The Bioscan 12(1): 177-181, 2017
AN INTERNATIONAL QUARTERLY JOURNAL OF LIFE SCIENCES
www.thebioscan.com

COMPARATIVE STUDY OF THE PERFORMANCE OF ARTIFICIAL NEURAL NETWORK AND MULTIPLE LINEAR REGRESSION TECHNIQUES FOR PREDICTING THE SOYBEAN YIELD USING ITS ATTRIBUTING CHARACTERS

MINAKSHI MISHRA[1*], BHAVYASHREE .S[1] AND H. L. SHARMA[2]
[1*]Department of Agricultural Statistics, BCKV, Nadia - 741 252, West Bengal, INDIA.
[2]Department of Mathematics & Statistics, JNKVV, Jabalpur - 482 004, Madhya Pradesh, INDIA
e-mail: mmishra02sep@gmail.com

KEYWORDS
ANN
MLR
Soybean
Yield.

Received on :
04.12.2016

Accepted on :
25.02.2017

***Corresponding author**

ABSTRACT

The present study was conducted at the Research Farm of Department of Entomology, Adhartal, JNKVV, Jabalpur, Madhya Pradesh to compare the performance of artificial neural network (ANN) and multiple linear regression (MLR) models for predicting the yield of soybean (JS-335) crop. In this study, the data concerning about yield and its attributing characters *i.e.* number of pod per 10 plant, number of grain per 10 plant, weight of grain per 10 plant and plant height were taken. ANN and MLR models were fitted using the results of experiments to predict total yield of soybean crop. ANN of 7 neurons is more accurate with high value of R-square (0.998) and low values of residual error than MLR prediction. Predicted soybean yield by ANN was found to be close to observed soybean yield and gave good results, so it can be recommended for predicting soybean yield.

INTRODUCTION

Glycine max, commonly known as soybean in North America or soya bean, is a species of legume native to East Asia, widely grown for its edible bean which has numerous uses. The plant, classed as an oilseed rather than a pulse by the UN Food and Agriculture Organization, produces significantly more protein per acre than most other uses of land.It has the highest protein (42%) and (20%) oil among the grain legumes (Gopalan *et al.*, 1994), rich in lysine and vitamins A, B and D.

An ANN, usually called neural network (NN), is a mathematical model or computational model that is inspired by the structure and/or functional aspects of biological neural networks. A neural network consists of an interconnected group of artificial neurons, and it processes information using a connectionist approach to computation. Each ANN of an input layer, one or more hidden layers and an output layer has been established as the following general model (Figure 1) is defined by Gevery *et al.* (2003).

There are important limitations, including the necessity of observing regression assumptions significant non-linear relationships and multipleco linearity between independent variables led to be predicting the MLR model non efficient (Molazem *et al.*, 2002). However, the prediction in an ANN method always takes place according to any data situation (without limitation) based on initial training prescribed by Adielsson, 2005. Other studies have compared two methods to show that in predicting the dependent variable, the ANN method results are more accurately than MLR (Adielddon, 2005; Diane *et al.*, 2007; Gail*et al.*, 2005; Pastor, 2005; Salt *et al.*, 1999 and Starett *et al.*, 1998). Although the ANN is considered as a powerful technique for non-linear models (Ozesmi and Ozesmi, 1999). But some researchers in this linear model have also applied and reported it better than the regression model (Manel *et al.*, 1999; Miao *et al.*, 2006; Ozesmi and Ozesmi, 1999 and Starett *et al.*, 1998).). According to the present constraints on regression techniques to predict yield ($\hat{Y}_i$) based on its attributing characters and its importance, this research on a real data from farm experiments of soybean to compare two methods of regression and artificial neural network approach were used. this research on a real data from farm experiments of soybean to compare two methods of regression and artificial neural network approach were used.

MATERIALS AND METHODS

Farm experiments and numerical resources (database)

An experiment in a randomized complete block design with variety JS-335 of soybean was planted in three replications of one year in the research farm of Department of Entomology, Adhartal, JNKVV, Jabalpur (M.P). Totally 21 samples was selected. In addition to yield per plot, important yield attributing characters such as number of pod per 10 plants, number of grain per 10 plants, weight of grain per 10 plants and plant height were measured.

