Australian Journal of Crop Science

AJCS

AJCS 10(6):895-903 (2016) ISSN:1835-2707
DOI: 10.21475/ajcs.2016.10.06.p7634

Prediction of biological and grain yield of barley using multiple regression and artificial neural network models

Marzieh Mokarram[1] and Ehsan Bijanzadeh[2*]

[1]Department of Range and Watershed Management, Agriculture College and Natural Resources of Darab, Shiraz University, Iran
[2]Department of Agroecology, Agriculture College and Natural Resources of Darab, Shiraz University, Iran

***Corresponding author:bijanzd@shirazu.ac.ir**

Abstract

Prediction of barley yield is an attempt to accurately forecast the outcome of a specific situation, using as input information extracted from a set of data features that potentially describe the situation. In this study, an attempt has been made to analyze and compare multiple linear regression (MLR), and artificial neural network (ANN) including multi-layer perceptron (MLP) and radial basis function (RBF) models to predicting biological yield (BY) and yield (Y) of barely. Data was collected from the literatures on the subject of barley production that was existed in http://sid.ir website. A total of 10563 data from 17 features were prepared in Excel software sheets. Then, the Matlab software was used to compare the models. Results of MLR model based on R^2 showed that Model 7, with 1000-kernel weight (gr), OC (%), grain/spike, soil pH, N applied (kg/ha), plant height (cm), and irrigation regime (according to FC) and Model 8 with 1000-kernel weight (gr), OC (%), soil pH, grain/spike, HI (%), plant height (cm), irrigation regime (according to FC), and plant density (plant/m^2), were the best models for prediction BY and Y of barley, respectively. The highest standardized coefficient (β) for prediction of BY was obtained in 1000-kernel weight (0.621), OC (0.396) and grain/spike (0.385). Also, for prediction of Y, 1000-kernel weight, OC, and grain/spike with 0.547, 0.403, and 0.347 had the highest β, respectively. Among the MLR, MLP and RBF models, MLP model had the highest R^2 values for prediction of BY (R^2=0.894) and Y (R^2=0.922). Overall, ANN models can be used to successfully estimate BY and Y from data.

Keywords: Correlation coefficient, data mining, barley grain yield, standardized.
Abbreviation: ANN_Artificial neural network; BY_biological yield; R^2_correlation coefficient ; MSE_mean square error; MLP_multi-layer perceptron, MLR_multiple linear regression; RBF_radial basis function; β_standardized coefficient; Y_yield.

Introduction

Artificial Neural network models (ANN) which emulate the central nervous system are part of theoretical neuroscience and computational neuroscience (Rumelhart and McClelland, 1986). Among various methods of ANN and learning algorithms, multi-layer perceptron networks (MLP), and radial basis function (RBF) are the most popular neural network models. One of ANN applications is in agriculture science (Heinzow and Richard, 2002; Memarian and Balasundram, 2012). In the past years there has been an increasing interest in ANN modeling in different fields of agriculture, particularly for some areas where conventional statistical modeling failed. Presently, many aimed prediction models of agriculture field such as crop yield have been divided into two classes of mechanistic and empirical models (Poluektov and Topaj, 2001). The prediction by a well-trained ANN is normally faster than the statistical models. In addition, it is possible to add or remove input and output variables in the ANN (Ghamari et al., 2010).
The applications of the ANN in agriculture include the prediction of crop yield, seeding dates, biomass production, physical and physiological damage to seeds, organic matter content in soils, estimation of sugar content in fruits, characterize crop varieties, and soil moisture estimation (Kaul et al., 2005; Park et al., 2005; Saberali et al., 2007; Khazaei et al., 2008). Some researchers used ANN models for precision agriculture. They examined the applicability of ANN for development of yield mapping and forecasting systems by satellite images vs. soil parameters (Uno et al., 2005; Park et al., 2005; Stathakis et al., 2006).
In a comprehensive review, application of ANN in predicting wheat crop yield considered by Khairunniza-Bejo et al. (2014) who concluded that the best model for prediction of wheat yield was ANN models compared to other models. ANN has become a good method because of its ability to prediction, forecasting and classification in biological science fields. Hosseini et al., (2007) used ANN and multi-variable regression models for dryland wheat yield prediction in a temperate climate in Ghorve of Kordestan Province, Iran. They showed that ANN model can estimate the crop yield with acceptable accuracy. Maximum and minimum air temperature, daily mean relative humidity, net radiation, precipitation, dew point temperature, and wind velocity were included as input data in their ANN models. Kaul et al., (2005) developed ANN for corn and soybean yield prediction. They used the historical yield data at numerous locations in Maryland, USA and concluded that ANN models had more accurate results than other models.Khashei-Siuki et al., (2011) used expert systems to predict the dryland wheat yield from meteorological data in Khorasan province, Iran. In this way, adaptive neuro-fuzzy inference system (ANFIS) and MLP

models were compared by statistical test indices. ANFIS model consistently produced more accurate statistical indices [R^2= 0.67; root mean square error (RMSE)= 151.9 kg ha^{-1}, and mean absolute error (MAE)= 130.7 kg ha^{-1}], when temperature data (max, min, and dew temperature) used as independent variables for prediction of dryland wheat yield. O'Neal et al., (2002) designed a fully connected back-propagation ANN to predict maize yield with five data coding schemes at three scales using local crop-stage weather data and yield and showed that, the best version of the networks came out with root means squared testing error which was better than quadratic regression. Barley (*Hordeum vulgare* L.) is one of the most important crops in Iran, which is one of the largest barley producing areas in Middle East (Emam, 2002). The aim of the study is providing a comparative analysis of multiple linear regressions (MLR), and two ANN models including multi-layer perceptron (MLP) and radial basis function (RBF), for predicting biological yield (By) and yield (Y) of barely using available data in Iran. This is the first report about prediction of biological yield and yield of barely where cultured in Iran using MLR and ANN models.

