Journal of Plant Protection Research ISSN 1427-4345

ORIGINAL ARTICLE

Artificial neural network potential in yield prediction of lentil (*Lens culinaris* L.) influenced by weed interference

Alireza Bagheri*, Negin Zargarian, Farzad Mondani, Iraj Nosratti

Department of Agronomy and Plant Breeding, Razi University, Kermanshah, Iran

Vol. 60, No. 3: 284–295, 2020

DOI: 10.24425/jppr.2020.133953

Received: February 29, 2020
Accepted: May 20, 2020

*Corresponding address:
alireza884@gmail.com

Abstract

This study was conducted to predict the yield and biomass of lentil (*Lens culinaris* L.) affected by weeds using artificial neural network and multiple regression models. Systematic sampling was done at 184 sampling points at the 8-leaf to early-flowering and at lentil maturity. The weed density and height as well as canopy cover of the weeds and lentil were measured in the first sampling stage. In addition, weed species richness, diversity and evenness were calculated. The measured variables in the first sampling stage were considered as predictive variables. In the second sampling stage, lentil yield and biomass dry weight were recorded at the same sampling points as the first sampling stage. The lentil yield and biomass were considered as dependent variables. The model input data included the total raw and standardized variables of the first sampling stage, as well as the raw and standardized variables with a significant relationship to the lentil yield and biomass extracted from stepwise regression and correlation methods. The results showed that neural network prediction accuracy was significantly more than multiple regression. The best network in predicting yield of lentil was the principal component analysis network (PCA), made from total standardized data, with a correlation coefficient of 80% and normalized root mean square error of 5.85%. These values in the best network (a PCA neural network made from standardized data with significant relationship to lentil biomass) were 79% and 11.36% for lentil biomass prediction, respectively. Our results generally showed that the neural network approach could be used effectively in lentil yield prediction under weed interference conditions.

Keywords: neural network, prediction models, pulses, weed interference, yield estimation

Introduction

Lentil is one of the most important pulses in rainfed and irrigated systems in different parts of the world (Mohamed *et al.* 1997), and it is one of the most important food products in the Middle East and South Asia (Sarker and Erskine 2006). The cultivated area and the production of lentil in Iran were reported as 138,739 ha and 83,329 tons, respectively (FAO 2017). Low heights, slow establishment, limited vegetative growth, and slow canopy closure of lentil gives low competition against weeds (Blackshaw *et al.* 2002; Elkoca *et al.* 2004; Erman *et al.* 2004). In addition, lentil is cultivated at a low density (the recommended density is between 80 and 100 plants per square meter) and has a slow growth rate, so it does not create a dense canopy in the early growing stages (Erman *et al.* 2008). The lentil yield reduction due to weed competition has been estimated at 20–84% (Yenish *et al.* 2009), depending on the infestation intensity and weed species (Knott and Halila 1988). Therefore, the control of lentil weeds is essential to prevent crop yield loss (Karimmojeni *et al.* 2015).

Prediction of crop yield during the growing season can optimize field management operations, such as optimum fertilizer application, suitable sowing density, and effective weed control by increasing the manager's awareness of the conditions on the farm. Different methods have been used to predict crop yield before harvest (Seyed Jalali *et al.* 2016). Bazgeer (2005) used several regression models to predict wheat yield in Punjab, India. Song *et al.* (2017) used the reciprocal hyperbolic model to predict soybean yield under the influence of single- and multiple-weed interference. Generally, crop losses caused by weeds are greater than those caused by disease and insect pests (Gnanavel and Natarajan 2014). Therefore, under normal farm conditions without biotic and abiotic stresses, modeling the relationship between weeds and crop yield can be effective in predicting crop yield (Ali *et al.* 2013).

Artificial neural networks are known as a preferred method in biosciences due to the predictive quality and simplicity compared to other empirical models (Joergensen and Bendoricchio 2001; Ozesmi *et al.* 2006). The neural network method is able to predict nonlinear and complex relationships and can show hidden connections between input variables (Batchelor *et al.* 1997). In fact, they are compromise-analytic methods based on the human brain's neuronal structure and have an ability to learn and process information (Torrecilla *et al.* 2004).

Neural networks generally consist of three layers and each layer is composed of some processor units called neurons (or cells, units, and nodes). The first layer of each network is called the input layer, which is the place where the data is entered. In this layer, independent variables of the model were interred and no processing was done. The last layer is the output layer in which the output data are deployed (Alvarez 2009; Cilimkovic 2015). Also, each network consists of some intermediate layers, called hidden layers, which compute the relationship between variables and also create the weights associated with each of the independent variables (Menhaj 2005). So far, artificial neural networks have a high potential in several issues such as estimation of leaf area (Movahedian *et al.* 1386), soil moisture (Chang and Islam 2000), water quality (Zhang *et al.* 2002), biomass (Jin and Liu 1997) and also crop yield (Liu *et al.* 2001; Drummond *et al.* 2003).

In recent years, research on the effectiveness of artificial neural network in predicting crop yield has increased. Alvarez (2009) used an artificial neural network for modeling the effects of soil factors and climatology on the average wheat yield in the Pampas, Argentina. He showed that an artificial neural network was better than regression in crop yield estimation. Sajadi and Sabouri (2014) used a multi-layer perceptron neural network based on meteorological data to predict rapeseed yield. They found that the neural network was able to predict canola yield. Irmak *et al.* (2006) also predicted the spatial patterns of soybean yield using an artificial neural network and evaluated the role of factors that trigger spatial variations of yield, including topography and soil fertility. The results showed that the back-propagation neural network model could potentially predict the spatial yield variability of soybean. In the study of Niazian *et al.* (2018) an artificial neural network and a multiple regression model were applied to predict the seed yield and seed yield components of ajowan (*Trachyspermum ammi* L.). The yield prediction of an artificial neural network was better than the multiple regression model. The findings of Niedbała *et al.* (2019) showed a practical possibility of using neural network models based on quantitative and qualitative data to predict the yield of winter rapeseed (*Brassica napus* L.). In recent years, new methods of machine learning with a high accuracy have been used for crop yield predictions such as; Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) (Sun *et al.* 2019; Khaki *et al.* 2020b), and Neural Collaborative Filtering Approach (NCF) (Khaki *et al.* 2020a).

Generally, no pest and weed management operations are carried out in most lentil rainfed fields in Kermanshah province, Iran. In fact, due to the semiarid climate and low rainfall in the study area, no outbreak of diseases or pests has been observed in rainfed lentil fields for many years. Therefore, weed infestation is the most important factor of crop reduction. It seems that, crop yield can be predicted by quantifying the characteristics of weeds as effective factors under common weather conditions using prediction models like an artificial neural network. Based on the authors' literature research, so far, no study has been done on the prediction of yield affected by weeds using an artificial neural network. In fact, measuring some weed traits such as weed height, density, and canopy cover as well as weed diversity and evenness could improve an understanding of the relationship between weeds and crop yield and consequently be used to predict crop yield. It seems that it is possible to predict crop yield by using some modeling methods such as an artificial neural network and multiple regression. Therefore, according to the significant effects of weeds in lentil yield reduction, in this research, we tried to compare the efficiency of an artificial neural network and multiple regression models in predicting lentil yield under multispecies weed competition.