Multiple linear regression (MLR) model

A multiple regression equation is used to describe linear relationships involving more than two variables. A multiple linear regression equation expresses a linear relationship between a response variable y and two or more predictors variable ($X_1, X_2, ..., X_K$) . The general form of a multiple regression equation is used by Tushar and Bhatt, 2013:

$Y = b_0 + b_1X_1 + b_2X_2 + ... + b_kX_k$ (I)

A multiple linear regression equation identifies the plane that gives the best fit to the data,

$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4$ (II)

Where,

Y : Predicted value of soybean yield (SY)

X_1: Number of pod per 10 plants (NOP)

X_2: Number of grain per 10 plants (NOG)

X_3: Weight of grain per 10 plants (WOG)

X_4: Plant height (Plht)

b_0: estimate value of y-intercept

b_1, b_2, b_3, b_4: estimate value of the independent variable coefficient.

Artificial neural network (ANN)

The ANN can handle continuous as well as discrete data and have good generalization capability as with fuzzy expert systems. An ANN is a computational model of the brain. They assume that the computation is distributed over several simple units called neurons, which are interconnected and operate in parallel thus known as parallel distributed processing systems. Implicit knowledge is built into a neural network by training it. Several types of ANN structures and training algorithms have been proposed. Figure 2 shows a flowchart for comparison of ANN and MLR model. Both models are compared on the basis of error (Patel and Bhatt, 2013 and Tushar and Bhatt, 2013).

RESULTS AND DISCUSSION

MLR model

Linear regression is an approach to modeling the relationship between a scalar dependent variable Y and one or more explanatory variables denoted X. The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, it is called multiple linear regressions (Tushar and Bhatt, 2013).

$A_j = 115.92 + 12.32\ NOP - 1.47\ NOG + 25.04\ WOG - 3.38 Plht$ (III)

$Y = 115.92 + 12.32\ X_1 - 1.47\ X_2 + 25.04\ X_3 - 3.38\ X_4$ (IV)

The Soybean yield estimation formula (Equation III) was calculated by using MLR (Table 1). There are four chosen independent variables with 21 cases. Where the potentials independent variables are X_1 = Number of pod per 10 plants, X_2 = Number of grain per 10 plants, X_3 = Weight of grain per 10 plants, X_4 = Plant height (Plht) and the dependent variable Y = Soybean yield.

Table 1 shows the highest R Square (0.763) and Adjusted R Square (0.703) values. Hence, as it is found that formula for the Soybean yield is ideal. From the calculation we can conclude that the Soybean yield estimation formula using MLR

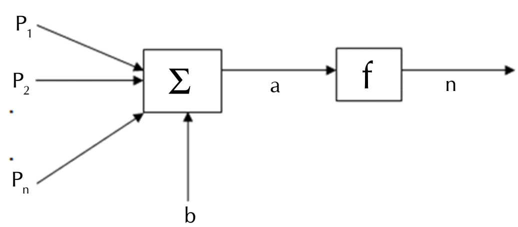

Figure 1: Artificial neural network model

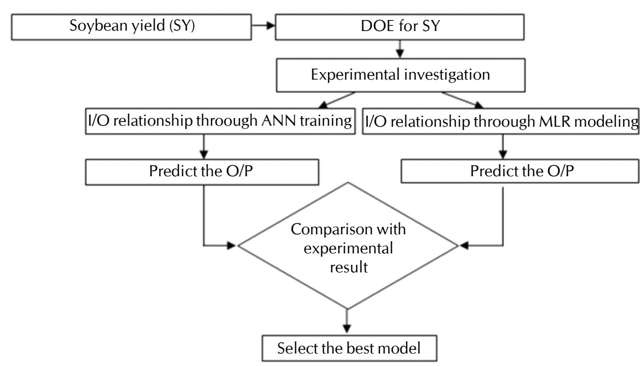

Figure 2: Flowchart for comparison of ANN and MLR model

Table 1: Multiple linear regression (MLR) analyzed data for soybean yield

Predictor	Coefficient	Standard error coefficient	T	P
Constant	115.92	288.88	0.40	0.69
NOP	12.32	4.05	3.04	0.01**
NOG	-1.47	0.50	-2.92	0.01**
WOG	25.04	7.65	3.27	0.00**
Plht	-3.38	9.02	-0.37	0.71