Results and Discussion

Data characteristics of barley

In order to prediction yield and BY, data subdivided into two sets: 70% of the data for training or calibrating, while the 30% of remaining data used for testing the performance of the prediction models (Khoshhal and Mokarram, 2012). The Matlab software separated and ranked the data, automatically. Minimum, maximum, average and standard deviation (STDEV) for the training and testing data showed in Table 1 and Table 2, respectively.

Correlation coefficient of biological yield and yield of barley

As shown in Table 3, a positive correlation was observed between biological yield of barley (BY) and 1000-kernel weight (r=0.700**), plant height (r=0.561**), N (r=0.439**), irrigation regime (r=0.413**), and OC (r=0.401**). In contrast, spike/m^2 (r=-0.540**), water EC (r=- 0.479**), and soil pH (r=-0.405**) had negative correlation with biological yield, significantly. Emam (2002) reported that plant height, 1000-kernel weight, and spike/m^2 had a key role in biological yield improvement. Afshari et al. (2011) reported that water deficit around anthesis to maturity may lead to a loss in yield by reducing 1000-kernel weight, especially if accompanied by high temperatures, hastening of whole-plant senescence, and reducing biological yield. Bijanzadeh et al. (2012) reported that when biological yield was as output, grain/spike and 1000-kernel weight had a strong relationship with biological yield, with values of 0.5 to 1.0 in various attribute-weighting algorithms. Ten different attribute-weighting models showed that while irrigation regime was an important attribute for improving biological yield, harvest index was less important in modern wheat genotypes in Iran, and was only selected by the Relief model when biological yield was as output. Likewise, Tambussi et al. (2002) reported that biological yield in wheat may be increased by improving grain yield at a given level of harvest index. In the current study, harvest index was found to be less important in BY improvement (Table 3). Potassium applied to the soil was not found to be important using correlation coefficient (Table 3). Malakoti (2003) found that soils in western and southern Iran were rich in available potassium ions, and farmers often did not apply potassium fertilizer in these areas. The calculated simple linear correlation coefficients (r) between barley grain yield (Y) and independent variables summarized in Table 4. It was found that there is a positive and highly significant correlation between grain yield with 1000-kernel weight (r=0.635**), HI (r=0.622**), OC (r=0.544**), and spike/m^2 (r=0.508**). On the other hand, water EC (r=-0.535**), soil pH (r=-0.476**), P (r=-0.324**) and K (r=-0.178**) had negative correlation with Y. In a comparison study, Bijanzadeh and Naderi (2014) used feature selection model and 10 attribute weighting models to find the important features contributed to barley grain yield. They found that from 20 features as input, 10 features including irrigation regime, nitrogen applied to the soil, rainfall amount, grain number per spike, spike number per unit area, and growing season length (with value more than 0.961) were the most important features related to barley grain yield. Additionally, soil organic content (0.941 value), electrical conductivity of water (0.911), harvest index (0.905), and plant density (0.904) had the marginal effect on barley grain yield. Similar to our results, Austin (1984) also reported that one alternative for grain yield improvement is increasing the HI produced by the crop. In a similar study in an arid climate, Etemadi et al. (2005) reported that 1000-grain weight in Reyhan was 33% less than Karoon barley cultivar. Also, Bijanzadeh and Emam (2012) reported that reduction in assimilates' remobilization to the grain caused to lower 1000-kernel weight and grain yield in Shiraz cultivars (sensitive to drought) under severe drought while controlled soil drying in Sistan and Pishtaz (tolerated to drought) might result in a better remobilization of pre-stored assimilates to the grain in arid areas. Rezaii et al., (2010) declared that grain yield in Nosrat barley cultivar was 361.8, 418.3, 587.2, and 618.5 g/m^2 for the treatments with 25%, 50%, 75%, and 100% FC, respectively. They concluded that irrigation regime, 1000-kernel weight, and grain number spike had a key roles in barley grain yield improvement.

Nikkhah et al. (2010) also reported that under terminal drought stress of two and six rowed barley genotypes, grain yield decreased from 32% to 43%, and 1000-grain weight in two rowed barley genotypes was not affected by severe drought stress. Afshari et al. (2011) reported that water deficit around anthesis to maturity may lead to a loss in yield by reducing 1000-kernel weight, especially if accompanied by high temperatures, hastening of whole-plant senescence, and reducing biological yield.. Bijanzadeh and Emam (2012) reported that water stress decreased grain yield of five wheat cultivars by decreasing 1000-kernel weight. Our results showed that Y more affected by 1000- kernel weight using correlation coefficient.

Multi linear regression model (MLR) of biological yield and yield of barley

For prediction of Y and BY using MLR, first the most essential input variables were selected using stepwise method and then linear interaction term of these basic input data properties were investigated. Finally, the linear interaction term of the basic weather properties were investigated using SPSS software (2013). The results based on R^2 showed that Model 7, with 1000-kernel weight(gr), OC (%), grain/spike, soil pH, N applied (kg/ha), plant height (cm), and irrigation regime (according to FC) and Model 8 with 1000-kernel weight(gr), OC (%), soil pH, grain/spike, HI(%), plant height (cm), irrigation regime (according to FC), and plant density (plant/m2), were the best models for

Table 1. Descriptive statistics of the training dataset.

