STM JOURNALS

Research & Reviews: Journal of Agricultural Science and Technology
ISSN: 2278-2206 (Online), ISSN: 2349-3682 (Print)
Volume 7, Issue 3
www.stmjournals.com

Predicting the Rice Production of Bangladesh by Machine Learning Technique

Shohel Mahmud*
Department of Statistics, Jahangirnagar University, Savar, Dhaka, Bangladesh

Abstract

*Bangladesh is an agricultural country and its economic condition largely depends on agriculture. The country produces much agricultural merchandise like rice, jute, wheat, onion, chilly, banana, garlic, ginger, pulse and so on. However, rice (Oryza Sativa) is produced most widely all over the country. Moreover, United States Department of Agriculture (USDA), 2017 estimate that the rice production of Bangladesh is 34.7 million metric tons for the period 2017–2018 and the position of Bangladesh is after China, India, and Indonesia. In addition, Bangladesh produces three types of rice which are Aus, Aman, and Boro. A huge portion of the population in Bangladesh immensely depends on rice as the main food. So, this paper attempts to predict the rice production of Bangladesh with the help of machine learning model like Artificial Neural Network (ANN). This paper considers a secondary data set of yearly rice production in Bangladesh over the period 1971–1972 to 2014–2015. This paper identifies the most suitable neural network model with architecture ANN 3*3*2*1 based on model selection criteria like MSE. Thus, this paper suggests an ANN model for envisaging the rice production of Bangladesh.*

Keywords: *Machine learning, prediction, rice prediction, neural network*

****Author for Correspondence*** E-mail: mahmud.shohel40@gmail.com

INTRODUCTION

Bangladesh lies in northeastern South Asia with a land mass of 144,000 km^2. The country is bounded by India on the west, north, and northeast; by Myanmar on the southeast; and by the Bay of Bengal on the south. Except for the hilly regions in the southeast and some in the northeast, and patches of highlands in the central and northwest regions, Bangladesh consists of low, flat, fertile land. About 230 rivers and their tributaries, with a total length of 24,140 km, flow across the country down to the Bay of Bengal. The alluvial soil is continuously enriched by heavy silt deposited by the rivers through frequent flooding during the rainy season. The country enjoys a subtropical monsoon climate. Summer, monsoon, and winter are the most prominent of six distinct seasons. Winter, which is pleasant, extends from November to February, with minimum temperature ranging from 7 to 13ºC; in summer, maximum temperature ranges from 24 to 41ºC [1].

The weather of Bangladesh is amicable for the rice production. About 75% of the agricultural land use for rice production and 28% of GDP comes from the exporting rice [2]. So, the total yield of rice in Bangladesh seeks a great importance to predict.

Ji et al. (2007) [3] investigated an artificial neural network (ANN) models could effectively predict Fujian rice yield for typical climate conditions of the mountains region which leads to comparing the effectiveness of multiple linear regression models with ANN models. There have been a number of reported studies that have used ANNs to set up a model for rice production. Jabjone and Jiamrun (2013) [4] developed an ANN model based on the result of RMSE and MAPE to predict the rice production of Phimai district, Thailand. Data from 2002 to 2007 were used as the training data to predict the rice yield between 2008 and 2012. The input data from six meteorological factors; rainfall, water distribution, evapotranspiration, temperature, humidity and wind speed were used.

Decision-making processes in agriculture often require reliable crop response models.

The Fujian province of China is a mountainous region where weather aberrations such as typhoons, floods and droughts threaten rice production. Agricultural management specialists need simple and accurate estimation techniques to predict rice yields in the planning process [5].

Sometimes meteorological factors; rainfall, water distribution, evapotranspiration, temperature, humidity, wind speed etc. are not available in third world country like us. So, here a model is developed to predict the production by only one variable. In this paper, we want to present an ANN model to predict the production of Aus, Aman, Boro in Bangladesh on the basis of some model selection criterion. Different activation functions were tested along with the different combination of hidden layers to get a suitable model.

