

PREDICTION OF STEM WEIGHT IN SELECTED ALFALFA VARIETIES BY ARTIFICIAL NEURAL NETWORKS, MULTIVARIATE ADAPTIVE REGRESSION SPLINES AND MULTIPLE REGRESSION ANALYSIS

Ş. Çelik^{1,*}, E. Çağan² and S. Yaryab³

¹Department of Genetic and Biometry, Faculty of Agriculture, Bingöl University, 12000, Bingöl, Türkiye

²Department of Plant and Animal Production, Vocational School of Genc, University of Bingol, Türkiye

³Department of Field Crops, Institute of Science, University of Bingol, Türkiye

*Corresponding author's email: senolcelik@bingol.edu.tr

ABSTRACT

In this study, artificial neural networks (ANNs), Multivariate Adaptive Regression Splines (MARS) algorithm and multiple regression analysis (MLR) were used for plant stem weight prediction. Stem length, stem diameter, number of lateral branch, branch length, leaf number, stipule length and distance between stipules have been selected as input variables in these mentioned methods. A total of 150 plants were examined. Fifty plants from each of Gea, Bilensoy and Basbag alfalfa cultivars were analyzed separately. Our alfalfa varieties in this study are Gea, Bilensoy and Başbağ. In the ANN method, 70% of the data were allocated for training 20% for validation and 10% for testing. ANN training data were used in MARS algorithm and MLR. To measure which models can predict better, the coefficient of determination (R^2) and mean square error (MSE) were compared each other. Correlation coefficients (r) of ANN, MARS and MLR in Stem Weight estimation were 0.801, 0.999 and 0.753 for Gea clover variety, respectively; 0.864, 0.997 and 0.711 for Bilensoy variety, respectively, and 0.781, 0.998 and 0.561 for the Basbag variety, respectively. In the same models, R^2 was 0.642, 0.998 and 0.567 for the Gea variety, respectively, 0.746, 0.994 and 0.505 for the Bilensoy variety, respectively, and 0.610, 0.997 and 0.315 for the Basbag variety, respectively. MSE values were 0.023, 0.008 and 2.498 for the Gea variety, respectively, 0.113, 0.014 and 1.409 for the Bilensoy variety, respectively, and 0.151, 0.017 and 4.641 for the Basbag variety, respectively. According to these criteria, the MARS algorithm provides a more realistic prediction than ANNs and MLR. The order of used algorithms in obtaining better prediction results in stem weight estimation in alfalfa plants was MARS> ANN> MLR.

Keywords: Artificial neural network, *Medicago sativa* L., Plant Characteristics, Regression

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Published first online April 15, 2023

Published final August 04, 2023

INTRODUCTION

Alfalfa (*Medicago sativa* L.) is an important perennial forage plant with deep and strong root system in the family of legumes (*Fabaceae*) (Davis, 1978). Its homeland is Asia, Iran, Turkmenistan and the surrounding regions (Bolton, 1962). There are about 60 alfalfa species (Lesins and Gillies, 1972). Alfalfa naturally grows in a wide area extending from South and Central Europe to the near east and Japan (FAO, 2013). Alfalfa (*Medicago sativa*), is successfully grown in Turkey cold regions such as Central Anatolia and Eastern Anatolia, as well as in Southern regions of Turkey (Saglamtimur *et al.*, 1990). Because of its strong root system (which can descend from 4 m to 9 m) alfalfa is a drought-resistant plant (Volaire, 2008). Besides being an easy produced yield, having a low cost, and containing a

rich source of crude protein for animal husbandry, alfalfa also has high digestibility (Radovic *et al.*, 2009). In addition, alfalfa is very rich in minerals and vitamins (Altinok and Karakaya, 2002). In terms of high yield potential and adaptation properties to different environmental conditions, alfalfa varieties also show a wide range of genetic variations (Hill *et al.*, 1988). Alfalfa is adaptable to different climatic and soil conditions can be grown in almost all regions in Turkey (Erisen, 2005). Since alfalfa plants show foreign pollination, alfalfa varieties generally consist of many parent plants and have a broad genetic base (Şengül and Sagsoz, 2004).

Alfalfa breeding programs have focused on forage yield and quality, resistance to biotic and abiotic stressors, and fall dormancy. Conventional breeding is typically based on simple phenotypic selection, in which

each plant must be phenotyped, and on pedigree-based methods such as BLUP (Piepho *et al.*, 2008) which attempt to predict individual breeding values based on pedigree.

The stem weight are forage production indicators in alfalfa (Matthew *et al.*, 1996). Authors obtained lower and higher weight per stem (0.27 and 0.45 g) at 30 and 50 d.

One of the best ways to increase yield profitability is to identify production-related problems at an early stage. ANNs are among the most accurate and widely used methods for data mining and forecasting (Rad, 2018). The main excellence of this approach is that ANNs use ordinary patterns for simulations, and they do not use complex mathematical calculations (Razavi *et al.*, 2003). One of the most important advantages of this method is modeling between input and output information with a nonlinear relationship. On the other and in traditional modeling methods, modeling of this information is done with many errors (Zhang *et al.*, 2012). In research on cultivation as well as yield production management forecasts, the ANN method has provided more accurate results than traditional methods (Solaimany-Aminabad *et al.*, 2013). According to Jiang *et al.* (2004)'s studies, estimating winter wheat harvest by the ANN method gave very realistic results. Uno *et al.* (2005) tried to estimate the corn harvest data with different statistical methods by choosing a large number of parameters and stated that the ANN method gave the best results. Higgins *et al.* (2010) reached the same conclusion in a classification based on repeatedly growing trees.

The literature review yielded no studies that predicted stem weight from the morphological characteristics of the alfalfa plant. The fact that it will be done in this study for the first time increases the importance of this study.

The purpose of this study is to determine the stem weight from various plant properties in different types of alfalfa plants using the MARS algorithm, ANNs and MLR.

MATERIALS AND METHODS

Plant characteristics and measurements used in this research were taken from the research project that was established to determine the yield and qualities of some alfalfa genotypes in 2019 at Bingöl-Genç Vocational School Application and Research Area in Turkey. Fifty plant samples belonging to Gea, Bilensoy-80 and Basbag alfalfa varieties were taken randomly from the trial area. The average length of the growing season is 150 days (June to October), and the average annual precipitation is 110 mm. The research site has no identifiable water tables. Alfalfa (*Medicago sativa*) was planted on 25 June 2019. The plot was irrigated uniformly at a rate of 10 mm per week to the end of the 2019 irrigation season (5 October) to insure uniform germination. The statistics of the various plant properties related to the Gea, Bilensoy and Başbağ varieties of alfalfa are presented in Table 1. Stem weight is output variable. Other variables are input variables.

Table 1. Descriptive Statistics

	Genotype	SW	SL	SD	LSN	LSL	TL	DBE	EL
N	Gea	50	50	50	50	50	50	50	50
	Bilensoy	50	50	50	50	50	50	50	50
	Basbag	50	50	50	50	50	50	50	50
\bar{X}	Gea	5.85	53	2.05	7.02	8.09	74	9.07	1.59
	Bilensoy	3.21	43.9	1.42	6.2	5.47	49.1	7.05	1.33
	Basbag	5.16	56.6	2.49	7.06	5.64	56.6	9.31	1.08
s_x	Gea	0.32	1.2	0.074	0.184	0.426	3.52	0.39	0.288
	Bilensoy	0.22	1.09	0.074	0.194	0.201	3.59	0.29	0.091
	Basbag	0.34	1.58	0.096	0.209	0.255	3.15	0.26	0.045
s	Gea	2.22	8.48	0.52	1.3	3.01	24.9	2.76	2.04
	Bilensoy	1.56	7.7	0.53	1.37	1.42	25.4	2.02	0.643
	Basbag	2.41	11.1	0.68	1.48	1.81	22.3	1.85	0.315
Minimum	Gea	2.63	33	0.88	5	4	28	0.3	0.1
	Bilensoy	1.14	23	0.4	4	3	14	2.3	0.6
	Basbag	2.15	30	0.8	4	3	21	5.3	0.4
Maximum	Gea	10.4	68	3.08	12	25.5	142	14.2	13.5
	Bilensoy	9.36	63	2.91	11	8.5	118	12	4.5
	Basbag	12.8	76	3.9	11	10.2	104	13	1.9

\bar{X} : Mean, s_x : Std. Error, s: Std. Deviation. SW: Stem weight (g), SL: Stem length (cm), SD: Stem diameter (cm), LSN: Lateral stem number, LSL: Lateral stem length (cm), TL: Total leaf, DBE: Distance between the ears (cm), EL: Ear length (cm).

