Original Research Article   https://doi.org/10.20546/ijcmas.2018.704.090

Yield Prediction in Brinjal (*Solanum melongena* CV MAHYCO-11) Across Different Growth Stages Using ANN Models

R. Hanumanthaiah[1*], R. Venugopalan[2], K. Padmini[2] and K.J. Yogeesh[1]

[1]Department.of Farm Engineering, IAS, Banaras Hindu University, Varanasi, Uttar Pradesh-221005, India
[2]Indian Institute of Horticultural Research, Bengaluru-560089, Karnataka, India

Corresponding author

A B S T R A C T

Keywords

Artificial neural network, Biometrical traits, Crop logging, Growth stages, Root mean square error, Precision farming

Article Info

Accepted:
07 March 2018
Available Online:
10 April 2018

An attempt has been made in the present investigation to assess the influence of various biometrical characters across different growth stages in Brinjal crop yield along with various yield attributing characters across four growth stages using ANN models. Results of seedling stage had indicated that plant height, girth and number of primary branches could together predict the crop yield to an extent of 83 % for training set and 62 % for validation of model accuracy. Normalized importance was 0.225, 0.142, and 0.131 respectively. In case of Vegetative stage plant height, girth and number of leaves could together predict the crop yield to an extent of 89 %, for training set and 85 % for validation of model accuracy and Normalized importance was 0.126, 0.098 and 0.133, respectively. Where as in Flowering stage both the plant spreads (east-west and north-south) could together predict the crop yield to an extent of 88 % for training set and 64 % for validation of model accuracy and Normalized Importance was 0.220 and 0.245 respectively. Finally for fruiting stage plant height, number of primary branches and plant spread (east-west) could together predict the crop yield to an extent of 68 % for training set and 64 % model accuracy. Normalized importance was 0.229, 0.134 and 0.227 respectively.

Introduction

Sustainable crop production is a key factor for achieving self-sufficiency of horticultural crop produce in our country.

Precision agriculture gives farmers the ability to use crop inputs more effectively. Precision farming is the need of the hour to achieve crop sustainability. For a successful precision farming practice in any crop research, researchers are interested to know explicitly by which stage of a standing crop, yield could be predicted more accurately and what are all the key indicators of crop yield (crop logging parameters) across all the stages, this could facilitate the horticulture production researchers to incorporate suitable agro-management measures at the early crop stages to ensure that desired yield is guaranteed and also crop improvement (breeding) research may also be benefited, as selection can be made in the early stages based on the identified significant crop-logging parameters.

Crop yield forecast before harvest, is likely to provide valuable information to the farmers/researchers/policy makers on sales, storage, export, price fixation, grading, marketing to horticultural industries and also to the government for advance planning so as to ensure sustainable crop production in the years ahead. On the other hand, researchers are interested to know explicitly by which stage of a standing crop, yield could be predicted more accurately and what are all the key indicators of crop yield (crop logging parameters) across all the stages, such information would facilitate for the horticulture production experts to incorporate suitable agro-management measures at the early crop stages to ensure that desired yield is guaranteed further crop improvement (breeding) research may also be benefited, as selection can be made in the early stages based on the identified significant crop-logging parameters. Role of biometrical factors in crop modeling is gaining importance in horticultural studies (Venugopalan, 2010). To achieve this, development of statistical models based on recorded biometrical traits across crop growth stages is highly essential.

ANN is extremely versatile approach which can handle non-linear functional relationship between response and predictor variables (Singh and Prajneshu, 2008). One of the main attractions of the ANN approach is that it does not require an explicit understanding of the mechanisms underlying the process in the phenomenon studies.

Materials and Methods

Experimental area

Material for investigation of this study is the secondary source data on Brinjal (Cv Mhyco-11) yield along with 8 yield attributing characters for four different growth stages (26 DAP: Seedling stage; 52 DAP: Vegetative phase; 72: Flowering stage; 89: Fruiting phase). This research is based on secondary data that was obtained from the Statistical unit of I.I.H.R., Bengaluru. Data were collected from farmer's field at Kolar. The aim of this study is to identify the significant crop-logging parameters influencing crop yield across growth stages using the data based on yield kg per plant and biometrical characters such as plant height, girth, plant spread (NS, EW), no of branches and leaves, leaf length, breadth and leaf area.