R-Sq = 76.3% and R-Sq (adjusted) = 70.3%, **Significant at 1%

Table 2: Analysis of variance

Source	Df	SS	MS	F	P
Regression	4	381203	95300.7	17.8482	0.00**
Residual Error	16	85432.3	5339.52		
Total	20	466635			

** Significant at 1%

Table 3: MLR and ANN Prediction Comparison Table

Sl.No.	No. of pod/10 plant	No. of grain / 10plant	Wt. of grain / 10 plant (gm)	Plant height (cm)	Soybean yield (gm)	MLR Predicted Soybean yield	MLR Error	ANN Predicted Soybean yield	ANN Error
1	38	743	52.70	34.50	777.70	691.20	86.50	777.70	0.00
2	34	764	54.70	37.10	654.70	644.80	9.90	654.70	0.00
3	34	743	58.10	36.60	733.10	778.67	-45.57	733.10	0.00
4	22	464	35.40	29.50	535.40	485.47	49.93	522.24	13.16
5	22	410	34.30	33.50	509.30	526.29	-16.99	508.54	0.76
6	18	451	34.10	36.70	384.10	400.91	-16.81	386.48	-2.38
7	16	301	22.70	33.10	372.70	332.12	40.58	372.70	0.00
8	20	469	30.80	33.20	230.80	329.50	-98.70	230.80	0.00
9	17	435	28.40	32.70	253.40	280.42	-27.02	253.40	0.00
10	28	344	24.70	34.90	349.70	460.77	-111.07	349.70	0.00
11	25	647	44.30	32.00	444.30	478.89	-34.59	444.30	0.00
12	24	520	36.40	31.80	436.40	454.92	-18.52	436.40	0.00
13	25	610	44.20	34.50	519.20	419.88	99.32	519.20	0.00
14	30	600	43.00	31.20	593.00	569.89	23.11	593.00	0.00
15	28	695	47.20	35.20	497.20	500.93	-3.73	497.20	0.00
16	18	750	48.90	32.90	348.90	248.41	100.49	348.90	0.00
17	18	344	24.00	31.70	274.00	323.43	-49.43	274.00	0.00
18	22	408	27.20	35.70	527.20	341.55	185.65	505.06	22.14
19	15	355	26.20	32.70	426.20	319.54	106.66	424.47	1.73
20	18	360	24.00	34.10	298.00	191.80	106.20	298.14	-0.14
21	19	439	30.70	34.10	305.70	353.28	-47.58	305.70	0.00