It was possible to access 50 plants (number of plants) for each possible plant species. Efforts were made to achieve the best possible result with the training and testing performed with ANNs. The data was obtained by measuring the samples' stem weight, stem length, stem diameter, number of lateral branches, length of lateral branches, leaf number, stipule length, and the distance between stipules.

Before training any network, the general practice is to first divide the data into three subsets training, validation, and testing. The “*dividerand*” function of the MATLAB program is used to divide the data randomly. Out of the whole database 70% (35 data)

were used for training, 20% (10 data) for validation, and the rest 10% (5 data) for testing. Training, validation and testing were done separately for each plant variety. The number of data in each group was kept balanced with the utmost care. Furthermore, for validation the data should cover the whole range.

An estimate of complex system output with different impressive input parameters is the capability of artificial intelligence system such as ANN systems. ANNs are composed of interconnected neurons that are placed in three different layers (Nowruzi and Ghassemi, 2016): Input layer, hidden layers, and output layer (Figure 1).

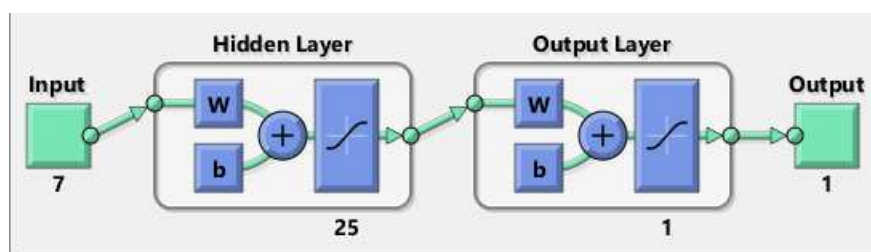


Figure 1. Two-layer feed forward network structure

In this study the activation function used to calculate the output was the Hyperbolic Tangent function (Spiegel *et al.*, 2009).

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

When the predicted is done by ANNs, two-layer feed forward neural networks containing hidden neurons and output neurons with provided data and hidden layers containing a sufficient number of neurons are well adapted to multi-dimensional imaging problems. The network was trained with the Levenberg-Marquardt back propagation algorithm (trainlm). A single hidden layer was used for this network. The number of neurons in a given layer was 25. The number of output neurons was determined by the number of elements in the target vector. The network was manually configured for better performance.

MARS (Multivariate adaptive regression splines), a nonlinear and nonparametric regression method, was first introduced by Friedman (1991) as a pliable process that replicates interactions between inputs and outputs with fewer variable.

In training set, the V three-fold cross-validation resampling technique was used to select the best predicting MARS model with degree = 1:4 and nprune = 5:40 as the number of selected terms. The V three-fold cross-validation resampling technique was used to find the best model. To maximize the predictive accuracy of the MARS algorithm, the penalty definition was taken as -1. The number of basis functions used in MARS algorithm was 38, 33, and 34 for Gea, Bilensoy and

Basbag varieties, respectively. The degree of interaction that is very important was a maximum 3.

The fit of the entire MARS model in the study was evaluated using statistics such as Coefficient of Determination (R^2), Standard Deviation Ratio (SD ratio), Root Mean square error (RMSE), Mean square error (MSE), and Mean absolute percent error (MAPE) (Eyduran *et al.*, 2019; Zaborski *et al.*, 2019).

Stem weight was estimated by using the leaf characteristics of the alfalfa plant with the ANN method in the Matlab R2016a program (MathWorks, Inc., 2016) and R Program (R Core Team, 2019).

RESULTS

As a result of the training of the network, a graphic showing the error values related to training, validation and test sets in each iteration was obtained. As can be seen in the graphic, the number of iterations in the training of the network was taken as 24, and the best performance was obtained at the 18th iteration (Figure 2).

In Figure 3, the regression values (0.86407, 0.92686, 0.73474 and 0.8541) of the data assigned for training, validation and testing are given. These results indicate that the learning process of the network was carried out successfully. Briefly, coefficients of correlation were found as 0.86407, 0.92686, 0.73474 and 0.8541, respectively for the training, validation, testing and all data in Figure 3. The mean square error (MSE) of the model has been determined as 0.335 for Gea alfalfa. The r , RMSE, MSE, R^2 , MAPE and Sdratio values

calculated for the ANN method are 0.854, 0.579, 0.335, 0.642, 2.426 and 0.729, respectively.