MATERIALS AND METHODS

To set the model, secondary data of the rice production in Bangladesh over the period 1970–1971 to 2014–2015 published by Bangladesh Bureau of Statistics (BBS) is used. Total 41 rice production data were analyzed, where three lag values were taken. 75% data were used as training data to train the network. To predict the production of the next year of Aus, Aman, Boro and total yield of rice, R 3.4.1 version and R Studio 1.0.35 were used to set up the model of the rice production.

ANN

Neural Network (NN) often referred to as Artificial Neural Network to distinguish from the Biological Neural Networks. The key element of this illustration is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements which is called neurons, working in unison to solve specific problems. ANNs like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons. This is true of ANNs as well.

The process of receiving and sending signal is done in particular manner like a neuron receives signals from other neuron through dendrites. The neuron send signals at spikes of electrical activity through a long thin stand known as an axon and an axon splits this signals through synapse and send it to the other neurons [6].

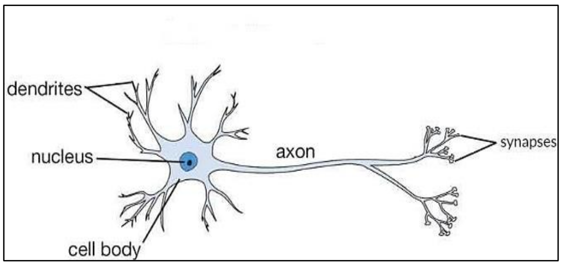

Fig. 1: *Structure of Human Neuron.*

Similar to biological neuron (Figure 1 [6]), Artificial Neural Network also have neurons which are artificial and they also receive inputs from the other elements or other artificial neurons and then after the inputs are weighted and added, the result is then transformed by a transfer function into the output [7].

In this study, model was developed on the basis of the multilayer perception (Figure 2 [8]) using backpropagation algorithm with weights. Tangent hyperbolic is used as the activation function of these models. The mathematical form of tangent hyperbolic function can be written as

$$tanh\,x = \frac{sinh\,x}{cosh\,x}$$

Hence, the function of tanhx can be defined as

$$tanh\,x = \frac{e^x - e^{-x}}{2} \div \frac{e^x + e^{-x}}{2} = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

And the graph of tanh x can be sketched as in Figure 3 [9].

Model Selection Criteria

Different model selection criteria have been used to select the appropriate model. Like AIC, BIC, MSE, MAPE. In this paper I have considered the MSE to select the best fitted model. MSE of an estimator measures the average of the squares of the errors or deviations-that is the difference between the estimator and what is estimated.