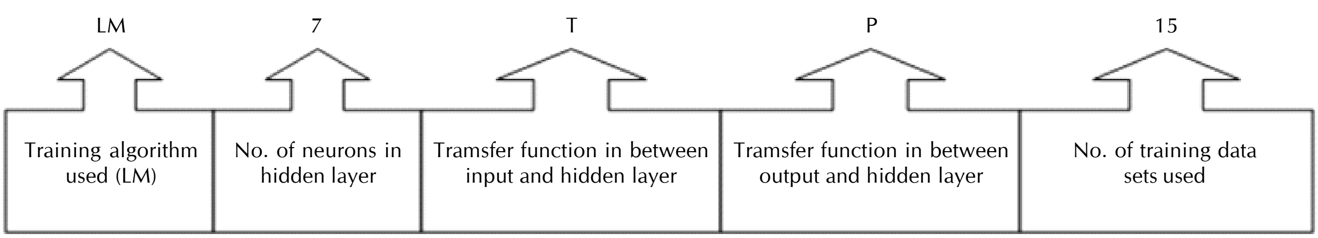

Figure 3: ANN Model designation

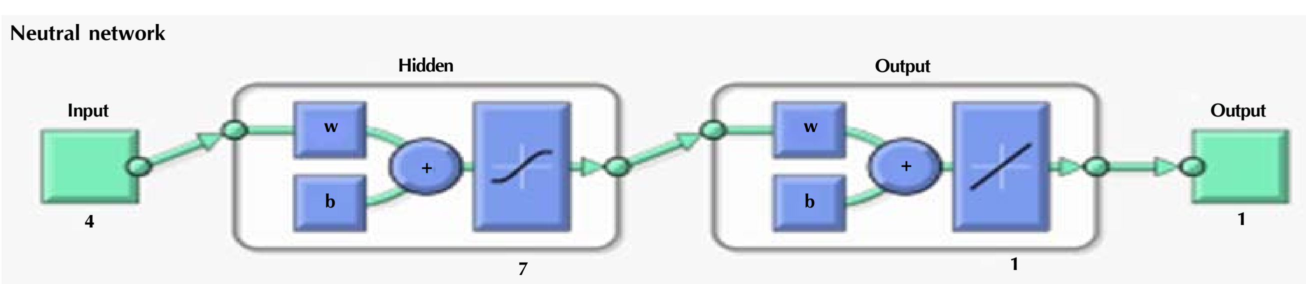

Figure 4: Abbreviated view of LM7TP Model in MATLAB window

is as shown in Equation (III) and its Analysis of variance was given in Table2.

ANN model

ANN is a massive parallel-distributed information processing system that has certain performance characteristics. Literature review shows that ANN models have better prediction capability than the regression models. ANN is found to be better than multiple regression models to find the weather based prediction for first arrival of aphid reported by Saxena and Murty (2014). So ANN models are also created for soybean yield prediction. This section describes preprocesses, model design and training, model simulation and post processes in the generation of ANN prediction models. All 21 experimental data sets are divided for training, validation and testing. Using GUI in Neural Network Toolbox in MATLAB, different network configuration with different number of hidden neurons is trained and their performance is checked. There are 15 data sets are used for training, 3 data sets for validation and 3 data sets for testing. A network is constructed each of them is trained separately, and the best network is selected based on the accuracy of the predictions in the testing phase (Diane *et al.*,

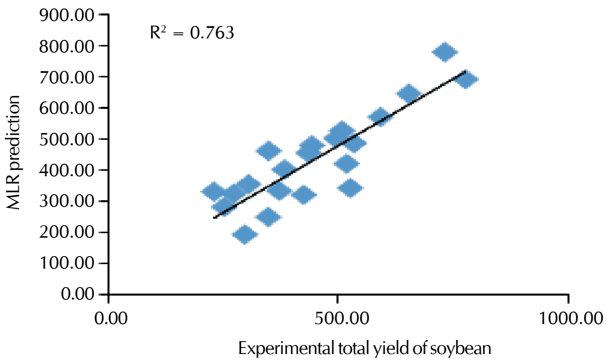

Figure 5: Regression model of MLR

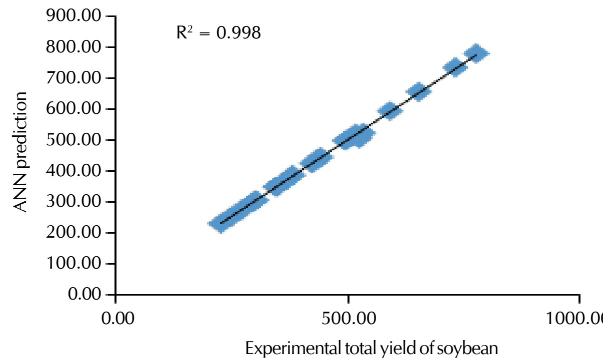

Figure 6: Regression Model of ANN

2007 and Starett *et al.*, 1998). Figure 3 suggests how ANN model is designated. In this study, ANN consists of three layers (Figure 4) where the first layer is input layer (four) which is triggered using the sigmoid activation function whereas the second layer is hidden layer (seven) and third layer is the output layer (one) which is triggered using the linear activation function. Diane *et al.*, 2007 and Lek *et al.*, 1996 used a network of two transfer function, where the first transfer function is tansig and the second transfer function is purelin, can be trained to approximate any function.

The network is trained using Levenberg-Marquardt (LM) algorithm. In the case of supervised learning, the network is presented with both the input data and the target data called the training set. The network is adjusted based on comparison of the output and target values until the outputs match the targets. After the data have been normalized, input data files and targets data files are created for training purpose. These input data files include file for training, validation and testing which contains input data sets in random order. Target data files include targets (normalized measured soybean yield values respectively of input data sets) for training, validation and testing data sets. The work in this paper included a function approximation or prediction problem that required the final error to be reduced to a very small value (Patel *et al.*, 2013 and Zahouily *et al.*, 2002).

