International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2020): 7.803

Intelligent Computational Techniques for Crop Yield Based On BP Neural Network

Wenwu Xie, Aguwa Dominic Ugochukwu, Jianhui Wu, Zhihe Yang*

Hunan Institute of Science and Technology, Department of Information and Communication Engineering, Hunan Province, China
Corresponding Author: Zhihe Yang, yangzhihe[at]hnist.edu.cn

Abstract: ***Predicting crop yield is one of the most difficult problems in precision agriculture and one of the largest areas of agriculture to improve the economy. Although, it is a new concept in agriculture. However, there are several challenges due to its dependence on multiple factors such as extreme weather, crop genotype, environmental factors, obtaining minimized production due to climate change, availability of water, management practices, and their environment. Therefore, we designed an efficient BP neural network to obtain crop yield prediction based on environmental data. Hence in this research, we focus on using archival data to predict corn yield across the US cornfield (including 39 states) from 1980 to 2018. The BP neural network model performed very well and achieved a mean-square-error (MSE), mean-absolute-error (MAE), and mean-absolute-percentage-error (MAPE) with an overall accuracy (OA) of 89% of their respective average yields, which is much better than the tested RNNs model, which shows that our BP neural network model has high accuracy in agricultural data prediction.***

Keywords: Big data, Crop yield prediction, BP neural networks

1. Introduction

Predicting crop yield is one of the most difficult problems in precision agriculture, and it is also one of the largest agricultural areas to improve economic benefits. Therefore, it is important to provide a user-friendly interface for farmers that enable yield analysis based on the available data to maximize crop productivity. Numerous studies have used machine learning technology to predict yield, including artificial neural networks, correlation rule mining, multiple regression, and random forest. Although the machine learning model regards yield as an implicit function of input variables (such as meteorological elements and soil environment), this may be very complex and nonlinear. Jeong et al. (2016) carried out random forest analysis and multiple regression analysis on wheat, corn, and potatoes. They found that random forest predicted crop yield very well and exceeded multiple regression analysis. However, in the past few decades, machine learning algorithms for yield prediction have been developed, such as genetic algorithm (GA), support vector machine (SVM), linear regression, artificial neural network (ANN), and Naive Bayes (NB) [6].

In addition, an intelligent agriculture system has been developed to solve the problems of traditional agricultural applications, although the previous applications are the basis of digital agriculture and need to spend a lot of energy to process the data from the buffer. Today, intelligent agricultural systems are being developed through the Internet of things (IoT) and wireless sensor networks (WSN) [8].compared with the above machine learning algorithms, the deep learning method has very powerful functions and can accurately show good performance in agriculture [9]. Deep learning is an intelligent computing method that can solve many problems such as scalability, image processing, reliability, optimization, and fault tolerance. Therefore, we developed a BP neural network modeling technique to predict yields based on environmental data and management techniques.

Although deep learning is a multi-level presentation technique. Each presentation layer has a nonlinear element to convert the representation of the current layer based on the original input into a simple abstraction layer (Lecun et al., 2015). It also provides general approximation architecture. In other words, no matter what operation we want to implement, we can use deep neural networks to represent such an operation (Hornik et al., 1989; Goodfellow et al., 2016). Deep learning techniques do not require a manual operation but implement its operation through data, which improves the accuracy of results (Lecun et al., 2015). However, higher models are very hard to train and require more sophisticated hardware and optimization techniques (Goodfellow et al., 2016).

2. Related Work

In recent years, many scholars have studied the methods of water-saving irrigation and crop yield in the process of agricultural production through big data analysis technology and machine learning algorithm. For example, in the Pitaya experimental field of Nanning irrigation experimental station, the decision tree algorithm is used to evaluate whether the environment of the experimental field is suitable for planting pitaya [10], so as to provide support for managers to make subsequent scientific management decisions. Literature [11], monitored the drug structure of pesticide application, such as variety, dosage, cost and application times, and evaluated the rationality of drug structure of open field vegetables. Literature [12] takes rice producing areas in North China, East China and Northeast China as examples. Through big data analysis, it is concluded that temperature and precipitation have a great impact on crop yield, which is convenient to monitor the rice cultivation process. The research shows that big data analysis and machine learning algorithm have great application space in the research field of water-saving irrigation.

