Group Method of Data Handling-Type Neural Network Prediction of Hazelnut Leaf Area Based On Length and Width of Leaf

Seyed Abolfazl Hassani[1], Ali Salehi Sardoei[1*], Fatemeh Sadeghian[1], Davood Bakhshi[2], Saeed Fallahi[2] and Sona Hossainava[2]

[1]*PhD student Horticultural Science, Faculty of Plant Production, Gorgan University of Agricultural Sciences and Natural Resources, Iran.*

[2]*Department of Horticulture, Faculty of Agriculture, University of Guilan, Rasht, Iran.*

***Corresponding Author:** alisalehisardoei@gau.ac.ir

ABSTRACT

Artificial neural network have been shown to be powerful tools for system modeling. One sub model of artificial neural network is the group method of data handling-type neural network (GMDH-type NN). The use of such self-organizing network leads to successful application in abroad range of areas. However, in some fields, such as horticultural science, the use of GMDH-type NN is still scare. Accurate and nondestructive methods to determine individual leaf areas of plants are a useful tool in physiological and agronomic research. Determining the individual leaf area (LA) of hazelnuts (*Corylus avellana*) involves measurements of leaf parameters including: length (L) and width (W) parameters. In this way, a genetic algorithm is deployed in a new approach to design the whole architecture of the GMDH-type NN. This study addressed the question of whether GMDH-type NN could be used to estimate leaf area (outputs) based on specified variables inputs (leaf with, leaf length). Results suggest that GMDH-type NN provide an effective means of efficiently recognizing the patterns in data and accurately predicting a performance index based on investigating inputs, and also can be used to prediction leaf area based on leaf width, leaf length factors.

Keywords: *Corylus avellana*, Leaf area, Leaf width, Leaf length, Modeling, Neural network.

INTRODUCTION

Plant leaf area is an important determinant of light interception and consequently of transpiration, photosynthesis and plant productivity [Goudriaan *et al.,* 1994]. Plant physiologist and agronomist have demonstrated the importance of this parameter in estimating crop growth, development rate, yield potential, radiation use efficiency, and water and nutrient use [Bhatt and Chanda 2003; Olivera and Santos, 1995; Williams, 1987; Williams and Martinson, 2003]. Leaf area can be measured by destructive or nondestructive measurement. Many methods have been devised to facilitate the measurement of leaf area. However, these methods, including those of tracing, blueprinting, photographing, or using a conventional planimeter, require the excision of leaves from the plants. It is therefore not possible to make successive measurement of the same leaf. Plant canopy is also damaged, which might cause problems to other measurement or experiments. Leaf area can be measured quickly, accurately, and nondestructively using portable scanning planimeter [Daughtry, 1990], but it is suitable only for small plants with few leaves [Nyakwende *et al.,* 2003]. An alternative method to measure leaf area is using image analysis with image measurement and analysis software. The capture of image by digital camera is rapid, and the analysis using proper software is accurate [Bignami and Rossini, 1996], but the processing is time consuming, and the facilities are generally expensive. Therefore, an inexpensive, rapid, reliable, and nondestructive method for measuring leaf area is required by the agronomists. If the mathematical relationships between leaf area and on or more dimensions of the leaf (length and width) could be clarified, in this way, the GMDH was used to circumvent the difficulty of having a priori knowledge of the mathematical model of the process being considered. Therefore, the GMDH can be used to model complex systems without having specific knowledge of the systems. The man idea of the GMDH is to build an analytical function in a feed forward network based on a quadratic node transfer function [Farlow, 1984] whose coefficients are obtained using the regression technique. In fact, the real GMDH algorithm in which the model coefficients are estimated by means of the least square method has been classified as complete induction and incomplete induction, which represent the combinatorial (COMBI) and multi-layered iterative algorithms (MIA), respectively [Mueller and Lemke, 2000]. In recent years, however, the use of such self-organizing networks has led to successful application of the GMDH type algorithm in abroad range of areas in engineering, science and economics [Iba *et al.,* 1996; Ivakhnenko, 1971; Nariman-Zadeh *et al.,* 2003; 2002a; 2002b].

MATERIAL AND METHOD

The experiment was conducted in 2016 at the Guilan University, Rasht, Iran. The hazelnut varieties (on eleven hazelnut genotypes) that used in this research are regarded from research garden of Karaj plant and Seed

Research Instituted. Leaves with different size used as samples for leaf area estimation obtained randomly from different levels of the canopy. Leaves from different hazelnut varieties were used for leaf area (LA), length (L) and width (W) measurement. Leaves were immediately placed in plastic bags after cutting and were transported to laboratory. Leaf length was measured from lamina tip to the intersection point of the lamina and the petiole, a long the midrib values of L and W were recorded to the nearest 0.01 mm. the area of each leaf (LA) was measured using a planimeter (A. OTT Kempten, Germany, Bayern). We used 82 data sets (input-output data). The collected data consisted of length (L) and width (W) as the input variables and leaf area (LA) as the system output. Fifty data lines (training set) and thirty two data lines (testing set) were randomly extracted from the data based to train and calibrate the GMDH-type NN.