SAS JMP package was used for developing ANN models by dividing the data set into three parts: 90% for training (to learn pattern present in the data); 10% for validation and rest for testing (to assess the performance of trained network) by using Multi-Layer Perceptron (MLP) ANN architecture.

Observations recorded

Yield (kg per plant), Plant height (cm), Plant girth (cm), Plant spread (NS, EW) (cm), Number of branches, Number of leaves, leaf length (cm), Leaf breadth (cm) and Leaf area (cm^2).

Artificial neural network models

Multiple linear regression (MLR) modeling is a very powerful technique and is widely used to estimate linear relationship between response variable and predictors. Its main limitation is that it is useful only when the underlying relation between response and predictor variables is assumed to be "linear". However, in a realistic situation, this assumption is rarely satisfied. Also, if there are several predictors, it is well-nigh impossible to have an idea of underlying non-linear functional relationship between response and predictor variables. Fortunately, to handle such a situation, an extremely versatile approach of "Artificial neural

networks" (ANN) is developed rapidly. Cheng and Tittering ton (1994) have reviewed the ANN methodology from a statistical perspective, while Warner and Misra (1996) have laid emphasis on understanding of ANN as a statistical tool. Recently, Pratheepa *et al.,* discussed the utility of ANN models in biological studies (Pratheepa, 2011).

A distinguish feature of ANNs that makes them valuable and attractive for a statistical task is that, as opposed to traditional model-based methods, ANNs are data-driven self-adoptive methods in that there are a few a-priori assumptions about the models for problems under study.

Preliminaries of ANN

ANN can be considered as an interconnected assembly of simple processing elements (or units/nodes/neurons). The processing ability of network is stored in the inter-unit connection strengths or weights obtained by a process of learning from a set of training patterns. A typical ANN consists of one input layer, one output layer and hidden layers. Each layer can have several units whose output is a function of weighted sum of their inputs. Input in to a node is a weighted sum of outputs from nodes connected to it. Thus, net input into a node is given by equation (1):

$$Netinput_i = \sum_{ij}(w_{ij} * output_j) + u_i \quad (1)$$

Where w_{ij} are weights connecting neuron j to neuron i;

Output ^{j} is the output from the unit j; and u_i is threshold-term is base line input to a node in the absence of any other inputs. If weights w_{ij} is negative, it is termed

'inhibitory' because it decreases net input; otherwise it is called 'excitatory'. Each unit takes its net input and applies an activation function to it. For example output of the jth unit, also called activation value of the unit,

is $g\left(\sum_{ij}^{n} w_{ji} x_i\right)$, where g (.) is activation function and x_i is output of the ith unit connected to unit j. Two important activation functions commonly used are:

Pure line: g (net input) = constant. (net input)

Sigmoid: g (net input) = $\dfrac{1}{1 + e^{-netinput}}$

With no hidden units, an ANN can classify only linearly separable problems (ones for which possible output values can be separated by global hyper planes). However, it has been shown by Cybenko (1989) that with one hidden layer, an ANN can describe any continuous function (if there are enough hidden units), and that with two hidden layers, it can describe any function.

Accordingly, approach of "Artificial neural networks" (ANN), which uses the inherent pattern in the data to express the functional relationship between response and predictors, were utilized. The potential use of ANN methodology has been highlighted for successfully tackling the realistic situation in which exact nonlinear functional relationship between response variable and set of predictors is not unknown. Most widely used ANN is multilayered feed forward artificial neural network (MLFANN).

Multilayered feed forward artificial neural network (MLFANN)

An MFLANN is one in which units in the next layer, and not to units in the preceding layer.

An MFLANN can have a number of hidden units per layer. When counting layers, it is common practice not to count input layer because it does not perform any computation, but simply passes data into the next layer, so an MFLANN with an output layer is termed as two layered MFLANN. The MFLANN is the most popular network architecture. It is the type of network in which units are arranged in a layered feed forward topology. The network thus has a simple interpretation as a form of input-output model, with weights and thresholds (biases) as free parameters of the model functions of almost arbitrary complexity, with the number of layers, and the number of units in each layer, determining the function complexity.