In a research paper [13], researchers mention that large amounts of data are collected and stored for analysis. Proper use of this data can often result in significant efficiency gains and consequent economic benefits. Although, there are several uses for data mining technology in agriculture. Researchers have implemented the K-Means algorithm to predict air pollution [14] and used the K-nearest neighbor method [15] to simulate daily precipitation and other meteorological variables and use support vector machines. And analyze the various possible changes in the weather scenario [16]. On the other hand, the researchers worked on analyzing the variability of rainfall and its impact on crop yields [17]. The results to obtained seasonal climatic conditions (such as precipitation and temperature changes) on crop yield prediction can be explained by experimental crop models [18].

Currently, deep learning techniques are used to predict crop yields. Khaki and Wang (2019) developed a deep neural network system to predict maize yields at 2, 247 sites between 2008 and 2016. This system was superior to other techniques such as multiple regression, flat neural networks, and lasso. On the other hand, you et al. (2017) predicted soybean yields based on a series of remote sensing images using CNNs and RNNs. Kim and associates. (2019) Designed a deep neural network system from 2006 to 2015 to predict crop yields using improved input variables from satellite products and weather datasets. Wang and colleagues. (2018) constructed a deep learning architecture for predicting soybean crops in Argentina, in which they used previous learning methods and obtained good results when predicting soybean yield with fewer data in Brazil. Yang et al. (2019) studied the ability of CNN to evaluate rice yield using remote sensing images and discovered that the CNN model gives reliable yield prediction in the whole mature phase. Meanwhile, Khaki and Khalil Zadeh (2019) utilized deep CNNs to forecast maize yield losses in 1, 560 regions of Canada and the United States.

3. Proposed Work

Our designed problem statement is to build a big data neural network system for the prediction of agricultural crop yield to increase agricultural production, which improves the economic growth of the country. Although, It has been shown that agricultural production generates a lot of data, and the type of data is diverse, and the quantitative indicators between different data are not the same, which makes the processing of agricultural production data challenging.

Therefore, this paper aims to solve the difficulties in processing agricultural production. As an intelligent algorithm, our design utilizes a BP neural network to process big data in the context of a significant amount of agricultural data, accurately predict agricultural production according to the big data generated by the production process, and then guide agricultural production.

3.1 BP Network Topology

A BP neural network with three layers is illustrated in Figure 1. There is an input layer, a hidden layer, and an output layer. There is an equal number of nodes in the input layer and in the output layer n. The output layer has the same number of nodes as the output module type q. Each node in the hidden layer p corresponds to a specific application, which is usually chosen during testing.

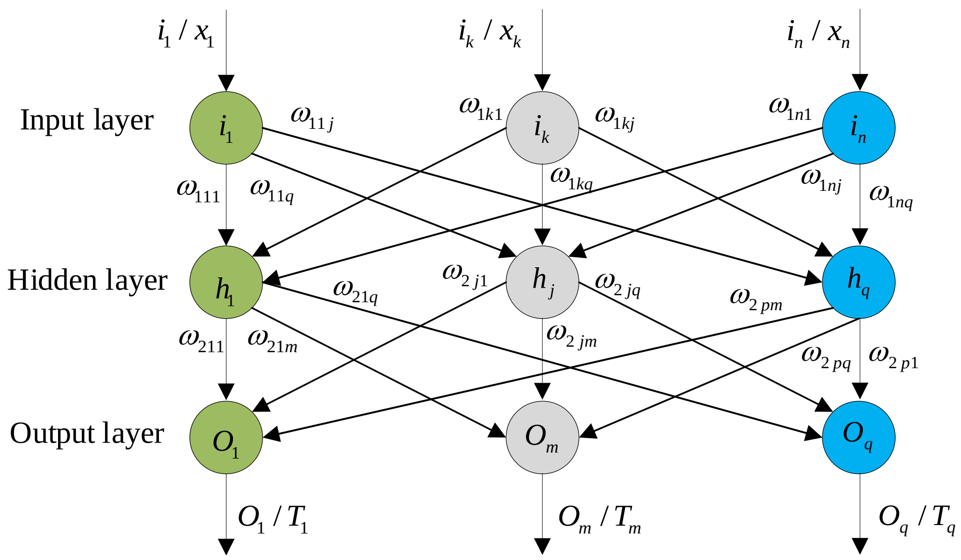


Figure 1: BP network topology

W_1 defines the connection weight matrix between input layer nodes and hidden layer nodes as formula (1). W_2 is an equivalent connection weight matrix that characterizes the connection between hidden layer nodes and output nodes. θ_1 and θ_2 are the threshold matrices of the nodes in the hidden layer and output layer, respectively:

$$W_1 = \begin{Bmatrix} w_{111} & \cdots & w_{11j} & \cdots & w_{11p} \\ \vdots & \cdots & \vdots & & \\ w_{1k1} & \cdots & w_{1kj} & \cdots & w_{1kp} \\ \vdots & \cdots & \vdots & & \\ w_{1n1} & \cdots & w_{1nj} & \cdots & w_{1np} \end{Bmatrix} \quad (1)$$

3.2 Forward Propagation

As a first step, enter the training samples X, T where X is the input vector and T is the output vector:

$$X = (x_1, x_2, ..., x_k, ..., x_n), 1 \le k \le n,$$
$$T = (T_1, T_2, ..., T_m, ..., T_q), 1 \le m \le q, \quad (2)$$

In this equation, n represents the number of input nodes; q represents the number of output nodes. Through the input layer and hidden layer, data is forwarded to the output layer. The learning result is the weight value of output pattern classification. In general, these steps should be followed:

Identify the output value of the hidden layer nodes, (Step 1). Nodes in the hidden layer have the following input value:

$$i_{1j} = \sum_{k=1}^{n} w_{1kj} x_k, 1 \le j \le p \quad (3)$$

In this case, n represents the nodes of the input layer, p represents the nodes of the hidden layer, w_{1kj} represents the weight of the connections, and x_k represents the component of input vector.

$$h_j = f(i_{1j} + \theta_{1j}) \quad (4)$$

Here, θ_{1j} is the threshold value for node j. We have used the sigmoid function f, which was given by Rumelia:

$$f(i_{1j}) = \frac{1}{1 + e^{-i_{1j}}} \quad (5)$$

(Step 2) The calculation node of output layer m and the calculation node of output value of m

The input values of nodes of output layer m

$$i_{2m} = \sum_{j=1}^{p} w_{2jm} h_j, 1 \le m \le q \quad (6)$$

The nodes of output value

$$O_m = f(i_{2m} + \theta_{2m}) \quad (7)$$

In addition, θ_{2m} is the threshold for the node of output layer m and f is the activation function.

3.3 BP Neural Network Back Operation

Determine the difference between the output value and the expected value of the output layer. The error travels backwards from the output layer to the hidden layer and to the hidden layer and to the input layer and modify the connection weight values. The following steps are necessary.

Step 1 of the process involves comparing the learning value O_m of node of output layer m and the output value of learning samples T_m (the calculation of the output error of nodes of the out layer), which is given as:

$$\varepsilon_m = |O_m - T_m| \quad (8)$$

During step 2 (testing learning error), ε_0 represents the maximum learning error, which falls within the range of 0 and 1. When $\max(\varepsilon_m) \le \varepsilon_0$ put the next learning sample. If not, modify the weights of the network and put the original one. The learning process is terminated once all samples meet the above criteria. Step 3 involves calculating the learning error for the nodes of the output layer m

$$d_{2m} = O_m(1 - O_m)(O_m - T_m) \quad (9)$$

In the process of step 4 we calculate the learning error of nodes of hidden layer j as

$$d_{1j} = h_j(1 - h_j) \sum_{j=1}^{p} w_{2jm} d_{2m} \quad (10)$$

While step 5 Set the weight at time t+1 as the adjusted new weight value; then update the connection weights matrix W_2. as shown below

$$w_{2jm}(t+1) = w_{2jm}(t) + \eta d_{2m} h_j + \alpha[w_{2jm}(t) - w_{2jm}(t-1)] \quad (11)$$

The common BP algorithm is based on two parameters, η and α. Both η and α are within scope. Using a can increase the learning speed and provide a way to deal with local minimums. In Step 6, we reformed the weights of the connection matrix W_2. below is the computation

$$w_{1kj}(t+1) = w_{1kj}(t) + \eta d_{1j} x_k + \alpha[w_{1kj}(t) - w_{1kj}(t-1)] \quad (12)$$