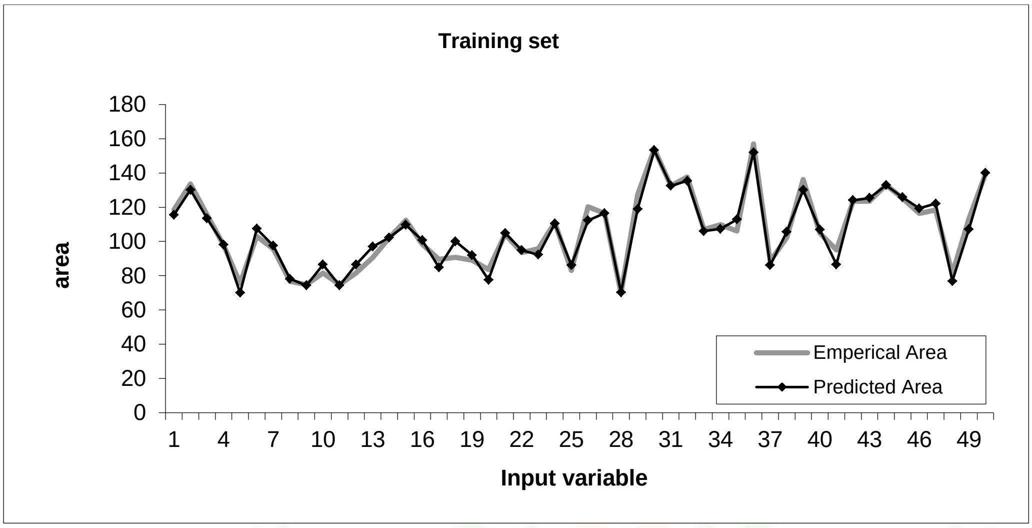

Figure 1. Neural network model-predicted performance in comparison with actual data for the training set (50 input-output data).

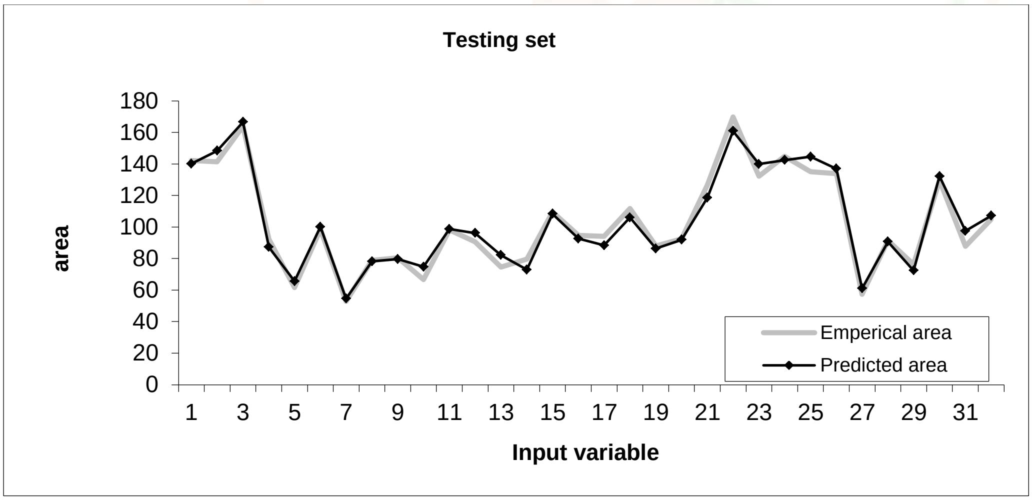

Figure 2. Neural network model-predicted performance in comparison with actual data for the training set (32 unforeseen input-output data).

Model development

A detailed description of GMDH-type terminology, development, and application is beyond the scope of this paper. Suggested references include [Nariman-Zadeh *et al.*, 2003; Nariman-Zadeh *et al.*, 2005]. Such a neural network identification process, in turn, needs some optimization method to find the best network architecture. In this way, genetic algorithm (GA) are deployed in a new approach to design the whole architecture of the GMDH-type NN, that is, the number of neurons in each hidden layer and their configuration of connectivity's, in combination with singular value the composition to find the optimal set of appropriate coefficients of quadratic expressions to model leaf area. The parameters interests in this multi-input, single-output system that affect the leaf area are length (mm) and width (mm). Fifty input-output actual data lines obtained were used to train the GMDH-type NN models. The testing set, which consisted of thirty two unpredictable input-output data lines during the training process, were used merely for testing to show the prediction ability of such evolved neural networks during the training process. Two hidden layers were considered for each model. To genetically design such neural networks, a population of 50 individuals with a cross over probability of 0.9, mutation probability of 0.01 and 300 generations was used. The accuracy of model was determined by using the

1) Mean absolute deviation (MAD), computed as
   MAD = $\frac{\sum_{i=1}^{n}|y_i-\hat{y}_i|}{n}$
2) The mean absolute percentage error (MAPE), computed as
   MAPE = $\frac{\sum_{i=1}^{n}\left|\frac{y_i-\hat{y}_i}{y_i}\right|}{n} \times 10$
3) The MS error (MSE), computed as
   MSE = $\frac{\sum_{i=1}^{n}|y_i-\hat{y}_i|^2}{n}$

Where y_i equals the actual value, $\hat{y}_i$ equals the predicted value, and n equals the number of observations (50 for training and 32 for testing).