Materials and Methods

Study area

This study was conducted in 2016 on a 2 ha lentil rainfed field, located on the Campus of Agriculture and Natural Resources of Razi University Kermanshah, Iran (34.32°N, 47.28°E, at 1374 m above sea level) with an average temperature of 13.4°C and an average

Table 1. Physical and chemical properties of the soil of the studied field

Soil texture	Clay [%]	Silt [%]	Sand [%]	K [ppm]	P [ppm]	N [%]	OC [%]	EC [ds · m^{-2}]	pH	Sampling depth
Silty clay	44.4	43.9	11.7	282	16.2	0.07	1.5	0.64	7.66	0–30

OC – organic carbon; EC – electrical conductivity

annual rainfall of 455 mm. The soil type was silty clay with 1.5% organic material and pH of 7.66 (Table 1). Two consecutive years before the experiment, the studied field was left fallow and under wheat cultivation, respectively. Soil preparation was done using a plow and disk. A local lentil landrace was planted using a row crop planter with a row spacing of 25 cm on March 6, 2016. Similar to most lentil rainfed fields in Kermanshah no pesticides or other methods for pest, disease or weed management were applied.

Sampling

Field sampling was done systematically, based on a network of points at a distance of 7 m from each other (at 7 to 7 m intervals), which were regularly located inside the field. The distance between the sampling points was according to the spatial range of weeds, which for most important weeds of farms was less than 7 m (Bagheri *et al.* 2014). A 1 m^2 was used at each sampling point to record lentil and weed characteristics. Sampling was done in two stages, so each point was sampled twice. Therefore, in order to do an accurate sampling, the geographic coordinates of each point in the first sampling stage were recorded with a GPS (Global Positioning System) device (GARMIN eTrex Summit). In addition, wooden nails were used to mark the points in order to easily identify them in the second sampling stage. In this study, 184 points were monitored and measured at each sampling stage.

Non-destructive sampling was done in the first sampling stage, at 8-leaf to the early-flowering phenological stage of lentil (May 3rd). Weed density and canopy cover, as well as lentil canopy cover were recorded. Weed density was recorded by counting the number of plants per square meter. To measure the canopy cover percentage, gridded quadrats were used so that the percentage of soil covered by crop and weeds was estimated visually and was considered as canopy cover percentage (Brim-DeForest *et al.* 2017). The height of the plants was measured using a ruler. The weed density data were used to calculate species richness, biodiversity indices of Shannon-Weiner (Equation 1), Simpson (Equation 2), and evenness indices of Smith-Wilson (Equation 3) and Camargo (Equation 4):

$$H' = -\sum_{i=1}^{s} p_i \ln p_i \,, \qquad (1)$$

where: H' – the Shannon-Weiner diversity Index, p_i – the proportion of species i relative to the total number of species which is defined as $p_i = n_i/N$ (n_i is the number of species i and N are the total number of species) and s – the total number of species (Shannon and Weaver 1964).

$$D = 1 - \sum_{i=1}^{s} p_i^2, \qquad (2)$$

where: D – the Simpson diversity index, p_i – the proportion of species i relative to the total number of species which is defined as $p_i = n_i/N$, and s – the total number of species (Simpson 1949).

$$E_{\mathrm{var}} = 1 - \left[\frac{2}{\pi \arctan \dfrac{\sum_{i=1}^{s} \left(\ln n_i \sum_{j=1}^{s} - \ln n_j / s \right)^2}{s}} \right], \qquad (3)$$

where: E_{var} – the Smith-Wilson homogeneity index, n_i – the number of species i in the sample, n_j is the number of species j in the sample and s – the total number of species (Smith and Wilson 1996).

$$E' = 1 - \left[\sum_{i=1}^{s} \sum_{j=i+1}^{s} \frac{|p_i - p_j|}{s} \right], \qquad (4)$$

where: E' – the Camargo's evenness index, p_i – the proportion of species i in the sample, p_j – the proportion of species j in the sample, and s – the total number of species (Camargo 1993).

In the second stage of sampling, destructive sampling was done at the time of the physiological maturity of lentils when 90% of pods were golden-brown (June 4th). The sampling was performed at exactly the same points as the first sampling stage. Thus, lentil plants in the 1 m^2 were harvested from the crown and transferred to the laboratory to measure the biomass dry weight and grain yield of lentil.

Data preparation

In this study, the data from the first sampling stage (lentil canopy cover, weed density, height, canopy cover and species richness, as well as Shannon-Weiner and Simpson biodiversity indices and Smith-Wilson and Camargo evenness indices for weed population) were used as input variables to predict lentil yield and biomass by studied models. In order to compare the accuracy of the models, another data series was prepared by obtaining the relationships of the measured variables of the first sampling stage with the yield and biomass of lentil. To this end, first, the statistical distribution of lentil biomass and yield data was tested by Kolmogorov Smirnov normality test. The statistical distribution of lentil yield was normal, hence the relationship between lentil yield and measured variables in the first sampling stage was investigated based on the stepwise regression method. The statistical distribution of lentil biomass was not normal and also no function was found to normalize this data based on the Johnson transformation method. Therefore, Spearman's correlation method was used to identify the significant relationship between lentil biomass and measured variables. In this way, two sets of input variables including overall collected data in the first sampling stage and significant variables extracted from stepwise and Spearman methods were considered as input variables.

Furthermore, to obtain the desired results in the lentil yield and biomass prediction, data standardization was performed. Accordingly, the Wittendorf linear standardization method (Equation 5) was used to standardize the input variable (Anysz *et al.* 2016). Thus, two sets of data including raw and standardized data were prepared from the overall collected data in the first sampling stage and significant variables extracted from the stepwise and Spearman methods. Finally, four data sets including total raw data, significant raw data extracted from the stepwise and Spearman methods, total standardized data and significant standardized data extracted from the stepwise and Spearman methods were considered as input variables to predict yield and biomass of lentil.

$$X_i = \frac{X - X_{\min}}{X_{\max} - X_{\min}}, \tag{5}$$

where: X_i – the standardized value of the data, $X_{\min}$ – the minimum amount of the data, and $X_{\max}$ – the maximum amount of the data of each input variable.