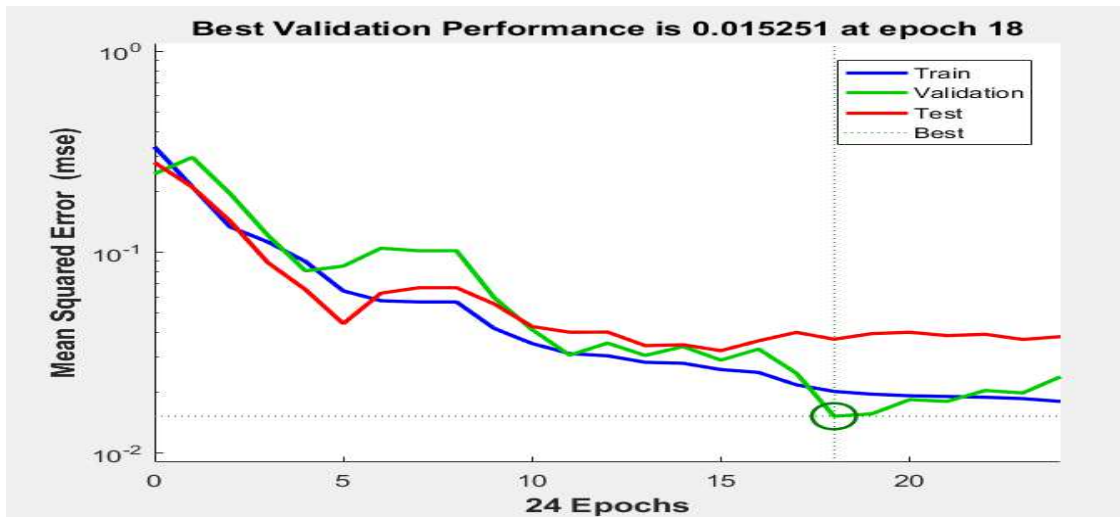


Figure 2. Performance of artificial neural network

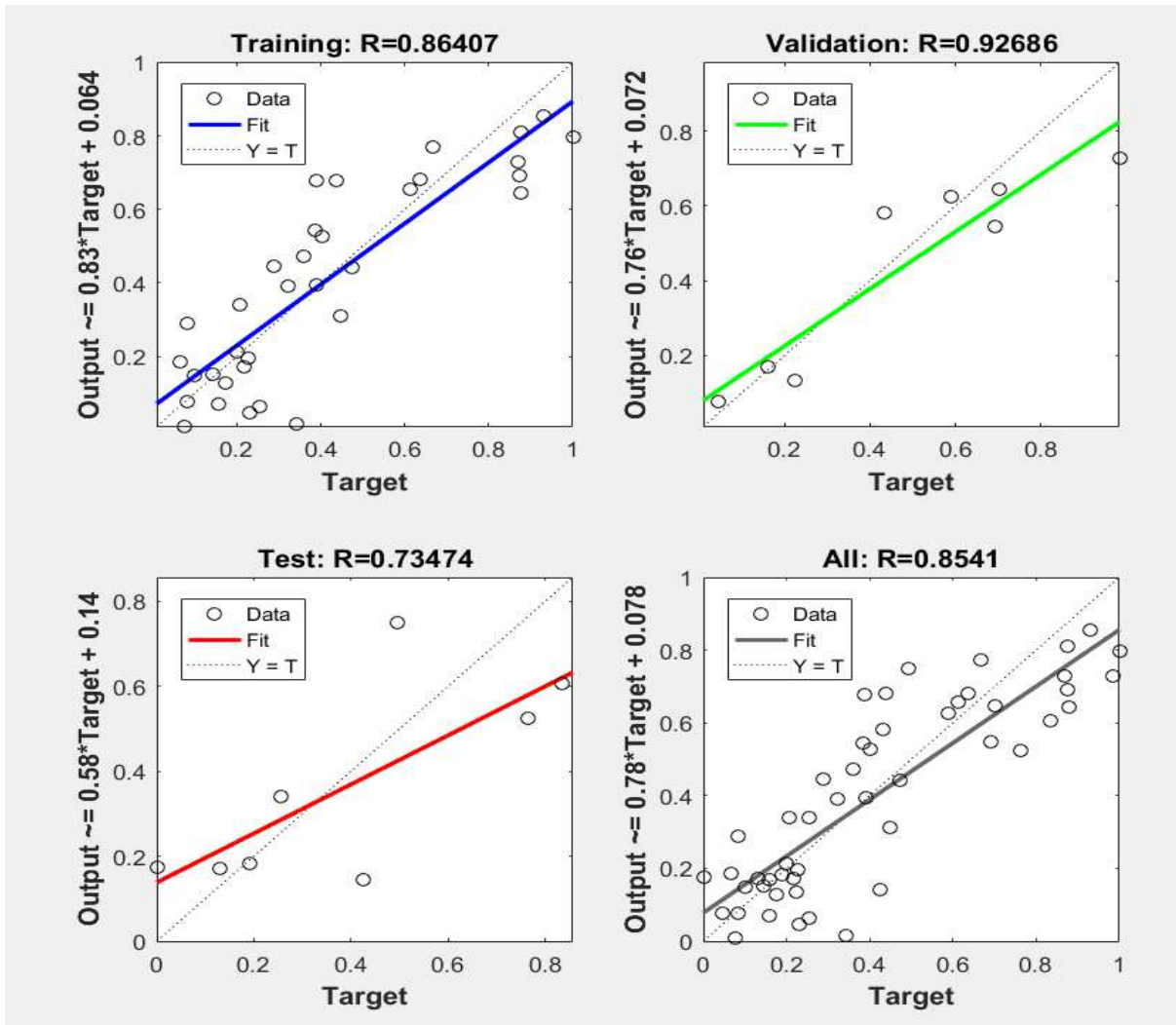


Figure 3. Regression graphics related to training, validation and test sets

Also the MARS algorithm was used to estimate stem weight (SW) in the Gea variety. The main functions, the coefficients and significance value of the MARS algorithm are presented in Supplementary Material (Table 2).

According to the results presented in Supplementary Table S2, the model created according to

the MARS algorithm had 38 basis functions. According to these results; the main functions and variables that had the greatest positive effect on stem weight (SW) prediction in the Gea variety of alfalfa were summarized as follows.

The variables that contribute to SW are listed in order of importance (Figure 4).

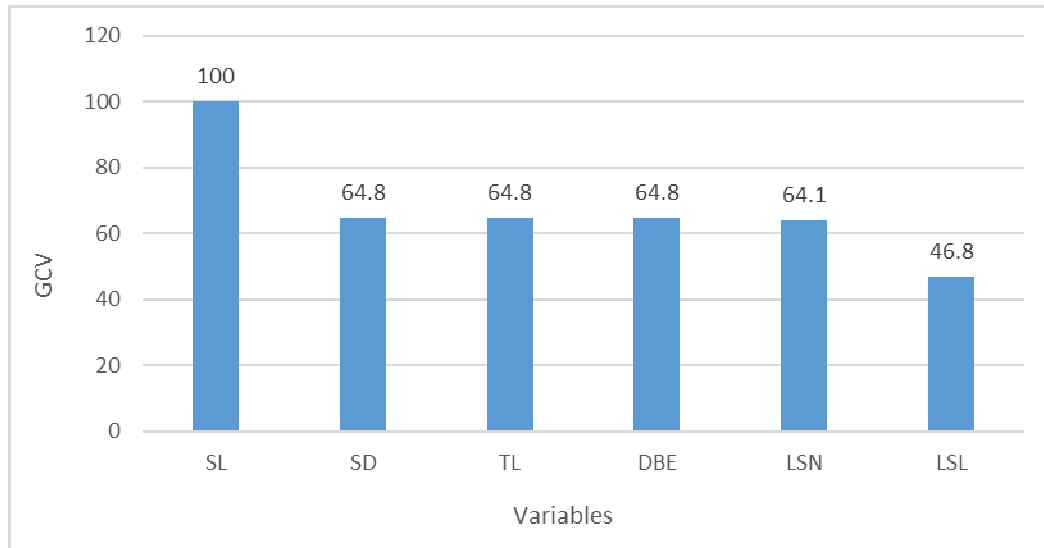


Figure 4. Most important variables for Gea variety of alfalfa

The correlation coefficient was $r = 0.999$, in addition $RMSE = 0.089$, $MSE = 0.008$, $R^2 = 0.998$, $MAPE = 1.284$ and $Sdratio = 0.04$ were obtained.

The ANN model results for the stem weight prediction of the Bilensoy variety of alfalfa plant are summarized as follows. As a result of the training of the

network, the change of error values related to training, validation and test sets in each iteration is shown graphically in Figure 5. As seen in the graphic, the number of iterations in the training of the network was taken as 122, and the best performance was obtained at the 116th iteration (Figure 5).

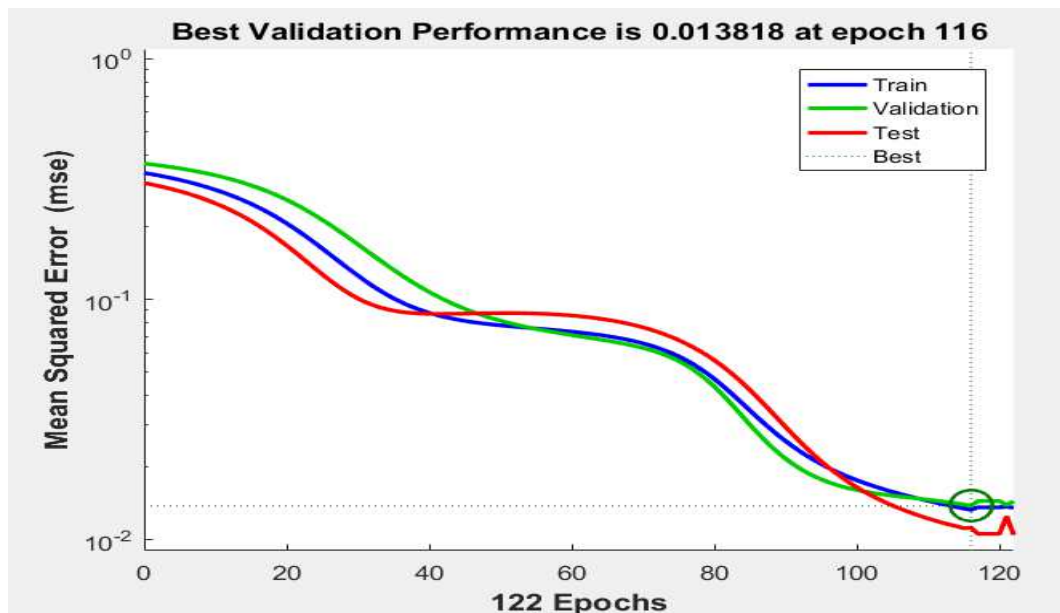


Figure 5. Performance of artificial neural network

The correlation values of the data used for training, verification and testing are given in Figure 6. The most important value is that for the test set and this value was 0.85. The mean square error (MSE) of the

model was calculated as 0.336 for training. The r , RMSE, MSE, R^2 , MAPE and Sdratio values calculated for the ANN method are 0.799, 0.580, 0.336, 0.638, 4.352 and 0.403, respectively.