RESULTS AND DISCUSSION

Equations (1-4) revealed the quantitative relation between input (L,W) and output (LA) variables under investigation. The corresponding polynomial equation representations of such a model were obtained as follows:

$$Y_1 = 30.519727402383559 - 7.884579584356982L + 5.036272926387118W + 0.85179742197188L^2 + 0.720822689179950W^2 \quad (1) - 0.785448872955065LW$$

$$Y_2 = 30.519727404673013 + 5.036272926141378W - 7.884579584461710L + 0.720822689186719W^2 + 0.857179742200752L^2 - 0.785448872944948LW \quad (2)$$

$$Y_3 = -11.704423764495218 + 3.845221199720372L + 0.761599400302094Y_1 - 0.298790397315835L^2 - 0.000992369545657{Y_1}^2 + 0.035377252294655LY_1 \quad (3)$$

$$LA = 8.109183283967390 - 16.850841596176970Y_2 + 17.672989788505149Y_3 - 9.674014961986641{Y_2}^2 - 9.875771751734593{Y_3}^2 + 19.550760272680083Y_2Y_3 \quad (4)$$

Table 1 summarizes the statistical results for the training and testing sets of GMDH-type NN models. These results indicate forecasting error measurements based on different between the model and actual values.

Table1. Model statistics and information for the group method of data handling-type neural network model for predicting the hazelnut leaf area

Statistic	Neural training	Neural testing
R^2	0.998634	0.997797
MSE	15.87519	25.74325
MAD	3.10742	4.159273
MAPE	3.0786	4.264
Number of hidden layers	2	
Hidden neurons	3	

REFERENCES

Goudriaan, J. and Van Laar, H. H. 1994. Modeling Potential Crop Growth Processes. Kluwer Academic Publishers, Dordrecht, the Netherlands. 180 pp.

Bhatt, M. and Chanda, S.V. 2003. Prediction of leaf area in Phaseolus vulgaris by nondestructive method. Bulg. J. Plant Physiol. 29: 96-100.

Olivera, M. and Santos, M. 1995. A semi-empirical method to estimate canopy leaf area of vineyards. Am. J. Enol. Viticult. 46: 389-391.

Williams, L. E. 1987. Growth of 'Thompson seedless' grapevines. I. Leaf area development and dry weight distribution. J. Am. Soc. Hort. Sci. 112: 325-330.

Williams, L. and Martinson, T.E. 2003. Nondestructive leaf area estimation of 'Niagara' and 'De Chaunac' grapevines. Sci. Hort. 98: 493-498.

Daughtry. C. 1990. Direct measurements of canopy structure, Remore Sens. Rev. 5: 45-60.

Nyakwende, E., Paull, C. J. and Atherton, J.G. 1997. Non-destructive determination of leaf area in tomato plants using image processing. J. Hort. Sci. 72: 225-262.

Bignami, C. and Rossini, F. 1996. Image analysis estimation of leaf area index and plant size of young hazelnut plants, J. Hort. Sci. 71: 113-121.

Farlow, S. J. 1984. Self-organizing method in modeling: GMDH type algorithm. Marcel, Dekker Inc.

Mueller, J. A. and Lemke, F. 2000. Self-organizing data mining: an intelligent approach to extract knowledge from data. Hamburg, Pub. Libri.

Iba, H., DeGaris, H. and Sato, T. 1996. A numerical approach to genetic programming for system identification. Evol Comput. 3(4): 417–52.

Ivakhnenko, A. G. 1971. Polynomial theory of complex systems. IEEE Trans Syst Man Cybern. 1: 364–78.

Nariman-Zadeh, N., Darvizeh, A. and Ahmad Zadeh, G. R. 2003. Hybrid genetic design of GMDH-type neural networks using singular value decomposition for modelling and prediction of the explosive cutting process. Proc Inst Mech Eng Part B: J Eng Manuf. 217: 779–90.

Nariman Zadeh, N., Darvizeh, A., Felezi, M. E. and Gharababaei, H. 2002. Polynomial modeling of explosive compaction process of metallic powders using GMDH-type neural network sand singular value decomposition. Modell Simul Mater Sci Eng. 10(6): 727–44.

Nariman Zadeh, N., Darvizeh, A., Darvizeh, M. and Gharababaei, H. 2002. Modeling of explosive cutting process of plates using GMDH-type neural network and singular value decomposition. J Mater Process Technol. 128(1): 80–7.

Nariman Zadeh, N., Darvizeh, A., Jamali, A. and Moieni, A. 2005. Evolutionary design of generalized polynomial neural networks for modelling and prediction of explosive forming process. J. Mater. Process. Technol. 165: 1561–1571.