Acta Physiologiae Plantarum (2023) 45:112
https://doi.org/10.1007/s11738-023-03587-8

ORIGINAL ARTICLE

Practicability of MARS and bagging MARS algorithms in prediction of plant length of grass pea (*Lathyrus sativus* L.) in Turkey

Muhammed İkbal Çatal[1] · Senol Celik[2] · Adil Bakoglu[3]

Received: 6 August 2021 / Revised: 1 June 2023 / Accepted: 11 August 2023
© The Author(s) under exclusive licence to Franciszek Górski Institute of Plant Physiology, Polish Academy of Sciences, Kraków 2023

Abstract

In this study, it is aimed to apply to estimate the plant height of grass pea (*Lathyrus sativus* L.) plant, data mining methods MARS (Multivariate Adaptive Regression Splines) and Bagging MARS (Bootstrap Aggregating Multivariate Adaptive Regression Splines) algorithms. Plant height, dry leaf, dry stalk, wet leaf and wet stalk were considered in different lines as plant characteristics. Plants consist of Leo, Ela, 504, Line-17, 481, 563, 528, Coloratus, Karadag, İptas, Elazig Pop. and Gurbuz lines. In the experiment carried out in 12 lines and 3 replication plots, 36 plants were examined. Belonging to the MARS algorithm that predicts plant height, r (correlation coefficient), R^2 (determination coefficient), Adjust R^2, Standard deviation ratio (SD_{ratio}), Root-mean-square error (RMSE), Relative root mean square error (RRMSE), Performance index (PI), Mean error (ME), Relative approximation error (RAE), Mean absolute percentage error (MAPE) and Mean absolute deviation (MAD) values, were found 0.867, 0.752, 0.701, 0.498, 2.459, 5.937, 3.180, 0, 0.058, 4.733 and 1.941, respectively. The same statistics of the Bagging MARS algorithm were obtained as 0.901, 0.811, 0.811, 0.435, 2.148, 5.212, 2.878, − 0.056, 0.050, 4.101 and 1.735, respectively. To compute the prediction value, the samples were split in two sets, a training one and a test one (70% and 30% of the total samples). An iterative process (1000 iterations) was run with each successive loop designing the matrices randomly. In MARS and Bagging MARS algorithms, no overfitting problem was observed under a set of independent variables consisting of plant height and plant morphological characteristics. It was concluded that both algorithms are a good statistical tool in revealing the farmers' trends in crop production at the observed location and produce important clues in increasing plant height. However, in the light of the findings of this study, the results of the Bagging MARS algorithm can be evaluated primarily.

Keywords Data mining · MARS model · Plant length · Line

Introduction

Grass pea (*Lathyrus sativus L.*, *Fabaceae*), which has the potential to be an alternative legume in many crop growing systems around the world, is primarily used as an animal feed and human consumption (Kumar et al. 2013). The plant is one of the important leguminous plants in different countries, such as India, China, Iran, Pakistan and Turkey (Barpete 2015).

Grass pea (*Lathyrus sativus* L.) is a legume plant widely grown in various parts of the world, especially in food-deficient countries, because it has a very cheap source of dietary protein (Kenicer et al. 2005). With its N2 fixation properties and growing global demand for food and feed, grass pea is considered a forage crop. (Tamburino et al. 2012; Makoi and Ndakidemi 2011).

Lathyrus species are generally valued in human nutrition (as legumes and vegetables for food), animal nutrition (as greengrass, straw and grain fodder), and soil structure improvement (as green manure crops). After *Lathyrus* species are grazed with animals for a certain period of time, seeds can be obtained from growing plants when grazing

Communicated by B. Zheng.

✉ Senol Celik
senolcelik@bingol.edu.tr

[1] Department of Field Crops, Faculty of Agriculture, University of Recep Tayyip Erdogan, Rize, Turkey

[2] Department of Animal Science, Faculty of Agriculture, University of Bingol, Bingol, Turkey

[3] Department of Plant and Animal Production, Vocational School of Pazar, University of Recep Tayyip Erdogan, Rize, Turkey

Published online: 31 August 2023

is interrupted. *Lathyrus* species are grown in small quantities in Turkey and are generally used in animal nutrition and very few are evaluated in human nutrition (Genç and Şahin 2001; Catal and Bakoglu 2018).

Grass pea has gained the potential to be an alternative legume in many crop growing systems around the world (Kumar et al. 2013). Because, grass pea (*Lathyrus sativus, Fabaceae*) has the ability to survive against extreme climatic conditions, such as drought, cold, water stagnation and heat stress (Vaz Patto et al. 2006). Therefore, it has been a highly tolerant plant against stress caused by various abiotic factors (Kumar et al. 2011). In addition, *L. sativus* grows easily and without the need for fertilization by pesticide application in many different soil and climatic conditions. It has an important role because of its high yield and its application in alternation and sustainable agriculture (Basaran et al. 2007).