Meanwhile, step7 reform threshold θ_2 of nodes of output layer

$$\theta_{2m}(t+1) = \theta_{2m}(t) + \eta d_{2m} h_j + \alpha[w_{2jm}(t) - w_{2jm}(t-1)] \quad (13)$$

In addition, step 8 reform threshold θ_1 of nodes of hidden layer j

$$\theta_{1j}(t+1) = \theta_{1j}(t) + \eta d_{1j} + \alpha[\theta_{1j}(t) - \theta_{1j}(t-1)] \quad (14)$$

3.4 Running BP Neural Network

To implement pattern classification, run the BP neural network after learning and only use the forward operation of the BP learning algorithm: The method is as follows.
The first method involves assigning revised values to connection weight matrices W_1 and W_2, and threshold matrices Q1 and Q2. The second method involves inputting the input vector I.

Then, apply formulas (3), (4), and (5) to calculate the output of the hidden layer. Then, use formulas (5), (6), and (7) to calculate the output of the output layer, which is the classification result of input vector I.

4. Data Description

We evaluate our model in the United States and choose corn as the target crop since it has been widely investigated in prior work (Shahhosseini et al.2019). The input data we use include yield performance, meteorological, and soil; (nitrogen, pH, solar radiation, coarse fragment, nitrogen mean 0-5cm, nitrogen mean 5-15cm, W1, and W2).

The yield performance data were obtained from the National Agricultural Statistics Service of the United State (USDA-NASS, 2019). 'Meteorological' attribute specifies the daily recording of meteorological variables, which includes precipitation, solar radiation. The meteorological data were obtained from Daymet (Thornton et al., 2018).

'Yield performance' attribute specifies the average yield for corn from 1980 to 2018 under 1, 176 corn counties in 39 Corn Belt states. 'Soil' attribute specifies the nitrogen, pH, and coarse fragments measured at depths 0-5, 5-10, 10-15, 15-30, 30-45, 45-60, 60-80, 80-100, and 100-120cm. The soil data were obtained from Gridded Soil Survey Geographic Database for United States (gSSURGO, 2019).

5. Methodology

Our model for predicting crop yield is a BP network. In the first step, we took sample weather and soil samples from 39 Corn Belt states and took the average of the samples from 1, 176 corn counties to create representative samples. Meanwhile, the corn yield over the past five decades has shown an increasing trend, as illustrated in **Figure 2**.

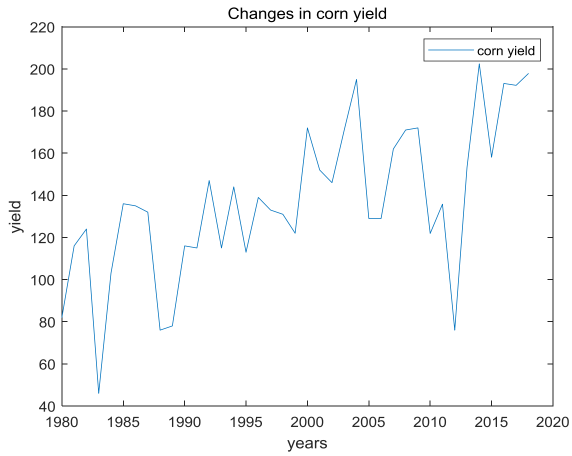


Figure 2: The average corn yield from 1980 to 2020

Table 1: 1980 to 2018 corn data

Year	CEF	NN	pH	RF	SR	W1	W2	NM1	NM2	Yield
1980	8.3	1461.9	60.8	2.5	321.3	193.4	223.9	3809.6	1728.2	32.5
1981	8.3	1461.9	60.8	3.4	319.4	218.9	227.5	3809.6	1728.2	36.0
1982	8.3	1461.9	60.8	3.5	317.2	196.2	182.6	3809.6	1728.2	37.0
1983	8.3	1461.9	60.8	2.6	318.9	204.5	233.1	3809.6	1728.2	23.0
1984	8.3	1461.9	60.8	3.2	316.3	163.2	190.1	3809.6	1728.2	28.5
. .	.	.	.	.	.	.	.	.	.	.
2014	8.3	1461.9	60.8	3.0	322.9	120.8	215.1	3809.6	1728.2	50.7
2015	8.3	1461.9	60.8	3.5	315.9	188.1	109.5	3809.6	1728.2	52.1
2016	8.3	1461.9	60.8	2.2	316.1	237.0	185.7	3809.6	1728.2	56.8
2017	8.3	1461.9	60.8	2.3	318.9	161.2	171.7	3809.6	1728.2	55.9
2018	8.3	1461.9	60.8	2.9	317.3	203.9	185.0	3809.6	1728.2	60.7