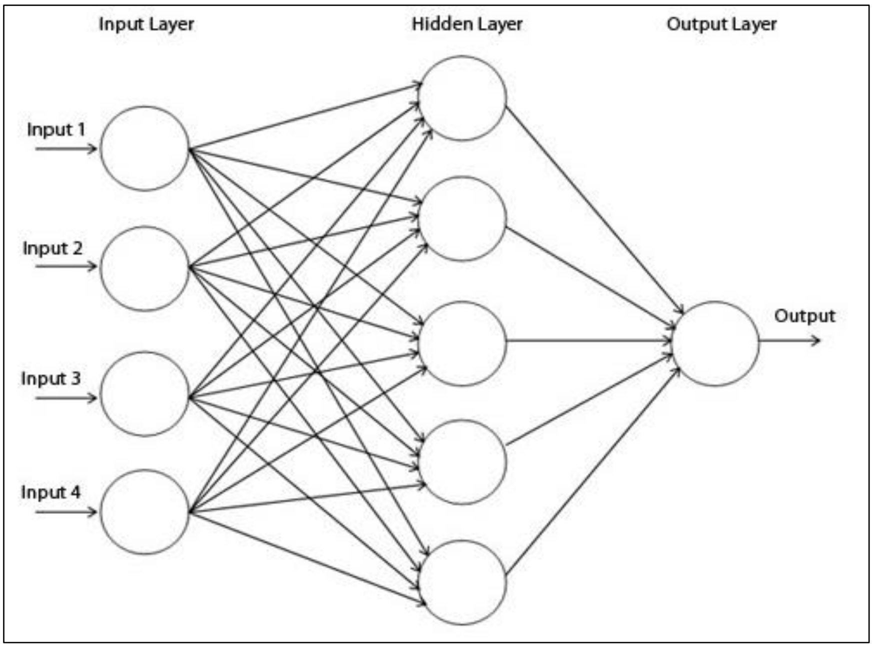

Fig. 2: *Multilayer ANN Architecture.*

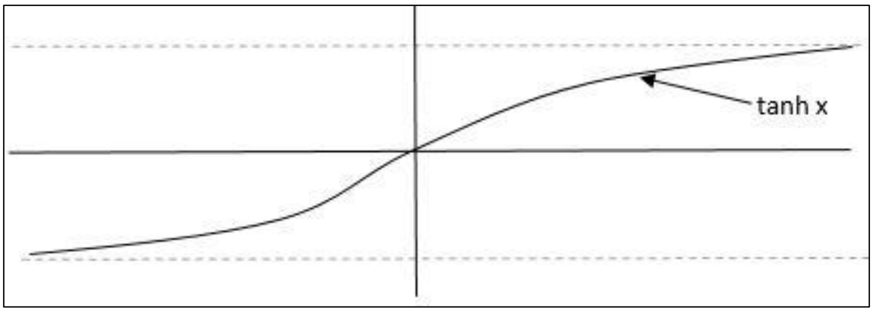

Fig. 3: *Curve of tanh Function.*

The MSE assesses the quality of an estimator or a predictor. The general from of MSE can be written as

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$$

Where, $\hat{Y}$is a vector of n predictors, Y is the vector of observed values of the variable is being predicted [10].

RESULTS AND DISCUSSION

Based on the multilayer perception, the architecture of the neural network has been made. Different activation functions were used for forecasting the output. Among them tanh activation function produced the minimum error. The data set was split into train and test dataset to set up the model and to estimate how well the model has been trained. Different seeds values, initial weights of the nods and the different number of hidden layer were used to set up the ultimate architecture. Different meteorological factors were used to predict production more precisely in Phiami district of Thailand [9]; those records are not available or up to date in Bangladesh. As a result, lag value of the production of the rice was used for predicting the future value. Though RMSE was used to select the best model [6, 5] here the best model was selected on the basis of the MSE.

Table 1 shows the MSE value for the three types of rice according to train, test dataset and for the overall model. Table 2 shows the final weighs for the different nods of the architecture for predicting the Aus. Figure 4 shows the ANN architecture for Aus. Similarly, Table 3 shows the final weights for Aman and Table 4 shows the weights for Boro. As well as, Figure 5 shows the ANN architecture of Aman and Figure 6 shows the ANN architecture of Boro.

Table 1: *Mean Sum of Square Error for Test, Training and Validation Three Sets of Data.*

	Test	**Train**	**Validation**
Aus	0.0238358	0.0038552	0.0087285
Aman	0.0123357	0.0076708	0.0088086
Boro	0.0300681	0.0077702	0.0132087

Table 2: *Estimated Weights of the Architecture for Aus.*

Predictor		**Predicted**					
		Hidden Layer 1			**Hidden Layer 2**		**Output Layer**
		H (1:1)	**H (1:2)**	**H (1:3)**	**H (2:1)**	**H (2:2)**	**Tt**
Input Layer	(Bias)	1.5362	1.31186	-0.3210			
	t	0.3491	0.01874	-0.2949			
	t1	-2.4720	-0.8752	-1.8236			
	t2	1.9000	-0.2610	0.0685			
Hidden Layer 1	(Bias)				2.7841	-1.1962	
	H (1:1)				1.2085	0.84858	
	H (1:2)				0.2252	-0.4419	
	H (1:3)				1.1961	-0.3918	
Hidden Layer 2	(Bias)						0.9024
	H (2:1)						0.3555
	H (2:2)						2.2080

