficult to operate at optimal efficiency, since many factors such as climate, pests, and disease can adversely influence crop management plans. Smart agriculture can help modify management systems to maximize crop yield while minimizing input requirements given current conditions.

* Corresponding author at: Department of Biosystems and Agricultural Engineering, Michigan State University, East Lansing, MI 48824, USA.
E-mail address: pouyan@msu.edu (A.P. Nejadhashemi).

https://doi.org/10.1016/j.aiia.2021.09.001

1.1. Crop modeling

The complexity of agricultural systems demands exploratory research to discover and validate system factors' interactions and influences on each other. At the systems level, biotic and abiotic factors interact nonlinearly and are very difficult to study in isolation. Meanwhile, crop simulation models can help researchers go around resource constraints by mimicking the physiological process in connection with relevant physical conditions (soil and weather). Crop models help synthesize complex systems through a reductionist approach, i.e., reducing the number of inputs to only include the components that significantly influence crop growth and development (De Wit and De Vries, 1983). System behavior and processes can be hypothesized through proper analysis of crop models. Historically, crop models have been used for yield gap analysis, understanding weather impacts, studying crop phenology and physiology, and developing management strategies (Jha et al., 2018). Most of the dynamic crop models produced to date have been developed from equations representing growth processes and the influence of abiotic factors like soil and weather (Hoogenboom et al., 2015).

Deterministic crop models can be categorized into three main groups: statistical models (Lobell and Asseng, 2017; Schlenker et al., 2006), mechanistic models (Arnold et al., 2012), and functional models (Arnold et al., 2012; Ballesteros et al., 2016). However, these deterministic models could not capture spatial and temporal variability in inputs (soil, climate, and other factors). Therefore, input uncertainty could lead to biases in model outputs that must be validated. Biophysical models (De Wit, 1965) and their continuous evolution (Bouman et al., 1996) have led to significant advancements in crop biomass/yield estimation with less risk of bias and uncertainty. Basic crop growth simulation models include Elementary CROp growth Simulator-ELCROS (Bouman et al., 1996), BAsic CROp growth Simulator-BACROS (De Vries, 1973; De, 1978; Goudriaan, 1977; van Keulen, 1975), Simple and Universal Crop growth Simulator-SUCROS (Spitters et al., 1989), WOrld FOod Studies-WOFOST (Van Diepen et al., 1989), Modules for Annual CRop Simulation-MACROS (Penning de Vries, 1989), rice crop model-ORYZA (Kropff, 1994), and PAPRAN (Seligman and Van Keulen, 1981; Van Keulen, 1982). These basic models were the precursors of the modern crop simulation models: Agricultural Production Systems sIMulator-APSIM (McCown et al., 1996), Decision Support System for Agrotechnology Transfer-DSSAT (Jones et al., 2003a, 2003b), Cropping Systems simulation model-CropSyst (Stöckle et al., 2003), InfoCrop (Aggarwal et al., 2006), and other dynamic models that assess biophysical impacts on crops.

Crop models play a significant role in interpreting field experiment results, assisting in a timely decision-making process for input management (Jha et al., 2018) and for climate change's impact on crop yield (Lobell and Asseng, 2017). Biophysical crop models estimate the processes and factor influence through parameterized equations (Wallach et al., 2018). Parameter estimation can be performed through sensitivity analysis and model calibration (Sehgal et al., 2017) as long as uncertainty is accounted for (Ahuja and Ma, 2011; He et al., 2009). For instance, many environmental variables, like precipitation, temperature, solar radiation, irrigation, and fertilizer applications, can directly impact crop growth and yield. Not only the quantity, but also the timing of environmental variables play a significant role in determining overall outcomes. The large number of variables and nonlinear system responses limit the techniques and algorithms that can be used to model biophysical systems.

1.2. Artificial intelligence application in crop modeling

A systematic approach to analyzing the complex and unpredictable behavior of agriculture can be used to meet smart farming requirements (Chi et al., 2016; Hashem et al., 2015). Heterogeneous data collection, processing, and analysis produce a vast network requiring real-time data synthesis and reclassification (Kempenaar et al., 2016). The application of new artificial intelligence (AI) techniques in agricultural fields is desirable, given their ability to analyze and use big data. Additionally, these techniques can be used for model development without extensive knowledge of the specific area of application (Angermueller et al., 2016; Latha and Mohana, 2016; Menger et al., 2018). One of the most popular AI techniques is Deep Learning (DL). DL is a large structure (multiple layers) of an Artificial Neural Network (ANN), which has been around since the seventies. Using DL techniques for modeling has only recently become possible through the advancements made in computer hardware technologies, such as high-performance computer clusters (HPCC), multicore computer processing units (CPUs), and powerful graphic processing units (GPUs). However, most agriculture-related DL experiments conducted thus far have employed Convolutional Neural Network (CNN) architectures, which vary in learning rates and efficiency (Amara et al., 2017; Prasad et al., 2020).

In principle, DL is similar to ANN, with higher performance capabilities and more than three layers. Both of these techniques are widely used (Chen et al., 2014; LeCun et al., 2015; Schmidhuber, 2015). However, a comparative study among these techniques showed the superiority of the DL-based model to ANN to estimate biochemical oxygen demand and total phosphorus loads at watershed scales (Song et al., 2016). In another study by Khaki and Wang (2019), a deep neural network (DNN) model was designed for predicting crop yield, and the results were compared against a shallow neural networks (SNN) model. Overall, the predictive model based on DNN outperformed the SNN model.

DL allows multiple levels of abstraction by hierarchical data processing. In agriculture, DL is a new but promising approach with immense potential, uncovering different areas and dimensions through image visualization and analysis. The most explored areas of DL application in agriculture are classification of land use and land cover, crop type, recognition or identification of plant type in weed management, counting of final produce (fruits/vegetables), and recognizing diseases in plants. Algorithms based on DL have even more potential to predict future farm parameters, such as soil moisture (Song et al., 2016) and weather (Sehgal et al., 2017). Variations in available data help to generate further training in DL models, which offers the ability to differentiate between characteristics and achieving greater accuracy in classification (Kamilaris and Prenafeta-Boldú, 2018). However, differentiation based on data is difficult to identify when assessing crop type (Dyrmann et al., 2017; Ienco et al., 2017; Kussul et al., 2017; Rebetez et al., 2016), crop stages (Chen et al., 2017; Minh et al., 2017; Namin et al., 2018; Yalcin, 2017), and crop conditions (Amara et al., 2017; Prasanna et al., 2016; Rahnemoonfar and Sheppard, 2017; Sladojevic et al., 2016). Complexities in layered structures which change gradually with phenology create a complex network of data, increasing the size of the dataset (Einheuser et al., 2012). As a result, the computational capability to develop and train the model is compromised by the number of variables and training performance is reduced.

In order to address the problems associated with the high number of input variables necessary to model environmental and agricultural systems, many techniques have been used such as the Bayesian variable selection (O'Hara and Sillanpää, 2009; Woznicki et al., 2015), the Spearman's rank correlation (Einheuser et al., 2012; Einheuser et al., 2013; Maret et al., 2010; Waite et al., 2010), and the Principle Component Analysis Feature Extraction methods (Khalid et al., 2014; Pearson, 1901). Other popular machine-learning-based variable reduction techniques have been applied to model portions of the biophysical system: the ant colony optimization (Dorigo and Di Caro, 1999) was used for selecting the best intrinsic mode functions in forecasting monthly solar radiation (Prasad et al., 2019), the bio-inspired Bat algorithm was used to improve features estimated for forecasting monthly rainfall (Ali et al., 2018b), the Simulated Annealing algorithm (FAO, 2017) was used in the development of a drought model (Ali et al., 2019), the singular value decomposition algorithm (Bretherton et al.,

1992) was used to develop a model for forecasting weekly solar radiation (Prasad et al., 2020), and the non-dominated sorting genetic algorithm (Deb et al., 2002) was used in the development of a long-term precipitation model (Ali et al., 2020).

In summary, crop models are widely used to predict plant growth, water input requirements, and yield. However, the existing models are very complex and require hundreds of variables to perform accurately. Due to these shortcomings, the large-scale applications of crop models are limited. Machine learning techniques, such as DL, can be used to address some of these limitations; however, their applications are currently limited to qualitative assessments, such as computer vision and speech recognition (Liu et al., 2017). Here, we tried to address the existing problems with crop models by not only developing a DL model for predicting yield and water requirements, but also by reducing the number of input variables from hundreds to only a few. This way, the DL crop models can be used for real-time and large-scale applications, which are not currently possible. In this study, we evaluated the reliability of the DL crop model using about 10 million tested scenarios, which to the best of our knowledge has not been attempted before. This effort will mainstream crop model applications for predicting yield and water use at the regional and national scale. Crop models can help guide policymakers toward achieving sustainable water and food security in the 21st century.

The paper is organized as follows: In section 2, first, the input variables (e.g., precipitation) and output variables (e.g., crop yield) for the development of DL crop models were described. Next, the architecture for the DL model, along with training and testing procedures, were explained. Due to the high number of the input variables, three different techniques were tested for variable reduction efficiency without compromising the model accuracy. Finally, the best DL models developed by the variable reduction techniques were compared with a newer feature selection method. Under section 3, a DL crop model was developed using all input variables (800). This model was used as a reference to examine the performance of variable reduction techniques. Systematically, the number of variables was reduced until the model accuracy was compromised. The smallest structure for crop DL models with comparable accuracy to the reference model were then identified and the performance of the variable reduction method was evaluated against other commonly used methods. Finally, under section 4, the results of the analysis were synthesized to identify the best approach for the development of accurate and reliable DL crop models.