6. Data Processing Result

We implemented our BP network in the most efficient way possible to predict corn yields. For 38 consecutive years, we predicted corn yields. In the validation year, we used data from 1980 to 2018. Above is a table that shows that corn bushels per acre are different for different soil pH, different soil coarse fragments, different nitrogen, different solar radiation, and different precipitation.

As a result of the huge amount of agricultural data, the data are presented into our BP neural network, where several predictions are carried out, the difference between prediction yield and actual yield is observed, and the accuracy of the neural network model is evaluated. The prediction results and actual results are shown in **Table 2** and **Figure3**.

Table 2: Corn yield prediction performance of the BP model from 1980 to 2018

Indicator	MAE	MSE	MAPE	OA
Train	0.066	0.007	42.32	88.89%
Test	0.053	0.003	25.04	

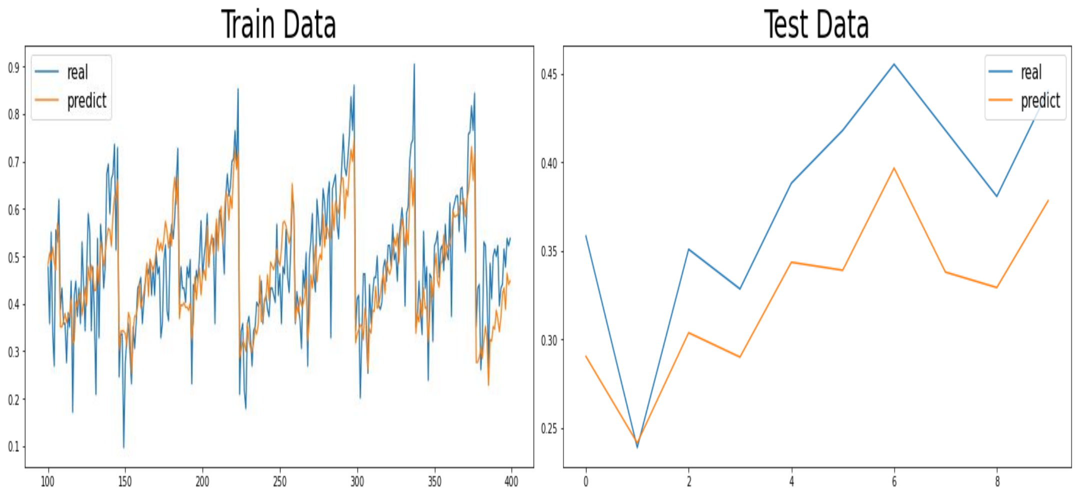


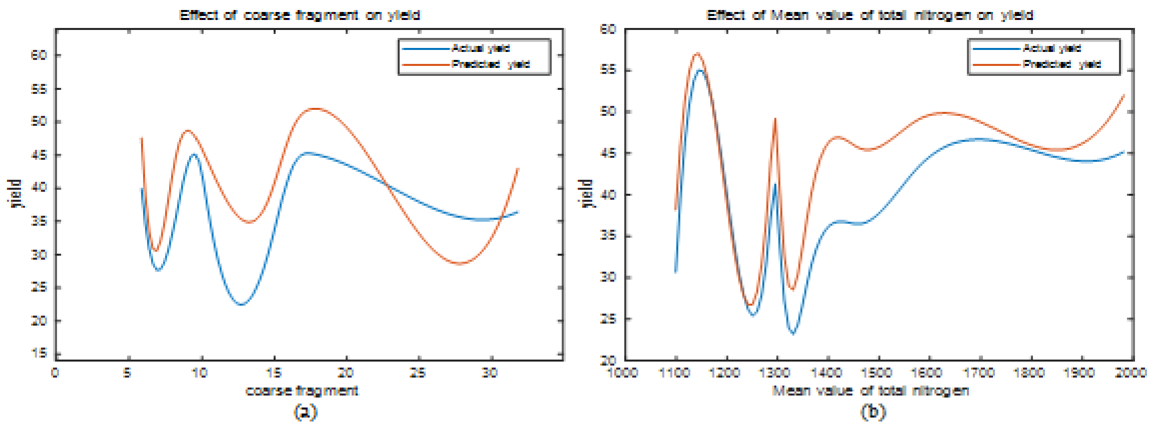
Figure 3: 1980 to 2018 train data and test data of BP neural network

