Prediction of Rice Yield via Stacked LSTM

Xiangyan Meng, College of Science, Northeast Agricultural University, Harbin, China

Muyan Liu, College of Engineering, Northeast Agricultural University, Harbin, China

Qiufeng Wu, College of Science, Northeast Agricultural University, Harbin, China

ABSTRACT

In order to guarantee the rice yield more effectively, the prediction of rice yield should be taken into account. Because the rice yield every year can be seen as a sequence of time series, many methods applied in prediction of time series can be considered. Long Short-Term Memory recurrent neural network (LSTM) is one of the most popular methods of time series prediction. In consideration of its own characteristics and the popularity of deep learning, an improved LSTM architecture called Stacked LSTM which has multiple layers is proposed in this article. It is based on the idea of increasing the depth of LSTM. The comparison among the Stacked LSTM architectures which have different numbers of LSTM layers and other methods including ARIMA, GRU, and ANN has been carried out on the data of rice yield in Heilongjiang Province, China, from 1980 to 2017. The results showed the superior performance of Stacked LSTM and the effectiveness of increasing the depth of LSTM.

INTRODUCTION

Rice is one of the most important food sources for more than half of the world's population (Jeon et al., 2011), it is the second most widely grown cereal crop worldwide (Hirooka et al., 2018) and the demand for rice is expected to grow because of the increasing population on earth (Daniela et al., 2018). In order to ensure the rice yield keeping up with the pace of population growth, some measures to guarantee sufficient rice supply, in other words, food security, are needed necessarily. Crop yield prediction is a representative measure which is vital for food security (Hutchinson, 1991). It can obtain the result whether the future crop yield can achieve the demand of population, therefore, it plays a key role in government's policy making and preparing production plan for following year. Proper decisions of government based on crop yield prediction can make more efficient management of crop production processes. On the other hand, crop yield prediction can provide a reference for farmers and enterprise, thus helping them increase outcome (Na-Udom & Rungrattanaubol, 2015), so rice yield prediction is a matter of importance.

Crop simulation models are used extensively, simulation of plant-growth stages and consequently forecasting the crop yield permits better planning (Inoue, Susan & Horie, 1998). Crop growth

simulation model are combined with remote sensing data to estimate rice yield (Abou-Ismail, Huang & Wang, 2004; Inoue, Susan & Horie, 1998). N. Pirmoradian et al. (2006) proposed a very simple model(VSM) to predict rice grain and biomass yields. But the application scope of crop growth models is a little limited, they may only adapt to a few crop species. Meanwhile, the development time and cost of these models are extremely large (Alberto, Juan & Waldo, 2014). Many statistical methods are also used for yield prediction in early stage, such as Multiple Linear regression, logistic regression (Cakir, Kirci & Gunes, 2014), which eliminate the limitation of application range. Liu Qin-pu (2011) proposed grain yield spatio-temporal regression prediction model of Henan Province; Umid Kumar Dey et al. (2017) applied Multiple Linear Regression AdaBoost to the task of predicting rice yield in Bangladesh. R. K. Pal (2012) developed the multiple regression equations for prediction of growth and yield attributes of wheat. The complicated relationships between crop production and interrelated factors cannot be described well by statistical models (Cakir, Kirci, & Gunes, 2014). Due to the aforementioned shortcomings of traditional models, many methods based on machine learning are inspired to solve yield prediction tasks recently. Ratchaphum Jaikla et al. (2008) used the Support Vector Regression method to predict rice yield and acquired comparable performance with Crop Simulation Model. Artificial neural networks (ANN) is a popular method used in crop prediction. For example, Yüksel Çakır et al. (2014) used ANN to predict wheat yield in south-east region of Turkey; Niketa Gandhi et al. (2016) applied ANN with multilayer perceptron to rice yield prediction from year 1998 to 2002 in India.