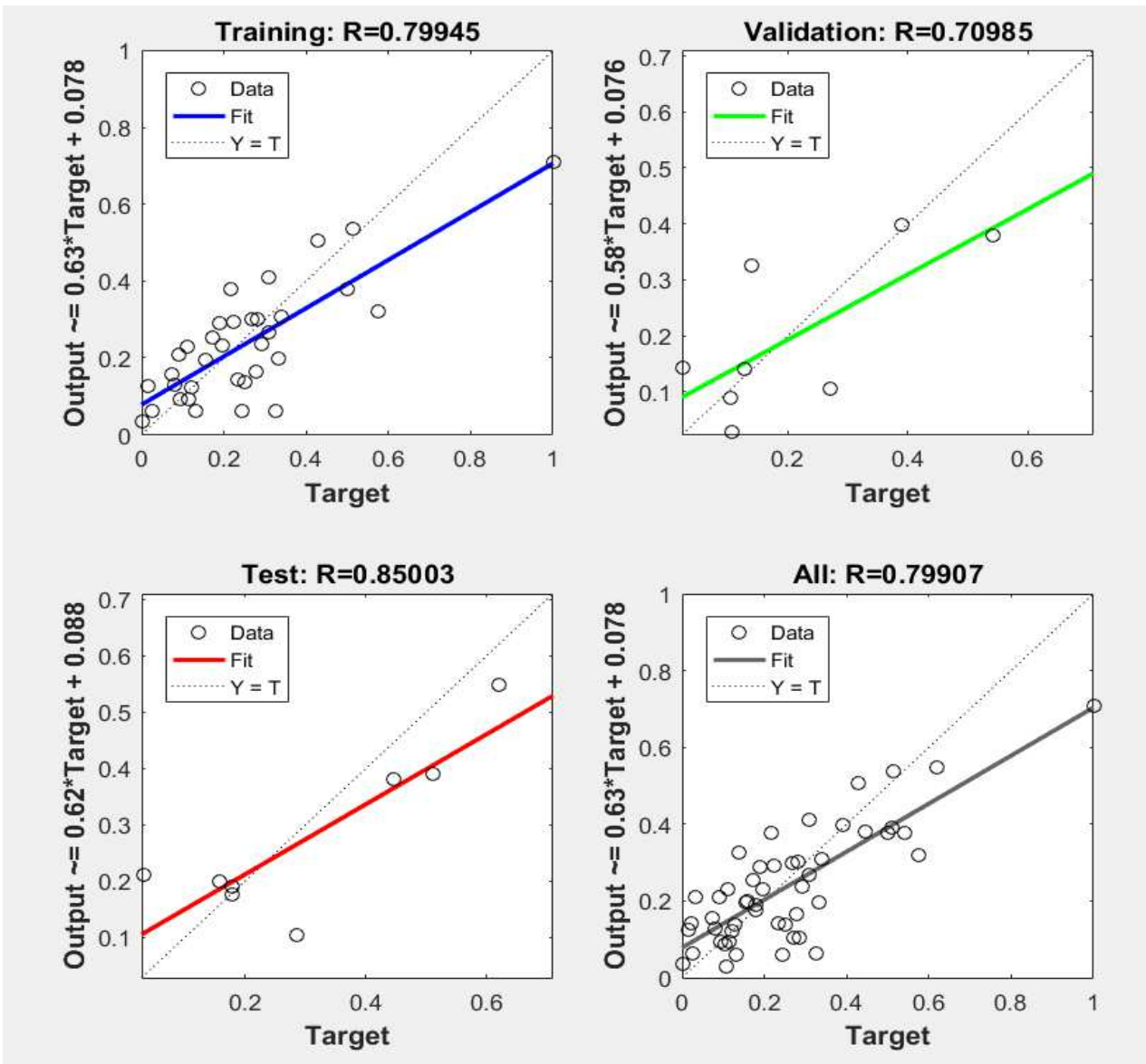


Figure 6. Regression graphics related to training, validation and test sets

SW estimations for Bilensoy variety were predicted by MARS algorithm were presented in Supplementary Material (Table S3).

According to the results presented in Table S3, the model created according to the MARS algorithm had 33 basic functions. According to these results, the main functions and variables that had the greatest positive effect on stem weight (SW) prediction in the Bilensoy variety of alfalfa were summarized as follows.

The variables that contributing to SW were listed in order of importance (Figure 7).

To determine the suitability of the model, $r = 0.997$, RMSE = 0.118, MSE = 0.014, $R^2 = 0.994$, MAPE = 3.517 and Sd ratio = 0.076 were extracted. These values are the goodness of fit statistics of the model.

ANN model results for the stem weight prediction of Basbag variety of alfalfa plant are summarized as follows. As a result of the training of the network, the change of error values related to training, validation and test sets in each iteration is shown graphically in Figure 10. As seen in the graphic, the number of iterations in the training of the network was

taken as 16, and the best performance was obtained at 10th iteration (Figure 8).

The correlation values of the data used for training, validation and testing are given in Figure 9. The

mean square error (MSE) of the model was calculated as 0.123. The r , RMSE, MSE, R^2 , MAPE and Sdratio values calculated for the ANN method were 0.7847, 0.35, 0.123, 0.616, 4.501 and 0.463, respectively.