Table 3: *Estimated Weights of the Architecture for Aman.*

Predictor		**Predicted**					
		Hidden Layer 1			**Hidden Layer 2**		**Output Layer**
		H (1:1)	**H (1:2)**	**H (1:3)**	**H (2:1)**	**H (2:2)**	**Tt**
Input Layer	(Bias)	0.2245	0.5861	-0.1822			
	t	1.8405	2.2992	-1.0315			
	t1	-0.1041	0.0606	-0.0676			
	t2	0.1236	-3.0318	0.2869			
Hidden Layer 1	(Bias)				0.0992	1.0377	
	H (1:1)				-0.4168	-0.2986	
	H (1:2)				-0.2604	-0.4409	
	H (1:3)				-1.3848	-0.7297	
Hidden Layer 2	(Bias)						-0.8280
	H (2:1)						1.4972
	H (2:2)						0.9490

Table 4: *Estimated Weights of the Architecture for Boro.*

Predictor		**Predicted**					
		Hidden Layer 1			**Hidden Layer 2**		**Output Layer**
		H (1:1)	**H (1:2)**	**H (1:3)**	**H (2:1)**	**H (2:2)**	**Tt**
Input Layer	(Bias)	-0.1130	0.8739	1.3248			
	t	1.2324	-1.3778	-0.4321			
	t1	0.0400	-0.1631	0.4675			
	t2	-0.3860	-0.1717	-0.0542			
Hidden Layer 1	(Bias)				-0.1210	-1.9308	
	H (1:1)				-0.7877	-1.1677	
	H (1:2)				-0.8795	-1.5916	
	H (1:3)				0.6009	0.0943	
Hidden Layer 2	(Bias)						-0.2091
	H (2:1)						1.5091
	H (2:2)						-0.5615

