Journal of Crop and Weed, 13(1) : 93-96 (2017)

Comparative study of feed-forward neuro-computing with multiple linear regression model for predicting mulberry leaf yield

M. MISHRA, S. BHAVYASHREE AND [1]G. C. GIRISHA

Department of Agricultural Statistics, Bidhan Chandra Krishi Viswavidyalaya, Mohanpur-741252, Nadia, West Bengal

[1]ITC ILTD-ABD, Hunsur, Mysore, Karnataka, India

Received: 03-01-2017; Revised:24-02-2017; Accepted : 28-02-2017

ABSTRACT

In this study, the data of 16 mulberry genotypes for a year was collected from Department of Sericulture, UAS Bangalore. The results from multi-regression analysis using stepwise method for leaf yield showed that 3 variables (fresh leaf weight, number of leaves per plant and moisture content) having significant impacts. The main objective of this work is to compare the accuracy of Artificial Neural Network (ANN) and Multiple Linear Regression (MLR) models for prediction of leaf yield. We have compared MLR with ANN of two layer feed-forward network with sigmoid hidden neurons (10) using Levenberg-Marquardt (LM) algorithm. The performance of ANN was found to be better than the MLR model for leaf yield prediction with maximum value of R-square and minimum value of root mean square error.

Keywords : ANN, LM, MLR, Mulberry genotypes

Sericulture has become an important cottage industry in many countries. Today, China and India are the two main producers, with more than 60 per cent of the world's annual production.

Sericulture provides gainful employment, economic development and improvement in the quality of life to the people in rural area and therefore it plays an important role in anti poverty programme and prevents migration of rural people to urban area in search of employment. Hence several developing nations have taken up sericulture to provide employment to the people in rural area.

Success of sericulture enterprise depends on successful rearing of silkworms. Among the different types of silk production practices in India, Mulberry silk is the most important which is contributing to the tune of 90 per cent of the total production in India. Karnataka is the pioneering state contributing about 40 per cent of Indian silk production. Mulberry leaf is a major economic component in sericulture since the quality and quantity of leaf produced per unit area have a direct bearing on cocoon harvest. Therefore it is essential to give much emphasis to leaf yield and its attributing characters.

Most of the researchers have employed regression models for prediction purposes in various disciplines. Due to the nature of linear relationship in the parameters, regression models may not provide accurate predictions in some complex situations such as non linear data and extreme values data. Thus, Artificial Neural Network (ANN) is highly suggested to present the complicated relations between different parameters and crop yield. The use of the ANN for modeling and prediction purposes has been increasingly becoming popular in the last decades. Researchers have been applying the ANN method successfully in various fields of mathematics, engineering, medicine, economics, meteorology, psychology, neurology, in electrical and thermal load predictions, in adaptive and robotic control and so on (Cam *et al*., 2005; Sahin, 2012).

Other studies have compared two methods to show that in predicting the dependent variable, the ANN method results are more accurately than MLR (Adielsson, 2005; Pastor, 2005; Gail *et al*., 2005 and Starett *et al*., 1998). ANN is also considered as a powerful technique for non-linear models (Lek *et al*., 1996). But some researchers in this linear model have also applied and reported it better than the regression model (Mio *et al*., 2006 and Manel *et al*., 1996). Thomas *et al*. (1996) and Zaefizadeh *et al*. (2011) were also compared MLR and ANN models.

This paper compares the accuracy performance of ANN and MLR models for the prediction of Mulberry leaf yield.

MATERIALS AND METHODS

For the study data of 16 mulberry genotypes were collected from the Department of Sericulture, UAS, GKVK, Bangalore for 1 year. The data concerning leaf yield (g) and its attributing characters viz., plant height (cm), number of leaves per plant, number of branches, fresh leaf weight (g), leaf area (cm^2), total shoot length (cm), moisture content (%) and stem girth (cm) for 16 genotypes were collected for the study.

Multiple Linear Regressions

A multiple linear regression equation expresses a linear relationship between a response variable y and two or more predictors variable (x1, x2... xk). The general form of a multiple regression equation is:

Email: mmishra02sep@gmail.com

$$\hat{y} = b0 + b1x1 + b2\ x\ 2 + ... + bkxk \quad (1)$$

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

$$\hat{y} = b0 + b1x1 + b2x2 + b3x3 + b4x4 + b5x5 + b6x6 + b7x7 + b8x8 \quad (2)$$

Where, $\hat{y}$ predicted value of leaf yield, x1: plant height, x2: number of leaves, x3: number of branches, x4: fresh leaf weight, x5: leaf area, x6: total shoot length, x7: moisture content, x8: stem girth, b0: estimate value of y-intercept, b1, b2, b3……b8: estimate value of the independent variable coefficient.