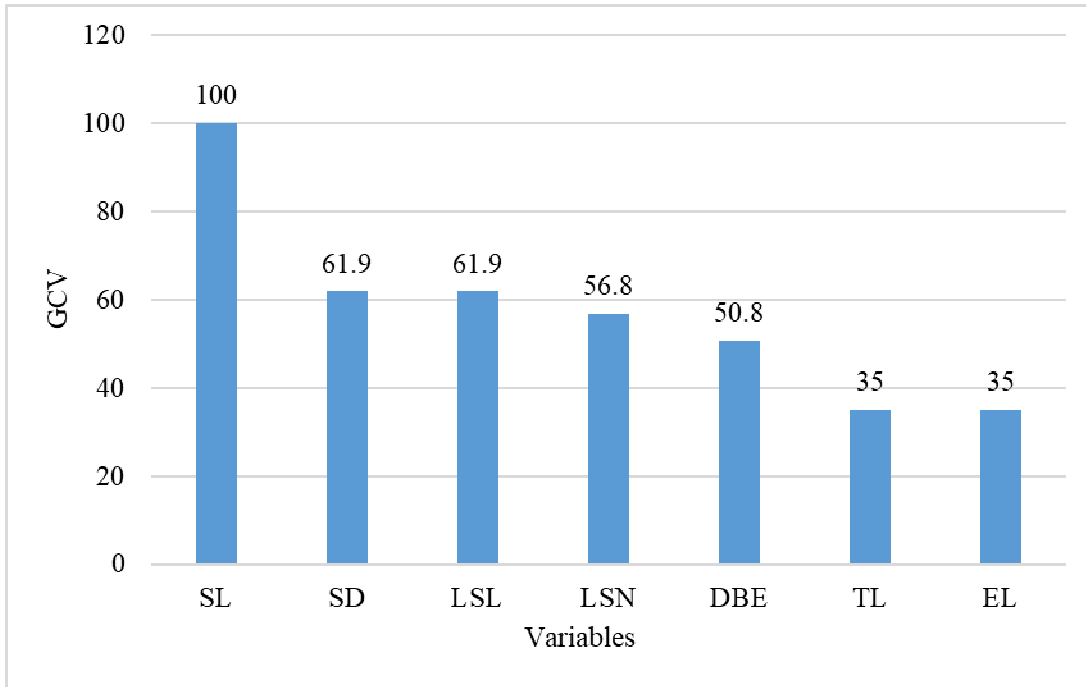


Figure 7. Most important variables for Bilensoy variety of alfalfa

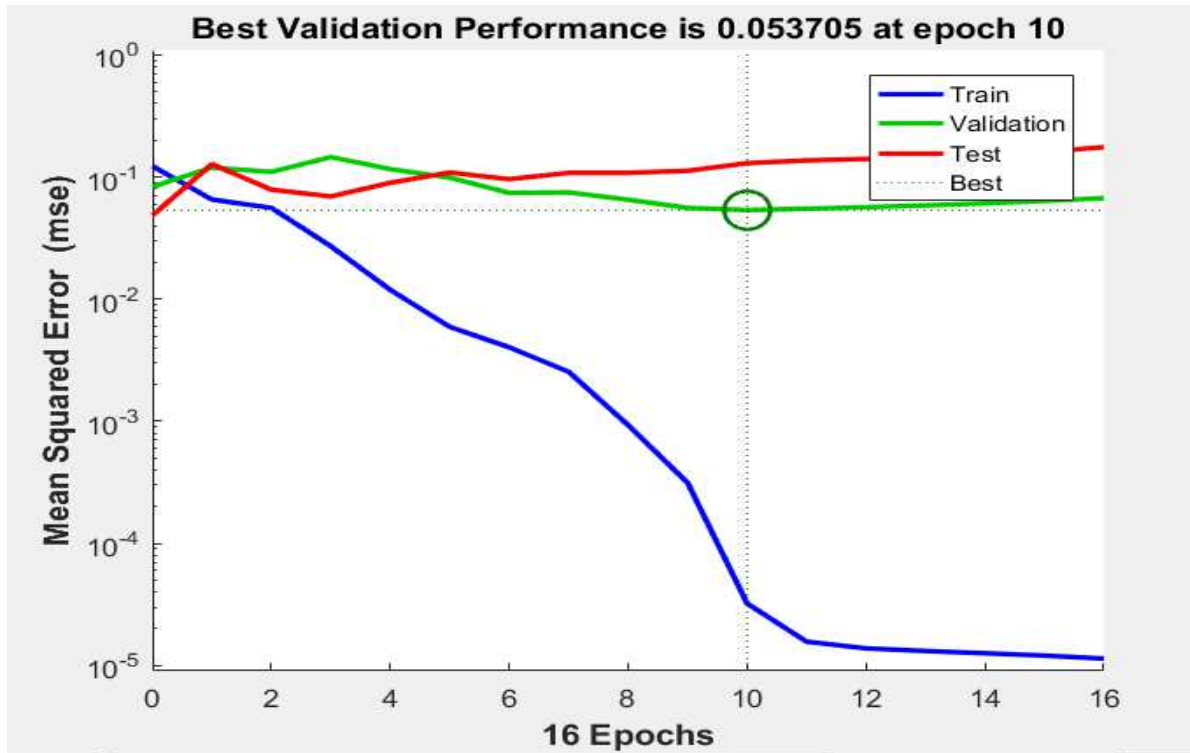


Figure 8. Performance of artificial neural network

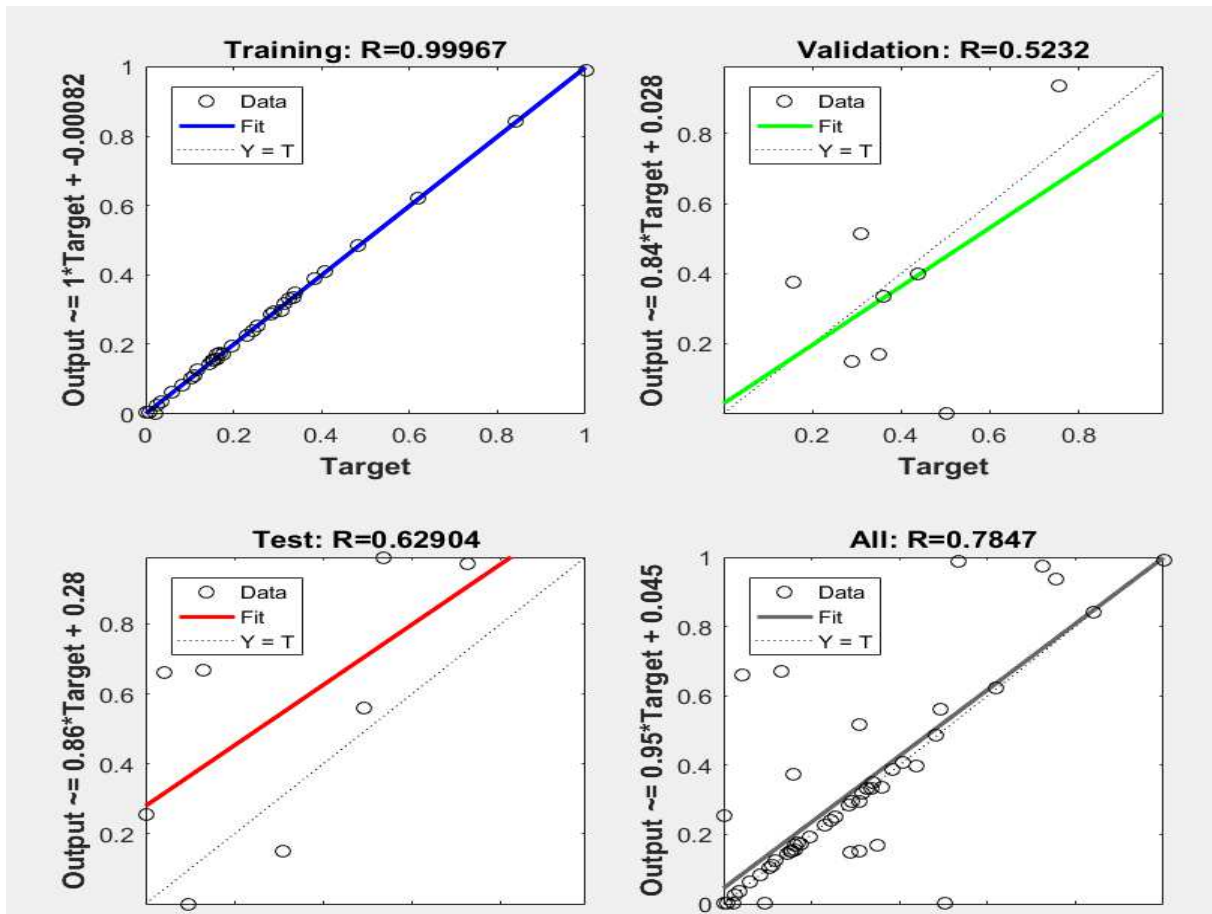


Figure 9. Regression graphics related to training, validation and test sets

The MARS algorithm was used to estimate stem weight (SW) in the Basbag alfalfa variety. The main functions, the coefficients, and the significance value of

the MARS algorithm are presented in Supplementary Material (Table S4). The variables contributing to SW are listed in order of importance for the Basbag variety (Figure 10).

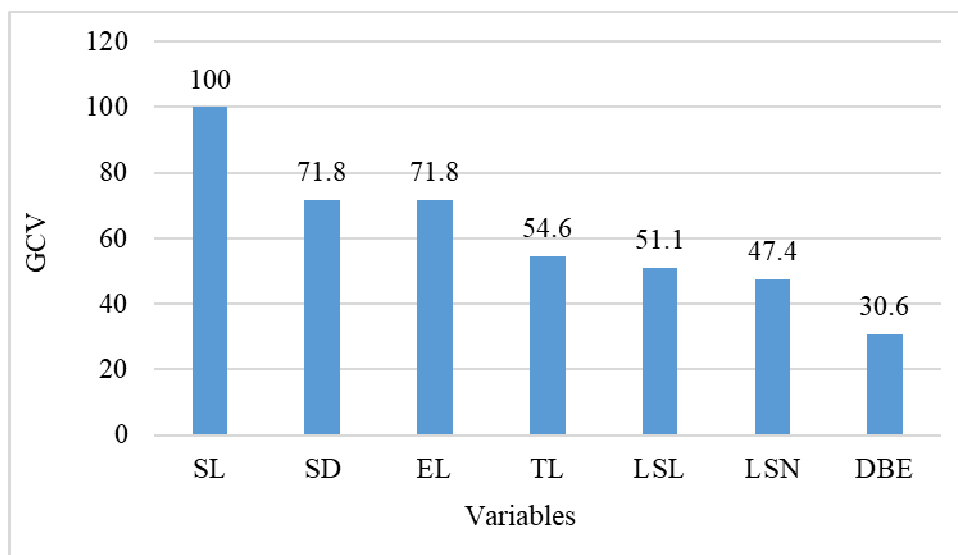


Figure 10. Most important variables for Basbag variety of alfalfa

To determine the suitability of the model, $r = 0.998$, $RMSE = 0.131$, $MSE = 0.017$, $R^2 = 0.997$, $MAPE=2.44$ and $Sd\ ratio=0.055$ have been extracted. According to these results, the MARS algorithm produced a very good estimate. It can be understood that the model fits very well.

To obtain a good fit or very good fit in the model, the standard deviation ratio of the constructed regression model should be in the range of 0.10-0.40 (Grzesiak and Zaborski, 2012; Eyduran *et al.*, 2019).

Regression analysis results plant properties of Gea, Bilensoy and Basbag alfalfa varieties that affect plant stem weight (SW) was presented in Table 8.

Table 8. Regression analysis results of plant properties in different alfalfa varieties

Gea				
Variables	β	$s(\beta)$	t	p
(Constant)	-5.923	2.450	-2.418	0.020
SL^{***}	0.183	0.027	6.820	0.001
SD	0.200	0.464	0.431	0.669
LSN	-0.020	0.180	-0.112	0.911
LSL	-0.094	0.081	-1.156	0.254
TL	0.008	0.010	0.819	0.417
DBE	0.188	0.104	1.817	0.076
EL	0.170	0.146	1.164	0.251

$r=0.753$, $RMSE=1.581$, $MSE=2.498$, $R^2=0.567$, $MAPE=20.91$, $Sdratio=0.573$, $***: (P<0.001)$.