In general, rice yield data can be considered as a sequence of time series, so the prediction of rice yield can be implemented by time series predicting methods (Chen, Qi, Yuan, & Li, 2018). Due to the intra-layer nets are connectionless in the architecture of traditional neural network, it results in bad performance in dealing with problems of time series. Long short-time memory recurrent neural network (LSTM) is a novel but effective method to deal with time series. LSTM is designed to overcome the problem of vanishing/exploding gradient in Recurrent Neural Network (Hochreiter & Schmidhuber, 1997) and can learn to store input information for a long time (Lecun, Bengio, & Hinton, 2015). It yields excellent performance on processing sequential data. To the best of our knowledge, there is little research about applying LSTM to crop yield prediction, so this paper attempts to explore potential of LSTM in processing crop yield prediction problems. Due to the recurrent nature in the architecture of LSTM, it is deep essentially. It inspires discussion about whether LSTM can get more effective performance by deepening the depth of the network architecture. In this paper, an improved LSTM architecture called Stacked LSTM which deepens the network by adding layers on the basis of 1-layer LSTM is proposed to address task of rice yield prediction. After comparing with 1-layer LSTM and several other methods of prediction, Stacked LSTM has been found that achieved superior prediction performance.

The remaining paper is structured as follows. Section 2.1 illustrates data acquisition and preprocessing, Section 2.2 describes the specific theory of LSTM and Stacked LSTM, the evaluation methods are also described in this section. Section 3 will discuss research results achieved by comparing Stacked LSTM with 1-layer LSTM, ARIMA, GRU, ANN. Finally, Section 4 will give conclusion of this paper.

MATERIALS AND METHODS

Data Acquisition and Preprocessing

The data of rice yield available in this work have been obtained from the year 1980 to 2017 in Heilongjiang Province, China, from the website of Heilongjiang provincial bureau of statistics. In total, there are 38 records, thereof none with missing values and none with outlier. Because the rice yield data follows a rising trend obviously, it was transformed into stationary time series through difference processing first. Note that, before using LSTM model, the data should be transformed into supervised learning problem. The data used here is shown in Table 1.

Table 1. Rice yield from 1980-2017 in Heilongjiang province

Year	1980-1991	1992-2003	2004-2015	2016-2017
Rice yield (units:10000t)	79.6	376.6	1120	2763.622
	55.7	388.3	1172.5	2819.334
	70.9	410.4	1360	
	91.5	469.9	1655.053	
	124	636	1851.45	
	162.9	860.9	1899.61	
	220.8	925.8	2277.473	
	225.7	944.3	2438.401	
	243.5	1042.2	2600.235	
	231.7	1016.3	2710.819	
	314.4	921	2797.219	
	316.2	842.8	2720.874	

Model Methods

LSTM

Long Short-Term Memory recurrent neural network (LSTM) is a kind of Recurrent Neural Network (RNN), which has gained lots of attentions in recent years with its application in many domains including time-series prediction. As mentioned in previous section, LSTM solves shortcomings of conventional RNN. In this section, more specific theory about LSTM will be given.

LSTM short for Long Short-Term Memory recurrent neural network, is skilled in "memory", especially long duration of time. LSTM unit is composed by memory cells in which information is stored. Each memory cell is controlled to update by three "gates" called forget gate, input gate, output gate. It is exactly the gate units make the remarkable advantages of LSTM, that it solves the problem about vanishing/exploding gradient in process of back-propagation and the long chronological lags (Cortez, Carrera, Kim & Jung, 2017). The output of each gate is a number between 0 and 1 calculated by sigmoid function. This number depicts the extent to which data is allowed to pass through each gate. For example, forget gate decides which information to discard and which information are necessary for prediction, if output equals 1, that means current information should be conserved completely, if output equals 0, that means current information should be discarded completely. Input gate decides which input information can flow into memory cell, that is to say, which values are worthy of remembering. Output gate decides which values are selected as output of each memory cell finally. Concrete computation process in memory cell is shown as follows:

The final output vector h_t is calculated as follows:

$$h_t = o_t * \tanh(c_t) \tag{1}$$

where o_t is output gate vector, c_t is cell state vector, and $\tanh$ is the hyperbolic tangent function, the equations of them are shown as follows:

$$o_t = \sigma\left(W_{xo}x_t + W_{ho}h_{t-1} + b_o\right) \tag{2}$$

$$c_t = f_t c_{t-1} + i_t \tanh\left(W_{xc}x_t + W_{hc}h_{t-1} + b_c\right) \tag{3}$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{4}$$

where x_t is input vector; f_t is forget gate vector, i_t is input gate vector, W_{xo} is the weight matrix from hidden layer input to output gate, W_{ho} is the weight matrix from previous cell output state to output gate, W_{xc} is the weight matrix from hidden layer input to input cell state, W_{hc} is the weight matrix from previous cell output state to input cell state, b_o, b_c are the variable biases. σ is gate activation function, sigmoid function is normally used. The equation of σ is shown as follows:

$$\sigma\left(x\right) = \frac{1}{1 + e^{-x}} \tag{5}$$

f_t and i_t are calculated as follows:

$$f_t = \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + b_f\right) \tag{6}$$

$$i_t = \sigma\left(W_{xi}x_t + W_{hi}h_{t-1} + b_i\right) \tag{7}$$

W_{xf} is the weight matrix from hidden layer input to forget gate, W_{xi} is the weight matrix from hidden layer input to input gate, W_{hf} is the weight matrix from previous cell output state to forget gate. W_{hi} is the weight matrix from previous cell output state to input gate, b_f, b_i are the variable biases.

The inner structure of a Long Short-Term Memory cell is shown in Figure 1.

Stacked LSTM

A Stacked LSTM is constructed for the task of rice yield prediction. Many researches have proved that increase of the depth in neural network can effectively improves the performance of models (Simonyan & Zisserman, 2014). More representative features can be learned by deeper neural networks (Farabet, Couprie, Najman, & LeCun, 2013). Stacked LSTM can be simply created by

Figure 1. Inner structure of a long short-term memory cell
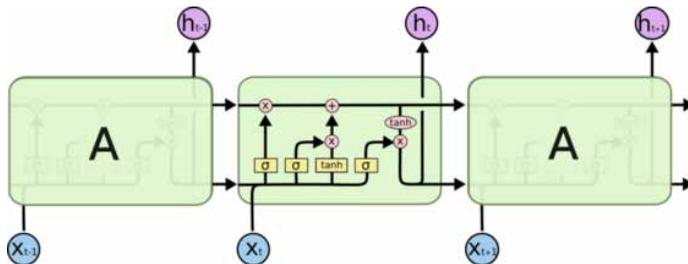
stacking multiple hidden layers on top of each other, the output of top layer is the input of the next. The units between the lower layer and the layer above it are connected by feedforward connection. The Stacked LSTM architecture used in this paper has multiple hidden layers, and the output layer is a fully connected layer. Dropout layers are introduced to prevent overfitting, and every LSTM layer is followed by a dropout layer. Input sequence flows through every LSTM layer: after calculation and state update process as mentioned above in the memory cells of each layer, then output to next LSTM layer, until the last LSTM layer feeds into the final fully connected layer and obtains the final output. Differing from some other stacked LSTM models, such as stacked bidirectional LSTM (Graves, Mohamed & Hinton, 2013) or stacked LSTM with two hidden layers (Xie, Wu, Liu, & Li, 2017) the proposed stacked LSTM model in this paper uses three original LSTM layers, which is more suitable for yield prediction, because the use of bidirectional LSTM may not make sense for certain time series prediction problems and deeper depth may give better performance. The loss function used in this experiment is MSE whose equation is shown as follows:

$$Loss = MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i^p - y_i\right)^2 \tag{8}$$

where n is the number of samples, y_i is the true values, y_i^p is the predicted values, $y_i^p - y_i$ is the residual of this model.