Yield and biomass prediction models

Two methods of artificial neural network (ANN) and multiple regression were used to predict the yield and biomass of lentil. In order to predict yield and biomass of lentil by ANN, the four input data sets were used as input layer of neural networks. The number of input neurons was considered nine (total raw and standardized data) and six (raw and standardized data derived from the stepwise and Spearman methods). The number of neurons in the output layer was the prediction values of yield or biomass of lentil. ANN performance depends on the choice of the number of hidden layers (Ramchoun *et al.* 2017). Therefore, while making the artificial neural networks, one to ten hidden layers were used and tested. The neuron number of the hidden layer of the best network was determined, after constructing the networks. To find the most accurate networks in lentil yield prediction, different types of neural networks such as Multilayer Perceptrons, Generalized feedforward, Modular and Principal Component Analysis were trained and tested (Bagheri *et al.* 2019). In addition, learning rules of Momentum, Levenberg Marquardt, Step, and Quickprop were tested. The transfer functions of TanhAxon, SigmoidAxon, TanhAxon Linear, SigmoidAxon Linear, SoftMaxAxon, Linear Axon, and Axon also were evaluated (Bagheri *et al.* 2019). Furthermore, the numbers of hidden layers and neurons in each hidden layer were manipulated to find the best neural networks. The above-mentioned steps to find the most accurate neural networks were done by trial and error (Niazian *et al.* 2018; Bagheri *et al.* 2019). Eighty percent of each data set was used to train the neural networks and the remaining 20% were used for accuracy testing. The training process was terminated based on the mean squared error below 0.01 threshold from one iteration to the next. The software NeuroSolution v. 5.00 was used to build neural networks. Sensitivity analysis was used for extracting the cause and effect relationship between the inputs and outputs of the networks. This provided feedback as to which input was the most effective on networks' output.

Given the normal distribution of lentil yield data and non-normal distribution of lentils biomass data, multiple regression models were applied only to predict lentil yield. Therefore, multiple regression models were constructed from the four input data series of lentil yield using IBM SPSS Statistics v. 26.

Validation of prediction precision

The correlation coefficient (R) and coefficient of determination (R^2) of the observed and predicted values were used to determine the accuracy of the neural network and multiple regression models in predicting lentil yield and biomass. The accuracy of different models could be evaluated through the two parameters of the correlation coefficient and coefficient of determination. However, these coefficients cannot describe solely the accuracy of the different models. Even though in a model there may be a large difference between the values of the observed and predicted data, the changing trend in these data is still the same. In this case, although the correlation coefficient and the coefficient of determination can accurately reflect the process of the changes in the observed and predicted data, they do not indicate the numerical matching between them (Rahmani *et al.* 2008). Therefore, in addition to the above-mentioned coefficients, the mean square error (*MSE*) (Equation 6), root mean square error (*RMSE*)

(Equation 7), and normalized root mean square error ($nRMSE$) (Equation 8) were also used for accuracy evaluation of the models (Haykin and Lippmann, 1994).

$$MSE = \frac{\sum_{i=1}^{n}\left(P_i - O_i\right)^2}{n}, \tag{6}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(P_i - O_i\right)^2}{n}}, \tag{7}$$

$$nRMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(P_i - O_i\right)^2}{n}} \times \frac{100}{\bar{O}}, \tag{8}$$

where: O_i – the observed value (lentil yield), P_i – the predicted value of the model (neural network), $\bar{O}$ – the average number of observations, and n – the number of observations or repetitions.

Normalized root mean square error expresses the percentage of the difference between the predicted values and the actual values. On this basis, the network prediction power is "excellent" if this value is less than 10%, if between 10 to 20% it is "good", if it is between 20 and 30% it is "moderate" and, if it is above 30% it is "weak" (Shirdeli and Tavassoli 2015).

The t-test was used for statistical comparison of neural network and multiple regression in lentil yield prediction results within the four sets of input data.

Results

Significant relationships between yield and biomass of lentils with measured variables

The results showed that the stepwise regression model was able to describe significantly ($p \leq 0.01$) the variations of the dependent variable affected by independent variables (Table 2). Weed density and canopy cover percentage had negative effects and lentil canopy cover percentage had significant positive effects on the lentil yield (Table 3).

The stepwise regression model of the effects of weed diversity and evenness indices as independent variables on lentil yield as the dependent variable was significant ($p \leq 0.01$) and was able to explain the changes of dependent variable influenced by independent variables (Table 4). Based on the results, the Smith and Wilson evenness index and Simpson diversity index had a significant positive effect on lentil yield (Table 5).

Biomasses of lentil had a significant negative correlation with weed canopy cover and density, while it

Table 2. Analysis of variance of the stepwise regression for the effect of weed and lentil parameters on lentil yield

Model	Sum of squares	*df*	Mean square	*F*	*p*
Regression	22,311.380	1	22,311.380	19.670	0.000
Residual	199,632.744	176	1,134.277	–	–
Total	221,944.124	177	–	–	–

Table 3. Significant weed and lentil parameters on lentil yield by stepwise regression

Model	Regression coefficient	Standard error	Beta coefficient	*p*-value
Constant coefficient	201.56	10.88	–	0.000
Weed density	–0.69	0.18	–0.25	0.000
Weed canopy	–0.40	0.17	–0.16	0.019
Lentil canopy	0.47	0.13	0.25	0.000

The weed height did not have a significant effect on lentil yield therefore, it was removed from the model

Table 4. Analysis of variance of the stepwise regression for the effect of weed diversity and evenness indices on lentil yield

Model	Sum of squares	*df*	Mean square	*F*	*p*
Regression	19,013.889	2	9,506.945	8.247	0.000
Residual	199,436.939	173	1,152.815	–	–
Total	218,450.828	175	–	–	–

Table 5. Significant weed diversity and evenness indices on lentil yield by stepwise regression

Model	Regression coefficient	Standard error	Beta coefficient	*p*-value
Constant coefficient	127.666	19.016	–	0.000
Smith and Wilson	76.488	18.840	0.436	0.000
Simpson	64.667	21.129	0.329	0.003

The Shannon and Camargo indices did not have a significant effect on lentil yield therefore, they were removed from the model

Table 6. The Spearman correlation between weed and lentil parameters with lentil biomass

	Weed canopy [%]	Weed density	Weed height	Lentil canopy [%]	Crop biomass
Crop biomass	–0.212*	–0.147**	0.075	0.586***	1.000

*, **, *** are significant at the 0.01, 0.05 and 0.1 probability level respectively

Table 7. The Spearman correlation between weed diversity and evenness indices with lentil biomass

	Richness	Shannon	Simpson	Camargo	Smith and Wilson	Crop biomass
Crop biomass	–0.199*	–0.026	0.038	0.021	0.020	1.000

*is significant at the 0.01 probability level respectively

had a significant positive correlation with lentil canopy cover (Table 6). It seems that, by increasing the weed canopy and density, the competition between weeds and lentils was intensified and resulted in lentil biomass reduction. In addition, increasing lentil canopy ultimately led to an increase in lentil biomass.