Bilensoy				
Variables	B	$s(\beta)$	t	p
(Constant)	-3.950	1.912	-2.066	0.045
SL^{***}	0.130	0.026	4.961	0.001
SD	-0.303	0.392	-0.773	0.444
LSN	0.124	0.127	0.974	0.336
LSL	0.051	0.133	0.385	0.702
TL	0.005	0.008	0.627	0.534
DBE	0.048	0.095	0.504	0.617
EL	0.197	0.291	0.678	0.501

$r=0.711$, $RMSE=1.187$, $MSE=1.409$, $R^2=0.505$, $MAPE=31.40$, $Sdratio=0.704$, $***: (P<0.001)$.

Başbağ				
Variables	β	$s(\beta)$	t	p
(Constant)	-1.915	3.359	-0.570	0.572
SL^{***}	0.112	0.029	3.860	0.000
SD	-0.051	0.511	-0.100	0.921
LSN	0.249	0.235	1.062	0.294
LSL	-0.033	0.186	-0.175	0.862
TL	0.006	0.016	0.396	0.694
DBE	0.001	0.171	0.008	0.994
EL	-1.006	1.034	-0.973	0.336

$r=0.561$, $RMSE=2.154$, $MSE= 4.641$, $R^2=0.315$, $MAPE=35.17$, $Sdratio=0.828$, $***: (P<0.001)$. β : Regression coefficients, $s(\beta)$: Standard error.

SW: Stem weight, SL: Stem length, SD: Stem diameter, LSN: Lateral stem number, LSL: Lateral stem length, TL: Total leaf, DBE: Distance between the ears, EL: Ear length.

As a result of multivariate regression analysis in Gea, Bilensoy and Basbag varieties, it was observed that only the SL (stem length) variable significantly affected the plant stem weight (SW) and the effects of other variables were insignificant. In regression analysis applied in Gea, Bilensoy and Basbag alfalfa varieties,

determination coefficients (R^2) were found as 0.567, 0.505 and 0.315, respectively. On the other hand, mean error squares (MSE) were 2.498, 1.409 and 4.641, respectively. Therefore, models that were created with the MARS algorithm were more suitable and preferred.

DISCUSSION

In recent years, various environmental and hydrological research studies have been carried out using the MARS method (Yang *et al.*, 2003). Simulation of pesticide transport tendencies in soils (Leathwick *et al.*, 2005) and estimation of fish distribution are among these studies that used the MARS method (Bera *et al.*, 2006). In another study, the MARS model and ANN algorithms were compared, and the MARS model gave better results (Zhang and Goh 2016). In this study, when the ANN, MLR and MARS algorithm of the alfalfa plant were examined comparatively in estimating the stem weight of the alfalfa plant by using plant characteristics, the stem weight was predicted better when considering the MARS algorithm goodness criteria. It is compatible with this study in terms of model performance of the statistical methods used. In another study, the MARS algorithm gave better results than ANN method in runoff forecasting in micro basins (Adamowski *et al.* 2012). Consistent with these studies, in our study, the MARS method yielded better results, as well.

Benabderrahim *et al.* (2009) investigated the diversity of different lucerne (*Medicago Sativa L.*) populations in South Tunisia. The authors found the alfalfa stem length in the 49.4-83.6 range and stem diameter in the 0.15-0.40 range in different populations. While the stem length was similar to the results in this study, the stem diameter differed from the results obtained in this study. This difference may be due to environment, genotype, and growing conditions.

The average plant height decreased from 11.08 to 10.66 cm, respectively, in control and treated plants in Tunisia. In contrast, the average number of leaves showed no significant increase (from 6.16 to 6.47). Regarding genotype by salinity interaction, there were no significant effects on the plant development and growth traits according to ANOVA two-way analysis (Benabderrahim *et al.*, 2020).

The average number of leaves in alfalfa grown in the Erzurum region, which was 692.6 in normal soil, was determined as 54 in saline-alkaline soils. The clover plant height, which was 42.4 cm in normal soil, was 24.3 cm in saline-alkaline soil. While the number of leaves was higher than the results obtained in this study, the plant height was slightly lower. The number of leaves and plant height differed depending on the soil type grown, the environment, and genotype differences (Tan *et al.*, 2002).

Since the alfalfa genotypes and growing conditions in this study were different, the results differed from the results obtained in other studies.

Conclusion: As a result of the application of ANNs and regression analysis and MARS algorithm that used for stem weight prediction research in alfalfa, it was seen that

MARS algorithm predicted variables more precise and with fewer errors. Given the performance criteria of MARS, ANN, and MLR methods, the highest correlation coefficient and the lowest MSE values were obtained in the MARS algorithm. Therefore, it can be suggested that the MARS algorithm provides more reliable and consistent results than the ANN method and MLR. Future studies measuring various properties of plants with the MARS algorithm are expected to prove the contributions of this algorithm to the field of agriculture.

REFERENCES

- Adamowski, J., H.F. Chan, S. O., Prasher and V.N. Sharda (2012). Comparison of multivariate adaptive regression splines with coupled wavelet transform artificial neural networks for runoff forecasting in Himalayan micro-watersheds with limited data. *J. Hydroinformatics*, 14(3):731-745. <https://doi.org/10.2166/hydro.2011.044>
- Altınok, S. and A. Karakaya (2002). Forage Yield of Different Alfalfa Cultivars under Ankara Conditions. *Turk J Agric.* 26:11-16. <https://journals.tubitak.gov.tr/agriculture/vol26/iss1/2>
- Benabderrahim, M.A., H. Mansour and F. Ali (2009). Diversity of lucerne (*Medicago Sativa L.*) populations in South Tunisia. *Pakistan J. Botany*, 41(6): 2851-2861.
- Benabderrahim, M.A., M. Guiza and M. Haddad (2020). Genetic diversity of salt tolerance in tetraploid alfalfa (*Medicago sativa L.*). *Acta Physiologiae Plantarum*, 42:5 <https://doi.org/10.1007/s11738-019-2993-8>.
- Bera, P., S.O. Prasher, R.M. Patel, A. Madani, R. Lacroix, J. Gaynor, C.S. Tan and S.H. Kim (2006). Application of MARS in simulating pesticide concentrations in soil. *American Society of Agricultural and Biological Engineers*, 49(1):297-307. DOI: 10.13031/2013.20228
- Bolton, J.L. (1962). Alfalfa botany, cultivation and utilization. In: *World Crops Books*. (Editor: N., Polunin). Leonard Hill Ltd, London; Interscience Publishers, New York, 474.
- Davis, P.H. (1978). *Flora of Turkey*. Edinburgh at the University Press 22, 6:527, Edinburgh.
- Erisen, S. (2005). Yonca (*Medicago sativa L.*)'da somatik embriyogenesis aracılığıyla bitki rejenerasyonu. *Ankara Üniversitesi Tarım Bilimleri Dergisi* 11(3):311-315. doi: 10.1501/Tarimbil_0000000570
- Eyduran, E., M. Akin and S.P. Eyduran (2019). Application of multivariate adaptive regression splines through R Software. Nobel Academic Publishing, Ankara.