The optimization algorithm used in experiment is adam (Kingma & Ba, 2014). The activation function is the default hyperbolic tangent activation function $\tanh$. The epochs are set to 100, batch_size is set to 1, and the neurons of each hidden LSTM layer is set to 30. Note that, because the initial conditions of LSTM are random, in order to evaluate stability of the proposed model, the experiment under same parameters was repeated fifty times to achieve the mean number as the final result in this paper.

Model Evaluation

To assess the effectiveness of Stacked LSTM and other methods used to compare, four common indicators are used here: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percent Error (MAPE), Average Relative Accuracy (Average_RA). They are defined as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i^p - y_i\right| \tag{9}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i^p - y_i)^2} \tag{10}$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i^p - y_i}{y_i}\right| \tag{11}$$

$$Average\_RA = \frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{y_i^p - y_i}{y_i}\right)\times 100\% \tag{12}$$

It can be learned from the MAE and RMSE equations that they are mainly used to measure the deviation between the predicted values and the true values. MAE and RMSE operate on the same scale as the data, which is the main reason that we use them rather than MSE. They both represent the absolute error of time sequence. The lower the value of MAE and RMSE, which means the residual of this model is lower, the better accuracy of prediction model can achieve. MAPE is the percentage of the deviation between the true values and the predicted values divided by the true values. It represents the percentage error of the time sequence. It is same as MAE and RMSE, that the lower the values of MAPE, the prediction model will be more accurate. Average Relative Accuracy (Average_RA) measures the degree of fitting between the predicted values and real values, differing from above three indexes, the higher the values of Average_RA, the prediction model will be more accurate.

RESULTS AND DISCUSSION

In this section, three methods of time series prediction including Auto regressive Integrated Moving Average Model (ARIMA), Gated Recurrent Unit (GRU), ANN (Artificial Neural Network) are used here to compare with Stacked LSTM in order to prove the state of art performance of it. On the other hand, we also compare Stacked LSTM with different number of hidden layers, and the final results proved the importance of depth in increasing the performance of neural network. The yields in the first 26 years are used as training set, the yields of remaining 12 years are used as testing set. Note that, because passing by the difference process and the transformation process from unsupervised learning problem to supervised learning problem, the final training set is 24 years (from 1982 to 2005).

The results of MAE, RMSE, MAPE, on the testing set and Average_RA on both training dataset and testing dataset under different models are shown in Table 2. It can be seen in Table 2 that, the values of MAE, RMSE, MAPE according to the 3-layer Stacked LSTM are significantly lower than the other models, and the Average_RA is higher than others. When using the ARIMA model, MAE is 782.18 higher than the 3-layer Stacked LSTM, RMSE is 817.97 higher than the 3-layer Stacked LSTM, MAPE is 31.07 higher than the 3-layer Stacked LSTM, Average_RA is 0.3101 lower than the 3-layer Stacked LSTM on testing dataset, Average_RA is 0.0252 lower than the 3-layer Stacked LSTM on training dataset. To a large extent, this is because ARIMA cannot handle the non-stationary time series well, nevertheless the rice yield showed an increasing tendency obviously with the development of Chinese agriculture. When using the GRU model, MAE is 6.27 higher than the 3-layer Stacked LSTM, RMSE is 5.97 higher than the 3-layer Stacked LSTM, MAPE is 0.24 higher than the 3-layer Stacked LSTM, Average_RA is 0.0016 lower than the 3-layer Stacked LSTM on testing dataset, Average_RA is 0.0135 lower than the 3-layer Stacked LSTM on training dataset. GRU is an improved variant of LSTM in fact, which only has two gates including update gate and reset gate. The model of

Table 2. MAE RMSE MAPE on the test dataset and Average_RA on training and test dataset

Prediction Models	MAE	RMSE	MAPE	Average_RA (Train/Test)
ARIMA	912.55	979.24	37.35	86.13%/62.65%
GRU	136.64	167.24	6.52	87.30%/93.50%
ANN	161.03	255.16	8.11	88.91%/91.87%
1-layer LSTM	273.18	297.14	11.77	88.66%/88.46%
2-layer stacked LSTM	161.33	188.40	7.46	88.21%/92.60%
3-layer stacked LSTM	130.37	161.27	6.28	88.65%/93.66%