Comparison of ANN and MLR

Table 3 shows MLR and ANN prediction comparison of soybean yield. Figure 5 shows a regression model of MLR and Figure 6 shows a regression model of ANN. The result shows that ANN technique is more feasible in predicting the soybean yield than the MLR technique. The ANN model could probably predict soybean yield with a better performance owing to their greater flexibility and capability to model nonlinear relationships. Therefore, in the case of data sets with a limited number of observations in which regression models fail to capture reliably, advanced soft computing approaches like ANN may be preferred (Patel and Bhatt, 2013 and Starett *et al.*, 1998).

REFERENCES

Adielsson, S. 2005. Statistical and neural networks analysis of pesticide losses to surface water in small agricultural catchments in Sweden. M.Sc. Thesis, Sweden University, Sweden.

Diane, M. L. and David, P. A. 2007. For predicting facial cal form concentrations.*Hydrological Sci. J.* **52:** 713-731.

Gail, B., Viswanthan, C., Nelakantan, T. R., Srinivasa, L., Girones, R., Lees, D., Allard, A. and Vantarakis, A. 2005. Artificial neural networks prediction of viruses in shellfish. *Appl. and Environ. Microbiol.***31:** 5244-5253.

Gevery, M., Dimopoulos, I. and Lek, S. 2003. Review and comparison of methods to study the contribution of variable in artificial neural network models. *Ecological Modeling.* **160:** 249-264.

Gopalan, C., Ramashastry, B. V. and Balasubramanian, S. C. 1994. Nutritive value of Indian foods.Indian Counc.Med. Res. pp.24-26.

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I. Lauga, J. and Aulagnier, J. 1996. Application of neural networks to modeling nonlinear relationships in Ecology.*Ecological Modeling.* **90:** 39-52.

Manel, S., Dias, S. M. and Ormerod, S. J. 1999. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river Bird.*Ecological Modeling.* **120:** 337-347.

Miao, Y., Mulla, D. and Robert, P. 2006. Identifying important factors influencing corn yield and grain quality variability using artificial neural networks.Springer.**7:**117-135.

Molazem, D., Valizadeh, M. and Zaefizadeh, M. 2002. North West of genetic diversity of wheat. *J. Agricultural Sci.* **20:** 353-431.

Ozesmi, S. L. and Ozesmi, U. 1999. An artificial neural network approach to spatial habitat modeling with interspecific interaction.*Ecological Modeling.***116:** 15-31.

Pastor, O. 2005.Unbased sensitivity analysis and pruning techniques in ANN for surface ozone modeling. *Ecological modeling.* **182:** 149-158.

Patel, T. M. and Bhatt, N. M. 2013. Shear Stress Prediction Using FEA-ANN Hybrid Modeling Of Eicher 11.10 Chassis Frame. *IOSR, J.f Mechanical and Civil Engineering.*pp. 22-32.

Salt, D. W., Yildiz, N. and Livingstone, D. J. 1999. The use of artificial neural networks in QSAR.*Pesticide Sci.* **36:** 161-170.

Saxena, S. and Murty, N. S. 2014. Weather based model development for outbreak of mustard aphid (*L.Erysimi.,Kalt*) using artificial neuralnetwork. *The Ecoscan.* **8(1&2):** 47-52.

Starett, S. K., Najjar, Y., Adams, S. G. and Hill, J. 1998. Modeling pesticide leaching from golf courses artificial neural networks.*Communications in Soil Science and Plant Analysis.***29:** 3093-3106.

Tushar, M. P. and Bhatt, N. 2013. ANN and MLR Model for Shear Stress Prediction of Eicher 11.10 Chassis Frame: A Comparative Study. *International Journal of Mechanical Engineering and Technology (IJMET).* **4(5):** 216 - 223.

Zahouily, M., Rhihil, A., Bazoui, H., Sebti, S. and Zakarya, D. 2002. Structure-toxicity relationships study of a series of organophosphorus insecticides.*Molecular modeling annual*. **8(5):** 168-172.