- FAOSTAT, (2013). Food and Agriculture Organization of the United Nations, Faostat.fao.org, Accessed to: 15.09.2013.
- Friedman, J.H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, 19:79-141.
- Grzesiak, W. and D. Zaborski (2012). Examples of the use of data mining methods in animal breeding. In: *Data mining applications in engineering and medicine* (ed. A. Karahoca). InTech, Rijeka, Croatia. pp. 303-324. doi: 10.5772/50893
- Hill, J. R., J.S. Shenk and R.F. Barnes (1988). Breeding for yield and quality. In Hanson, A.A., Barnes, D.K. and Hill, R. R. (eds). *Alfalfa and Alfalfa Improvement*. Agronomy Monograph 29, ASA, Madison, USA, pp. 809-825.
- Higgins, A., D. Prestwidge, D. Stirling and J. Yost (2010). Forecasting maturity of green peas: An application of neural networks. *Computers and Electronics in Agriculture* 70(1):151-156. doi:10.1016/j.compag.2009.09.018
- Jiang, D., X. Yang, N. Clinton and N. Wang (2004). An artificial neural network model for estimating crop yields using remotely sensed information. *International J. Remote Sensing*, 25(9):1723-1732. <https://doi.org/10.1080/0143116031000150068>
- Leathwick, J.R, D. Rowe, J. Richardson, J. Elith and T. Hastie (2005). Using multivariate adaptive regression splines to predict the distributions of New Zealand's freshwater diadromous fish. *Fresh Biology*, 50(12):2034-2052. doi:10.1111/j.1365-2427.2005.01448.x
- Lesins, K.A. and C.B. Gillies (1972). *Taxonomy and Cytogenetics of Medicago in Alfalfa Science and Technology*, (Editor: C.H. Hanson). American society of agronomy. Madison, 53-86. <https://doi.org/10.2134/agronmonogr15.c3>
- Matthew, C., A. Hernández-Garay and J. Hodgson (1996). Making sense of the link between tiller density and pasture production. *Proc. New Zeal. Grassland Assoc.* 57: 83-87. DOI: <https://doi.org/10.33584/jnzg.1995.57.2190>
- MathWorks (2016). Math Works Announces Release 2016 of the MATLAB and Simulink Product Families, USA.
- Nowruzi, H. and H. Ghassemi (2016). Using artificial neural network to predict velocity of sound in liquid water as a function of ambient temperature, electrical and magnetic fields, *J. Ocean Engineering and Science* 1(3):203-211. <http://dx.doi.org/10.1016/j.joes.2016.07.001>.
- Piepho, H.P., J. Möhring, A.E. Melchinger and A. Büchse (2008). BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica* 161: 209–228. doi: 10.1007/s10681-007-9449-8
- R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, R version: 3.6.2.Vienna, Austria. URL: <https://www.R-project.org/>.
- Rad, M. R. N. (2018). Artificial Neural Networks and its Role in Plant Breeding Under Drought Stress. *Current Investigations in Agriculture and Current Research* 1(2):33-34. DOI: 10.32474/CIACR.2018.01.000106
- Radovic, J., D. Sokolovic and J. Markovic (2009). Alfalfa-most important perennial forage legume in animal husbandry. *Biot. in Anim. Husb.* 25(5-6):465-475. <https://doi.org/10.2298/BAH0906465R>
- Razavi, M.A., A. Mortazavi and M. Mousavi (2003). Dynamic modelling of milk ultrafiltration by artificial neural network. *J. Membrane Sci.*, 220: 47–58. doi:10.1016/S0376-7388(03)00211-4
- Saglamtimur, T., V. Tansı and H. Baytekin (1990). *Yem Bitkileri Yetiştirme. Çukurova Üniversitesi Ziraat Fakültesi Ders Kitabı*, No:74. p. 238.
- Solaimany-Aminabad, M., A. Maleki and M. Hadi (2013). Application of artificial neural network (ANN) for the prediction of water treatment plant influent characteristics. *J. Advances in Environmental Health Research* 1(2):89-100. doi: 10.22102/JAEHR.2013.40130
- Spiegel, M.R., S. Lipschutz, J.J. Schiller and D. Spellman (2009). *Complex Variables with an introduction to Conformal Mapping and its applications*. Schaum's Outline Series, McGraw-Hill Companies, Inc., New York.
- Şengül, S. and S. Sağsöz (2004). Evaluation of some biometric parameters of dry matter and seed yield components in alfalfa ecotypes (*Medicago sativa* L.) as criterion for selection. *Atatürk University J. Agricultural Faculty*, 35(1-2):5-10.
- Tan, M., A. Koç and H. İ. Erkovan (2002). The Determination of Suitable Forage Species In Saline Alkaline Soil of Dumlu (Erzurum). *Atatürk Üniv. Ziraat Fak. Derg.*, 33(3): 277-281.
- Uno, Y., S. Prasher, R. Lacroix, P.K. Goel, Y. Karimi, A. Viau and R.M. Patel (2005). Artificial neural networks to predict corn yield from compact airborne spectrographic imager data. *Computers and Electronics in Agriculture*, 47(2):149-161. doi:10.1016/j.compag.2004.11.014
- Yang, C.C., S.O. Prasher, R. Lacroix and S.H. Kim (2003). A multivariate adaptive regression splines model for simulation of pesticide transport in soils. *Biosystems Engineering* 86(1):9-15. doi:10.1016/S1537-5110(03)00099-0
- Voltaire, F. (2008). Plant traits and functional types to characterize drought survival of pluri specific perennial herbaceous swards in Mediterranean

- areas. *Eur. J. Agron.*, 29:116–124. doi:10.1016/j.eja.2008.04.008
- Zaborski, D., M. Ali, E. Eyduran, W. Grzesiak, M. M. Tariq, F. Abbas, A. Waheed and C. Tirink (2019). Prediction of Selected Reproductive Traits of Indigenous Harnai Sheep under the Farm Management System via various Data Mining Algorithms. *Pakistan J. Zool.*, 51(2):421-431. doi: <http://dx.doi.org/10.17582/journal.pjz/2019.51.2.421.431>
- Zhang, H., T. Song, K.L. Wang, G.X. Wang, H. Hu and F.P. Zeng (2012). Prediction of crude protein content in rice grain with canopy spectral reflectance. *Plant Soil and Environment*, 58(11):514-520. doi:10.17221/526/2012-PSE
- Zhang, W. and A.T.C. Goh (2016). Multivariate adaptive regression splines and neural network models for prediction of pile drivability. *Geoscience Frontiers* 7:45-52. doi:10.1016/J.GSF.2014.10.003

SUPPLEMENTARY MATERIAL

Table S2. The main functions, coefficients and significance value of the MARS algorithm for stem weight prediction in Gea variety of alfalfa

Coefficients:	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-6.500e-01	4.649e-01	-1.398	0.187410
bx[, -1]h(SL-47)	-2.810e-01	3.827e-02	-7.343	8.94e-06 ***
bx[, -1]h(47-SL)	-5.487e-02	3.972e-02	-1.382	0.192303
bx[, -1]h(DBE-10.8)	2.216e+01	4.575e+00	4.844	0.000403 ***
bx[, -1]h(10.8-DBE)	-1.216e+00	1.217e-01	-9.990	3.62e-07 ***
bx[, -1]SD*h(DBE-10.8)	-6.657e+00	2.196e+00	-3.031	0.010444 *
bx[, -1]h(SL-47)*TL	1.541e-02	1.087e-03	14.181	7.38e-09 ***
bx[, -1]h(SD-2.47)	2.181e+01	2.905e+00	7.508	7.16e-06 ***
bx[, -1]h(2.47-SD)	5.676e-01	2.574e-01	2.205	0.047685 *
bx[, -1]h(TL-66)	-5.939e-01	5.562e-02	-10.677	1.76e-07 ***
bx[, -1]h(66-TL)	-1.125e-01	2.583e-02	-4.356	0.000934 ***
bx[, -1]h(LSL-9.2)	5.170e+00	5.250e-01	9.848	4.22e-07 ***
bx[, -1]h(9.2-LSL)	3.506e+00	4.018e-01	8.726	1.53e-06 ***
bx[, -1]h(LSL-7.9)	-8.931e-01	5.513e-01	-1.620	0.131168
bx[, -1]h(LSL-7)	3.384e+00	2.788e-01	12.137	4.26e-08 ***
bx[, -1]h(9.2-LSL)*EL	-7.533e-01	2.028e-01	-3.714	0.002959 **
bx[, -1]SD*h(66-TL)	1.799e-01	1.639e-02	10.975	1.30e-07 ***
bx[, -1]h(LSN-7)	1.675e+00	1.657e-01	10.105	3.20e-07 ***
bx[, -1]h(7-LSN)	7.696e-01	1.194e-01	6.444	3.19e-05 ***
bx[, -1]SD*h(9.2-LSL)	-6.966e-01	9.092e-02	-7.662	5.83e-06 ***
bx[, -1]SD*LSL*h(DBE-10.8)	-2.645e-01	3.804e-02	-6.952	1.53e-05 ***
bx[, -1]SD*LSN*h(DBE-10.8)	2.110e+00	3.063e-01	6.890	1.68e-05 ***
bx[, -1]LSN*h(DBE-10.8)	-4.894e+00	6.316e-01	-7.749	5.21e-06 ***
bx[, -1]h(LSL-8.6)	-7.279e+00	9.696e-01	-7.507	7.17e-06 ***
bx[, -1]h(9.2-LSL)*DBE	-1.072e-01	2.382e-02	-4.499	0.000728 ***
bx[, -1]h(TL-56)	3.353e-01	3.425e-02	9.789	4.51e-07 ***
bx[, -1]h(LSN-8)	-1.586e+00	2.065e-01	-7.678	5.71e-06 ***
bx[, -1]h(TL-71)	2.185e-01	3.465e-02	6.305	3.92e-05 ***
bx[, -1]h(9.2-LSL)*TL*EL	-8.755e-03	1.481e-03	-5.911	7.13e-05 ***
bx[, -1]SL*h(9.2-LSL)*EL	1.861e-02	2.786e-03	6.681	2.26e-05 ***
bx[, -1]SD*h(66-TL)*DBE	-1.390e-02	1.567e-03	-8.871	1.29e-06 ***
bx[, -1]h(SD-2.47)*DBE	-2.154e+00	2.750e-01	-7.834	4.65e-06 ***
bx[, -1]h(SL-47)*TL*DBE	-7.766e-04	9.739e-05	-7.974	3.89e-06 ***
bx[, -1]h(7-LSN)*EL	3.463e-01	7.515e-02	4.607	0.000603 ***
bx[, -1]h(EL-0.7)	1.226e-01	9.836e-02	1.246	0.236445
bx[, -1]h(0.7-EL)	3.888e+00	5.952e-01	6.532	2.80e-05 ***
bx[, -1]h(47-SL)*EL	2.172e-01	4.124e-02	5.266	0.000199 ***
bx[, -1]SD*h(66-TL)*EL	2.095e-02	5.280e-03	3.967	0.001869 **