From **Table 2** and **Figure 3**, it is evident that the prediction accuracy of our BP neural network model is high when it comes to agricultural data. As a result of the prediction of BP neural networks, advanced processing of the data can be performed, and the growth trend of the crops can be predicted, thereby preventing problems from occurring in the production process of the crops. Additionally, measures are taken to improve the efficiency of agricultural production and the yield of crops during the growing process.

The weather is one of the most important factors in crop yield prediction, but it is unknown a priori. Therefore, weather prediction is an integral part of crop yield prediction. We examined the impact of weather prediction on the performance of our BP neural network model as shown in **Figure 4.**

7. Conclusion

Our paper presented a new BP neural network approach for predicting crop yields that correctly predicted corn across the entire Corn Belt in the United States based on environmental data and soil data such as soil coarse fragments, rainfall, nitrogen, soil pH, nitrogen mean 0-5cm, nitrogen mean 5-15cm, sun radiation, yield performance, and meteorological factors. We then directly import the data into our BP neural network for further processing. BP performs significantly better than other methods like RNNs. Furthermore, by comparing the prediction results with the actual data, it was found that the overall accuracy (OA) of the model designed in this article was within 89%, which indicates a high level of predictability in agricultural data, which also has a positive impact on production efficiency. Moreover, our methodology offered the results that allowed us to explain yield prediction and neural networks rather than just predict them.