Feature	Units	min	max	average	STDEV
Irrigation regime	(according to FC)	100.00	100.00	100.00	0.00
Water EC	(dS/m)	1.00	1.50	1.26	0.16
Nitrojen applied (N)	(kg/ha)	50.00	138.00	98.51	28.06
Phosphours applied (P)	(kg/ha)	0.00	125.00	24.26	26.38
Potassium applied (K)	(kg/ha)	0.00	33.00	9.10	14.75
Plant density	(plant/m^2)	100.00	500.00	249.29	152.08
Growing season length	(day)	201.00	263.00	226.03	15.37
Organic content	%	0.23	0.61	0.47	0.08
Soil pH	-	7.40	7.90	7.59	0.19
Rianfall amount	(mm)	53.00	456.00	374.19	120.52
Plant height	(cm)	54.90	142.30	84.20	17.58
Grain/spike	-	19.30	65.33	39.08	9.82
Spike/m^2	-	118.00	1315.44	482.86	316.50
1000-kernel weight	(gr)	25.24	63.80	40.89	10.12
%HI	%	19.03	58.59	30.89	6.01

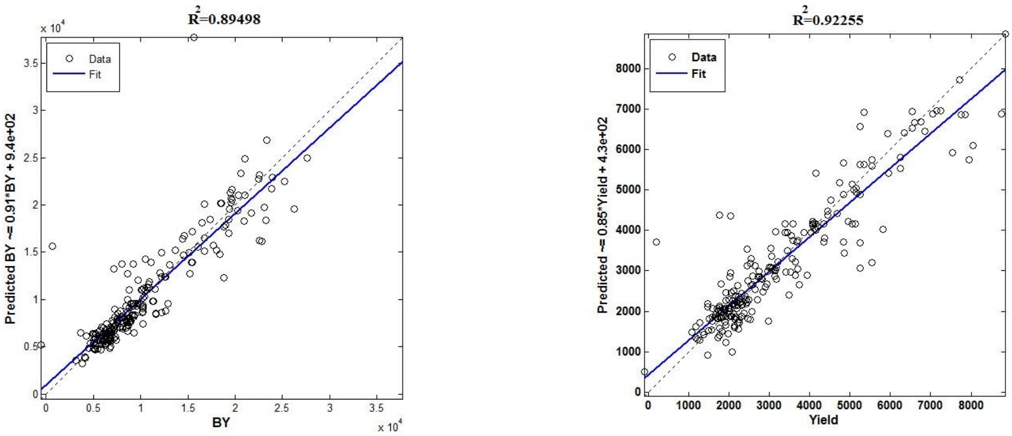

Fig 1. The scatter plot of the measured versus predicted the biological yield (BY) and grain yield of barley (Y) using multi-layer perceptron model (MLP).

prediction BY and Y of barley, respectively (Tables 5 and 6). Also, in these models, the R^2 values have been obtained 0.768 and 0.784, for BY and Y, respectively.

In Table 7 and Table 8, standard errors refer to the standard errors of the regression coefficients, which can be used for hypothesis testing and constructing confidence intervals. The standardized coefficients (β) are what the regression coefficients would be if the model were fitted to standardized data, that is, if from each observation we subtracted the sample mean and then divided by the sample standard deviation (SD). Also, the t statistic tests the hypothesis that a population regression coefficient β is 0, that is, H_0: β = 0. It is the ratio of the sample regression coefficient β to its standard error. According to Table 7, the highest standardized coefficients (β) for prediction of BY was obtained in 1000-kernel weight (0.621), OC (0.396) and grain/spike (0.385) while soil pH with -0.649 and HI with -0.138 had the lowest β (Table 7). For prediction of Y, 1000-kernel weight, OC, and grain/spike with 0.547, 0.403, and 0.347 had the highest β, respectively (Table 8). In contrast, soil pH with -0.690 had the lowest β.

Dreccer et al. (1997) and Alqudah and Schnurbusch (2014) reported that grain yield of barley is mainly determined by the of grain number per unit area of land. However, the mean grain weight can differ significantly between genotypes and environments. Bijanzadeh and Naderi (2014) used attribute weighting models to barley grain yield improvement and declared that four important features included spike number per unit area, genotype, grain number per spike, and organic content had weights more than 0.7 by principle component analysis models. Also, three features including spike number per unit area, grain number per spike and 1000-kernel weight were selected by three models as the most important attributes. Overall, in the present study, 1000-kernel weight was important feature selected by MLR model, when BY or Y were as output.

Prediction of biological yield and yield of barley using MLR, MLP and RBF models

The optimal architecture for each model was determined based on R^2 values of the trained data sets (Table 9). Among the MLR, MLP and RBF models, MLP model had the highest R^2 values for prediction of BY (R^2=0.894) and Y (R^2=0.922). In contrast, the lowest value of R^2 for prediction of BY (R^2=0.781) and Y (R^2=0.750) was obtained in MLR and RBF models, respectively (Table 9). The scatter plots between measured and predicted BY and Y using MLP and RBF models for testing stage with acceptable accuracy, indicated in Figures 1 and 2, respectively. According to Fig. 1, by MLP model R^2 value was 0.89 and 0.92 for prediction of BY and Y respectively. The equations for prediction of BY and Y was as below:

Table 2. Descriptive statistics of the testing dataset.