2. Materials and methods

2.1. Overview of methodology

Fig. 1 presents an overview of this study. First, 100 weather scenarios were combined with 100,000 random irrigation applications to create 10,000,000 scenarios in which crop production could be examined using a crop model. Next, a DL model was trained and tested based on these 10 million scenarios. Eight hundred input variables were introduced within the DL model, including four varying environmental variables during a 200-day crop growing season. Three commonly used variable reduction techniques were then used to develop additional deep learning models based on the reduced number of input parameters and smaller structures. These models were tested and compared with the original deep learning model to identify the best new models with the lowest number of input variables. The impact of variable reduction methods on the performance of DNN models with various architectures were discussed at length. In each round of the experiment, the number of input variables was reduced (400, 200, 100, 50, 40, 30, 20, and 10 variables), and the model's DNN structure was downsized (combinations of 600, 400, 200, 100, 50, 40, 30, 20, 10, 8, 6, and 4 neurons by 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2, and 1 layers). The process was iterated until a minimum number of architectures with comparable accuracy to the original model was identified. Each computed model was identified by three numbers in the form of Inputs-Neurons-Layers. For example, model 50–40-6 indicates a DNN with 50 input variables, 40 neurons per layer, and 6 layers. Finally, the best DNN model was compared against a new DNN model developed using a recent feature selection method.

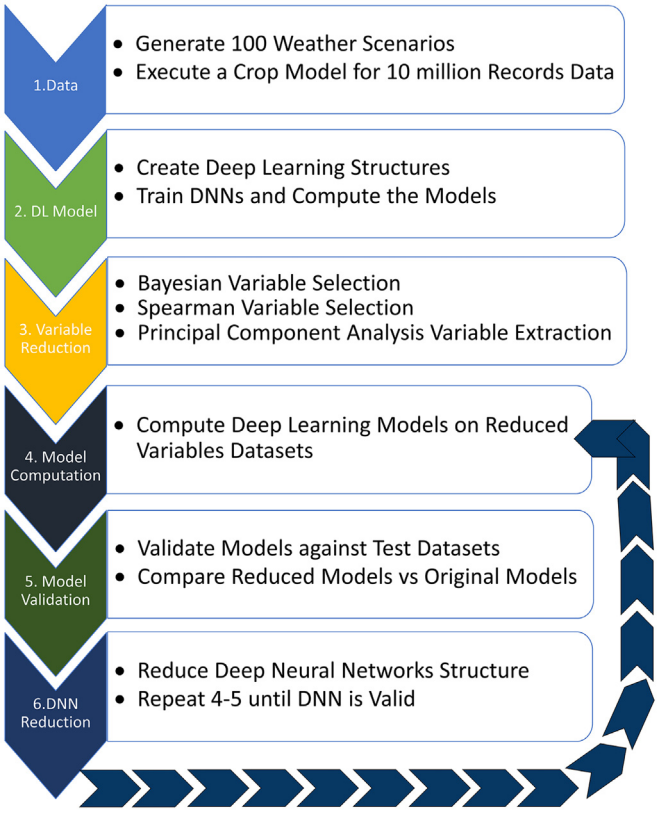


Fig. 1. Overview of the procedures for generating reduced deep neural network structures for crop yield estimations.

2.2. Model input variables

In this study, several environmental input variables were considered, including precipitation and irrigation, maximum temperature, minimum temperature, and solar radiation. 800 variables were introduced for a 200-day growing season. Irrigation was added in combination with precipitation to generate the total water applied in mm per day. To generate the model's training data, 100 different climate scenarios were considered. These climate scenarios were combined with 100,000 random irrigation scenarios and applied to a crop model, which generated 10,000,000 records of maize yields. Data related to climate variabilities were produced by weather generators as described below.

2.2.1. Generating weather data

Providing daily precipitation, temperature, and solar radiation in addition to irrigation amounts, is essential for assuring the accuracy of crop model outputs. One-year field weather data was collected and used to generate different weather realization scenarios. The monthly temperature, solar radiation, and precipitation were collected from the nearest weather station to the study site (PRISM, 2011). A weather stochastic disaggregation tool was used to generate daily weather information from the environmental variables' monthly historical records (Hansen and Ines, 2005). This tool generated daily weather information by disaggregating average monthly data from historical records. To accurately simulate weather in the study area, 30 years of data were

used in the weather generator. The weather generator uses stochastic approaches to generate daily information for the locations of interest. In regards to precipitation, three conditions above the average, the average, and below the average levels of precipitation are considered. Precipitation occurrence is modeled by the Markov chain and the amount is sampled from hyperexponential distribution or probability density function of the random variable x (precipitation occurrence) given by Eq. (1).

$$f(x) = \alpha \frac{\exp\left(-\frac{x}{\beta_1}\right)}{\beta_1} + (1-\alpha)\frac{\exp\left(-\frac{x}{\beta_2}\right)}{\beta_2} \quad (1)$$

where: α is the mixing probability of hyperexponential distribution, and β_1, β_2 are means of the i^{th} component of a hyperexponential rainfall intensity distribution.

In addition, maximum and minimum temperatures are sampled from a Gaussian distribution function conditioned on the occurrence of precipitation. To generate solar radiation, the weather generator used the *logit* function transformations of daily clearness rescaled between upper and lower limits of field solar radiation Eq. (2):

$$logit(p) = \ln(p/(1-p)) \quad (2)$$

where, p is a probability of daily clearness.

To consider statistical behavior of daily sequences, the amount was shifted to match monthly average values of solar radiation and temperatures from historical records.

2.2.2. Data preparation

After removing duplicate data from the 10,000,000 generated records, 8,970,685 unique maize production data remained (Table 1). These data were used for further developing the DL model. Despite the fact that two computers were used in the study, a 24 core of Intel® Xeon® CPU E5-2680 v3 @ 2.50GHz with Quadro M6000 GPU and an Intel® Core™ i7-4770 CPU @ 3.40GHz with GeForce GTX 1080 GPU, the GPU memory did not allow for all available records to be used in training the DNN. To reduce the number of records and maintain diversity, the data were categorized by maize yield into 12 categories, each 1000 kg apart, then sampled to create 10 similar population distribution datasets. Table 1 shows the number of data records in each dataset.

2.3. Study area

Irrigation Research Park in Gainesville, Florida, was selected as the study area for analyzing the effect of irrigation on maize yield. The Irrigation Research Park is located at the experimental station of the University of Florida (29° 37′ 8″ N, 82° 22′ 22″ W). The humid subtropical climate with abundant rainfall after the growing season makes this location suitable for quantitative studies of water management (Lascody and Melbourne, 2002). Less than 30% of the annual rainfall occurs from February through May, which is the main growing season (Fig. 2). The 30-year minimum and maximum temperatures of the location were 14.3 °C and 26.7 °C, respectively. On average, the highest temperature ranged from 32.7 °C in July to a low of 19 °C in January. Millhopper fine sand was the major soil type found in this region, known to drain moderately well.

2.4. Crop modeling

The crop model selected for this study is the Decision Support System for Agrotechnology Transfer (DSSAT). This model is designed to dynamically model over 40 different crops and has been widely used in the past 30 years by researchers and academic institutes worldwide (Hoogenboom et al., 2015; Jones et al., 2003a, 2003b; Nurudeen, 2011). According to the DSSAT website (Hoogenboom et al., 2019), the model has been used by more than 14,000 researchers, educators, consultants, extension agents, growers, and policy/decision-makers in over 150 countries. DSSAT's software application package includes soil, weather, crop management tools, and experimental data. DSSAT simulates and models crop growth, development, and yield as a function of the soil, weather, and plant dynamics.

To obtain data for the purposes of DNN model training, a maize irrigation experiment was setup in DSSAT and calibrated based on the results from the experimental study site at the Irrigation Research Park. The growing season comprises 200 days (Hoogenboom et al., 2015). The DSSAT input files were setup using maize cultivar (McCurdy 84aa) with planting and harvesting dates of February 16 and May 7, respectively. Ten irrigation applications were selected to generate random scenarios within the growing season to further train the DNN model irrigation applications. Irrigations occurred within the growing season, with amounts ranging from 10 mm to 250 mm of water per day.

2.5. Deep learning architecture for crop modeling

Extensive knowledge of climate, geology, and agricultural management practices is needed to accurately operate typical crop models. In addition, the application of these types of models on a large-scale is limited by model complexity. To address these limitations, this study was aimed to show the potential for applying DL techniques to crop models. The DL architecture that was used in this study was a DNN with a Multi Layers Perceptron (MLP) architecture (Fig. 3). MLP is a feed-forward DNN and was selected for this study, since it has been shown to successfully generate solutions for classification problems (Deng and Yu, 2014).

In this large-scale analysis, we were interested in estimating the production class rather than the actual yield due to the high level of uncertainty for individual fields. To make the production estimation more reliable, the maize production level was grouped into 12 classes (ranges from 0 to 12,000 kg/ha), with each class of input representing a range of 1000 kg/ha yield. For example, class 0 represents 0 to 1000 kg/ha, and

Table 1
Nonduplicated maize yield datasets for different climatological and irrigation scheduling that were used for deep learning crop model development.