Neural networks are constructed by learning from repeated presentation of inputs (the x's) and outputs (y's) and adjusting internal parameters so as to minimize error between fitted and desired y's. The utility and flexibility of neural network arise from the application of learning algorithms that allow the network to construct correct weights, and hence, desired function, for a given set of observations. Model performance was determined in the training and testing procedure, by comparing the RMSE between observed and predicted values. Statistical models were developed using MLFANN for first stage (seedling stage 26DAP), second stage (vegetative phase 52DAP), third stage (flowering stage 79DAP) and fourth stage (fruiting stage 89DAP) separately, and results are discussed below.

Results and Discussion

Artificial neural network models of first stage (seedling stage 26DAP)

To select the best ANN architecture, root mean square error (RMSE) was used as performance criteria. Performance of model on training data set and testing set are shown in Table 1. Optimal connection weights may be obtained by using gradient descent algorithm or conjugate gradient algorithm with sigmoid transfer function a view to minimizing sum of the squared error functions of the network output and also high R^2.

It was observed that the total length of record was 31, out of which 84% was taken for model training, 16% for model testing and 100% for model evaluation initially (Table 1). However independent variable (plant height, plant girth, number of leaves, number of primary branches, plant spread east-west and plant spread north- south) importance was found to be 0.225, 0.142, 0.167, 0.131, 0.159 and 0.176 respectively (Table 2). In the next step, for optimized ANN model total length of record was 31, out of which 73.1% was taken for model training, 26.9% for model testing and 100% for model evaluation with exclusion of 5 variables. Independent variables (plant height and plant spread east-west) importance were found to be 0.387 and 0.613 respectively. Further R^2 for training and validation are $R^2 = 0.68$ and 0.64 respectively with Root mean square error RMSE=0.007 and 0.074 respectively. A graphical representation of ANN diagram for the optimized model is also appended (Fig. 1).

Artificial neural network models of second stage (vegetative phase 52DAP)

It is observed that the total length of record was 34, out of which 89.3% was taken for model training, 10.7% for model testing and 100% for model evaluation initially (Table 1). However independent variable (plant height, plant girth, number of leaves, number of primary branches, plant spread east-west, plant spread north- south, leaf length and leaf breadth) importance was observed to be 0.126, 0.098, 0.133, 0.066, 0.129, 0.159,0.148 and 0.093 respectively (Table 2).

Table.1 Results of ANN for optimized model of 1, 2, 3 and 4 {seedling (26DAP), vegetative (52DAP), flowering (72 DAP) and fruiting (89DAP)} stages

Stage 1 (seedling stage 26 DAP)	optimized		
		Training	**Validation**
	R²	0.83	0.62
	RMSE	0.29	0.37
Stage 2 (vegetative phase 52DAP)	**R²**	0.89	0.85
	RMSE	0.14	0.18

Stage 3(flowering stage 79DAP)	optimized		
		Training	**Validation**
	R²	0.88	0.64
	RMSE	0.13	0.23
Stage 4 (fruiting stage 89 DAP)	**R²**	0.68	0.64
	RMSE	0.007	0.074

Table.2 Normalized Importance among crop-logging parameters (stage 1 and 2) for crop yield

Crop-logging variable Importance		
Trait	**Importance**	**Normalized Importance**
Pt. ht.	0.225	100.0%
Pt. girth	0.142	63.4%
No. of leaves	0.167	74.4%
No. of primary branches	0.131	58.5%
Pt. SP NS	0.159	70.7%
Pt. Sp EW	0.176	78.3%

Independent Variable Importance		
Trait	Importance	Normalized Importance
Pt. ht.	0.126	79.2%
Pt. girth	0.098	61.8%
No. of leaves	0.133	83.7%
No. of primary branches	0.066	41.7%
Pt. SP EW	0.129	81.2%
Pt. SP NS	0.159	100.0%
No. of flowers	0.046	29.1%
Leaf length	0.148	92.9%
Leaf breadth	0.093	58.3%