Feature	Units	min	max	average	STDEV
Irrigation regime	(according to FC)	75.00	100.00	97.35	7.70
Water EC	(dS/m)	1.00	1.30	1.23	0.12
Nitrojen applied (N)	(kg/ha)	80.00	150.00	95.30	27.59
Phosphours applied (P)	(kg/ha)	0.00	51.00	37.09	22.71
Potassium applied (K)	(kg/ha)	0.00	0.00	0.00	0.00
Plant density	(plant/m^2)	100.00	250.00	140.91	66.80
Growing season length	(day)	201.00	259.00	220.83	12.95
Organic content	%	0.50	1.75	0.65	0.36
Soil pH	-	7.10	7.90	7.72	0.32
Rianfall amount	(mm)	225.00	582.00	418.39	102.75
Plant height	(cm)	42.50	86.18	69.41	12.86
Grain/spike	-	27.00	76.27	49.47	12.70
Spike/m^2	-	340.00	1925.28	886.15	289.69
1000-kernel weight	(gr)	22.79	52.53	32.25	5.93
%HI	%	23.00	46.00	32.66	5.35

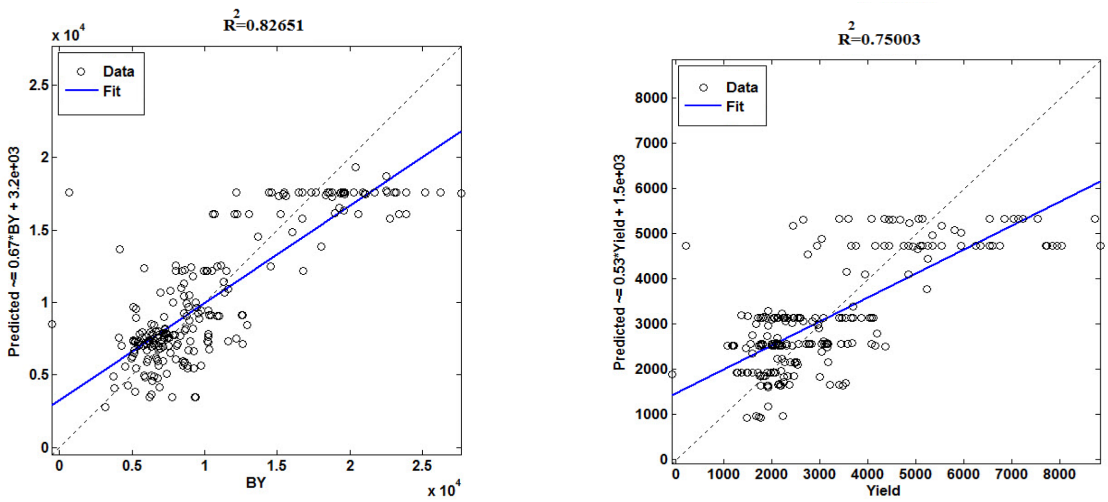

Fig 2. The scatter plot of the measured versus predicted the biological yield (BY) and grain yield of barley using radial basis function model (RBF).

Predicted BY= 0.91×BY+9.4e+02 (1)

Predicted Y = 0.85×Y + 4.3e+02 (2)

Also, using RBF model, R^2 value was 0.82 and 0.75 for prediction of BY and Y, respectively (Fig. 2). The equations for prediction of BY and Y was in the following using RBF model.

Predicted BY= 0.67×BY+32e+03 (3)

Predicted Y = 0.53×Y + 1.5e+03 (4)

Overall, according to R^2, the best model for prediction of BY and Y was MLP model using the equations of 3 and 4. This finding was in agreement with Memarian and Balasundram (2012) who reported that the MLP model showed a slightly better output than RBF network model in predicting suspended sediment discharge, especially in the training process.

Materials and Methods

Data Collection for prediction of biological yield and yield of barley

Data presented in this study was collected from the literatures on the subject of barley production in Iran that was existed in http://sid.ir website. A total of 10563 data was extracted from the literatures, including irrigation regime (according to FC), Water EC (dS/m), nitrogen applied to the soil (N; kg/ha), phosphorus applied to the soil (P; kg/ha), potassium applied to the soil (K; kg/ha), plant density (plant/m^2), growing season length (day), organic content (OC %), soil pH, rainfall amount (mm), plant height (cm), grain/spike, spike/m^2, 1000- kernel weight (gr), harvest index (HI%), crop yield (kg/ha) and biological yield (kg/ha) were prepared in Excel software sheets. Then, the Matlab software version 2014 was used for prediction of BY and Y. Characteristics of each feature parameters including average, maximum value, minimum value and standard deviation (STDEV) as inputs shown in Tables 1 and 2.

Multiple linear regression models (MLR)

The general purpose of multiple regressions is to learn more about the relationship between several independent or predictor variables and a dependent or criterion variable. The general form of the regression equations is according to Eq. 5 (Gunst and Mason, 1980; Kahane, 2008):

$$Y = A0 + A1X1 + A2X2 + \ldots + bnXn \quad (5)$$

Table 3. Simple linear correlation coefficient (r) among biological yield of barley (BY) and independent variables.