Artificial Neural Network

ANN captures the domain knowledge; it 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 and it is composed of an input layer, one or more hidden layer, and an output layer. 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. The number of hidden layer and the neuron number in the hidden layer are determined by trials. In the hidden layer, inputs and relevant weights are multiplied, and then the results are transmitted to transfer function (Yongjae and Sehun, 2005).

RESULTS AND DISCUSSION

Multiple Linear Regressions

The step-wise regression analysis results revealed the significant differences among the 16 genotypes for all the character under study (Table 1). The result shows that few traits are highly significant like fresh leaf weight (X4), number of leaves per plant (X2) and moisture content (X7). Therefore, more emphasis should be laid on these characters during selection for further improvement of mulberry leaf yield (Y).

Multi-regression analysis of stepwise method for leaf yield showed that three variables having significant impacts and those that maintained in the equation as,

$$Y= -361.70 + 5.\ * X4 + 2.45 * X2 - 14.87 * X7 \quad (3)$$

Table 1: Multiple linear regression analyzed data for Mulberry leaf yield

Variables	Coefficients	Standard Error	T- value	Sig.
Constant	-361.703	443.3864	-0.8158	0.43
LW	5. 0359	0.228809	23.1650	0.00*
NOL	2.45416	0. 1059	8.1517	0.00*
MC	-14.8676	5.938395	-2.5036	0.03*

*Note: LW- fresh leaf weight, NOL- number of leaves per plant, MC- moisture content, *significant at 5%*

Table 2: Comparison on ANN and MLR

Leaf Yield	LW	NOL	MC	ANN	ANN error	MLR	MLR error
3256.82	527.86	619	54.58	3256.82	0.00	3143.19	113.63
2192.83	529.13	413	72.98	2192.83	0.00	2371.94	-179.11
369.79	83.89	437	60.98	369.79	0.00	248.78	121.01
2715.24	457.25	592	66.23	2871.65	-156.41	25 .37	184.87
2081.38	424.28	489	64.79	2081.38	0.00	2123.74	-42.37
1386.56	186.58	741	72.41	1386.56	0.00	1369.56	17.00
1346.23	201.90	664	66.21	1346.23	0.01	1354.18	-7.95
1102.43	175.58	626	63.88	1102.43	0.00	1154.64	-52.21
2037.39	391.89	538	69.98	1990.67	46.72	1996.25	41.14
1403.79	351.29	349	64.52	1403.79	0.00	1396.93	6.86
1883.37	361.87	578	62.09	1883.37	0.00	2050.61	-167.24
2247.81	447.55	504	69.03	2247.81	0.00	2220.56	27.25
1042.53	155.03	616	62.44	954.71	87.82	1043.99	-1.46
1110.95	187.60	590	58.34	1110.95	0.00	1213.03	-102.08
1528.84	207.23	736	62.01	1528.84	0.00	1620.94	-92.10
1444.97	256.85	562	71.81	1263.70	181.27	1310.21	134.76
	R^2			**0.992**		**0.970**	
	RMSE			**64.82**		**101.92**	

Note: LW- fresh leaf weight, NOL- number of leaves per plant, MC- moisture content

Artificial Neural Network

All 16 experimental data sets are divided for training, validation and testing. There are 12 data sets are used for training, 2 data sets for validation and 2 data sets for testing using MATLAB software. It is clear that more data sets in training reduces processing time in ANN learning and improves the generalization capability of models, so large number of data sets are used to train the models.

A feed-forward neural network with back propagation was used and the network consists of three layers: first the input layer (three selected variables) is triggered using the sigmoid activation function whereas the second layer is hidden layer (10 neurons) and the third layer is the output layer (one) which is triggered using the linear activation function as shown in Figure 1.

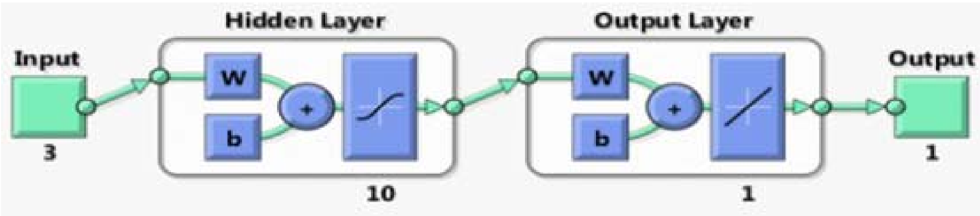

Fig. 1 : Abbreviated view of ANN Model in MATLAB Window

Comparison of MLR and ANN Model

For comparative study, ANN of two layer feed-forward network and regression analysis were used. Both models are compared on the basis of error and the results revealed that the ANN has the highest R-square and the lowest root mean square error (RMSE) for predicting leaf yield as compared to regression analysis (Table 2). It is found that the ANN model could probably predict leaf yield of mulberry with a better performance owing to their greater flexibility and capability to model linear/ nonlinear relationships. Figure 2 shows a regression model of MLR and Figure 3 shows a regression model of ANN. They show that ANN technique is more feasible in predicting the leaf yield than the MLR technique.