Table.3 Normalized Importance among crop-logging parameters (stage 3 and 4) for crop yield

Independent Variable Importance		
	Importance	Normalized Importance
Pt. ht.	0.222	90.6%
Pt. girth	0.031	12.8%
No. of leaves	0.196	80.2%
No. of primary branches	0.085	34.8%
Pt. SP. EW	0.220	89.7%
Pt. SP. NS	0.245	100.0%

Independent Variable Importance		
	Importance	Normalized Importance
Pt. ht.	0.229	100.0%
Pt. girth	0.020	8.6%
No. of leaves	0.227	99.0%
No. of primary branches	0.134	58.7%
Pt. SP. NS	0.164	71.6%
Pt. SP. EW	0.227	99.2%

Fig.1 Graphical representation of optimized artificial neural network for Stage 1 and 2 (Seedling Stage 26DAP and Vegetative stage 52DAP)
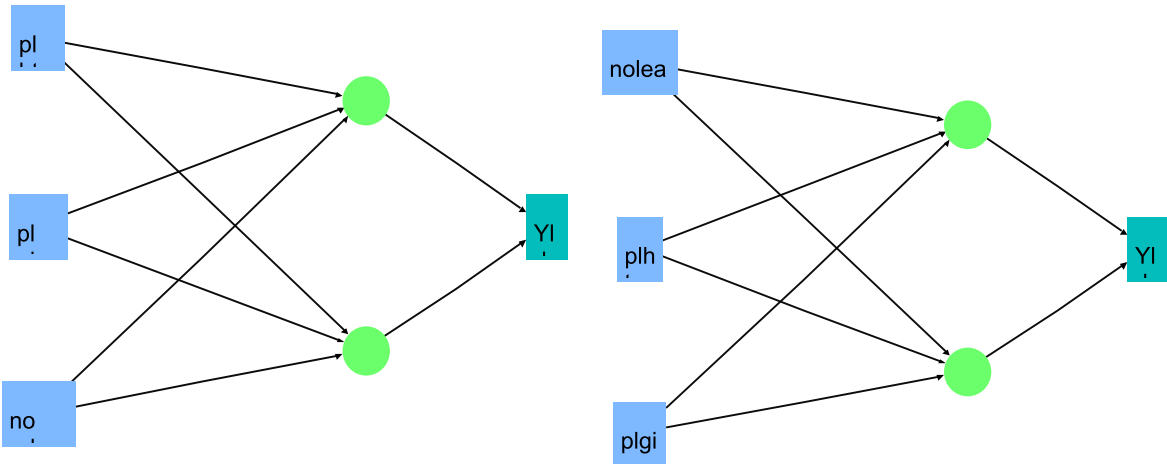

Fig.2 Graphical representation of optimized artificial neural network for stage 3 and 4 (flowering 79DAP and fruiting 89DAP)
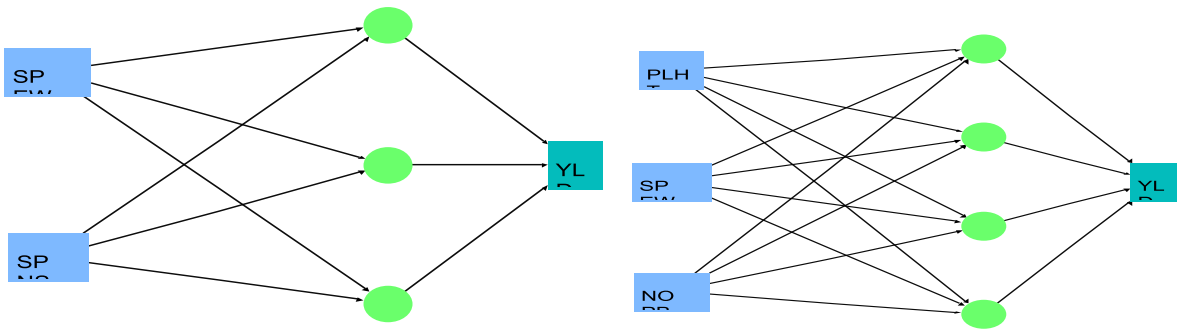