feature	Biological yield	Irrigation regime (according to FC)	Water EC (dS/m)	N(kg/ha)	P(kg/ha)	K(kg/ha)	Plant density (plant/m^2)	Growing Season length (d)	Organic content (%)	Soil pH	Rianfall amount (mm)	Plant height (cm)	Grain/spike	Spike/m^2	1000-kernel weight(gr)	HI (%)
Biological yield	1	0.413**	-0.479**	0.106	-0.295**	-0.159*	-0.072	-0.033	0.401**	-0.405**	0.097	0.561**	-0.248**	-0.540**	0.700**	-0.106
Irrigation regime (according to FC)		1	0.280**	-0.320**	0.185**	0.082	-0.040	0.156*	-0.147*	0.368**	0.239**	0.280**	0.191**	0.077	0.004	-0.232**
Water EC(dS/m)			1	-0.877**	0.470**	-0.673**	-0.304**	-0.414**	-0.139*	0.587**	-0.003	0.630**	0.325**	0.143*	0.281**	-0.149*
N(kg/ha)				1	-0.563**	0.717**	0.358**	0.359**	-0.015	-0.795**	-0.123	-0.527**	-0.507**	-0.389**	-0.101	0.129
P(kg/ha)					1	-0.538**	-0.565**	-0.391**	-0.264**	0.754**	0.178**	0.221**	0.495**	0.547**	-0.041	-0.042
K(kg/ha)						1	0.635**	0.614**	-0.168*	-0.464**	0.041	-0.405**	-0.321**	-0.276**	-0.299**	-0.092
Plant density(plant/m^2)							1	0.260**	0.003	-0.509**	-0.455**	-0.102	-0.480**	-0.432**	-0.247**	0.062
Growing season length(d)								1	-0.153*	-0.234**	0.102	-0.134*	-0.100	-0.152*	-0.111	-0.065
OC (%)									1	-0.051	0.283**	-0.365**	0.049	0.136*	-0.081	0.094
Soil pH										1	0.463**	0.166*	0.731**	0.737**	-0.196**	-0.211**
Rianfall amount(mm)											1	-0.271**	0.525**	0.423**	0.056	-0.307**
Plant height (cm)												1	-0.095	-0.290**	0.489**	-0.224**
Grain/spike													1	0.637**	-0.306**	-0.037
Spike/m^2														1	-0.511**	0.035
1000-kernel weight(gr)															1	-.095
HI(%)																1

*and ** are significant correlation at the 0.05 and 0.01 probability levels, respectively.

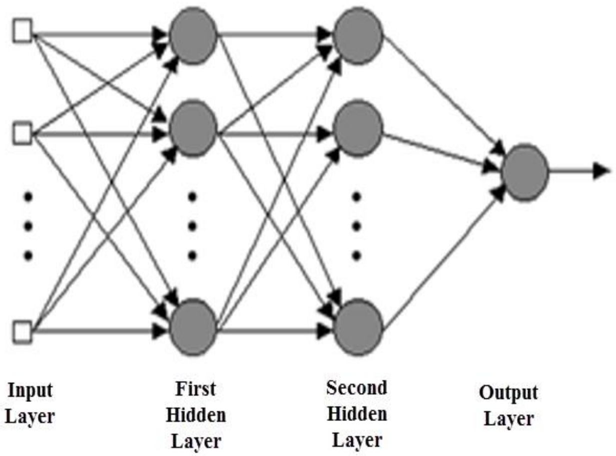

Fig 3. A graphical representation of multi-layer perceptron model (MLP).

Table 4. Simple linear correlation coefficient (r) between barley grain yield (Y) and independent variables.

feature	Grain yield	Irrigation regime (according to FC)	Water EC (dS/m)	N(kg/ha)	P(kg/ha)	K(kg/ha)	Plant density (plant/m^2)	Growing Season length (d)	Organic Content (%)	Soil pH	Rianfall amount (mm)	Plant height (cm)	Grain/spike	Spike/m^2	1000-kernel weight(gr)	HI (%)
Grain yield	1	0.197	-0.535**	0.168*	-0.324**	-0.178**	-0.054	-0.048	0.544**	-0.476**	0.111	0.156*	0.258*	0.508**	0.635**	0.622**
Irrigation regime (according to FC)		1	0.280**	-0.320**	0.185**	0.082	-0.040	0.156*	-0.147*	0.368**	0.239**	0.280**	0.191**	0.077	0.004	-0.232**
Water EC(dS/m)			1	-0.877**	0.470**	-0.673**	-0.304**	-0.414**	-0.139*	0.587**	-0.003	0.630**	0.325**	0.143*	0.281**	-0.149*
N(kg/ha)				1	-0.563**	0.717**	0.358**	0.359**	-0.015	-0.795**	-0.123	-0.527**	-0.507**	-0.389**	-0.101	0.129
P(kg/ha)					1	-0.538**	-0.565**	-0.391**	-0.264**	0.754**	0.178**	0.221**	0.495**	0.547**	-0.041	-0.042
K(kg/ha)						1	0.635**	0.614**	-0.168*	-0.464**	0.041	-0.405**	-0.321**	-0.276**	-0.299**	-0.092
Plant density(plant/m^2)							1	0.260**	0.003	-0.509**	-0.455**	-0.102	-0.480**	-0.432**	-0.247**	0.062
Growing season length (d)		.*						1	-0.153*	-0.234**	0.102	-0.134*	-0.100	-0.152*	-0.111	-0.065
OC (%)									1	-0.051	0.283**	-0.365**	0.049	0.136*	-0.081	0.094
Soil pH										1	0.463**	0.166*	0.731**	0.737**	-0.196**	-0.211**
Rainfall amount(mm)											1	-0.271**	0.525**	0.423**	0.056	-0.307**
Plant height (cm)												1	-0.095	-0.290**	0.489**	-0.224**
Grain/spike													1	0.637**	-0.306**	-0.037
Spike/m^2														1	-0.511**	0.035
1000-kernel weight(gr)															1	-0.095
HI(%)																1

*and ** are significant correlation at the 0.05 and 0.01 probability levels, respectively.

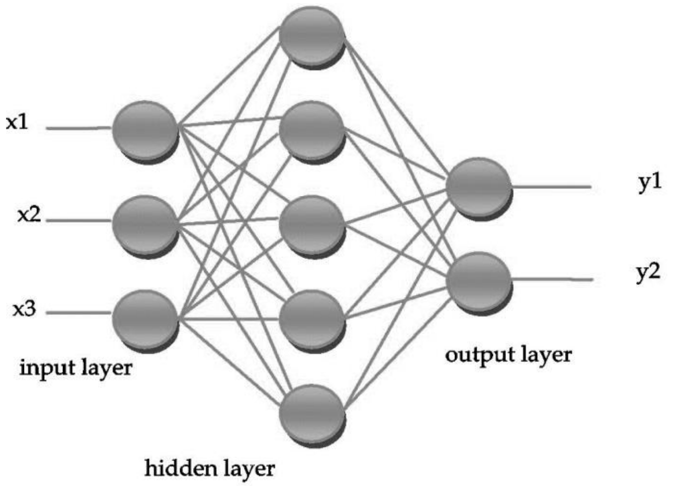

Fig 4. A graphical representation of radial basis function model (RBF).