The MARS algorithm is a non-parametric regression technique applied to detect the linear, nonlinear and interactive effects of the influential factors affecting the response variable, without requiring the assumption of distribution on all variables used (Akin et al. 2020). Bagging (Bootstrap aggregating) MARS algorithm, on the other hand, bootstraps among resampling techniques (Kunn and Johnson 2013).

MARS and Bagging MARS algorithms, which are among the data mining methods, have become very popular and have recently been applied in agriculture. These methods have some disadvantages such as being "susceptible to overfitting" and "more difficult to understand and interpret than other methods". However, these methods have significant advantages, such as "work well with a large number of predictive variables", "automatically detect interactions between variables", and "robust against outliers".

The feasibility of the MARS and Bagging MARS data mining algorithms, which is a nonparametric regression method that discloses the high dimensional relationship between sets of dependent and explanatory variables without require distributional and functional assumptions of the variables, has not yet been adopted in prediction of plant length of grass pea.

Plant height is one of the important parameters in the correlation of other parameters for high yield in grass pea plant. In the study of Gül et al. (2004), biological yield, plant height and number of seeds per pod made the positive and significant effects on grain yield for grass pea plant. In another study, a positive and significant correlation was determined between plant length and green grass yield, and green grass yields of tall plant species were also high (Özyiğit and Bilgen 2003).

The aim of this study is to estimate plant height based on quantitative characteristics of pea genotypes using MARS and Bagging MARS techniques.

Materials and methods

Materials

In this study, the seeds of 12 grasses pea genotypes (i.e., Leo, Ela, 504, Line-17, 481, 563, 528, Colorado, 60 Karadag, İptas, Elazığ Pop, and Gürbüz) were collected from different regions of Turkey.

This study was conducted in Recep Tayyip Erdoğan University Faculty of Agriculture Research and Application field, in 2019. The long-years of the province of Rize has been determined as average temperature 14.3 °C, the precipitation amount 2296 mm and the relative humidity 80% (Anonim 2019). Soil in a depth of 20 cm was sampled before the start of the experiment. Soil was very lowed in organic matter (0.91%), nötr (the hydrogen ion (H^+) and hydroxyl ion (OH^-) concentrations in the soil solution are in equilibrium) (pH: 6.69), in calcium carbonate (1.158%) and very low in P (10.6 kg ha-1 P_2O_5) and medium amount of in K_2O (264.4 kg ha-1) contest.

Seeds belonging to each genotype were sown in March with 3 replications according to the randomized complete block design (RCBD), with 5 m length plot, 30 cm row spacing and 5 cm intra-row. Observations were taken weekly for 5 weeks, 1 month after the seeds emerged. Measurements were made by collecting three plants from each plot. The characters of plant height (cm), fresh leaf weight (g), fresh stalk weight (g), dry leaf weight (g) and dry stalk weight (g) were investigated in the observation.

Multivariate adaptive regression spline (MARS)

Multivariate Adaptive Regression Spline (MARS) algorithm as a new method that automates the development of accurate predictive models for continuous and discrete response variables was introduced. The MARS method was introduced as a new method that automates the development of accurate prediction models for binary response variables (Friedman 1991). In the MARS algorithm, both continuous and categorical estimators can be used in practice and give successful results. However, the basic MARS algorithm predicts that the variables are continuous, and it achieves more successful results in ordinal variables than variables that can take only two values (Akyol 2011). MARS is one of the methods for modelling high-dimensional regression data. MARS is a form of extension of the Basis Spline Functions, where the number of basis function is the parameters of the model (Frank 1995). The current MARS predictive model with interaction term was conducted based on the lowest GCV (Akin et al. 2020).

A tenfold cross validation was took into accounted as a resampling technique in the MARS modeling.

MARS algorithm is formulated as follows:

$$f_M(x) = \beta_0 + \sum_{m=1}^{M} \beta_m B_m(x)$$

where β_0 and β_m are the basis function parameters of the MARS algorithm used on the basis of the least squares criterion. The spline basis function, $B_m(x)$, can be implemented as

$$B_m(x) = \prod_{k=1}^{k_m} \left[s_{\mathrm{km}} \left(x_{v(k,m)} - t_{k,m} \right) \right]$$

where km is the number of knots, s_{km} takes either 1 or − 1 and presents the right/left regions of the related step function, v(k,m) is the label of the input variable and $t_{\mathrm{k,m}}$ is the knot location (Friedman 1991).

The Generalized Cross Validation error is a measure of the goodness of fit that takes into account both the residual error and the model complexity as well. It is presented by (Kornacki and Ćwik 2005).

$$\mathrm{GCV}(\lambda) = \frac{\sum_{i=1}^{n} \left(Y_i - \widehat{Y}_i \right)^2}{\left[1 - \frac{M(\lambda)}{n} \right]^2}$$

where n is the number of sample in the study, Y_i is the observed value of the response variable, $\widehat{Y}_i$ is the predicted value of the response variable, $M(\lambda)$ is the penalty function for the complexity of the model containing λ terms. MARS predictive model with interaction term was constructed based on the lowest GCV (Akin et al. 2020).