Fig.3 Importance of the crop-logging parameters (stage 1 and 2) in relation to crop yield
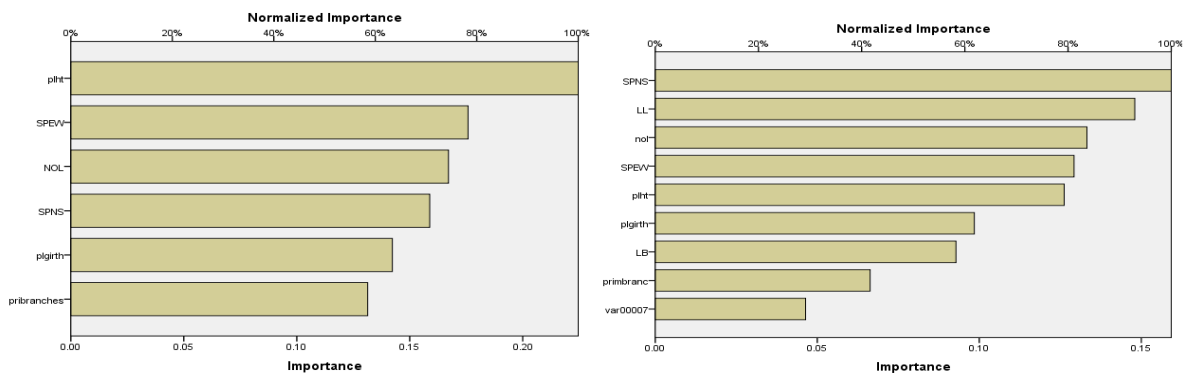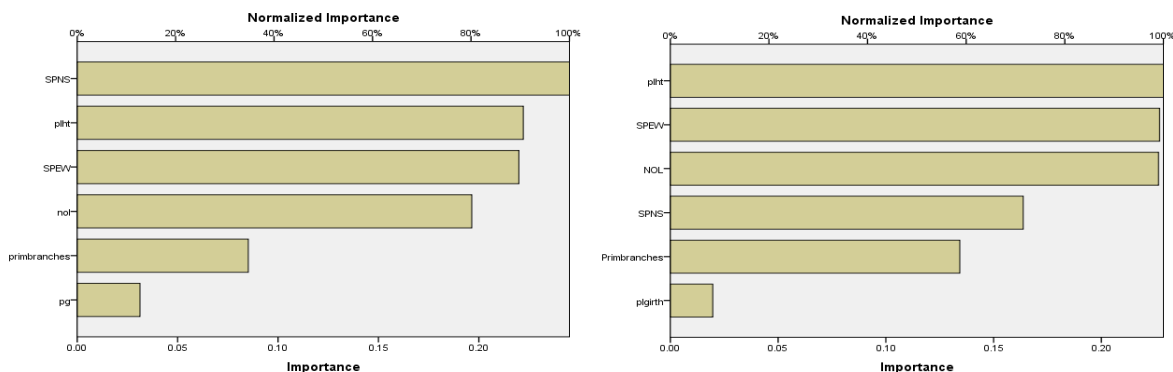
Fig.4 Importance of the crop-logging parameters (stage 3 and 4) in relation to crop yield

In the next step, for optimized ANN model total length of record was 34, out of which 90.9% was taken for model training, 9.1% for model testing and 100% for model evaluation with exclusion of 12 variables. Also, independent variables (number of leaves, plant spread north-south and leaf length) importance was worked out to be 0.344, 0.403 and 0.253 respectively. Further R^2 for training and validation are R^2= 0.88 and 0.64 respectively with Root mean square error RMSE=0.0.13 and 0.23 respectively. A graphical representation of ANN diagram for the optimized model is also appended (Fig. 1).

Artificial neural network models of third stage (flowering stage 79DAP)

It is observed that the total length of record was 32, out of which 95.8% was taken for model training, 4.2% for model testing and 100% for model evaluation initially (Table 1).