Table 5. Model summary for MLR models for BY and other parameters value.

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	0.700[a]	0.490	0.487	3995.3451	0.490	207.352	1	216	.000
2	0.786[b]	0.618	0.615	3463.8904	0.128	72.365	1	215	.000
3	0.826[c]	0.683	0.678	3165.4561	0.064	43.451	1	214	.000
4	0.836[d]	0.699	0.694	3088.8386	0.017	11.748	1	213	.001
5	0.855[e]	0.731	0.724	2929.7662	0.031	24.758	1	212	.000
6	0.866[f]	0.750	0.743	2830.0109	0.019	16.209	1	211	.000
7	0.877[g]	0.768	0.761	2730.6066	0.018	16.642	1	210	.000

a. Predictors: (Constant), 1000-kernel weight(gr)

b. Predictors: (Constant), 1000-kernel weight(gr), OC (%)

c. Predictors: (Constant), 1000-kernel weight(gr), OC (%), grain/spike

d. Predictors: (Constant), 1000-kernel weight(gr), OC (%), grain/spike, soil pH

e. Predictors: (Constant), 1000-kernel weight(gr), OC (%), grain/spike, soil pH, N applied (kg/ha)

f. Predictors: (Constant), 1000-kernel weight (gr), OC (%), grain/spike, soil pH, N applied (kg/ha), plant height (cm)

g. Predictors: (Constant), 1000-kernel weight(gr), OC (%), grain/spike, soil pH, N applied (kg/ha), plant height (cm), irrigation regime (according to FC)

Table 6. Model summary for MLR models for Yield and other parameters value.

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	0.635[a]	0.404	0.401	1351.3047	0.404	146.201	1	216	.000
2	0.749[b]	0.561	0.557	1161.9994	0.157	77.111	1	215	.000
3	0.819[c]	0.671	0.667	1008.0154	0.110	71.704	1	214	.000
4	0.852[d]	0.727	0.721	921.3503	0.055	43.152	1	213	.000
5	0.865[e]	0.748	0.742	886.2982	0.022	18.181	1	212	.000
6	0.876[f]	0.768	0.761	853.4668	0.019	17.624	1	211	.000
7	0.883[g]	0.780	0.772	833.1450	0.012	11.419	1	210	.001
8	0.885[h]	0.784	0.776	826.7479	0.004	4.262	1	209	.040

a. Predictors: (Constant), 1000-kernel weight(gr)
b. Predictors: (Constant), 1000-kernel weight (gr), OC (%)
c. Predictors: (Constant), 1000-kernel weight(gr), OC (%), soil ph
d. Predictors: (Constant), 1000-kernel weight(gr), OC (%), soil pH, grain/spike
e. Predictors: (Constant), 1000-kernel weight(gr), OC (%), soil pH, grain/spike, %HI
f. Predictors: (Constant), 1000-kernel weight(gr), OC (%), soil pH, grain/spike, HI(%), plant height (cm)
g. Predictors: (Constant), 1000-kernel weight (gr), OC (%), soil pH, grain/spike, HI(%), plant height (cm), irrigation regime (according to FC)
h. Predictors: (Constant), 1000-kernel weight (gr), OC (%), soil pH, grain/spike, HI(%), plant height (cm), irrigation regime (according to FC), Plant density (plant/m^2)

Table 7. Performance indices and coefficients of variables for different MLR models for BY value.

Model	Standardized Coefficients		
	SE	β	t
(Constant)	9071.37	0.112	8.806
Irrigation regime(according to fc)	49.33	0.125	3.461
OC %	949.95	0.396	11.278
Soil pH	1232.80	-0.649	-12.094
Plant height (cm)	14.29	0.179	4.103
Grain/spike	24.77	0.385	7.605
1000-kernel weight(gr)	22.81	0.621	15.633
HI(%)	34.99	-0.138	-4.018

SE: Standard error, β: standardized coefficients, t: the t statistic tests the hypothesis that a population regression coefficient β is 0, that is, H_0: β = 0 and it is the ratio of the sample regression coefficient β to its standard error.

Table 8. Performance indices and coefficients of variables for different MLR models for Y value.

Model	Standardized Coefficients		
	Std. Error	β	t
(Constant)	3063.21	0.101	8.016
Irrigation regime (according to fc)	15.60	0.136	3.721
OC %	295.85	0.403	11.548
Soil pH	412.24	-0.690	-12.021
Plant height (cm)	4.50	0.175	3.972
Grain/spike	7.99	0.347	6.629
1000-kernel weight (gr)	8.22	0.547	11.947
HI (%)	10.89	0.180	5.277
Plant density (p/m2)	0.55	-0.093	-2.065

SE: Standard error, β: standardized coefficients, t: the t statistic tests the hypothesis that a population regression coefficient β is 0, that is, H_0: β = 0 and it is the ratio of the sample regression coefficient β to its standard error.

Table 9. Correlation coefficient (R^2) for training using different models.

Model	Parameter	R^2
MLR	Yield	0.784
	BY	0.781
MLP	Yield	0.922
	BY	0.894
RBF	Yield	0.750
	BY	0.826

MLP: multi-layer perceptron, MLR: multiple linear regression; RBF: radial basis function.

