Industrial Crops & Products 117 (2018) 224–234

Contents lists available at ScienceDirect

Industrial Crops & Products

journal homepage: www.elsevier.com/locate/indcrop

Modeling the seed yield of Ajowan (*Trachyspermum ammi* L.) using artificial neural network and multiple linear regression models

Mohsen Niazian[a,b,*], Seyed Ahmad Sadat-Noori[a,*], Moslem Abdipour[c]

[a] Department of Agronomy and Plant Breeding Science, College of Aburaihan, University of Tehran, Tehran-Pakdasht, Iran
[b] Department of Tissue and Cell Culture, Agricultural Biotechnology Research Institute of Iran (ABRII), Agricultural Research, Education and Extension Organization (AREEO), 3135933151 Karaj, Iran
[c] Kohgiluyeh and Boyerahmad Agricultural and Natural Resources Research and Education Center, Agricultural Research, Education and Extension Organization (AREEO), Yasuj, Iran

ARTICLE INFO

Keywords:
Ajowan
Artificial intelligence
Medicinal plant
Multi-layer perceptron
Selection criteria

ABSTRACT

Ajowan is a medicinal plant with useful pharmaceutical compounds in its seeds. Seed yield improvement in ajowan through a better understanding of the relationship between seed yield and its components is one of the most important goals of any breeding program. In the present study, artificial neural network (ANN) along with multiple regression model (MLR) were applied to predict the seed yield of ajowan through seed yield components. According to the simple correlation analysis, four characters (number of secondary branches, shoot dry weight, number of umbellets in an inflorescence, and biological yield) were selected as input variables in both artificial neural network and multiple linear regressions models. The network with SigmoidAxon transfer function, Levenberg-Marquart learning algorithm, one hidden layer with four neurons, 1000 training epochs and with a root mean square error (RMSE) of 0.147, a mean absolute error (MAE) of 0.127 and a determination coefficient (R^2) of 0.932 was selected as the final ANN model. The performance of ANN was better than MLR with a RMSE of 0.210 and a r^2 of 0.792. Biological yield and shoot dry weight were the most important yield components traits that affect the seed yield of ajowan and assigned as selection criteria using both ANN and MLR models.

1. Introduction

Ajowan (*Carum copticum* L.) is an industrial medicinal plant belongs to *Apiaceae* family and endemic of Egypt (Boskabady et al., 2014; Niazian et al., 2017a). This plant is mainly grows in arid and semi-arid regions of Egypt, East of India, and northwest, central and eastern parts of Iran (Ashraf and Orooj 2006; Joshi 2000; Moosavi et al., 2015; Niazian et al., 2017a; Noori et al., 2017). The active substances of seeds make this plant valuable for medicinal purposes (Dalkani et al., 2011). Ajowan seeds contain an essential oil with about 50% content of thymol, which has a strong germicide, anti-spasmodic and fungicidal effect (Ashraf and Orooj 2006). Carvacrol, γ-terpinene, and p-cymene are the main components of Iranian and African ajowan, whereas thymol is the main component of south Indian's ajowan (Boskabady et al., 2014). Many of the medicinal and aromatic plants do not have stable production in their growing areas and usually gathered in conventional methods to meet demands (Dalkani et al., 2012). Hence, attention to stable quality and quantity production of medicinal plants is important to meet the growing needs of pharmaceutical products. Seed yield is the most important part of ajowan but it is a quantitative and complicate trait that controlled with many genes and is mainly influenced by environmental conditions (Dalkani et al., 2011), and therefore has low heritability (Ghanshyam et al., 2015). In this situation, plant breeders prefer to use indirect selection through yield component traits with high heritability and high correlation with seed yield. Morphological evaluation of plant's agronomic characteristics is easier and cheaper than other evaluation methods (Dalkani et al., 2012). There are several methods for analysis of yield components that researchers can use them according to objective of project. Techniques such as analysis of variance, simple correlation coefficient, multiple regression and path analysis usually used by researchers to analyze the yield components (Fraser and Eaton, 1983). One of the simplest

Abbreviations: ANN, Artificial neural network; BY, Biological yield; IL, Average internodal length; LL, Leaf length; MAE, Mean absolute error; MLP, Multi layered perception; MLR, Multiple linear regressions; MSE, Mean square error; NB, Number of branches; NSB, Number of secondary branches; NU, Number of umbles; NUI, Number of umbellets in an inflorescence; PH, Plant height; RMSE, Root mean square error; SDW, Shoot dry weight; SY, Seed yield

* Corresponding authors at: Department of Agronomy and Plant Breeding Science, College of Aburaihan, University of Tehran, Pakdasht-Tehran, Iran.
E-mail addresses: mniazian@ut.ac.ir (M. Niazian), noori@ut.ac.ir (S.A. Sadat-Noori).

https://doi.org/10.1016/j.indcrop.2018.03.013
Received 14 November 2017; Received in revised form 3 March 2018; Accepted 5 March 2018

Table 1
The geography profile of the origins of 23 Iranian ajowan ecotypes.

No.	Ecotype code	Province	City	Geographical state	No.	Ecotype code	Province	City	Geographical state
1	943	Tehran	Tehran	Center	13	7893	Ghoom	Ghoom	Center
2	906	Alborz	Karaj	Center	14	15484	Yazd	Shahedie	East
3	13299	Ardabil	Ardabil	North West	15	15864	Yazd	Sadugh	East
4	31831	Yazd	Yazd	East	16	12313	Fars	Shiraz	South
5	37483	South Khorasan	Birjand	East	17	–	Khorasan Razavi	Sabzevar	North east
6	23011	Kerman	Rafsanjan	East	18	1085	Ardabil	Ardabil	North West
7	22079	Sistan & Baluchestan	Iranshahr	South East	19	14322	Hamedan	Hamedan	Center
8	17902	Fars	Marvdasht	South	20	10583	Ardabil	Ardabil	North West
9	10563	Ardabil	Ardabil	North West	21	33683	Yazd	Sadugh	East
10	37492	South Khorasan	Birjand	East	22	4077	Esfahan	Felaverjan	Center
11	14492	Markazi	Arak	Center	23	37529	South Khorasan	Ghaen	East
12	23023	Kerman	Rafsanjan	East					

Table 2
Neuron activate function.