The lentil biomass had a significant negative correlation with weed species richness, while the other indices showed no significant correlation with lentil biomass (Table 7). The significant measured variables of weeds and lentils on lentil yield and biomass were used in the construction of neural networks and multiple regression.

Artificial neural network and multiple regression efficiency for predicting lentil yield

The results showed that the PCA neural network with the least MSE value of the training process, had the highest accuracy among the tested neural networks. This neural network consisted of hidden layers and the TanhAxon transmission function according to the Step-Learning Rules. The evolution of the PCA neural network output based on four input data sets (total raw and standardized input data as well as raw and standardized data on the stepwise regression output) showed that the neural network made from total standardized data could describe 0.80 of yield variations. However, the networks made from total raw data, raw data extracted from stepwise regression, and standardized data extracted from stepwise regression could describe 0.65, 0.58, and 0.67 of lentil yield variations, respectively. In addition, the results of the comparison of the RMSE and square normalized root mean error of the networks also showed that the neural network made from total standardized data had the smallest values (12.54 and 5.85, respectively) (Table 8). Investigating the sensitivity of the output of networks made from different input data sets showed that when using raw data, the diversity and evenness indices of weeds had the most impact on the output of the model. However, using standard data increased the impact of other inputs, so that weed density and canopy as well as lentil canopy had an obvious effect on lentil yield. In fact, the effect of standard input data was more homogeneous than the raw input data on the model output. (Table 9).

The results of lentil yield prediction by multiple regression models showed that multiple regression models were significant in the prediction of lentil yield. Nevertheless, comparison of multiple regression with neural networks showed that the accuracy of the neural networks for all four data series was significantly higher than the multiple regression models (Table 10).

As it was shown, the correlation coefficient between observed data with the predicted data by neural network made from total raw and standardized data was 0.80 and 0.89, respectively. These coefficients for raw and standardized stepwise regression data were 0.76 and 0.82, respectively. It showed that the input and output layers of the network had a relatively high correlation (Table 8). Accordingly, the charts of observed and predicted yield also showed a high correlation

Table 8. Neural network accuracy in predicting lentil yield using total raw and standardized data and raw and standardized stepwise regression data

Network input	*R*	R^2	*MSE*	*RMSE*	*nRMSE* [%]
			Total data		
Raw data	0.805	0.648	308.334	17.559	8.392
Standardized data	0.894	0.799	157.348	12.544	5.855
			Stepwise regression data		
Raw data	0.762	0.580	285.162	16.887	8.160
Standardized data	0.820	0.672	334.259	18.283	8.637

R – correlation coefficient; R^2 – coefficient of determination; *MSE* – mean square error; *RMSE* – root-mean-square error; *nRMSE* – normalized root mean square error

Table 9. The effect of each of the inputs on the output in the networks constructed by different input data sets

Inputs	Sensitivity about the mean			
	total raw data	total standardized data	regression raw data	regression standardized data
Weed canopy	0.12	10.02	0.18	15.82
Weed density	0.03	3.07	0.17	18.25
Lentil height	0.14	3.49		
Lentil canopy	0.19	15.22	0.20	18.32
Richness	0.81	8.32		
Shannon index	5.70	13.90		
Simpson index	4.69	4.14	0.44	1.68
Smith and Wilson index	11.03	8.78	13.23	12.46
Camargo index	0.46	4.62		

Table 10. Multiple regression models accuracy in predicting lentil yield using total raw and standardized data and raw and standardized stepwise regression data

Input	*R*	R^2	*MSE*	*RMSE*	*nRMSE* [%]
			Total data		
Raw data	0.549	0.301	3,976.995	63.063	30.980
Standardized data	0.601	0.362	3,773.594	61.429	29.553
			Stepwise regression data		
Raw data	0.594	0.359	5,515.969	74.269	35.730
Standardized data	0.593	0.352	5,512.358	74.245	35.718

R – correlation coefficient; R^2 – coefficient of determination; *MSE* – mean square error; *RMSE* – root mean square error; *nRMSE* – normalized root mean square error

between these values and the similarity of the tendency of changes. There was little difference between the actual values of yield and the values predicted by the neural network (Fig. 1).

Evaluation of artificial neural network efficiency for lentil biomass

The results of the lentil biomass prediction by the neural networks showed that a PCA neural network with two hidden layers and the TanhAxon transfer function according to Step's learning rule, with the least MSE of the training process was the best and most accurate network. The analysis of the PCA neural network based on the four input data sets showed that the neural networks made from the total raw and standardized data, as well as the raw and standardized data in correlation output could describe biomass variation by 0.77, 0.69, 0.63 and 0.62, respectively. The correlation coefficients between the observed and predicted data in the neural networks built by the total raw and standardized data, as well as the raw and standardized data in correlation