Where Y is the dependent variable, A_0 is the intercept, A_1. . .b_n are regression coefficients, and X_1– X_n are independent variables referring to basic soil properties.

Multi-layer perceptron networks model (MLP)

Multi-layer perceptron (MLP) network models are the most common network architectures used in most of the study applications in medicine, engineering, mathematical modeling, agriculture, etc. In MLP, the weighted sum of the inputs and bias term are passed to activation level through a transfer function to produce the output, and the units are arranged in a layered feed-forward topology called Feed Forward Neural Network (Venkatesan and Anitha, 2006). The goal of this type of network is to create a model that correctly maps the input to the output using historical data so that the model can then be used to produce the output when the desired output is unknown. A graphical representation of an MLP was shown in Fig. 3. Based on Fig. 3, there were three inputs layers so that, two hidden layers can predict once output layer.
MLP is perhaps the most popular ANN architecture (Dawson and Wilby, 1998). It is a network formed by simple neurons called perceptron. The perceptron computes a single output from multiple real-valued inputs by forming a linear combination according to input weights and then possibly subjecting the output to some nonlinear transfer function (Fig. 3). Mathematically this can be represented as (Sudheer et al., 2002):

$$y = f\left(\sum_{i=1}^{n} w_i p_i + b\right) \quad (6)$$

where, w_i represents the weight vector, p_i is the input vector (*i=1,2, …, n*), *b* is the bias, *f* is the transfer function, and *y* is the output. The transfer function used in this study was the tangent sigmoid function defined for any variable *s* as (Sudheer et al., 2002):

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \quad (7)$$

MLP is usually trained using the back error propagation algorithm. This popular algorithm works by iteratively changing a network's interconnecting weights such that the overall error (i.e., between observed values and predicted outputs by ANNs) is minimized (Sudheer et al., 2002).

Radial basis function model (RBF)

Radial Basis Function (RBF) neural network is based on supervised learning. RBF networks were independently proposed by many researchers and are a popular alternative to the MLP. RBF networks are also good that can be trained in one stage rather than using an iterative process as in MLP and also learn the given application quickly (Venkatesan and Anitha, 2006). Using the neural network concepts, one may refer to the RBF net as a very simple three-layer feed forward network. A typical architecture was shown in Fig. 4.
The RBF model, developed by Powell (1987) and Broomhead and Lowe (1988), consists of an input layer, a single hidden layer, and an output layer. Figure 4 shows a typical RBF model. The number of input and output nodes is similar to the MLP neural networks, determined by the nature of actual input and output variables. However, RBF networks tend to learn much faster than a MLP. The output of RBF was calculated as:

$$Y = \sum_{p=1}^{p} w^{p} \theta(\| x - x_p \|) \quad (8)$$

Where X is the input value, Y is the output value, θ is the radial basis function, *w* is the weight connecting the hidden and output nodes, X_p represents the center of each hidden node (depends on the observed input data), and $\| x - x_p \|$ is the Euclidean distance between input and hidden nodes.

Performance evaluation criteria

In this study, correlation coefficient (R^2, Eq. 5) and mean square error (MSE, Eq. 6) indices were calculated to control the performance of the prediction capacity of predictive models which developed and used by Finol et al., (2001), and Yilmaz and Kaynar (2011).

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)}{\sum_{i=1}^{N}(y_i - \bar{y}_i)}^2 \quad (9)$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \quad (10)$$

Where N is the number of data, y_i is the measured value of each variable, $\hat{y}_i$ is the predicted value of each variable and $\bar{y}_i$ is average of predicted value of each variable.

Conclusion

This study focused to prediction of BY and Y through multiple linear regression (MLR) and artificial neural networks (ANNs) with two algorithms (MLP and RBF) and comparison the predictive performance of these models by means of some statistical indicator such as mean square error (MSE) and correlation coefficient (R^2) in Iran country. Overall, in MLR model Y and BY more affected by 1000- kernel weight. It concluded that, for prediction of BY, the best model was MLP (R^2=0.89). Also, the R^2 value (0.92) for MLP model was higher than MLR and RBF for prediction of Y. Generally, ANNs can be used successfully to predict the BY and Y from a dataset and this study opened a new vista in barley production using ANNs models.

Acknowledgments

We would like to thank the Research Council of Shiraz University in Iran for financial support of this research.

References

Afshari N, Askari L, Hedaiati S (2011) Effect of terminal drought stress on yield and yield components of barley (*Hordeum vulgare* L.) cultivars. J New Physio Sci. 8: 3-8. (In Persian).

Alqudah AM, Schnurbusch T (2014) Awn primordium to tipping is the most decisive developmental phase for spikelet survival in barley. Funct Plant Biol. 41:424-436.

Austin RB (1984) Crop characteristics and the potential of wheat yield. J Agric Sci Camb. 98: 447-453.

Bijanzadeh E, Emam Y, Ebrahimie E, Ebrahimie M (2012) Application of unsupervised weighting algorithms for identifying important attributes and factors contributing to grain and biological yields of wheat. Crop Breed J. 2:111-117.

Bijanzadeh E, Emam Y (2012) Evaluation of assimilate remobilization and yield of wheat cultivars under different irrigation regimes in an arid climate. Arch Agron Soil Sci. 58: 1243-1259.

Bijanzadeh E, Naderi R (2014) Application of supervised and unsupervised algorithms to find the important features related to barley (*Hurdeum vulgare* L.) grain yield: A new vista in data mining. Aust J Crop Sci. 8:1590-1596.

Broomhead DS, Lowe D (1988) Multivariable functional interpolation and adaptive networks. Comp Sys. 2:321-355.