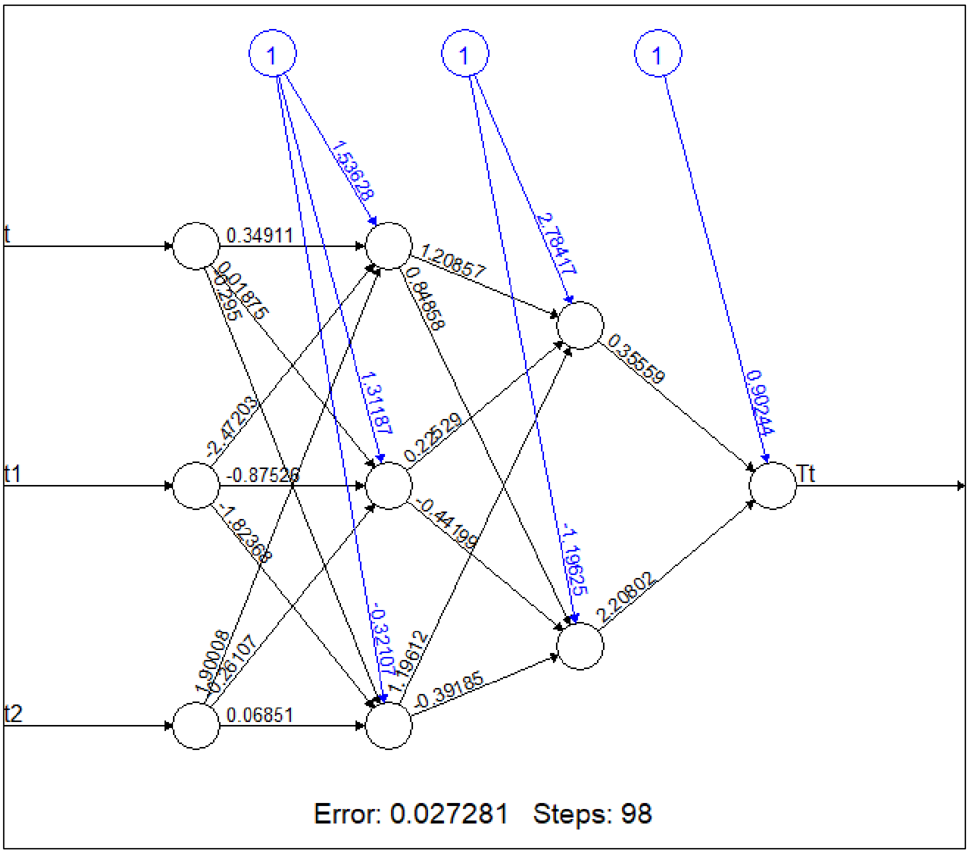

Fig. 4: *A 3*3*2*1 Architecture for Aus Prediction.*

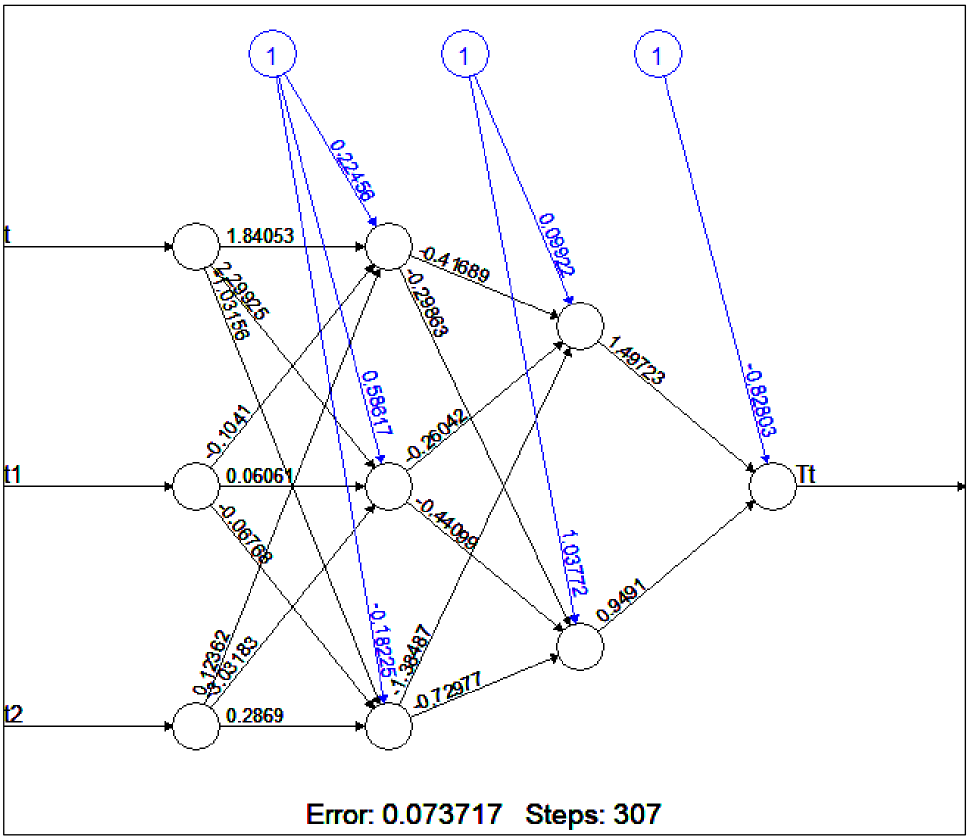

Fig. 5: *A 3*3*2*1 Architecture for Aman Prediction.*

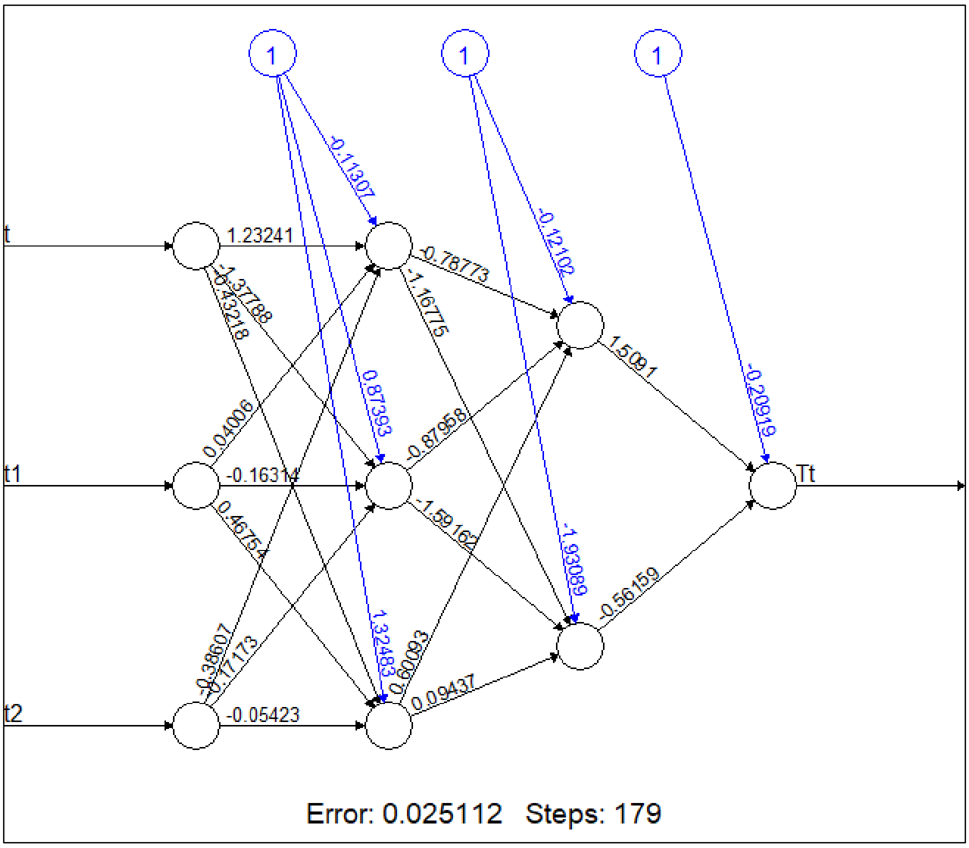

Fig. 6: *A 3*3*2**1 Architecture for Boro Prediction.*

CONCLUSIONS

Predicting the production of three types of rice in Bangladesh using ANNs has been the focus here in. Data from 1970–1971 period to 2014–2015 period were used to obtain the desire model. Models with the minimum MSE were selected. Though different forecasting methods are available, on the basis of the data and information availability ANN is used here to predict the rice production. Furthermore, ANNs is more flexible and effective in case of correlation among the data set and it does not require as much assumption as the time series model.

ACKNOWLEDGEMENTS

I'm grateful to Bangladesh Bureau of Statistics (BBS), from where I have collected the dataset. Besides, I would like to thank to my honorable teachers of my department to guide me and for their valuable suggestions.

REFERENCES

1. Maclean J, Hardy B, Hettel G. *Rice Almanac*. 4th ed. International Rice Research Institute. 1993.
2. Heitzman, James, Worden, Robert L. Bangladesh: A Country Study: Rice. Washington, D.C: Library of Congress; 1988.