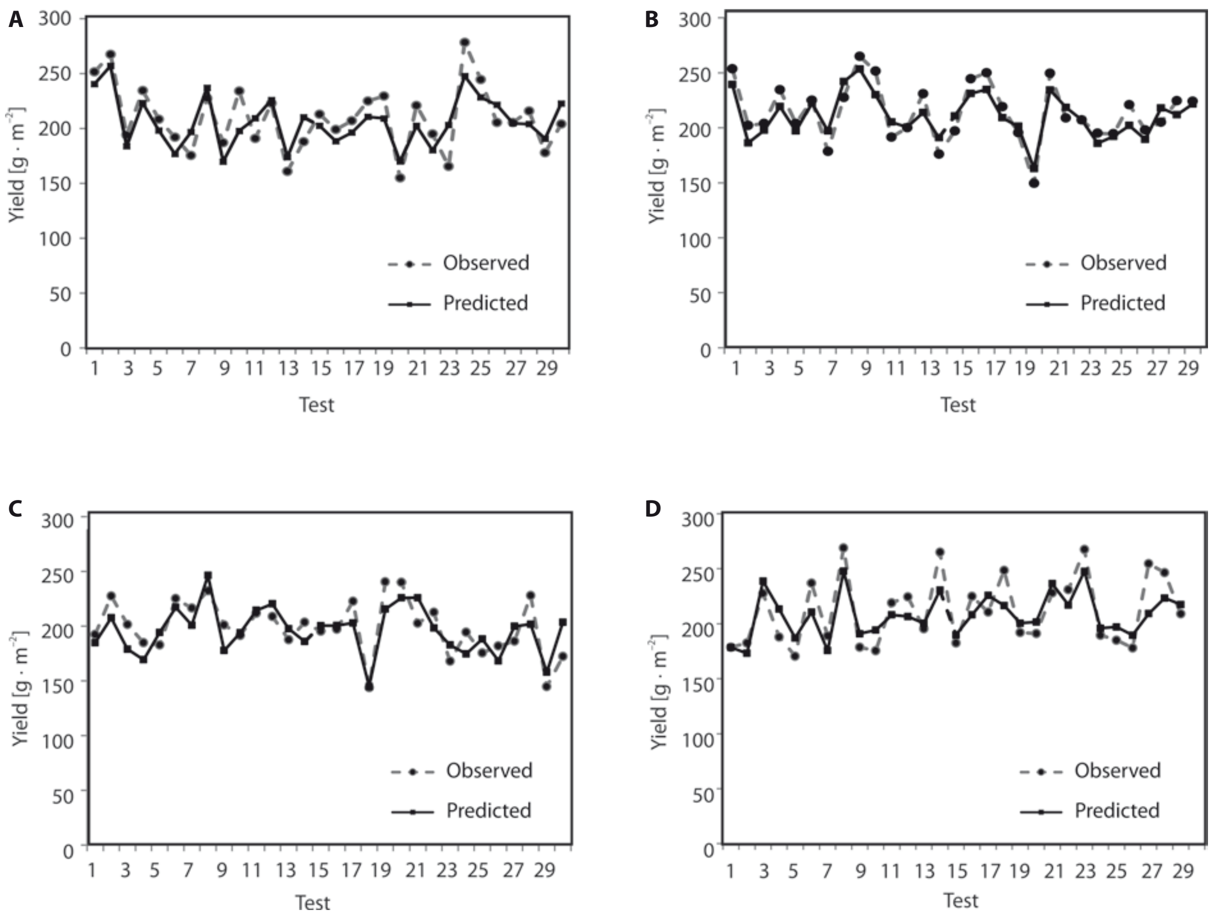

Fig. 1. Comparison between observed and predicted lentil yield by the neural network using: A – total raw data, B – total standardized data, C – raw stepwise regression data and D – standardized stepwise regression data

output were 0.88, 0.83, 0.79 and 0.79, respectively. The results of the comparison of RMSE between the constructed networks showed that the neural network made by the standardized correlation data was the most accurate with the lowest RMSE of 65.75 and lowest nRMSE of 11.36 (Table 11). The lentil biomass sensitivity to the different model inputs showed that the effect of weed species richness was greatest when using raw data, and the effect of other inputs increased when using standard data, as seen in lentil yield (Table 12).

Evaluation of observed and predicted biomass values in Figure 2 indicated that despite the higher correlation in total raw and standardized data than the correlation standardized data, the numerical resemblance of them was lower. It seems that that is why RMSE of the networks built from total raw and standardized data is higher and the neural network accuracy is lower than the network built from the correlation standardized data.

Table 11. Neural network accuracy in predicting lentil biomass using all raw and standardized data and raw and standardized correlation data

Network input	*R*	*R*^{2}	*MSE*	*RMSE*	*nRMSE* [%]
			Total data		
Raw data	0.88	0.77	5,111.66	71.49	12.01
Standardized data	0.83	0.69	5,304.87	72.83	12.5
			Correlation data		
Raw data	0.79	0.63	5,490.7	74.1	12.22
Standardized data	0.79	0.62	4,324.21	65.76	11.36

R – correlation coefficient; *R*^{2} – coefficient of determination; *MSE* – mean square error; *RMSE* – root mean square error; *nRMSE* – normalized root mean square error

Table 12. The effect that each of the inputs on the output in the networks constructed by different input data sets

	Sensitivity about the mean			
Inputs	total raw data	total standardized data	correlation raw data	correlation standardized data
Weed canopy	0.62	37.16	0.39	21.13
Weed density	0.16	12.57	1.63	25.83
Lentil height	0.93	17.68		
Lentil canopy	1.62	103.82	1.20	150.10
Richness	2.78	18.91	5.48	106.64
Shannon index	6.56	7.94		
Simpson index	2.27	0.85		
Smith and Wilson index	0.74	1.75		
Camargo index	0.78	1.92		

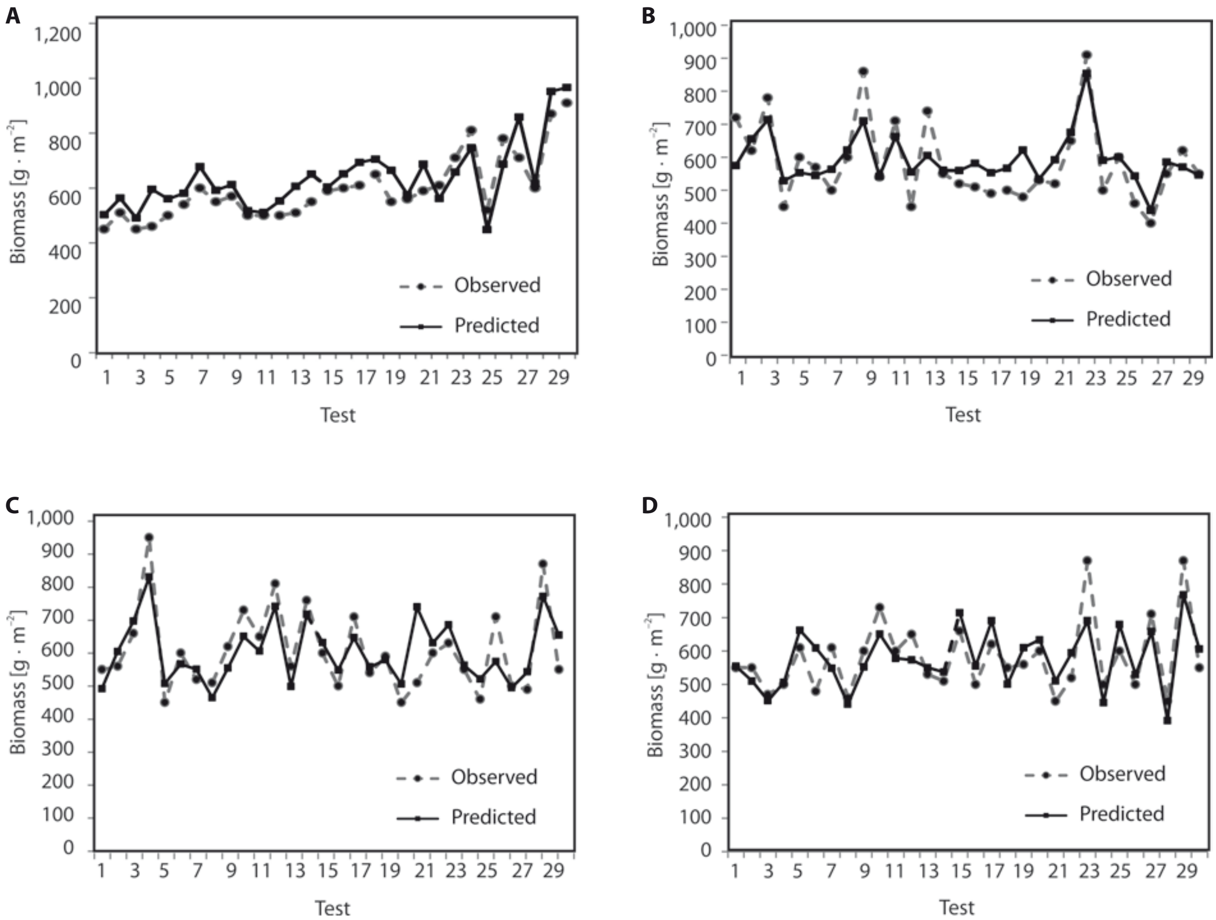

Fig. 2. Comparison between observed and predicted biomass by the neural network using: A – total raw data, B – total standardized data, C – raw correlation data and D – standardized correlation data