GRU used single layer here, it can be indicated that it performed better than 1-layer LSTM. However, when the layers of LSTM increased to three, GRU was shallower than the 3-layer Stacked LSTM, the performance of GRU was inferior to the 3-layer Stacked LSTM, which can be a powerful evidence that increasing depth can be a good idea for improving the performance of neural network. And GRU can be the next research point in the future learning. When using ANN, MAE is 30.66 higher than the 3-layer Stacked LSTM, RMSE is 93.89 higher than the 3-layer Stacked LSTM, MAPE is 1.83 higher than the 3-layer Stacked LSTM, Average_RA is 0.0179 lower than the 3-layer Stacked LSTM on testing dataset. The shortcoming of ANN that its intra-layer nets are connectionless caused the unsatisfied performance. Additionally, by comparing the results of different stacked LSTM which has various number of hidden LSTM layers, the remarkable advantage of the depth in architecture of LSTM is shown obviously. With the increase of number of LSTM layers from one to three, the MAE decreases from 273.18 to 130.37, the RMSE decreases from 297.14 to 161.27, the MAPE decreases from 11.77 to 6.28. The Average_RA increases from 88.46% to 93.66%. The predicted values which used 3-layers Stacked LSTM and real values on the training dataset and testing dataset are plotted in Figure 2, where the red line means the true yield, the green line means the predicted yield on training dataset, the blue line means the predicted yield on testing dataset. It can be seen in Figure 2 that the fitting degree of true values curve and predicted values curve is high, and according to the comparison results with other models, the higher effectiveness of 3-layers Stacked LSTM has been proved as well. Meanwhile, the experimental results also show that the Stacked LSTM architecture can be successfully applied in the task of rice yield prediction.

Furthermore, the diversity tests between GRU and 3-layer Stacked LSTM, between ANN and 3-layer Stacked LSTM are carried out so as to prove the difference between 3-layer stacked LSTM and the other comparative methods. Note that, ARIMA is not neural network, the result is invariable, so it is not considered in the diversity tests here. The P-values of diversity tests results are shown in Table 3.

The P-values are all less than 0.05, so it can be indicated obviously that the difference is significant.

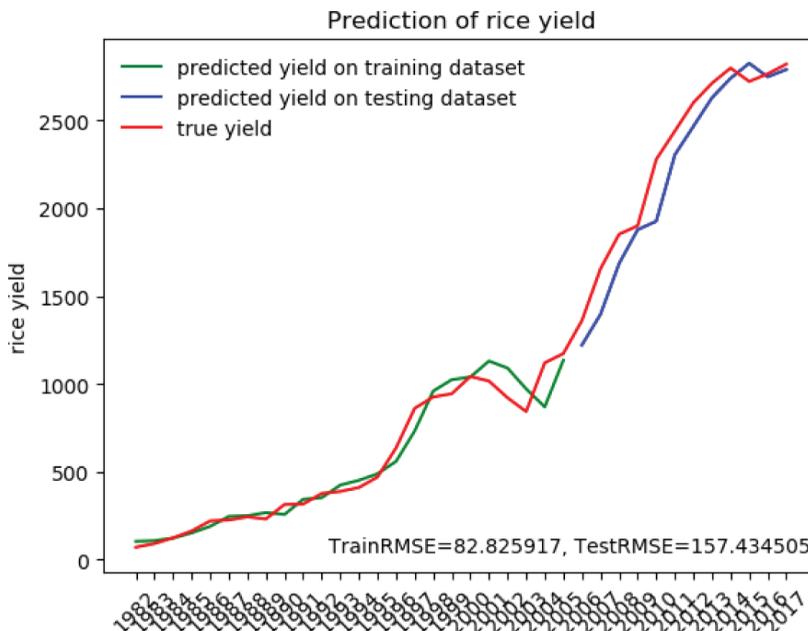**Figure 2. Comparison between true yield and predicted yield via 3-layer stacked LSTM**