Yield class	Data set 1	Data set 2	Data set 3	Data set 4	Data set 5	Data set 6	Data set 7	Data set 8	Data set 9	Data set 10
0–1000	46,872	46,919	47,131	46,985	46,664	46,718	46,680	46,861	46,707	44,936
1000–2000	99,136	99,251	99,783	100,049	99,829	99,882	99,719	99,821	100,204	96,568
2000–3000	102,770	102,178	102,101	102,235	102,288	102,187	102,320	102,286	101,961	99,032
3000–4000	167,257	168,053	167,320	167,595	167,256	167,508	167,786	166,691	167,349	161,942
4000–5000	114,608	114,090	114,052	113,762	114,278	114,607	113,753	114,485	114,531	110,320
5000–6000	85,792	86,493	86,807	86,388	86,571	86,460	86,144	86,361	86,183	83,332
6000–7000	99,090	99,033	99,108	99,439	99,650	98,712	99,751	99,675	99,294	96,510
7000–8000	62,626	62,559	61,742	62,081	61,887	62,443	62,292	62,502	62,201	60,065
8000–9000	61,518	60,911	61,190	61,082	60,970	60,798	61,037	60,995	61,254	59,276
9000–10,000	24,199	24,258	24,296	24,222	24,551	24,302	24,178	24,181	24,305	23,421
10,000–11,000	30,280	30,394	30,753	30,377	30,232	30,574	30,539	30,385	30,303	29,622
11,000–12,000	5852	5861	5717	5785	5824	5809	5801	5757	5708	5661
Total	900,000	900,000	900,000	900,000	900,000	900,000	900,000	900,000	900,000	870,685

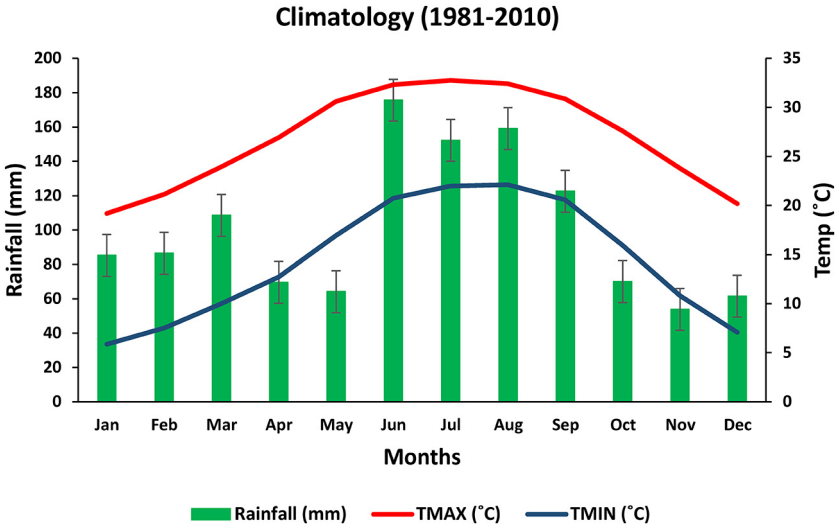


Fig. 2. The 30-year (1981–2010) average weather conditions at Gainesville, Florida.

class 1 represent 1000 to 2000 kg/ha. Different combinations of hidden layers with neurons (computation nodes) within each layer were used in the DNN to form the crop models. Activation functions of the same type were assigned for all neurons. The TanH (tangent hyperbolic) activation function was used for all hidden layers and the SoftMax (normalization constraint on the total output probability function) activation function was used for the output layer (Costa, 1996). Once the DNN structure was created, training is necessary and was performed as detailed in the next section.

2.6. Training and validation of the models

All models were trained using the first dataset of 900,000 records and 10% of the dataset (90,000 records) were used for validation throughout the training process. The backpropagation (Liu et al., 2017) method and the gradient descent (Baldi, 1995) algorithm were also used to train the network by minimizing a defined cost (Baldi, 1995). In this study, the negative log-likelihood equation Eq. (3) was used as a cost function (Friedman, 2002). The learning rate was fixed at 0.01, and the mini-batch size was 1000 records in all training runs. After reaching the lowest validation error, training was continued for 100 epochs to ensure that it was not trapped in the local minimum.

$$Cost = \frac{1}{n_{bz}} + \sum_{k=0}^{n} log\left(P(Y = y_k|\ \cos\frac{n\pi x}{L} + b_n\ \sin\frac{n\pi x}{L}\right) \tag{3}$$

where, n_{bz} is the mini-batch size, n is the number of output classes (12 in this case), and P is a function of likelihood probability calculated by Softmax (normalization constraint on the total output probability function) (Costa, 1996).

To compute the required time for each epoch in the DNN structure with l layers and n neurons per layer, it was assumed the total calculation time of each layer in parallel for the feed-forward and feed-backward process on the GPU is T_l second. To feed the next layer, the neuron's output between GPU memory and machine main memory was transferred in T_n second. Considering the training process, Eq. (4) calculated the time consumption for one epoch.

$$T_{epoch} = l \times T + n \times l \times T_n \tag{4}$$

Eq. (4) demonstrated that training time had a direct relationship with the number of layers and the number of neurons per layer. By reducing the number of layers and the number of neurons per layer, the training process became faster.

2.7. Variable reduction

To improve training performance and reduce the computational power required for developing the DL models. Among the numerous variable reduction techniques, some of the most commonly used techniques in environmental and agricultural studies were applied (Woznicki et al., 2015). The following three variables reduction methods were evaluated for preprocessing the data.

(1) Bayesian Variable Selection (O'Hara and Sillanpää, 2009)

In theory, a Bayesian model explains a response variable (output) with a (large) number of explanatory variables (inputs). The Bayesian Variable Selection method selects a small subset of variables that can be inferred and used to explain a large fraction of the variation present in the response. In many cases, the variable selection is done by specifying the variables; the variable selection task is to estimate the marginal posterior probability of whether the variable should be included in the model or not (O'Hara and Sillanpää, 2009). Several Bayesian Variable Selection software tools are currently available, such as BayesFactor, BayesVarSel, and BMS (Forte et al., 2018); however, none of them is suitable to work with a large dataset. As a result, we identified the Bayesian Generalized Linear Regression (BGLR) software (Pérez and de los Campos, 2014) that can work with big data. The BGLR is an R-based statistical package based on the Gibbs sampler technique with scalar updates to reduce the number of input variables (Casella and George, 1992).

(2) Spearman Variable Selection (Zwillinger and Kokoska, 1999)

Spearman Rank Correlation method calculates the ranking correlation between each variable and the output (Eq. (5)). In this method, it was assumed that the variables with higher correlations would have a greater impact on outputs and should be considered first in the model.

$$r_s = \frac{SS_{uv}}{\sqrt{SS_{uu} \times SS_{vv}}} \tag{5}$$

where, r_s presents the Spearman Rank Correlation coefficient, SS_{uv} is the covariance of the input and the output variables, respectively, SS_{uu} and SS_{vv} are standard deviations of input and output variables.

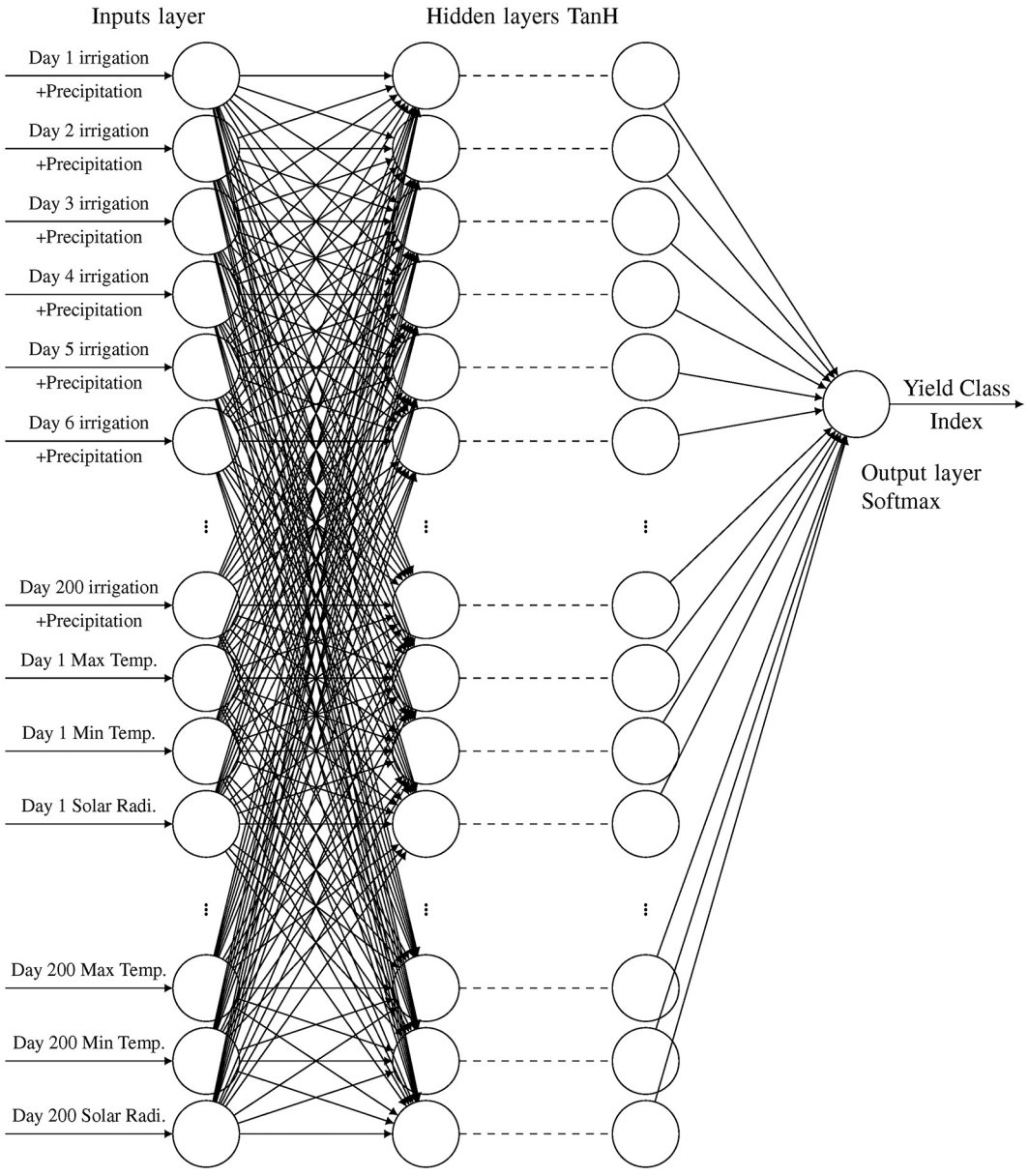


Fig. 3. Schematic representation of the deep neural network with multi-layer perceptron architecture that was used for crop yield estimations.