Activation function	Formula
SigmoidAxon	$f(x_i, w_i) = \frac{1}{1+\exp[-x_i^{lin}]}$
LinearSigmoidAxon	$f(x_i, w_i) = \begin{cases} 0x_i^{lin} < 0 \\ 1x_i^{lin} > 1 \\ x_i^{lin}\ else \end{cases}$
TainhAxon	$f(x_i, w_i) = \tanh[x_i^{lin}]$
LinerTanhAxon	$f(x_i, w_i) = \begin{cases} -1x_i^{lin} < -1 \\ 1x_i^{lin} > 1 \\ x_i^{lin}\ else \end{cases}$

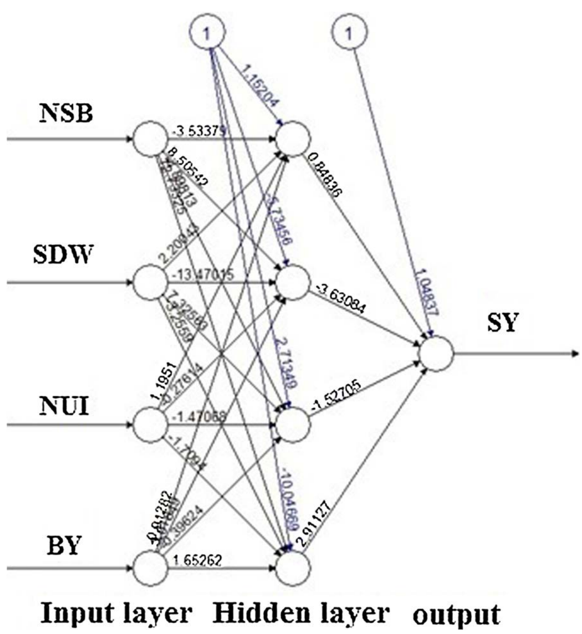

Fig. 1. Applied structure of multi layered perception model to predict seed yield of ajowan. (BY = Biological yield; NSB = Number of secondary branches; NUI = Number of umbellets in an inflorescence, SDW = Shoot dry weight; SY = Seed yield).

methods that can help to better understanding of yield component and assists in effective selection is correlation coefficient analysis (Mishra et al., 2015). The path coefficient analysis is also widely used to determine the nature of the relationships between yield and yield components (Dalkani et al., 2011). This method has been used frequently even in medicinal plants to testing cause/effect relationships among correlated variables (Bhandari and Gupta, 1991; Cosge et al., 2009; Dalkani et al., 2011; Lal, 2007). Multiple linear regression (MLR) refers to linear relationship of a dependent variable as function of the multiple independent variables (Quirk, 2016). One of the standard procedures of MLR for variable selection is stepwise regression. In this method, predictors introduce into the model sequentially and one at a time (Chong and Jun, 2005).

The main problem of regression-based models is that they cannot explain the highly nonlinear and complex relationship between seed yield and its components (Emamgholizadeh et al., 2015). To overcome this problem, in the recent years, agricultural scientists have been attracted to use of the artificial intelligence (AI) models such as artificial neural networks (ANN), genetic expression programming (GEP), and adaptive neuro-fuzzy inference system (ANFIS) (Azamathulla and Ghani, 2011; Emamgholizadeh et al., 2013a,b; Iquebal et al., 2014; Mansouri et al., 2016; Samadianfard et al., 2014; Shahinfar et al., 2012; Silva et al., 2014). Artificial neural network is an intelligence model that acts like the human brain (Tufail et al., 2008). Artificial neural network models are classified according to their structure, neurons type and etc. Furthermore, according to the training convergence in an ANN, different algorithm can be used (Govindaraju, 2000a,b). One of the most popular types of ANN that used in biological researches called the multi-layer perceptron (MLP) (Emamgholizadeh et al., 2015; Naroui Rad et al., 2015; Safa et al., 2015). A MLP is a feed-forward ANN model that consists of an input layer, hidden layer (s) and an output layer. In each MLP, multiple layers of nodes in a directed graph fully connected to the next one and each node (except for the input nodes) is a neuron with a nonlinear activation function (Rosenblatt, 1961).

Because of this fact that most of the herbal medicines are free of side effects, the interest in plant products has considerably increased all over the world (Ashraf and Orooj, 2006), but this increasing demand requires stable quantity and quality production of medicinal plants (Niazian et al., 2017b).

Direct selection for improve a quantitative trait such as seed yield is not effective, so indirect selection via simple morphological traits can be more helpful. Some conventional statistical approaches such as correlation coefficient, multiple linear regression analysis and path analysis can used for indirect improvement of quantitative and complex traits, but this methods have some shortage that complicated methods such as artificial neural network method can help to better understand these complex and unpredictable developmental phenomenons of biological systems. The objective of the present study were (a) predict the seed yield of ajowan medicinal plant using artificial neural network method, (b) compare the predicted results with the results of conventional regression-based method, and (c) find the most important selection criteria(s) for seed yield of Iranian ajowan ecotypes using ANN and MLR models.