Table S3. Coefficient of MARS algorithm for SW estimation in Bilensoy variety

Coefficients:	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.317391	0.413050	3.189	0.005368 **
bx[, -1]h(SL-43)	0.261255	0.014653	17.830	1.94e-12 ***
bx[, -1]h(43-SL)	0.571235	0.056547	10.102	1.33e-08 ***
bx[, -1]h(SD-1.4)	-2.110090	3.796550	-0.556	0.585595
bx[, -1]h(1.4-SD)	-17.529695	1.166877	-15.023	3.02e-11 ***
bx[, -1]h(1.4-SD)*LSL	2.581890	0.179217	14.407	5.87e-11 ***
bx[, -1]h(43-SL)*LSL	-0.106135	0.011020	-9.631	2.68e-08 ***
bx[, -1]h(LSN-6)	-0.699508	0.341979	-2.045	0.056605 .
bx[, -1]h(6-LSN)	3.686296	0.345631	10.665	5.97e-09 ***
bx[, -1]h(LSN-5)	2.315298	0.215123	10.763	5.22e-09 ***
bx[, -1]h(DBE-6.5)	0.629765	0.046598	13.515	1.60e-10 ***
bx[, -1]h(6.5-DBE)	1.410073	0.230389	6.120	1.13e-05 ***
bx[, -1]SD*h(6-LSN)	1.222755	0.383739	3.186	0.005403 **
bx[, -1]h(LSN-6)*LSL	-0.337181	0.037335	-9.031	6.75e-08 ***
bx[, -1]h(EL-1)	0.856890	0.277010	3.093	0.006598 **
bx[, -1]h(1-EL)	-10.252076	6.658700	-1.540	0.142052
bx[, -1]h(SD-1.4)*LSL	1.484999	0.190183	7.808	5.07e-07 ***
bx[, -1]h(1.4-SD)*TL	0.056612	0.009416	6.012	1.40e-05 ***
bx[, -1]h(TL-58)	-0.019795	0.005463	-3.623	0.002100 **
bx[, -1]h(58-TL)	-0.030445	0.005262	-5.786	2.19e-05 ***
bx[, -1]h(LSL-6.4)	0.901915	0.318917	2.828	0.011599 *
bx[, -1]h(6.4-LSL)	0.003498	0.066213	0.053	0.958480
bx[, -1]LSL*h(1-EL)	-1.984923	0.385269	-5.152	7.98e-05 ***
bx[, -1]SL*SD*h(6-LSN)	-0.067086	0.010565	-6.350	7.26e-06 ***
bx[, -1]h(SD-1.4)*LSL*TL	-0.006008	0.001430	-4.203	0.000598 ***
bx[, -1]TL*h(EL-1)	-0.031651	0.005623	-5.629	3.00e-05 ***
bx[, -1]LSN*h(6.5-DBE)	-0.173744	0.032279	-5.383	4.96e-05 ***
bx[, -1]h(SD-1.03)	-6.900332	1.371863	-5.030	0.000103 ***
bx[, -1]SL*h(SD-1.4)	0.136241	0.051420	2.650	0.016856 *
bx[, -1]h(LSL-5.7)	-1.207905	0.271128	-4.455	0.000348 ***
bx[, -1]h(SD-1.5)	-7.099597	2.392875	-2.967	0.008642 **
bx[, -1]SL*h(1-EL)	0.398469	0.164212	2.427	0.026653 *
bx[, -1]h(SD-1.25)	5.546908	2.304949	2.407	0.027758 *

Table S4. Coefficient of MARS algorithm for SW estimation in Başbag alfalfa variety

Coefficients:	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	12.084861	0.514202	23.502	7.85e-14 ***
bx[, -1]h(SL-60)	1.447345	0.079939	18.106	4.41e-12 ***
bx[, -1]h(60-SL)	0.078911	0.011959	6.599	6.12e-06 ***
bx[, -1]h(SD-2.6)	-4.317909	1.260283	-3.426	0.003464 **
bx[, -1]h(2.6-SD)	-3.192811	1.156994	-2.760	0.013957 *
bx[, -1]h(SD-2.6)*EL	-12.319141	1.276258	-9.653	4.48e-08 ***
bx[, -1]SL*h(SD-2.6)*EL	0.158609	0.018001	8.811	1.55e-07 ***
bx[, -1]h(EL-1.3)	41.447121	2.350644	17.632	6.61e-12 ***
bx[, -1]h(1.3-EL)	-2.292951	0.499038	-4.595	0.000299 ***
bx[, -1]h(EL-1)	3.518132	1.251984	2.810	0.012578 *
bx[, -1]h(TL-48)	0.445234	0.033284	13.377	4.20e-10 ***
bx[, -1]h(48-TL)	-0.558391	0.058473	-9.550	5.20e-08 ***
bx[, -1]h(LSL-5.8)	-3.074617	1.541815	-1.994	0.063465 .
bx[, -1]h(5.8-LSL)	0.236663	0.113340	2.088	0.053136 .
bx[, -1]h(SL-60)*SD	-0.478032	0.050833	-9.404	6.41e-08 ***
bx[, -1]h(LSN-6)	-1.278961	0.152880	-8.366	3.09e-07 ***
bx[, -1]h(6-LSN)	2.094692	0.250263	8.370	3.07e-07 ***
bx[, -1]h(TL-36)	-0.475593	0.032398	-14.679	1.06e-10 ***
bx[, -1]SL*h(2.6-SD)	-0.020325	0.020116	-1.010	0.327342
bx[, -1]h(LSL-6.5)	9.854424	0.666089	14.794	9.40e-11 ***
bx[, -1]h(EL-1.2)	-37.984168	2.815740	-13.490	3.71e-10 ***
bx[, -1]h(DBE-8.9)	0.962807	0.255196	3.773	0.001666 **
bx[, -1]h(8.9-DBE)	-0.527028	0.066141	-7.968	5.84e-07 ***
bx[, -1]h(SL-60)*SD*DBE	0.004375	0.004816	0.908	0.377109
bx[, -1]h(TL-48)*EL	0.006243	0.012531	0.498	0.625108
bx[, -1]h(LSL-5.3)	2.807360	0.568444	4.939	0.000148 ***
bx[, -1]LSL*h(48-TL)	0.064796	0.007461	8.684	1.88e-07 ***
bx[, -1]LSL*h(TL-48)	0.017535	0.002176	8.058	5.05e-07 ***
bx[, -1]h(SD-2.3)	-3.790567	1.747819	-2.169	0.045518 *
bx[, -1]h(LSN-7)	1.439002	0.224429	6.412	8.60e-06 ***
bx[, -1]h(LSL-6)	-10.051413	1.664044	-6.040	1.72e-05 ***
bx[, -1]SL*h(DBE-8.9)	-0.025363	0.005134	-4.940	0.000148 ***
bx[, -1]h(SD-2.4)	8.810575	2.352757	3.745	0.001767 **
bx[, -1]h(6-LSN)*TL	-0.015532	0.004916	-3.159	0.006074 **