A python library from scipy.org (scipy.stats.spearmanr) was used to calculate the Spearman Rank Correlation coefficient (Zwillinger and Kokoska, 1999) between inputs and outputs for variable selection purposes.

(3) Principal Component Analysis Feature Extraction (Khalid et al., 2014)

Principal Component Analysis (PCA) feature extraction is an orthogonal transformation for converting correlated variables to a smaller set of uncorrelated variables. PCA feature extraction method uses eigenvalues of $X^T X$ to calculate linear transformations between these two sets. This method is called feature/variable extraction (Khalid et al., 2014).

In this study, a python library (sklearn.decomposition.PCA) was used to perform this analysis. This library used the LAPACK implementation of the full Singular Value Decomposition (SVD) or a randomized truncated SVD, a method introduced by Halko et al. (2011).

2.8. Feature (variable) selection base on max-relevance and min-redundancy

Min-redundancy and max-relevance (mRMR) (Menger et al., 2018) is a novel and popular method that was initially introduced by Peng et al. (2005) and then improved by Bugata and Drotar (2020). The mRMR algorithm selects a set of explanatory variables with the highest relevancy and the lowest redundancy level to describe the output variable. Consequently, the most dependent variables are identified in a large set of variables, which ultimately results in better classification. Through reduced input variable redundancy, a smaller model with equal or better performance can be obtained.

Research by Bugata and Drotar (2020) showed that the max-redundancy is not always equivalent to max-dependency as was assumed by Peng et al. (2005). Therefore, Bugata and Drotar (2020) suggested that by adding an objective to the algorithm to maximize dependency, the overall performance of the mRMR algorithm can be improved. In fact, applying the revised mRMR algorithm resulted in

better feature selection under a real-world scenario (Bugata and Drotar, 2020).

3. Results of discussion

3.1. Deep neural network structure analysis

3.1.1. Evaluation of the original deep learning model's accuracy with 800 input variables

The original DL model was trained and tested using 800 inputs. The inputs consisted of four environmental variables varying over a 200-day crop growing season and all models were trained on the first 900,000 recordsets. Fig. 4 and Table S1 (Supplementary Materials) present the accuracy of the original DL models with different structures. The prediction accuracy results are shown in the "Training Set" columns and the average accuracy rates from the nine test sets are shown in the "Test Sets" columns. The results showed that in smaller DNN structures (50 to 100 neurons per layer), more layers resulted in lower model accuracy. This behavior was previously observed by Schmidhuber (2015), who noticed decreasing accuracy from exploding/vanishing gradients. The accuracy rates in the DNNs with small structures were significantly lower (30% to 70%) than those with larger structures (400 to 1000 neurons per layer) (75% to 80%), which can be seen in Fig. 4. This behavior shows that a DL model based on a large dataset with 800 inputs variables needs a large structure to perform reasonably well.

In this study, the accuracy reduced at a lower rate in DNNs with more than 400 neurons per layer compared to smaller DNNs with an increasing number of layers. However, larger structures require more computational time and more powerful hardware to perform at an acceptable level. It should be noted that accuracy rates in training sets and test sets were identical, indicating that the models were not overfitted.

3.1.2. The number of input variables for deep learning models was reduced to 400

Table S2 and Fig. 5 show the accuracy training and test sets of the DNN model predictions with 400 input variables. Evidently, the result with the Bayesian and Spearman Rank Correlation methods in the training set and the average of the nine test sets were almost identical. The similarity indicated that the models were not overfitted. Furthermore, the PCA feature extraction method displayed greater accuracy with the training set but poor performance with the test sets. As an unsupervised feature extraction method, PCA identifies and extracts the least correlated features from the input dataset (Calesella et al., 2021). Meanwhile, the developed DL models based on this extraction method were overfitted during the training process.

Additionally, Fig. 5 shows that the 400 neuron DNN with many layers had a lower accuracy with the Bayesian Variable Selection method. Similar results were observed with the Spearman Rank Correlation, but the Bayesian Variable Selection method showed a decrease in accuracy when the number of layers increased, while the Spearman Rank Correlation method did not. This indicated that the Spearman Rank Correlation method was less affected by the vanishing/exploding gradients issue. The PCA feature extraction method showed almost no sensitivity to an increasing number of layers. Although it reached a significant prediction accuracy (more than 97%) with the training set, poor performance on test sets showed the model was highly overfitted. Bayesian Variable Selection and Spearman Rank Correlation methods had the same prediction accuracy on both training and test sets of DNNs with 600 neurons per layer. For the PCA feature extraction method, the prediction accuracy on training sets were increased on 600 neurons per layer, demonstrating how the method became less reliable due to overfitting with larger DNN structures.

3.1.3. The numbers of input variables for deep learning models were reduced to 200,100 and 50 number

Table S3 and Fig. S1 (Supplementary Materials) presented the DNN models' accuracy with 200 input variables. Table S4 and Fig. S2 presented the DNN models' accuracy with 100 input variables. Almost identical behaviors to 400 input variables were observed with 200 and 100 variables. One hundred input variables (Fig. S2) indicated a flat accuracy amount for all DNN structures because the number of neurons was twice that of the number of inputs. As a result, the DNN was more flexible and compensated for vanishing/exploding gradient issues. The results from the Bayesian Variable Selection and the Spearman Rank Correlation methods showed comparable accuracy levels for the DNN models with similar architectures (79%). Meanwhile, the DNN models using the PCA feature extraction method had the highest accuracy for the training set (over 98%) and significantly lower accuracy for the test sets (less than 39%). This finding indicated the model was overfitted.

Fig. S1 shows a decrease in accuracy in relation to an increase in the number of layers in the DNN with 200 neurons per layer. However, for 400 and 600 neurons per layer, the accuracy rate remained almost constant for all three variable reduction methods. It can be observed in Fig. S2 that 100 input variables' accuracy rates were constant for all the different DNN structures except for the PCA feature extraction method, which had a comparable accuracy rate to the 400 inputs models. This illustrates that improvements in accuracy were established by increasing the number of layers in the training set without improving test set performance.

To study the effect of additional input variable reductions, some new models with only 50 input variables were computed. Table S5 and

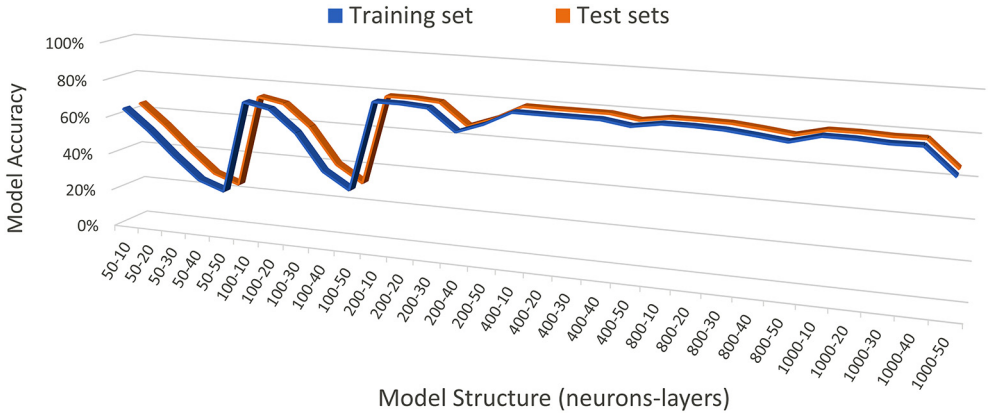


Fig. 4. Deep neural network original models' accuracy with 800 input variables under different model structures.