3. Ji, B., Sun, Y., Yang, S. and Wan, J. Artificial neural networks for rice yield prediction in mountainous regions. *The Journal of Agricultural Science*. 2007; 145, pp.249-261.
4. Jabjone S. Artificial Neural Networks for Predicting the Rice Yield in Phimai District of Thailand. *Int J Electr Energy*. 2013; 177–181p.
5. Stergiou C, Siganos, D. *Neural Networks* (n.d.). [online] Doc.ic.ac.uk. Available from: https://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html.
6. Medium. *Suspended – Medium* (2017). [online] Available from: https://medium.com/@xenonstack/overview-of-artificial-neural-networks-and-its-applications-2525c1addff7.
7. Li, E. Artificial neural networks and their business applications. *Information & Management*. 1994; 27, pp.303-313.

8. GitHub. (2018). cazala/synaptic. [online] Available at: https://github.com/cazala/synaptic/"
9. Hyperbolic functions. (2006). [ebook] p.7. Available from: http://www.mathcentre.ac.uk/resources/workbooks/mathcentre/hyperbolicfunctions.pdf.
10. En.wikipedia.org. *Mean squared error* (n.d.). [online] Available from: https://en.wikipedia.org/wiki/Mean_squared_error.

Cite this Article
Shohel Mahmud. Predicting the Rice Production of Bangladesh by Machine Learning Technique. *Research & Reviews: Journal of Agricultural Science and Technology*. 2018; 7(3): 7–13p.