However independent variable (plant height, plant girth, number of leaves, number of primary branches, plant spread east-west and plant spread north- south) importance was observed to be 0.222, 0.0131, 0.196, 0.085, 0.220 and 0.245 respectively (Table 3). In the next step for optimized ANN model total length of record was record was 23, out of which 88.5% was taken for model training, 11% for model testing and 100% for model evaluation with exclusion of 6 variables. Also, independent

variables (plant height, plant spread north-south and plant spread east-west) importance was computed to as 0.312, 0.297 and 0.391 respectively. Further R^2 for training and validation are R^2= 0.89 and 0.85 respectively with Root mean square error RMSE=0.14 and 0.18 respectively. A graphical representation of ANN diagram for the optimized model is also appended (Fig. 2).

Artificial neural network models of fourth stage (fruiting stage 89DAP)

It is observed that the total length of record was 37, out of which 96.4% was taken for model training, 3.6% for model testing and 100% for model evaluation initially (Table 1).

However independent variable (plant height, plant girth, number of leaves, number of primary branches, plant spread north- south and plant spread east-west and) importance was found to be 0.229, 0.20, 0.227, 0.134, 0.164 and 0.227 respectively (Table 3). In the next step for optimized ANN model total length of record was record was 37, out of which 92.6% was taken for model training, 7.4% for model testing and 100% for model evaluation. Also, the independent variable (plant height, number of leaves and plant spread east-west) importance was 0.293, 0.377 and 0.329 respectively. Further R^2 for training and validation are R^2= 0.83 and 0.37 respectively with Root mean square error RMSE=0.29 and 0.37 respectively.

A graphical representation of ANN diagram for the optimized model is also appended (Fig. 2).

ANN model results of seedling stage had indicated that plant height, girth and number of primary branches could together predict the crop yield to an extent of 83 % for training set and 62 % for validation of model accuracy. ANN captured the inherent nonlinearity among biometrical variables to predict the yield 12 % more accurately.

Also the ANN model produced RMSE values for both the training and validation sets were least. Results of vegetative stage had indicated that plant height, girth and number of leaves could together predict the crop yield to an extent of 89 % for training set and 85 % for validation of model accuracy. ANN captured the inherent nonlinearity among biometrical variables to predict the yield 11 % more accurately.

Also, the ANN model produced RMSE values for both the training and validation sets were least. Flowering stage had indicated that both the plant spreads (east-west and north-south) could together predict the crop yield to an extent of 88 % for training set and 64 % for validation of model accuracy. ANN captured the inherent nonlinearity among biometrical variables to predict the yield 17 % more accurately (Fig. 3 and 4).

Also, the ANN model produced RMSE values for both the training and validation sets were least. Fruiting stage had indicated that the plant height, number of primary branches and plant spread (east-west) could together predict the crop yield to an extent of 68 % only for training set and 64 % for validation of model accuracy.

References

Cheng, B., and Tittering ton, D.M., 1994. Neural networks: A review from a statistical perspective. *Statistical science*, 9:2-54.

Cybenko, G., 1989. Approximation by superposition's of a sigmoid function. *Mathematics of control, signals and systems*, 2:303-14.

Meena, M and Singh, P.K., 2013. Crop Yield Forecasting Using Neural Networks. *Swarm, Evolutionary, and Memetic Computing, 319-339.*

Pratheepa, M., Meena, K., Subramaniam, K.R., Venugopalan, R., Bheemanna, H., 2011. Neural-network Classifier for the Prediction of Occurrence of *Helicoverpa armigera* (hübner) and its Natural Enemies. *Journal of Biological Control*, 25(2):134-42.

Singh, R. K. and Prajneshu. 2008. Artificial neural network methodology for modeling and forecasting maize crop yield. *Agricultural Economics Research Review*, 21, 5-10.

Venugopalan, R., 2010. Application of statistical principles for evolving crop-logging models in banana (Cv. Ney Poovan). *Tropical Agriculture*, 87(1), pp. 29-32.

Warner, B., AND Misra, M., 1996. Understanding neural networks as statistical tools. *The American Statistician*, 50: 284-93.

William W. Guo and Heru Xue., 2014. Crop Yield Forecasting Using Artificial Neural Networks: A Comparison between Spatial and Temporal Models. *Mathematical Problems in Engineering, 1-7.*