Table 3. P-values of diversity tests

P-Value According to 3-Layer Stacked LSTM	MAE	RMSE	MAPE	Average_RA (Train)	Average_RA (Test)
GRU	1.5893e-06	1.9667e-05	8.6604e-05	4.9766e-24	0.0222
ANN	1.2550e-83	7.7578e-40	8.1517e-52	0.0015	9.8305e-46

CONCLUSION

In this paper, a stacked LSTM architecture was proposed. The architecture is constructed by multiple LSTM layers which are stacked on top of each other. After comparing with the other methods ARIMA, GRU, ANN on the prediction of rice yield in Heilongjiang Province, the stacked LSTM performed better than them on four indexes including MAE, RMSE, MAPE, Average_RA and the results of diversity test showed the existence of significant difference between the proposed method and the other comparison methods, which proved the efficiency of stacked LSTM consequently. More importantly, the idea about improving the performance of LSTM via increasing the depth of its architecture was proved to be effective. Furthermore, this paper provides a new method to cope with the tasks of rice yield prediction which can help government make proper decisions about management of rice production processes. On the other hand, the definition of time series is extremely simple, any sequence of data which is arranged in the order of time can be considered as time series. Therefore, yield of various crops in Heilongjiang Province or in any other provinces can also be considered as time series, the Stacked LSTM proposed here can be applied to them as well. Not only in the field of yield prediction, it can also be generalized to any fields as long as the data to be researched is time series.

ACKNOWLEDGMENT

REFERENCES

Abou-Ismail, O., Huang, J. F., & Wang, R. C. (2004). Rice yield estimation by integrating remote sensing with rice growth simulation model. *Pedosphere*, *14*(4), 519–526.

Alberto, G. S., Juan, F. S., & Waldo, O. B. (2014). Attribute selection impact on linear and nonlinear regression models for crop yield prediction. *The Scientific World Journal*, *2014*, 1–10. PMID:24977201

Cakir, Y., Kirci, M., & Gunes, E. O. (2014). Yield prediction of wheat in south-east region of Turkey by using artificial neural networks. *Proceedings of the Third International Conference on Agro-Geoinformatics*, Beijing, China. Academic Press. doi:10.1109/Agro-Geoinformatics.2014.6910609

Chen, W., Qi, W. Y., Yuan, F. X., & Li, Z. M. (2018). Comparative analysis of rice yield prediction in jilin province based on time series and cross section data. *China Agricultural Informatics*, *30*(5), 91–101.

Cortez, B., Carrera, B., Kim, Y. J., & Jung, J. Y. (2017). An architecture for emergency event prediction using lstm recurrent neural networks. *Expert Systems with Applications*, *97*, 315–324. doi:10.1016/j.eswa.2017.12.037

Daniela, R. C. et al.. (2018). Impacts of variable soil drying in alternate wetting and drying rice systems on yields, grain arsenic concentration and soil moisture dynamics. *Field Crops Research*, *222*, 101–110. doi:10.1016/j.fcr.2018.02.026

Dey, U. K., Masud, A. H., & Uddin, M. N. (2017). Rice yield prediction model using data mining. *Proceedings of the International Conference on Electrical, Computer and Communication Engineering (ECCE),* Cox's Bazar, Bangladesh. Academic Press. doi:10.1109/ECACE.2017.7912925

Farabet, C., Couprie, C., Najman, L., & LeCun, Y. (2013). Learning Hierarchical Features for Scene Labeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(8), 1915–1929. doi:10.1109/TPAMI.2012.231 PMID:23787344

Gandhi, N., Petkar, O., & Armstrong, L. J. (2016). Rice crop yield prediction using artificial neural networks. In *IEEE Technological Innovations in ICT for Agriculture and Rural Development*. Chennai, India: TIAR. doi:10.1109/TIAR.2016.7801222