Discussion

According to the results, increasing the weed density and canopy percentage had a significant negative effect on lentil yield. Song *et al.* (2017) in a study on the effects of single and multiple weed interactions reported that soybean yield was significantly affected by weed density. In fact, plants are affected by increasing the level of the adjacent canopy (Cressman *et al.* 2011). As a result, we can say that increasing the weed canopy can reduce the active photosynthesis of crops, resulting in reduced photosynthesis and ultimately loss of yield.

The Smith and Wilson evenness index expresses the contribution of each species in an ecosystem. The range of this index is between 0 to 1 with 1 representing maximum species evenness in an ecosystem.

Evenness indicates the distribution of plant species within an occupied niche space (Archibald 2019). In an ecosystem where the share of a species is approximately similar, dominance will decrease (Jost 2010). In fact, high evenness limits the abundance of dominant and competitive weeds inducing yield losses (Storkey and Neve 2018). The significant positive effect of the evenness index on lentil yield indicates that the more homogeneous presence of a species in the field not only has no negative effect but also could have a positive effect on the yield. Greater evenness of weed community reduces the intensity niche overlap of weeds with the crop (Adeux *et al.* 2019). In this situation competition between weed species can reduce the severity of the negative effect of weed dominant species, thereby it could have an indirect positive effect on yield. The results of our study were in accordance with Cierjacks *et al.* (2016) and Adeux *et al.* (2019) who reported a significant positive relationship between the crop yield with weed diversity.

The Simpson diversity index takes into account the number of species present, as well as the relative abundance of each species. With this index, 1 represents infinite diversity and 0, no diversity. That is, the larger the value of this index, the higher the diversity of a plant population. The positive effect of the Simpson diversity index on lentil yield revealed that in plots with larger values of this index, lentil yield was higher than in plots with lower values. Increasing weed diversity could reduce the probability of dominant and competitive weeds through inter-specific competition and consequently limit the negative impacts of weed competition on crop yield (Hooper *et al.* 2005). In the study of Adeux *et al.* (2019) higher weed diversity limited yield losses through reduced weed biomass production. Cierjacks *et al.* (2016) found a significant positive correlation between weed Simpson diversity and crop yield of coconut and banana. Ferrero *et al.* (2017) demonstrated a positive increase of soybean yield with increasing weed diversity.

Generally, constructed neural networks were able to predict lentil yield with more accuracy than multiple regression models. Niazian *et al.* (2018) compared the accuracy of neural network and linear regression in predicting seed yield of ajowan. They found that the neural network was more accurate. The artificial neural network could explore nonlinear and complex relationships between input variables and their connection with output variables (Batchelor *et al.* 1997) which is why the results showed more accuracy of the neural network than multiple regression models. The neural network generated by total standardized data had the highest precision. In the study of Kaul *et al.* (2005) neural network was able to predict maize and soybean yield with the precision of 0.77 and 0.81. Rahmani *et al.* (2008) also predicted oat yield using the Multilayer Perceptrons neural network with four hidden layers, based on meteorological parameters and drought indicators in different parts of East Azerbaijan, with the accuracy of 93 and 91%.

In this study, the most accurate network to predict lentil biomass was a network made from the standardized input data. However, network validation showed that the correlation between the observed data with predicted output data in this network was less than the other built networks. The correlation between the observed and predicted output of the network is not a complete criterion for determining the accuracy of the neural network. It is possible that there is a large difference between the observed and predicted data; however, the trend of variations is consistent with each other. In this case, although the correlation coefficient and the coefficient of determination can show the coordination of variation between the observed and predicted data well, numerical matching between observed and predicted data may not be acceptable (Rahmani *et al.* 2008). Figure 2 shows the numerical match between observed and predicted data. As can be seen, numerical matching of the network made from the standardized correlation data compared to other networks is more appropriate. Based on the nRMSE of the neural networks made for lentil biomass, it can be concluded that these networks had a relatively high predictive power (Table 9). The nRMSE is expressed as the percentage of the difference between the observed and predicted values, so that values <10%, between 10 and 20%, 20 to 30%, and >30%, respectively, indicated excellent, good, moderate and poor network performance (Shirdeli and Tavassoli 2015). Comparison of the neural network made for yield and lentil biomass showed that the neural network built for lentil yield was more accurate than the neural network built for lentil biomass prediction (Tables 8 and 9). Jin and Liu (1997) used one hidden layer of the Multilayer Perceptrons neural network to predict the wheat and oat biomass and results showed an appropriate correlation between predicted and observed data.

Conclusions

A lack of understanding the effects of weeds on lentil yield has led to inadequate weed management in Iran (personal observation). Weeds are one of the reasons that the average yield in Iran is lower than the world average. Different methods have been used to predict crop yield. So far, the neural network has not been used to predict crop yields using weed traits as input data. Given the fact that in the studied area, many rainfed farms are abandoned after sowing until harvest, and there are no pest and weed management operations, the experiment was conducted to understand the effect of weeds on predicting lentil yield. The results showed that the artificial neural network method constructed from weed traits as input data for predicting lentil yield and biomass was significantly more accurate than multiple regression. The neural network was an acceptable method to predict yield before harvesting under common environmental conditions without outbreaks of diseases and pests (in the rainy years) and environmental stress (e.g. drought stress in dry years). In general, understanding the effects of weeds on crop yield and predicting the yield during the growing season and before harvest can increase the awareness of the farmer and help him to make appropriate weed management decisions in order to prevent yield reduction.