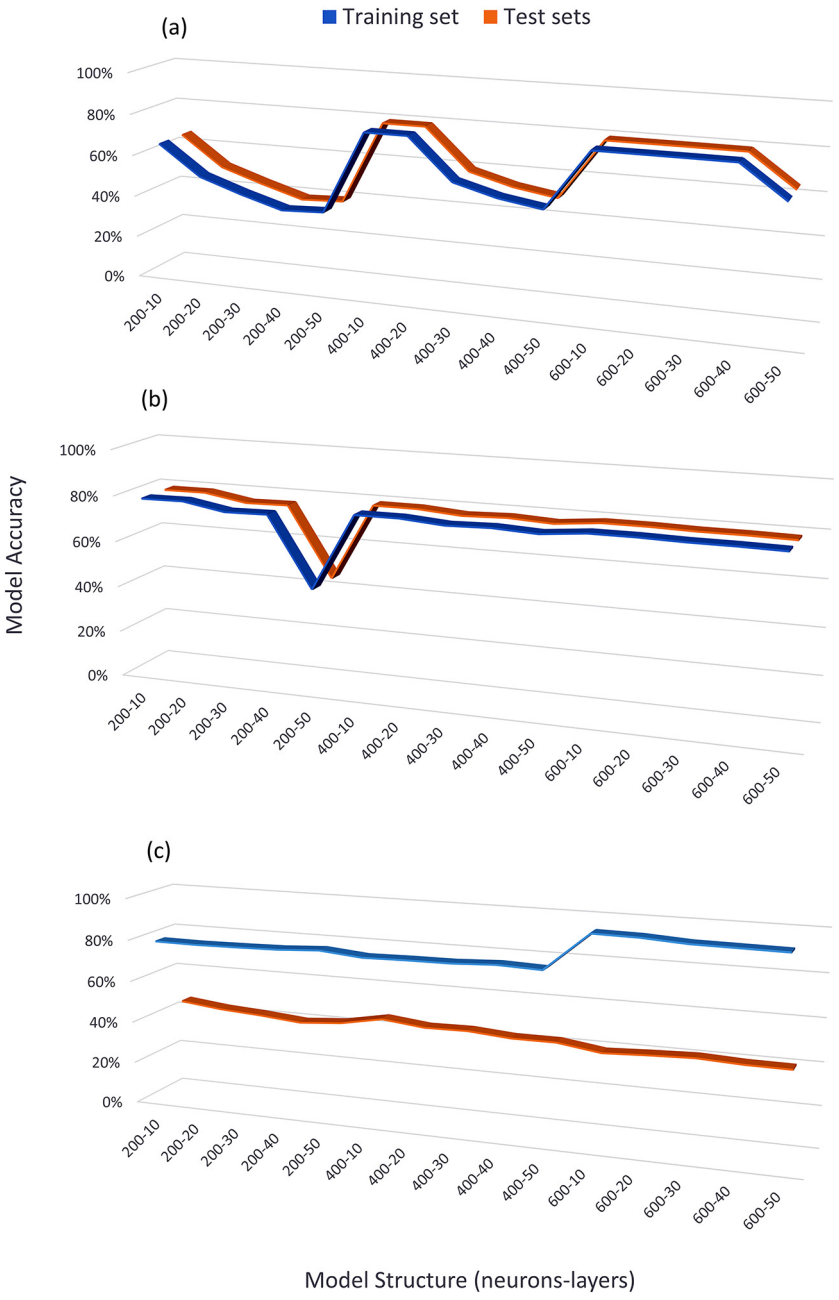


Fig. 5. Deep neural network models' accuracy with 400 input variables (a) Bayesian (b) Spearman (c) PCA.

Fig. S3 show the accuracy validation result for these models on the training sets and test sets with 50 input variables, which turned out to be identical to the results of 100 input variables. This result shows that the three variable reduction methods successfully reduced input variables and can be expected to produce models with small DNN structures at acceptable prediction accuracy levels. In the next step, the DNN structure was reduced in order to find out if these methods would work with smaller DNN structures.

3.1.4. Evaluating the performance of deep learning models with limited layers

To understand the impact of the number of hidden layers and variable reductions on the DNN model's accuracy, the number of layers was limited to less than 10 (1 to 9) for three different sets of neurons (i.e., 200, 400, and 600) and 200 input variables (Table S6).

Fig. S4 indicates that with 200 inputs and the proper number of neurons per layer, the accuracy with 1 hidden layer is almost the same as the large DNN structures with several layers. This shows the robustness of all three chosen variable reduction methods on the system. In other words, a shallow DNN (with less than 10 layers) with a higher number of neurons per layer (200, 400, and 600) performed as well as a deep DNN (with more than 10 layers) with several hidden layers (10, 20, 30, 50 hidden layers). However, similar to previous results, the models developed based on the PCA feature extraction method had excellent accuracy on the training sets and poor performance on the test sets.

3.1.5. Evaluate the impact of the number of neurons and layers on deep learning model accuracy

In this stage, the performance of several models was computed with a small number of neurons per layer. Table S7 and Fig. S5 show the accuracy result for the DNN models with 50 input variables. The number

of neurons per layer in these models was 4, 6, 8, and 10, while the number of layers was 1 through 9. In general, the overall accuracy of the DNN models with a low number of neurons (e.g., ten neurons in Fig. S5) was lower than the accuracy of the models with more neurons (e.g., 200 neurons in Fig. S4). Meanwhile, the DL models' accuracy fluctuated when the number of layers increased for the same number of neurons. However, when the data was rearranged (Fig. S6), the fluctuation patterns became more apparent in which the accuracy increased as the number of neurons increased for the same number of layers.

Meanwhile, DL models based on the Bayesian and Spearman variable reductions methods are more consistent in performance between testing and training sets, while the DL models developed using the PCA variable reduction method are less robust. For example, in section C of Fig. S5 the test set shows a considerable drop in accuracy compared to the training set.

Finally, regardless of the variable reduction methods and the number of layers, the accuracy of the DL model is higher as the number of neurons increases (Fig. S5).

3.2. Identifying the smallest DNN model with good accuracy

Large DNN structures require significant time for training their network (Sun et al., 2019). Therefore, the goal of this section was to identify the minimum number of input variables that can be used to develop a DL model with similar accuracy as the original model, with 800 input variables. Here we considered four models with 10, 20, 30, and 40 variables. Table 2 shows models by 10 to 40 inputs and 1 to 9 layers prediction accuracy results. The highest accuracy for all three variable reduction methods was obtained using the DNN with 10 neurons per layer, as highlighted in Table 2. Among these three methods, only the DL model based on the Bayesian variable reduction method was successful in producing reliable data comparable to the original model, with 800 input variables.

The model with 30 inputs, 10 neurons, and 5 layers (50 computational units) created by the Bayesian Variable Selection method had the same accuracy as a model with 800 inputs, 400 neurons, and 10 layers (Table S1 with 4000 computational units). This suggests that the DNN structure developed using the Bayesian Variable Selection method, which was 80 times smaller, with 1/27 the number of input variables, could have the same accuracy as the larger model. However, it took three weeks to calculate the posterior probability vector using 900,000 simulated crop yield datasets. This hurdle makes the Bayesian Variable Selection method less useful for fast applications.

The Spearman variable selection method, with more than 200 input variables, was acceptable with larger DNN structures since they had the same accuracy as the Bayesian Variable Selection model with 200 neurons per layer. However, the Spearman produced model reduced the prediction accuracy when the number of neurons per layer was reduced. This method was found to work well with larger input variables and DNN structures compared to the Bayesian Variable Selection method. In addition, the processing time for variable selection was much faster (~10 min) than the Bayesian Variable Selection method.

The PCA model consistently had the best results on the training set and poor performance on test sets with larger DNN structures showed inferior performance on training sets and test sets with small DNN structures. The accuracy of the smallest model's DNN structure for each method of variable reduction compared to the original 800 input variable models can be seen in Table 3.

One of the main benefits of variable reduction is minimizing the structure of DNN and, as a result, reducing prediction run time. To

Table 2
Deep neural network models' architectures and accuracy for the training and test sets.

Inputs	Neurons	Layers	Bayesian		Spearman		PCA	
			Training Set	Test Sets	Training Set	Test Sets	Training Set	Test Sets
10	10	1	53.73%	53.72%	65.82%	65.88%	27.54%	22.41%
20	10	1	71.86%	71.83%	69.15%	69.11%	32.31%	18.34%
30	10	1	77.84%	77.85%	54.28%	54.32%	35.63%	21.18%
40	10	1	78.52%	78.52%	71.02%	70.94%	38.64%	16.81%
10	10	2	68.89%	68.86%	60.08%	60.12%	29.61%	23.20%
20	10	2	76.88%	76.87%	56.80%	56.82%	35.45%	18.92%
30	10	2	78.53%	78.52%	61.18%	61.18%	36.88%	20.64%
40	10	2	78.55%	78.54%	69.46%	69.40%	39.54%	17.21%
10	10	3	68.95%	68.91%	59.19%	59.28%	31.24%	23.83%
20	10	3	76.81%	76.80%	54.05%	54.02%	35.75%	19.60%
30	10	3	78.57%	78.56%	54.24%	54.16%	37.54%	21.04%
40	10	3	78.57%	78.56%	52.14%	52.08%	40.70%	17.16%
10	10	4	69.20%	69.18%	61.28%	61.33%	32.44%	24.79%
20	10	4	77.03%	77.01%	56.49%	56.55%	36.28%	19.95%
30	10	4	78.58%	78.57%	54.11%	54.15%	38.96%	20.67%
40	10	4	78.57%	78.57%	56.78%	56.74%	40.35%	17.77%
10	10	5	69.93%	69.91%	48.66%	48.65%	33.21%	25.41%
20	10	5	76.94%	76.92%	51.80%	51.76%	37.44%	19.67%
30	10	5	78.59%	78.58%	45.41%	45.43%	39.61%	20.99%
40	10	5	78.57%	78.57%	50.89%	50.82%	41.34%	17.07%
10	10	6	67.81%	67.84%	50.39%	50.41%	33.44%	24.93%
20	10	6	77.11%	77.10%	55.79%	55.76%	37.42%	19.32%
30	10	6	78.57%	78.56%	41.60%	41.69%	39.83%	21.10%
40	10	6	78.56%	78.55%	48.09%	48.08%	41.11%	18.24%
10	10	7	67.00%	66.95%	42.95%	42.96%	33.77%	26.10%
20	10	7	77.02%	76.98%	47.74%	47.75%	37.56%	19.85%
30	10	7	78.57%	78.57%	50.02%	49.96%	39.64%	20.68%
40	10	7	78.57%	78.56%	48.63%	48.60%	41.91%	16.88%
10	10	8	67.02%	67.03%	45.74%	45.80%	33.98%	26.14%
20	10	8	77.03%	77.01%	49.00%	48.96%	38.26%	20.14%
30	10	8	78.58%	78.57%	43.69%	43.70%	40.12%	21.12%
40	10	8	78.54%	78.54%	51.99%	52.00%	41.38%	17.57%
10	10	9	67.27%	67.24%	45.49%	45.61%	34.05%	25.69%
20	10	9	76.96%	76.94%	39.34%	39.34%	38.38%	19.46%
30	10	9	78.58%	78.57%	37.95%	37.97%	40.09%	21.54%
40	10	9	77.70%	77.70%	55.71%	55.70%	42.21%	16.98%

Table 3
The accuracy of the smallest deep neural network structure for the three variable reduction methods compare to the original model with 800 inputs.