Dawson CW, Wilby R (1998) An artificial neural network approach to rainfall-runoff modeling. J Hydrol Sci. 43:47-66

Dreccer MF, Grashoff C, Rabbinge R (1997) Source-sink ratio in barley (*Hordeum vulgare* L.) during grain filling: effects on senescence and grain protein concentration. Field Crops Res. 49:269–277.

Emam Y (2002) Effect of planting density and chlormequat chloride on morphological and physiological characteristics of winter barley (*Hordeum vulgare* L.) cultivar. J Agric Sci Tech. 2:75-83.

Etemadi R, Naghavi A, Jafarii S, Sedaghat L (2005) Investigation of grain yield in some cultivars and genotypes of barely (*Hordeum vulgare* L.) using drought tolerance indices. J Crop Breed. 4:75–83. (In Persian)

Finol J, Guo YK, Jing, XD (2001) A rule based fuzzy model for the prediction of petrophysical rock parameters. J Pet Sci Eng. 29: 97-113.

Ghamari S, Borghei AM, Rabbani H, Khazaei J, Basati F (2010) Modeling the terminal velocity of agricultural seeds with artificial neural networks. Afr J Agr Res. 5: 389-398.

Gunst, RF, Mason RL (1980) Regression analysis and its application. Marcel Dekker Inc., New York.

Heinzow T, Richard T (2002) Prediction of crop yields across four climate zones in Germany: an artificial neural network approach. Working paper FNU-34 Centre for marine and climate research, Hamburg University, Germany.

Hosseini SM, Seyosemorde M, Fathi P, Seyosemorde A (2007) Application of ANN and multi-variable regression models on yield of dryland wheat in Gorve of Kordestan. Agric Res J. 7: 41-54. (In Persian)

Kahane LH (2008) Regression Basics, 2nd Ed. SAGE Publications Inc. Los Angeles.

Kaul M, Hill RL, Walthall C (2005) Artificial neural networks for corn and soybean yield prediction. Agric Sys. 85: 1-18.

Khairunniza-Bejo S (2014) Samihah M, Wan I (2014) Application of artificial neural network in predicting crop yield: a review. J Food Sci Eng. 4: 1-9.

Khashei-Siuki A, Kouchakzadeh M, Ghahraman, B (2011). Predicting dryland wheat yield from meteorological data using expert system, Khorasan Province, Iran. J Agr Sci Tech. 13: 627-640

Khazaei J, Chegini GR, Bakhshiani M (2008) A novel alternative method for modeling the effects of air temperature and Slice thickness on quality and drying kinetics of tomato slices. Int Dry Tech J. 26: 759-775.

Khoshhal J. Mokarram, M (2012). Model for prediction of evapotranspiration Using MLP neural network. Int J Envi Sci. 3: 1000-1009.

Malakoti MJ (2003) Recommendations and applications of chemical fertilizers in fields of Iran. Tarbiat Modarres University Press.

Memarian H, Balasundram SK (2012) Comparison between multi-layer perceptron and radial basis function networks for sediment load estimation in a tropical watershed. J Water Res Prot. 4: 870-876.

Nikkhah HR, Saberi MH, Mahlouji M (2010) Study of effective traits on grain yield of two and six row barley genotypes (*Hordeum vulgare* L.) under terminal drought stress condition. Iran J Crop Sci. 12:170-184. (In Persian).

O'Neal MR, Engel BA, Ess DR, Frankenberger JR (2002) Neural network prediction of maize yield using alternative data coding algorithms. Bio Eng. 83: 31-46.

Park SJ, Hwang CS, Vlek PL (2005) Comparison of adaptive techniques to predict crop yield response under varying soil and land management conditions. Agric Sys. 85: 59–81.

Poluektov RA Topaj AG (2001) Crop modeling: nostalgia about present or reminiscence about future. Agron J. 93: 653-659.

Powell MJD (1987) Radial basis functions for multivariable interpolation: a review. In: Mason JC, Cox MG (eds) Proceedings of IMA Conference on Algorithms for Approximation. Oxford UniversityPress, New York, pp 143-167

Rezaii S, Naseri F, Namazi B (2010) Evaluation of some physiological and morphological characteristic correlated to yield of barley cultivars. Weed Crop Ecophys. 21:43-51. (In Persian).

Rumelhart DE, McClelland JL (1986) Parallel distributed processing: Explorations in the microstructure of cognition. Foundations, Cambridge, MIT Press.

Saberali SF, Sadat Noori SA, Khazaei J, Hejazi A (2007) Artificial neural network modelling of common lambs quarters biomass production response to corn population and planting pattern. Pak J Biol Sci. 10: 326-334.

Stathakis D, Savina I, Negrea T (2006) Neuro-fuzzy modeling for crop yield prediction. Rem Sen Spa Inf Sci. 34:18-26.

Sudheer KP, Gosain AK, Ramasastri KS (2002) A data driven algo-rithm for constructing artificial neural network rainfall-runoff models. J Hydr Proc. 16:1325-1330.

Tambussi EA, Nogues S, Ferrio JP, Voltas J, Araus JL (2005) Does a higher yield potential improve barley performance under Mediterranean stresses conditions? A case study. Field Crop Res. 91: 149-160.

Uno Y, Prasher, SO, Lacroix R, Goel PK, Karimi, Y, Viau A, and Patel RM (2005) Artificial neural networks to predict corn yield from compact airborne spectrographic imager data. Comput Electron Agric. 47: 149-161.

Venkatesan P, Anitha S (2006) Application of a radial basis function neural network for diagnosis of diabetes mellitus. Cerr Sci. 91: 10-16.

Yilmaz I, Kaynar O (2011) Multiple regression, ANN (RBF, MLP) and ANFIS models for prediction of swell potential of clayey soils. Exp Sys App. 38: 5958–5966.