References

Adeux G., Vieren E., Carlesi S., Bàrberi P., Munier-Jolain N., Cordeau S. 2019. Mitigating crop yield losses through weed diversity. Nature Sustainability 2 (11): 1018–1026. DOI: https://doi.org/10.1038/s41893-019-0415-y

Ali A., Streibig J.C., Andreasen C. 2003. Yield loss prediction models based on early estimation of weed pressure. Crop Protection 53: 125–131. DOI: http://dx.doi.org/10.1016/j.cropro.2013.06.010

Alvarez R. 2009. Predicting average regional yield and production of wheat in the Argentine Pampas by an artificial neural network approach. European Journal of Agronomy 30 (2): 70–77. DOI: https://doi.org/10.1016/j.eja.2008.07.005

Anysz H., Zbiciak A., Ibadov N. 2016. The influence of input data standardization method on prediction accuracy of artificial neural networks. Procedia Engineering 153: 66–70. DOI: https://doi.org/10.1016/j.proeng.2016.08.081

Archibald S. 2019. From the Ground Up: Herbaceous Community Diversity and Management in Coffee Agroforestry Systems. Master of Science. University of Toronto, Toronto, Canada, 78 pp.

Bagheri A., Eghbali L., Sadrabadi Haghighi R. 2019. Seed classification of three species of amaranth (*Amaranthus* spp.) using artificial neural network and canonical discriminant analysis. The Journal of Agricultural Science 157 (4): 333–334. DOI: http://dx.doi.org/10.1017/S0021859619000649

Bagheri A.R., Rashed Mohasel M.H., Rezvani Moghaddam P., Nasiri Mhalati M. 2014. Crop rotation effect on spatial dynamic of *Fumaria vaillantti* and *Polygonum aviculare*. Iranian Journal of Field Crops Research 12 (2): 178–188. (in Persian) DOI: https://doi.org/10.22067/gsc.v12i2.39143

Batchelor D.W., Yang B.X., Tschanz T.A. 1997. Development of a neural network for soybean rust epidemics. Transactions of the ASAE 40 (1): 247–252. DOI: https://doi.org/10.13031/2013.21237

Bazgeer S. 2005. Land Use Change Analysis in the Submountainous Region of Punjab Using Remote Sensing, GIS, and Agrometerological Parameters. Departement of Agricultural Meteorology, Punjab Agricultural University, Ludhiana, India, 128 pp.

Blackshaw R.E., O'Donovan J.T., Harker K.N., Li X. 2002. Beyond herbicides: new approaches to managing weeds. Proceedings of the International Conference on Environmentally Sustainable Agriculture for Dry Areas: 305–312.

Brim-DeForest W.B., Al-Khatib K., Fischer A.J. 2017. Predicting yield losses in rice mixed-weed species infestations in California. Weed Science 65 (1): 61–72. DOI: http://dx.doi.org/10.1614/WS-D-16-00079.1

Camargo J. 1993. Must dominance increase with the number of subordinate species in competitive interactions? Journal of Theoretical Biology 161 (4): 537–542. DOI: http://dx.doi.org/10.1006/jtbi.1993.1072

Chang D.-H., Islam S. 2000. Estimation of soil physical properties using remote sensing and artificial neural network. Remote Sensing of Environment 74 (3): 534–544. DOI: http://dx.doi.org/10.1016/S0034-4257(00)00144-9

Cierjacks A., Pommeranz M., Schulz K., Almeida-Cortez J. 2016. Is crop yield related to weed species diversity and biomass in coconut and banana fields of northeastern Brazil? Agriculture, Ecosystems & Environment 220 (17): 175–183. DOI: https://doi.org/10.1016/j.agee.2016.01.006

Cilimkovic M. 2015. Neural Networks and Back Propagation Algorithm. Institute of Technology Blanchardstown, Blanchardstown Road North Dublin, Ireland, 12 pp.

Cressman S.T., Page E.R., Swanton C.J. 2011. Weeds and the red to far-red ratio of reflected light: Characterizing the influence of herbicide selection, dose, and weed species. Weed Science 59 (3): 424–430. DOI: http://dx.doi.org/10.1614/WS-D-10-00166.1

Drummond S.T., Sudduth K.A., Joshi A., Birrell S.J., Kitchen N.R. 2003. Statistical and neural methods for site-specific yield prediction. Transactions of the ASAE 46 (1): 5–14. DOI: http://dx.doi.org/10.13031/2013.12541

Elkoca E., Kantar F., Zengin H. 2004. Effects of chemical and agronomical weed control treatments on weed density, yield and yield parameters of lentil (*Lens culinaris* L. cv. Erzurum-89). Asian Journal of Plant Science 3 (2): 187–192. DOI: https://doi.org/10.3923/ajps.2004.187.192

Erman M., Tepe I., Bukun B., Yergin R., Taskesen M. 2008. Critical period of weed control in winter lentil under non-irrigated conditions in Turkey. African Journal of Agricultural Research 3 (08): 523–530. DOI: https://doi.org/10.3906/tar-1007-956

Erman M., Tepe I., Yazlik A., Levent R., Ipek K. 2004. Effect of weed control treatments on weeds, seed yield, yield components and nodulation in winter lentil. Weed Research 44 (4): 305–312. DOI: https://doi.org/10.1111/j.1365-3180.2004.00404.x

FAO. 2017. FAOSTAT. Available on: http://www.fao.org/faostat/en/#data/QC [Accessed: 23 September, 2019]

Ferrero R., Lima M., Davis A.S., Gonzalez-Andujar J.L. 2017. Weed diversity affects soybean and maize yield in a long term experiment in Michigan, USA. Frontiers in Plant Science 8: 236. DOI: https://doi.org/10.3389/fpls.2017.00236

Gnanavel I., Natarajan S. 2014. Eco-friendly weed control options for sustainable agriculture – a review. Agricultural Reviews 35 (3): 172–183. DOI: https://doi.org/10.5958/0976-0741.2014.00904.0

Haykin S., Lippmann R. 1994. Neural networks, a comprehensive foundation. International Journal of Neural Systems 5 (3): 363–364. DOI: https://doi.org/10.1142/S0129065794000372

Hooper D.U., Chapin F.S., Ewel J.J., Hector A., Inchausti P., Lavorel S., Lawton J.H., Lodge D.M., Loreau M., Naeem S., Schmid B., Setälä H., Symstad A.J., Vandermeer J., Wardle D.A. 2005. Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. Ecological Monographs 75 (1): 3–35. DOI: https://doi.org/10.1890/04-0922

Irmak A., Jones J., Batchelor W., Irmak S., Boote K., Paz J. 2006. Artificial neural network model as a data analysis tool in precision farming. Transactions of the American Society of Agricultural and Biological Engineers 49 (6): 2027–2037. DOI: https://doi.org/10.13031/2013.22264

Jin Y.Q., Liu C. 1997. Biomass retrieval from high-dimensional active/passive remote sensing data by using artificial neural networks. International Journal of Remote Sensing 18 (4): 971–979. DOI: https://doi.org/10.1080/014311697218863

Joergensen S.E., Bendoricchio G. 2001. Fundamentals of Ecological Modelling. Elsevier, Oxford, UK, 530 pp.