DNN structure			Maximum accuracy		
Models	Inputs	Neurons	Layers	Training Set	Test Sets
800 inputs	800	400	10	78.58%	78.54%
Bayesian	30	10	5	78.59%	78.58%
Spearman	40	10	1	71.02%	70.94%
PCA	40	10	9	42.21%	26.14%

measure this benefit, the original model (800 inputs, 400 neurons, 10 layers) and the smallest DNN model (30 inputs, 10 neurons, and 5 layers) that show the same level of accuracy have been run on 90,000 records on the same CPU platform to measure their execution time. As a result, the original model with 800 inputs took 166 s to run and the small model based on the Bayesian Variable Selection method took 0.86 s, which is 193 times faster. A decrease in the overall runtime of the DNN crop model can help with the application of these types of models at the large scale that is necessary for policymakers to make an informed decision at the national and international levels.

3.3. Comparing by mRMR

In this section, the mRMR method was used to identify a set of variables from the 800 inputs to develop the DNN models comparable with the most accurate model developed by the Bayesian variable selection method. Fig. 6 shows the comparison between these two methods. The DL models computed by mRMR are usually around 10% less accurate than the similar models developed based on the Bayesian variable selection method. The lower accuracy established a clear superiority of the Bayesian variable selection methodology to the mRMR.

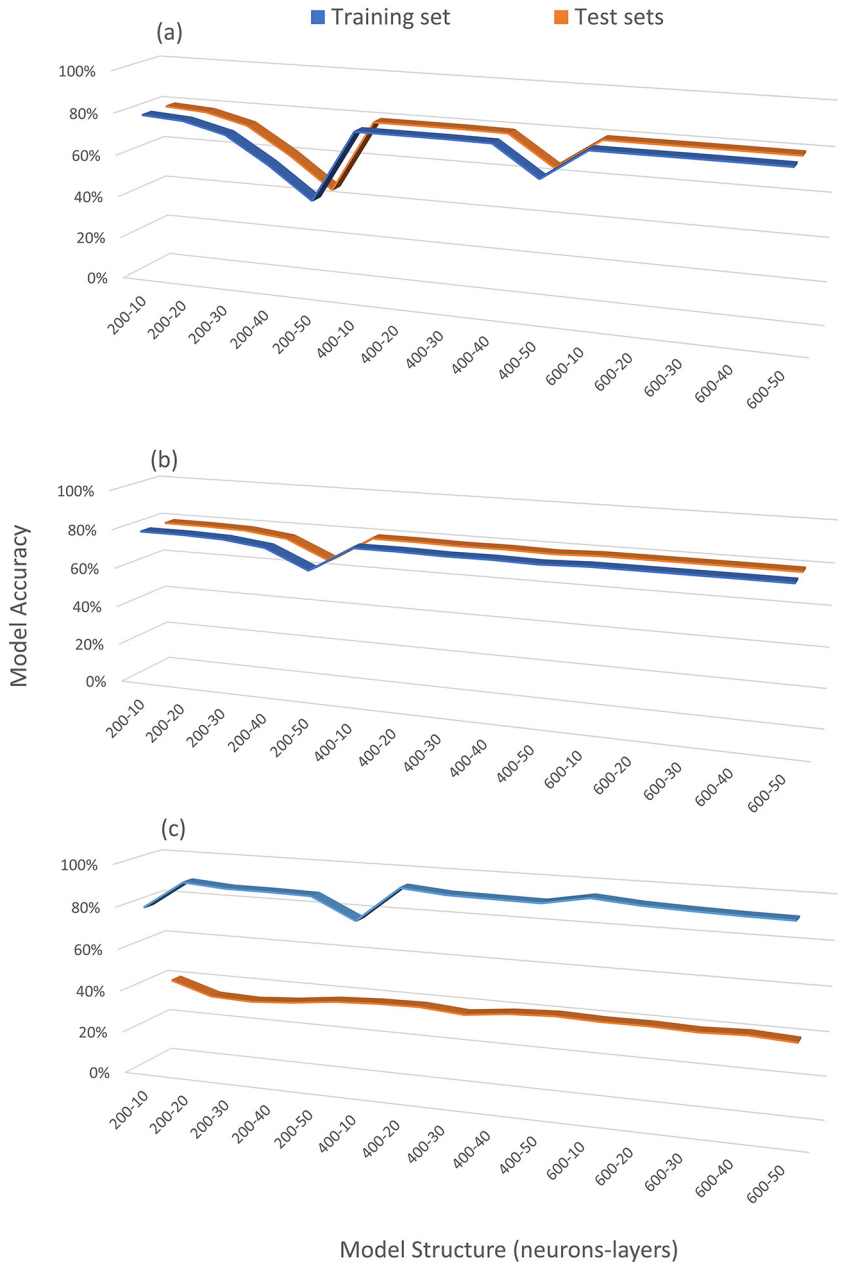


Fig. 6. Comparison between min-redundancy max-relevance and Bayesian model accuracy.

4. Conclusion

A cropping system modeled by DL with many input variables produces a large DNN structure, which is very computationally intensive. Additionally, the training process required for large DNNs can be very time-consuming. This shortcoming has limited their usage for large-scale applications. However, the utilization of an efficient deep learning crop model can be a game-changer. For example, a regional-scale irrigation scheduling system can be developed using a hybrid system comprised of DL-based crop models, a weather forecasting system, and an optimization algorithm.

In this study, we examined the possibility of developing a simpler deep learning model with comparable accuracy through the application of different variable reduction methods (i.e., Bayesian Variable Selection, Spearman Rank Correlation, and PCA variable extraction method).

The Bayesian Variable Selection method was identified as the most robust method of the three evaluated in this study. However, calculating the posterior probability for each variable is very time-consuming. The Spearman Rank Correlation was ranked the second best with similar accuracy to the Bayesian Variable Selection method. The performance of these models were also examined against the recently improved mRMR technique. In general, the models based on the mRMR technique are less accurate than the ones based on the Bayesian and the Spearman Rank Correlation techniques. Finally, the DNN models that were developed based on the PCA feature extraction method had the highest accuracy during the training tests, but the lowest levels during testing. Unfortunately, almost all DNN models developed by this technique were overfitted.

Even though the Bayesian Variable Selection method was selected as the best for this study; however, future studies are necessary for analyzing the robustness of this method and reduce uncertainty by utilizing ensembles of crop simulation models under different irrigation schemes, climatological conditions, and crop management strategies. Ultimately, the results of this work can then be compared to determine the best selection method for a wide variety of crops/regions.

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Acknowledgments

The authors would like to thank Ms. Anna Raschke for her review and constructive feedback. This work was supported by the USDA National Institute of Food and Agriculture, Hatch project 1019654.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aiia.2021.09.001.

References

Abbasi, A.Z., Islam, N., Shaikh, Z.A., 2014. A review of wireless sensors and networks' applications in agriculture. Comput. Stand. Inter. 36 (2), 263–270.

Aggarwal, P.K., Kalra, N., Chander, S., Pathak, H., 2006. Infocrop: a dynamic simulation model for the assessment of crop yields, losses due to pests, and environmental impact of agro-ecosystems in tropical environments. i. Model description. Agric. Syst. 89 (1), 1–25.

Ahuja, L., Ma, L., 2011. A synthesis of current parameterization approaches and needs for further improvements. Methods of Introducing System Models into Agricultural Research. 2, pp. 427–440.

Ali, M., Deo, R.C., 2020. Modeling wheat yield with data-intelligent algorithms: artificial neural network versus genetic programming and minimax probability machine regression. Handbook of Probabilistic Models. Butterworth-Heinemann, pp. 37–87.

Ali, M., Deo, R.C., Downs, N.J., Maraseni, T., 2018a. Cotton yield prediction with Markov Chain Monte Carlo-based simulation model integrated with genetic programing algorithm: a new hybrid copula-driven approach. Agric. For. Meteorol. 263, 428–448.

Ali, M., Deo, R.C., Downs, N.J., Maraseni, T., 2018b. Multi-stage hybridized online sequential extreme learning machine integrated with Markov Chain Monte Carlo copula-Bat algorithm for rainfall forecasting. Atmos. Res. 213, 450–464.