Graves, A., Mohamed, A. R., & Hinton, G. (2013). Speech Recognition with Deep Recurrent Neural Networks. *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, Canada. IEEE Press. doi:10.1109/ICASSP.2013.6638947

Hirooka, Y., Homma, K., Shiraiwa, T., Makino, Y., Liu, T., Xu, Z., & Tang, L. (2018). Yield and growth characteristics of erect panicle type rice (Oryza sativa, L.) cultivar, Shennong265 under various crop management practices in Western Japan. *Plant Production Science*, *21*(1), 1–7. doi:10.1080/1343943X.2018.1426993

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. doi:10.1162/neco.1997.9.8.1735 PMID:9377276

Hutchinson, C. F. (1991). Uses of satellite data for famine early warning in subSaharan Africa. *International Journal of Remote Sensing*, *12*(6), 1405–1421. doi:10.1080/01431169108929733

Inoue, Y. M., Susan, M., & Horie, T. (1998). Analysis of spectral measurements in paddy field for predicting rice growth and yield based on a simple crop simulation model. *Plant Production Science*, *1*(4), 269–279. doi:10.1626/pps.1.269

Jaikla, R., Auephanwiriyakul, S., & Jintrawet, A. (2008). Rice yield prediction using a Support Vector Regression method. *Proceedings of the 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*. Krabi, Thailand. Academic Press. doi:10.1109/ECTICON.2008.4600365

Jeon, J. S., Jung, K.-H., Kim, H.-B., Suh, J.-P., & Khush, G. S. (2011). Genetic and Molecular Insights into the Enhancement of Rice Yield Potential. *Journal of Plant Biology*, *54*(1), 1–9. doi:10.1007/s12374-011-9144-0

Kingma, D., & Ba, J. (2014). Adam: A method for stochastic optimization. *Computer Science*.

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. doi:10.1038/nature14539 PMID:26017442

Liu, Q. P. (2011). Grain yield prediction of Henan Province based on spatio-temporal regression model. *Asian Journal of Agricultural Research*, *3*(8), 58–60.

Na-udom, A., & Rungrattanaubol, J. (2015). A Comparison of Artificial Neural Network and Regression Model for Predicting the Rice Production in Lower Northern Thailand. In *Information Science and Applications* (pp. 745–752). Springer; . doi:10.1007/978-3-662-46578-3_88

Pal, R. K., Tripathi, P., & Rao, M. M. N. (2012). Prediction of growth and yield attributes of wheat using statistical regression model in eastern Uttar Pradesh. *Environment and Ecology*, *30*(4A), 1439–1444.

Pirmoradian, N., & Sepaskhah, A. R. (2006). A very simple model for yield prediction of rice under different water and nitrogen applications. *Biosystems Engineering*, *93*(1), 25–34. doi:10.1016/j.biosystemseng.2005.09.004

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *Computer Science*.

Xie, X., Wu, D., Liu, S., & Li, R. (2017). Iot data analytics using deep learning.

Xiangyan Meng received her B.Sc. degree in computational mathematics and applied software from the Harbin Institute of Technology University, China, in 1997, her M.Sc. degree in ecological mathematics from Northeast Agricultural University, China, in 2008, and the Ph.D. degree in agricultural water and soil engineering from Northeast Agricultural University, China, in 2014. Currently, she is an associated professor in the College of Science, Northeast Agricultural University, China. Her research interests include agricultural water and soil engineering and agricultural data analysis.

Muyan Liu received her B.Admin degree in Engineering Management from the Northeast Petroleum University, China, in 2018. Currently, she is a graduate student in Management Science and Engineering from Northeast Agricultural University, China. Her research interests include agricultural data analysis.

Qiufeng Wu received his Ph.D in computer application technology from Harbin Institute of Technology, China, in 2014. He is working as associated professor in College of Science in Northeast Agricultural University. He is CCF member. His research interests include machine learning, computer vision and smart agriculture.