Jost L. 2010. The relation between evenness and diversity. Diversity 2 (2): 207–232. DOI: https://doi.org/10.3390/d2020207

Karimmojeni H., Yousefi A.R., Kudsk P., Bazrafshan A.H. 2015. Broadleaf weed control in winter-sown lentil (*Lens culinaris*). Weed Technology 29 (1): 56–62. DOI: https://doi.org/10.1614/wt-d-13-00184.1

Kaul M., Hill R.L., Walthall C. 2005. Artificial neural networks for corn and soybean yield prediction. Agricultural Systems 85 (1): 1–18. DOI: https://doi.org/10.1016/j.agsy.2004.07.009

Khaki S., Khalilzadeh Z., Wang L. 2020a. Predicting yield performance of parents in plant breeding: a neural collaborative filtering approach. PLoS ONE 15: e0233382. DOI: https://doi.org/10.1371/journal.pone.0233382

Khaki S., Wang L., Archontoulis S.V. 2020b. A cnn-rnn framework for crop yield prediction. Frontiers in Plant Science 10: 1750. DOI: https://doi.org/10.3389/fpls.2019.01750

Knott C.M., Halila H.M. 1988. Weeds in food legumes: problems, effects and control. p. 535–548. In: "World Crops: Cool Season Food Legumes: A Global Perspective of the Problems and Prospects for Crop Improvement in Pea, Lentil, Faba Bean and Chickpea" (R.J. Summerfield, ed.). Springer Netherlands, Dordrecht, The Netherlands.

Liu J., Goering C., Tian L. 2001. A neural network for setting target corn yields. Transactions of the American Society of Agricultural and Biological Engineers 44 (3): 705–713. DOI: https://doi.org/10.13031/2013.6097

Menhaj M.B. 2005. Principles of Artificial Neural Network. Amir Kabir University, Tehran, Iran, 716 pp. (in Persian)

Mohamed E.S., Nourai A.H., Mohamed G.E., Mohamed M.I., Saxena M.C. 1997. Weeds and weed management irrigated lentil in northern Sudan. Weed Research 37 (4): 211–218. DOI: https://doi.org/10.1046/j.1365-3180.1997.d01-35.x

Movahedian M., Hosseini S.E., Ghorbanzadeh M. 1386. Estimation of leaf area using neural networks. Third Conference on Information and Knowledge Technology, November 27–29, 2007, Ferdowsi University of Mashhad, Mashhad, Iran, 819 pp. (in Persian)

Niazian M., Sadat-Noori S.A., Abdipour M. 2018. Modeling the seed yield of ajowan (*Trachyspermum ammi* L.) using artificial neural network and multiple linear regression models. Industrial Crops and Products 117: 224–234. DOI: https://doi.org/10.1016/j.indcrop.2018.03.013

Niedbała G., Piekutowska M., Weres J., Korzeniewicz R., Witaszek K., Adamski M., Pilarski K., Czechowska-Kosacka A., Krysztofiak-Kaniewska A. 2019. Application of artificial neural networks for yield modeling of winter rapeseed based on combined quantitative and qualitative data. Agronomy 9 (12): 781. DOI: https://doi.org/10.3390/agronomy9120781

Ozesmi S.L., Tan C.O., Ozesmi U. 2006. Methodological issues in building, training, and testing artificial neural networks in ecological applications. Ecological Modelling 195 (1): 83–93. DOI: https://doi.org/10.1016/j.ecolmodel.2005.11.012

Rahmani E., Liaghat A., Khalili A. 2008. Estimating barley yield in eastern Azerbaijan using drought indices and climatic parameters by artificial neural network (ANN). Iranian Journal of Soil and Water Research 39 (1): 47–56. (in Persian)

Ramchoun H., Idrissi M.A.J., Ghanou Y., Ettaouil M. 2017. New modeling of multilayer perceptron architecture optimization with regularization: an application to pattern classification. IAENG International Journal of Computer Science 44 (3): 261–269.

Sajadi S.J., Sabouri H. 2014. Application of artificial neural networks in canola crop yield prediction. Journal of Crop Production and Processing 3 (10): 157–164. (in Persian)

Sarker A., Erskine W. 2006. Recent progress in the ancient lentil. The Journal of Agricultural Science 144 (1): 19–29. DOI: https://doi.org/10.1017/S0021859605005800

Seyed Jalali S.A.R., Sarmadian F., Shorafa M., Mohammad Esmaiel Z. 2016. Application of kriging and cokriging in predicting wheat yield using principle component analysis. Electronic Journal of Crop Production 9 (2): 213–224. DOI: https://dx.doi.org/10.22069/ejcp.2016.3124 (in Persian)

Shannon C.E., Weaver W. 1964. The Mathematical Theory of Communication. University of Illinois Press, USA, 125 pp.

Shirdeli A., Tavassoli A. 2015. Predicting yield and water use efficiency in saffron using models of artificial neural network based on climate factors and water. Saffron Agronomy and Technology 3 (2): 121–131. DOI: https://dx.doi.org/10.22048/jsat.2015.10381 (in Persian)

Simpson E.H. 1949. Measurement of diversity. Nature 163 (4148): 688. DOI: https://dx.doi.org/10.1038/163688a0

Smith B., Wilson J.B. 1996. A consumer's guide to evenness indices. Oikos 76 (1): 70–82. DOI: https://doi.org/doi:10.2307/3545749

Song J.-S., Kim J.-W., Im J.-H., Lee K.-J., Lee B.-W., Kim D.-S. 2017. The Effects of single- and multiple-weed interference on soybean yield in the far-eastern region of Russia. Weed Science 65 (3): 371–380. DOI: https://dx.doi.org/10.1017/wsc.2016.25

Storkey J., Neve P. 2018. What good is weed diversity? Weed Research 58 (4): 239–243. DOI: https://doi.org/10.1111/wre.12310

Sun J., Di L., Sun Z., Shen Y., Lai Z. 2019. County-level soybean yield prediction using deep CNN-LSTM model. Sensors 19 (20): 4363. DOI: https://doi.org/10.3390/s19204363

Torrecilla J.S., Otero L., Sanz P.D. 2004. A neural network approach for thermal/pressure food processing. Journal of Food Engineering 62 (1): 89–95. DOI: http://dx.doi.org/10.1016/S0260-8774(03)00174-2

Yenish J.P., Brand J., Pala M., Haddad A. 2009. Weed management in lentil. p. 326–342. In: "The Lentil: Botany, Production and Uses" (W. Erskine, F.J. Muehlbauer, A. Sarker, B. Sharma, eds.). CABI, Wallingford, UK.

Zhang Y., Pulliainen J., Koponen S., Hallikainen M. 2002. Application of an empirical neural network to surface water quality estimation in the Gulf of Finland using combined optical data and microwave data. Remote Sensing of Environment 81 (2–3): 327–336. DOI: http://dx.doi.org/10.1016/S0034-4257(02)00009-3