The present study aimed at the comparison of ANN and MLR models for mulberry leaf yield prediction. ANN model proved to be a superior methodology for accurate prediction of mulberry leaf yield with maximum value of R-square and minimum value of RMSE rather than regression analysis.

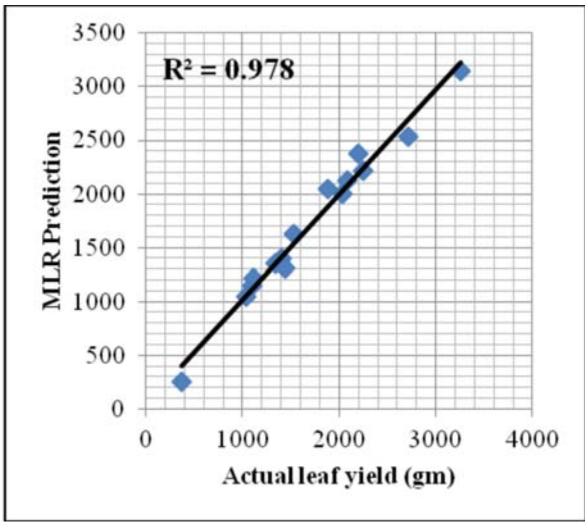

Fig. 2 : Regression model of MLR

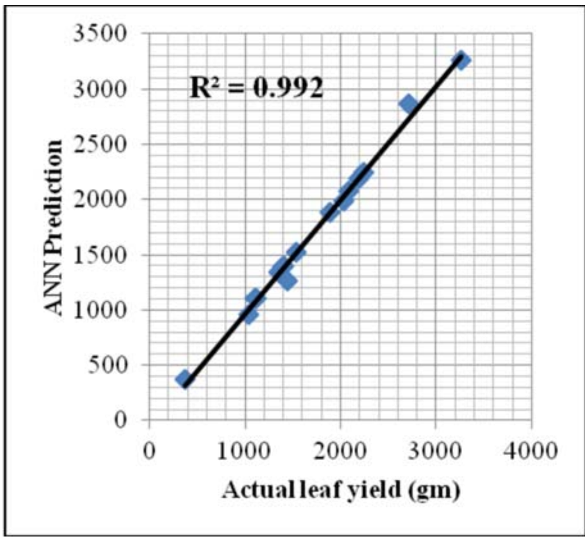

Fig. 3 : Regression model of ANN

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.

Thomas, B., Jarre-Teichmann, A. and Borlich, O. 1996. "Artificial neural network versus multiple linear regression: predicting P/B ratios from empirical data." Marine ecology progress series. *Oldendorf* **140**(1) : 251-56.

Cam, E., Arcakhoglu, E. C., Avusoglu, A., Akbiyik, B. A. 2005. Classification mechanism for determining average wind speed and power in several regions of Turkey using artificial neural networks. *Renew. Energy* : 227–39.

Gail, B., C. Viswanthan, T.R. Nelakantan, L. Srinivasa, R.Girones,D. Lees, A. Allard and A. Vantarakis. 2005. Artificial neural networks prediction of viruses in shellfish. *Appl. and Environ. Microbiol*. **31** : 5244-53.

Lek, S., M. Delacoste, P., Baran, I., Dimopoulos., Lauga, J and Aulagnier, A. 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** : 337-47.

Pastor, O. 2005. Unbased sensitivity analysis and pruning techniques in ANN for surface ozone modeling. *Ecological Modeling* **182**:149-58.

Sahin, M., Yýldýz, B.Y., Senkal, O., and Pestemalc, V. 2012. Modelling and remote sensing of land surface temperature in Turkey. *J. Ind. Soc. Remote Sens.* **40**:399–409.

Starett, S.K., Najjar, Y., Adams, S. G and Hill, J. 1998. Modelingpesticide leaching from golf courses using artificial neural networks. *Communications in Soil Science and Plant Analysis*, **29**: 93-3106.

Yongjae, K., and Sehun, R. 2005. Arc sensor model using multiple-regression analysis and a neural network. *ProQ. Sci. J*. **219**:431–47.

Zaefizadeh, M., Khayatnezhad, M and Gholamin, R. 2011. "Comparison of Multiple Linear regressions and Artificial Neural Network in Predicting the Yield Using its Components in the Hassle Barley," *American-Eurasian J. Agric. Environ. Sci.* **10** (1):60-64.