Ali, M., Deo, R.C., Maraseni, T., Downs, N.J., 2019. Improving SPI-derived drought forecasts incorporating synoptic-scale climate indices in multi-phase multivariate empirical mode decomposition model hybridized with simulated annealing and kernel ridge regression algorithms. J. Hydrol. 576, 164–184.

Ali, M., Deo, R.C., Xiang, Y., Li, Y., Yaseen, Z.M., 2020. Forecasting long-term precipitation for water resource management: a new multi-step data-intelligent modelling approach. Hydrol. Sci. J. 65 (16), 2693–2708.

Amara, J., Bouaziz, B., Algergawy, A., 2017. A deep learning-based approach for banana leaf diseases classification. Datenbanksysteme für Business, Technologie und Web (BTW 2017)-Workshopband, pp. 79–88.

Angermueller, C., Pärnamaa, T., Parts, L., Stegle, O., 2016. Deep learning for computational biology. Mol. Syst. Biol. 12 (7), 878.

Arnold, J.G., Moriasi, D.N., Gassman, P.W., Abbaspour, K.C., White, M.J., Srinivasan, R., Santhi, C., Harmel, R., Van Griensven, A., Van Liew, M.W., Kannan, N., 2012. SWAT: model use, calibration, and validation. T. ASABE 55 (4), 1491–1508.

Baldi, P., 1995. Gradient descent learning algorithm overview: a general dynamical systems perspective. IEEE Trans. Neural Netw. 6 (1), 182–195.

Ballesteros, R., Ortega, J.F., Moreno, M.Á., 2016. FORETo: new software for reference evapotranspiration forecasting. J. Arid Environ. 124, 128–141.

Basso, B., Ritchie, J., Pierce, F., Braga, R., Jones, J., 2001. Spatial validation of crop models for precision agriculture. Agric. Syst. 68 (2), 97–112.

Bouman, B., Van Keulen, H., Van Laar, H., Rabbinge, R., 1996. The 'school of de wit' crop growth simulation models: a pedigree and historical overview. Agric. Syst. 52 (2–3), 171–198.

Bretherton, C.S., Smith, C., Wallace, J.M., 1992. An intercomparison of methods for finding coupled patterns in climate data. J. Clim. 5 (6), 541e560.

World agriculture: towards 2015/2030: an FAO perspective. In: Bruinsma, J. (Ed.), Earthscan. http://www.fao.org/3/y4252e/y4252e00.htm (Accessed 12 January 2021).

Bugata, P., Drotar, P., 2020. On some aspects of minimum redundancy maximum relevance feature selection. Sci. China Inf. Sci. 63 (1), 1–15.

Calesella, F., Testolin, A., De Grazia, M.D.F., Zorzi, M., 2021. A comparison of feature extraction methods for prediction of neuropsychological scores from functional connectivity data of stroke patients. Brain Inform. 8 (1), 1–13.

Casella, G., George, E.I., 1992. Explaining the GIBBS sampler. Am. Stat. 46 (3), 167–174.

Chen, Y., Lin, Z., Zhao, X., Wang, G., Gu, Y., 2014. Deep learning-based classification of hyperspectral data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7 (6), 2094–2107.

Chen, S.W., Shivakumar, S.S., Dcunha, S., Das, J., Okon, E., Qu, C., Taylor, C.J., Kumar, V., 2017. Counting apples and oranges with deep learning: a data-driven approach. IEEE Robot Autom. Lett. 2 (2), 781–788.

Chi, M., Plaza, A., Benediktsson, J.A., Sun, Z., Shen, J., Zhu, Y., 2016. Big data for remote sensing: challenges and opportunities. Proc. IEEE 104 (11), 2207–2219.

Connor, D.J., Loomis, R.S., Cassman, K.G., 2011. Crop Ecology: Productivity and Management in Agricultural Systems. Cambridge University Press.

Costa, M., 1996. Probabilistic interpretation of feedforward network outputs, with relationships to statistical prediction of ordinal quantities. Int. J. Neural Syst. 7 (5), 627–637.

De, W., 1978. Simulation of assimilation, respiration and transpiration of crops. Simulation Monographs. PUDOC, Wageningen.

De Vries, F.P., 1973. Substrate Utilization and Respiration in Relation to Growth and Maintenance in Higher Plants. Ph.D. dissertationWageningen University.

De Wit, C., 1965. Photosynthesis of leaf canopies. Agricultural Research Report no 663. Center for Agricultural Publication and Documentation, Wagenin-gen, The Netherlands.

De Wit, C., De Vries, F.P., 1983. Crop growth models without hormones. Neth. J. Agri. Sci. 31, 313–323.

Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.A., 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans. Evol. Comput. 6 (2), 182–197.

Deng, L., Yu, D., 2014. Deep learning: methods and applications. Found. Trend Signal Process. 7 (3–4), 197–387.

Dorigo, M., Di Caro, G., 1999. Ant colony optimization: a new meta-heuristic. Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406). Vol. 2. IEEE, pp. 1470–1477.

Dyrmann, M., Jørgensen, R.N., Midtiby, H.S., 2017. RoboWeedSupport - detection of weed locations in leaf occluded cereal crops using a fully convolutional neural network. Adv. Anim. Biosci. 8 (2), 842–847.

Einheuser, M.D., Nejadhashemi, A.P., Sowa, S.P., Wang, L., Hamaamin, Y.A., Woznicki, S.A., 2012. Modeling the effects of conservation practices on stream health. Sci. Total Environ. 435, 380–391.

Einheuser, M.D., Nejadhashemi, A.P., Wang, L., Sowa, S.P., Woznicki, S.A., 2013. Linking biological integrity and watershed models to assess the impacts of historical land use and climate changes on stream health. Environ. Manag. 51 (6), 1147–1163.

FAO, Food and Agriculture Organization of the United Nations, 2017. The State of Food and Agriculture: Leveraging Food Systems for Inclusive Rural Transformation. Food and Agriculture Organization of the United Nations.

Forte, A., Garcia-Donato, G., Steel, M., 2018. Methods and tools for bayesian variable selection and model averaging in normal linear regression. Int. Stat. Rev. 86 (2), 237–258.

Friedman, J.H., 2002. Stochastic gradient boosting. Comput. Stat. Data Anal. 38 (4), 367–378.

Gebbers, R., Adamchuk, V.I., 2010. Precision agriculture and food security. Science 327 (5967), 828–831 12.

Goudriaan, J., 1977. Crop Micrometeorology: A Simulation Study. Doctoral dissertationPudoc. Wageningen University.