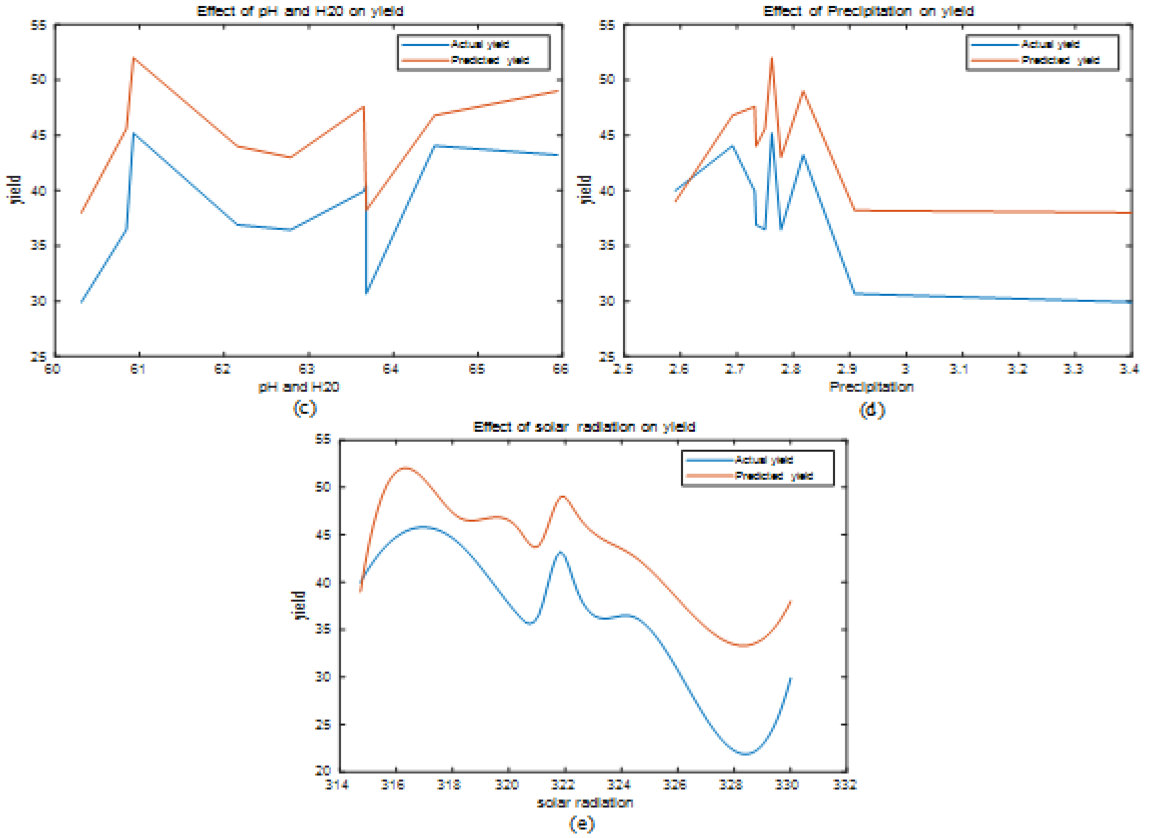


Figure 4: Possible influences on yield: (a) variation of yield in the presence of a coarse fragment, (b) change of yield at total mean of nitrogen, (c) change of yield at a certain pH, (d) variation of yield in the face of precipitation, and (e) variation of yield in the presence of sun radiation.

Acknowledgment

I would like to express my sincere thanks to our Project Guide Dr. Xie for his guidance during the project. Bringing this task to a successful conclusion would be difficult without your help. In addition, we would like to thank Ph. D. professor Jianhui Wu for providing all the necessary facilities. Supported by Hunan Social Science Fund Project, No.19YBA182, and Hunan Provincial Department of Education Innovation Platform Open Fund Project (No.19K037).

Reference

[1] Majumdar, J., Naraseeyappa, S., &Ankalaki, S. (2017). Analysis of agriculture data using data mining techniques: application of big data. Journal of Big Data, 4 (1).
[2] Ruan, J., Jiang, H. Q., Li, X., Shi, Y., Chan, F. T., &Rao, W. (2019). A Granular GA-SVM Predictor for Big Data in Agricultural CyberPhysical Systems.
[3] Pujari, S., &Bogiri, N. (2017). Precision Agriculture for Banana Using Wireless Sensor Network.2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA).
[4] Rehman, A., Deyuan, Z., Hussain, I., Iqbal, M. S., Yang, Y., & Jingdong, L. (2018). Prediction of Major Agricultural Fruits Production in Pakistan by Using an Econometric Analysis and Machine Learning Technique. International Journal of Fruit Science, 1-17.
[5] De Souza, A. V., Bonini Neto, A., Cabrera Piazentin, J., Dainese Junior, B. J., Perin Gomes, E., dos Santos Batista Bonini, C., & Ferrari Putti, F. (2019). Artificial neural network modeling in the prediction of bananas‟ harvest. Scientia Horticulturae, 257, 108724
[6] Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., et al. (2016). Random forests for global and regional crop yield predictions. PloS One 11, e0156571. doi: 10.1371/journal. pone.0156571
[7] Liu, J., Goering, C., and Tian, L. (2001). A neural network for setting target corn yields. Trans. ASAE 44, 705. doi: 10.13031/2013.6097
[8] Ransom, C. J., Kitchen, N. R., Camberato, J. J., Carter, P. R., Ferguson, R. B., Fernández, F. G., et al. (2019). Statistical and machine learning methods evaluated for incorporating soil and weather into corn nitrogen recommendations. comput. Electron. Agric.164, 104872. doi: 10.1016/j.compag.2019.104872
[9] Komal Bodake, Rutuja Ghate, Himanshi Doshi, Priyanka Jadhav, and Balasaheb Tarle, "Soil based Fertilizer Recommendation System using Internet of Things", MVP Journal of Engineering Sciences, 1 (1), 1-7, 2018
[10] Kale, S. S., & Patil, P. S. (2018). Data Mining Technology with Fuzzy Logic, Neural Networks and Machine Learning for Agriculture. Advances in Intelligent Systems and Computing, 79-87
[11] Khaki, Saeed & Wang, Lizhi. (2019). Crop Yield Prediction Using Deep Neural Networks. Frontiers in Plant Science.10.
[12] LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature 521, 436. doi: 10.1038/nature14539
[13] Goodfellow, I., Bengio, Y., and Courville, A. (2016).