Halko, N., Martinsson, P.G., Tropp, J.A., 2011. Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions. SIAM Rev. 53 (2), 217–288.
Hansen, J.W., Ines, A.V., 2005. Stochastic disaggregation of monthly rainfall data for crop simulation studies. Agric. For. Meteorol. 131 (3–4), 233–246.
Hashem, I.A.T., Yaqoob, I., Anuar, N.B., Mokhtar, S., Gani, A., Khan, S.U., 2015. The rise of "big data" on cloud computing: review and open research issues. Inf. Syst. 47, 98–115.
He, J., Dukes, M., Jones, J., Graham, W., Judge, J., 2009. Applying glue for estimating ceres-maize genetic and soil parameters for sweet corn production. T. ASABE 52 (6), 1907–1921.
Hoogenboom, G., Jones, J., Wilkens, P., Porter, C., Boote, K., Hunt, L., Singh, U., Lizaso, J., White, J., Uryasev, O., 2015. Decision Support System for Agrotechnology Transfer (Dssat) Version 4.6. Dssat Foundation, Prosser, Washington.
Hoogenboom, G., Porter, C.H., Boote, K.J., Shelia, V., Wilkens, P.W., Singh, U., White, J.W., Asseng, S., Lizaso, J.I., Moreno, L.P., Pavan, W., 2019. The DSSAT crop modeling ecosystem. Adv. Crop. Model. Sustain. Agri. 173–216.
Ienco, D., Gaetano, R., Dupaquier, C., Maurel, P., 2017. Land cover classification via multitemporal spatial data by deep recurrent neural networks. IEEE Geosci. Remote Sens. Lett. 14 (10), 1685–1689.
Jha, P.K., Kumar, S.N., Ines, A.V., 2018. Responses of soybean to water stress and supplemental irrigation in upper indo-gangetic plain: field experiment and modeling approach. Field Crops Res. 219, 76–86.
Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L., Wilkens, P.W., Singh, U., Gijsman, A.J., Ritchie, J.T., 2003a. The DSSAT cropping system model. Eur. J. Agron. 18 (3–4), 235–265.
Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L.A., Wilkens, P.W., Singh, U., Gijsman, A.J., Ritchie, J.T., 2003b. The DSSAT cropping system model. Eur. J. Agron. 18 (3–4), 235–265.
Kamilaris, A., Prenafeta-Boldú, F.X., 2018. Deep learning in agriculture: a survey. Comput. Electron. Agric. 147, 70–90.
Kempenaar, C., Lokhorst, C., Bleumer, E., Veerkamp, R., Been, T., van Evert, F., Boogaardt, M., Ge, L., Wolfert, J., Verdouw, C., 2016. Big Data Analysis for Smart Farming: Results of Two Project in Theme Food Security. Tech. Rep.Wageningen University & Research.
Khaki, S., Wang, L., 2019. Crop yield prediction using deep neural networks. Front. Plant Sci. 10, 621.
Khalid, S., Khalil, T., Nasreen, S., 2014. A survey of feature selection and feature extraction techniques in machine learning. Proceedings of 2014 Science and Information Conference. 2014. SAI, pp. 372–378.
Kitzes, J., Wackernagel, M., Loh, J., Peller, A., Goldfinger, S., Cheng, D., Tea, K., 2008. Shrink and share: humanity's present and future ecological footprint. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 363, 467–475.
Kropff, M.J., 1994. Oryza1-an ecophysiological model for irrigated rice production. SARP Research Proceedings. DLO-Research Institute for Agrobiology and Soil Fertility.
Kussul, N., Lavreniuk, M., Skakun, S., Shelestov, A., 2017. Deep learning classification of land cover and crop types using remote sensing data. IEEE Geosci. Remote Sens. Lett. 14 (5), 778–782.
Lascody, R., Melbourne, N., 2002. The Onset of the Wet and Dry Seasons in East Central Florida, a Subtropical Wet-dry Climate. National Weather Service Weather Forecast Office Melbourne, FL.
Latha, C.P., Mohana, P., 2016. A review on deep learning algorithms for speech and facial emotion recognition. Aptikom 1 (3), 88–104.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.
Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., Alsaadi, F.E., 2017. A survey of deep neural network architectures and their applications. Neurocomputing 234, 11–26.
Lobell, D.B., Asseng, S., 2017. Comparing estimates of climate change impacts from process-based and statistical crop models. Environ. Res. Lett. 12 (1), 015001.
Lobell, D.B., Burke, M.B., 2010. On the use of statistical models to predict crop yield responses to climate change. Agric. For. Meteorol. 150 (11), 1443–1452.
Maret, T., Konrad, C., Tranmer, A., 2010. Influence of environmental factors on biotic responses to nutrient enrichment in agricultural streams. J. Am. Water Resour. Assoc. 46 (3), 498–513.
McCown, R.L., Hammer, G.L., Hargreaves, J.N.G., Holzworth, D.P., Freebairn, D.M., 1996. APSIM: a novel software system for model development, model testing and simulation in agricultural systems research. Agric. Syst. 50 (3), 255–271.
Menger, V., Scheepers, F., Spruit, M., 2018. Comparing deep learning and classical machine learning approaches for predicting inpatient violence incidents from clinical text. Appl. Sci. 8 (6), 981.
Minh, D.H.T., Ienco, D., Gaetano, R., Lalande, N., Ndikumana, E., Osman, F., Maurel, P., 2017. Deep Recurrent Neural Networks for Mapping Winter Vegetation Quality Coverage Via Multi-temporal Sar Sentinel-1 arXiv preprint arXiv,1708.03694.
Namin, S.T., Esmaeilzadeh, M., Najafi, M., Brown, T.B., Borevitz, J.O., 2018. Deep phenotyping: deep learning for temporal phenotype/genotype classification. Plant Meth. 14 (1), 66.
Nurudeen, A.R., 2011. Decision Support System for Agro-technology Transfer (DSSAT) Model Simulation of Maize Growth and Yield Response to NPK Fertilizer Application on a Benchmark Soil of Sudan Savanna Agro-ecological Zone of Ghana (Doctoral Dissertation). Kwame Nkrumah University of Science and Technology Kumasi.
O'Hara, R.B., Sillanpää, M.J., 2009. A review of Bayesian variable selection methods: what, how and which. Bayesian Anal. 4 (1), 85–117.
Pathak, R., Barzin, R., Bora, G.C., 2018. Data-driven precision agricultural applications using field sensors and Unmanned Aerial Vehicle. IJPAA 1 (1).
Pearson, K., 1901. On lines and planes of closest fit to systems of points in space. Lond. Edinb. Dublin. Philos. Mag. J. Sci. 2 (11), 559–572.
Peng, H., Long, F., Ding, C., 2005. Feature selection based on mutual information criteria of maxdependency, max-relevance, and min-redundancy. IEEE Trans. Pattern Anal. Mach. Intell. 27 (8), 1226–1238.
Penning de Vries, F.W.T., 1989. Simulation of ecophysiological processes of growth in several annual crops. Int. Rice Res. Inst. 29.
Pérez, P., de los Campos, G., 2014. BGLR: a statistical package for whole genome regression and prediction. Genetics 198 (2), 483–495.
Prasad, R., Ali, M., Kwan, P., Khan, H., 2019. Designing a multi-stage multivariate empirical mode decomposition coupled with ant colony optimization and random forest model to forecast monthly solar radiation. Appl. Energy 236, 778–792.
Prasad, R., Ali, M., Xiang, Y., Khan, H., 2020. A double decomposition-based modelling approach to forecast weekly solar radiation. Renew. Energy 152, 9–22.
Prasanna, Mohanty, S., Hughes, D., Salathe, M., 2016. Using Deep Learning for Image-based Plant Disease Detection (arXiv e-prints).
PRISM, 2011. Prism Climate Data. http://prism.oregonstate.edu/ (Accessed 12 January 2021).
Rahnemoonfar, M., Sheppard, C., 2017. Deep count: fruit counting based on deep simulated learning. Sensors 17 (4), 905.
Rao, R.N., Sridhar, B., 2018. IoT based smart crop-field monitoring and automation irrigation system. 2018 2nd International Conference on Inventive Systems and Control (ICISC). IEEE, pp. 478–483.
Rebetez, J., Satizábal, H.F., Mota, M., Noll, D., Büchi, L., Wendling, M., Cannelle, B., Pérez-Uribe, A., Burgos, S., 2016. Augmenting a Convolutional Neural Network With Local Histograms-A Case Study in Crop Classification From High-resolution UAV Imagery. InESANN.
Sacks, W.J., Kucharik, C.J., 2011. Crop management and phenology trends in the us corn belt: impacts on yields, evapotranspiration and energy balance. Agric. For. Meteorol. 151 (7), 882–894.
Schlenker, W., Hanemann, W.M., Fisher, A.C., 2006. The impact of global warming on us agriculture: an econometric analysis of optimal growing conditions. Rev. Econ. Stat. 88 (1), 113–125.
Schmidhuber, J., 2015. Deep learning in neural networks: an overview. Neural Netw. 61, 85–117.
Sehgal, G., Gupta, B., Paneri, K., Singh, K., Sharma, G., Shroff, G., 2017. Crop planning using stochastic visual optimization. 2017 IEEE Visualization in Data Science (VDS). IEEE, pp. 47–51.
Seligman, N., Van Keulen, H., 1981. Papran: a simulation model of annual pasture production limited by rainfall and nitrogen, simulation of nitrogen behavior of soil-plant systems: papers of a workshop. Models for the Behavior of Nitrogen in Soil and Uptake by Plant, Comparison Between Different Approaches, Wageningen, the Netherlands, January 28–February 1, 1980. Centre for Agricultural Publishing and Documentation, Wageningen, Netherlands, p. 1981.
Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., Stefanovic, D., 2016. Deep neural networks based recognition of plant diseases by leaf image classification. Comput. Intell. Neurosci. 1–11 3289801.
Slavin, P., 2016. Climate and famines: a historical reassessment. Wiley Interdiscip. Rev. 7 (3), 433–447.
Song, X., Zhang, G., Liu, F., Li, D., Zhao, Y., Yang, J., 2016. Modeling spatio-temporal distribution of soil moisture by deep learning-based cellular automata model. J. Arid Land. 8 (5), 734–748.
Spitters, C., Van Keulen, H., Van Kraalingen, D., 1989. A simple and universal crop growth simulator: Sucros87. Simulation and Systems Management in Crop Protection. Pudoc, pp. 147–181.
Stöckle, C.O., Donatelli, M., Nelson, R., 2003. Cropsyst, a cropping systems simulation model. Eur. J. Agron. 18 (3–4), 289–307.
Sun, P., Wen, Y., Han, R., Feng, W., Yan, S., 2019. Gradientflow: Optimizing Network Performance for Large-Scale Distributed DNN Training. IEEE Trans. Big Data.
Tyagi, A.C., 2016. Towards a second green revolution. Irrig. Drain. 65 (4), 388–389.
Van Diepen, C.V., Wolf, J., Van Keulen, H., Rappoldt, C., 1989. Wofost: a simulation model of crop production. Soil Use Manag. 5 (1), 16–24.
van Keulen, H., 1975. Simulation of Water Use and Herbage Growth in Arid Regions. Pudoc, Wageningen University.
Van Keulen, H., 1982. Crop production under semi-arid conditions, as determined by nitrogen and moisture availability. Simulation of Plant Growth and Crop Production. Pudoc, pp. 234–249.
Waite, I., Brown, L., Kennen, J., May, J., Cuffney, T., Orlando, J., Jones, K., 2010. Comparison of watershed disturbance predictive models for stream benthic macroinvertebrates for three distinct ecoregions in western US. Ecol. Indic. 10 (6), 1125–1136.
Wallach, D., Makowski, D., Jones, J.W., Brun, F., 2018. Working with Dynamic Crop Models: Methods, Tools and Examples for Agriculture and Environment. Academic Press.
Woznicki, S.A., Nejadhashemi, A.P., Ross, D.M., Zhang, Z., Wang, L., Esfahanian, A.H., 2015. Ecohydrological model parameter selection for stream health evaluation. Sci. Total Environ. 511, 341–353.
Yalcin, H., 2017. Plant phenology recognition using deep learning: deep-pheno. 6th International Conference on Agro-geoinformatics. IEEE, pp. 1–5.
Zwillinger, D., Kokoska, S., 1999. CRC Standard Probability and Statistics Tables and Formulae. Crc Press.