Deep Learning Vol.1 (Cambridge: MIT Press).
[14] Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feed forward networks are universal approximators. Neural Networks 2, 359-366. doi: 10.1016/0893-6080 (89) 90020-8
[15] Luo Wei. Big data analysis of pesticide application in open field vegetables in 2015 D] Zhejiang: Zhejiang University, 2016, 7-9
[16] Cui yuan Research on agricultural climate and crop yield change based on big data analysis [J] China Agricultural Resources and planning, 2017, 38 (2): 112-117
[17] G Ruß, "Data Mining of Agricultural Yield Data: A Comparison of Regression Models", Conference Proceedings, Advances in Data Mining -Applications and Theoretical Aspects, P Perner (Ed.), Lecture Notes in Artificial Intelligence 6171, Berlin, Heidelberg, Springer, 2009, pages: 24-37.
[18] Jorquera H, Perez R, Cipriano A, Acuna G, "Short Term Forecasting of Air Pollution Episodes", In: Zannetti P (eds) Environmental modeling, WIT Press, UK, 2001.
[19] Rajagopalan B, Lall U, "A K-Nearest Neighbor Simulator for Daily Precipitation and Other Weather Variables", Wat Res Res 35 (10), 1999, pages: 3089-3101.
[20] Tripathi S, Srinivas V V, Nanjundiah R S, "Downscaling of Precipitation for Climate Change Scenarios: A Support Vector Machine Approach", J Hydrol, 2006, pages: 621-640.
[21] Mehta D R, Kalola A D, Saradava D A, Yusufzai A S, "Rainfall Variability Analysis and Its Impact on Crop Productivity-A Case Study", Indian Journal of Agricultural Research, Volume 36, Issue 1, 2002, pages: 29-33.
[22] M Trnka, “Projections of Uncertainties in Climate Change Scenarios into Expected Winter Wheat Yields”, Theoretical and Applied Climatology, vol.77, 2004, pages: 229-249.
[23] Khaki, S., and Wang, L. (2019). Crop yield prediction using deep neural networks. Front. In Plant Sci.10, 621. doi: 10.3389/fpls.2019.00621.
[24] You, J., Li, X., Low, M., Lobell, D., and Ermon, S. (2017). “Deep gaussian process for crop yield prediction based on remote sensing data, ”in Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, California, USA.4559-4566.
[25] Kim, N., Ha, K.-J., Park, N.-W., Cho, J., Hong, S., and 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 Int. J. Geo-Information 8, 240. doi: 10.3390/ijgi8050240
[26] Wang, A. X., Tran, C., Desai, N., Lobell, D., and Ermon, S. (2018). “Deep transfer learning for crop yield prediction with remote sensing data, ”in Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies (San Jose CA, USA: ACM), 50.
[27] Yang, Q., Shi, L., Han, J., Zha, Y., and Zhu, P (2019). Deep convolutional neural networks for rice grain yield estimation at their opening stage using uav-based remotely sensed images. Field Crops Res.235, 142-153. doi: 10.1016/j. fcr.2019.02.022
[28] Khaki, S., and Khalilzadeh, Z. (2019). Classification of crop tolerance to heat and drought: A deep convolutional neural networks approach. arXiv preprint arXiv: 1906.00454.
[29] Lu GY, Ren L, Kolagunda A, et al. Representing 3D shapes based on implicit surface functions learned from RBF neural networks. J Vis Commun Image R 2016; 40: 852-860.
[30] Qi HX, Lin ZX and Lan YB. The application of big data in precision agriculture. China Agric Sci Technol Her 2019; 21 (1): 1-10.
[31] Roca Barcelo´ F, Jae´n del Hierro P, Ribes Llario F, et al. Development of an ultrasonic weld inspection system based on image processing and neural networks. Non destruct Test Eva 2018; 33 (2): 229-236.
[32] Jiang W and Xu FY. Design of intelligent agricultural internet of things monitoring system based on M2M platform. J Jinling Inst Technol 2018; 34 (4): 52-55.
[33] Guo ZY, Yang SF and Wang J. Global synchronization of memristive neural networks subject to random disturbances via distributed pinning control. Neural Networks 2016; 84: 67-79.