Bagging MARS algorithm

Bagging MARS algorithm is formulated as follows (Breiman 1996):

$$\widehat{f}_{\mathrm{BaggingMARS}} = \frac{1}{M} \sum_{m=1}^{M} f_M(x)$$

here, $f_M(x)$: MARS model.

Bagging predictor is a method to generate multiple versions of predictors and use them for aggregate predictors. Multiple versions are formed by a bootstrap replication of a data set. In some cases, bagging on real data sets can improve accuracy. If the changes in the data sets lead to significant changes then bagging can evolve (Hasyim et al. 2018).

Determination of the amount of replication B is very varied, since the size of B can give different results at each phase of analysis. It commits 25 or fifty times replications (Sutton 2005), nevertheless, (Hastie et al. 2001) suggests that increased accuracy will be formed if the number of replications is increased from fifty to one hundred times, whereas if the number is increased to more than a hundred times it will produce no greater accuracy of the replication accuracy of a hundred times.

R software was used for analyses, taking cross validation as 10 (R Core Team 2020). Results were obtained by MARS and Bagging MARS algorithms "earth" packages. Data mining algorithms' model evaluation performance criteria were evaluated using the "ehaGoF" package (Eyduran 2020). The multicollinearity problem between the independent variables was performed at the outset of the analysis, and it was detected that there was none.

Model evaluation criteria is presented in Table 1 (Willmott and Matsuura 2005; Liddle 2007; Takma et al. 2012; Grzesiak and Zaborski 2012; Eyduran et al. 2019). The definitions of RRMSE and ρ are based on Gandomi and Roke (2013).

$\mathrm{SD}_{\mathrm{ratio}}$ for ensuring a good fit should be less than 0.40 and, for a very good fit SD ratio should be less than 0.10 (Grzesiak and Zaborski 2012). The closer the R^2 value is to 1 (generally 0.80 and above), the higher the efficiency of the model will be (Mendeş 2013).

Results

The descriptive statistics of plant characteristics in the grass pea plant are given in Table 2.

Correlation coefficients between plant characteristics in grass pea are given in Table 3.

When Table 3 is examined, the correlation coefficients between all plant characteristics are statistically significant ($p < 0.01$ and $p < 0.001$). The correlation coefficients between the variables ranged from 0.502 to 0.971. The highest coefficient correlation was between DL–WL ($r = 0.971$), followed by DH–WH ($r = 0.952$), and WH–WL ($r = 0.755$). Briefly, high and significant correlations were found between plant characteristics. Model performance was determined by goodness-of-fit tests for MARS and Bagging MARS algorithms applied to predict plant height (Table 4).

The desirable predictive attribute of the MARS equation applied here was founded with providing the smallest GCV = 12.50. The observed values in grass pea (*Lathyrus sativus L.)* were established strongly with those predicted by the MARS algorithm as a plant length estimation modeling. For prediction equation of MARS model with 6 terms, was found R^2 estimate (0.752). The Adj. R^2 of 0.701, standard deviation ratio of 0.498, RMSE of 2.459 and MAPE of 4.733 represented that the MARS model capturing two influential factors in other words live weight in grass pea had a good

Table 1 Model performance criteria

Goodness-of-fit statistics	Formula
Coefficient of determination (R^2)	$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}$
Adjusted coefficient of determination ($\overline{R}^2$)	$\text{Adj.}R^2 = 1 - \frac{\frac{1}{n-k-1}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{\frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}$
Standard deviation ratio	$SD_{\text{ratio}} = \sqrt{\frac{\frac{1}{n-1}\sum_{i=1}^{n}(\varepsilon_i - \overline{\varepsilon})^2}{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2}}$
Root-mean-square error (RMSE)	$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}$
Relative root mean squared error (RRMSE)	$\text{RRMSE} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}}{\frac{1}{n}\sum_{i=1}^{n}\hat{Y}_i^2} \times 100$
Mean absolute percentage error (MAPE)	$\text{MAPE} = \frac{1}{n}\sum_{1}^{n}\left\lvert\frac{Y_i - Y_{ip}}{Y_i}\right\rvert.100$
Mean absolute deviation (MAD)	$\text{MAD} = \frac{1}{n}\sum_{i=1}^{n}\left\lvert Y_i - Y_{\text{ip}}\right\rvert$
Performance index (ρ)	$\rho = \frac{\text{RRMSE}}{1+r}$

fit. In the MARS algorithm, the Pearson correlation coefficient between real and predicted values was 0.867 ($p < 0.01$). The coefficients in the prediction equation were founded significant ($p < 0.05$, $p < 0.001$). The prediction equation for MARS algorithm was found as and parameter coefficients are displayed in Table 5.

The equation of the MARS model is as follows:

$$\begin{aligned}\text{PL} = {} & 43.3 + 4.26 \times \text{Lines Karadag} - 7.1 \times \max(0, 2.62 - \text{WH}) \\ & - 0.116 \times \max(0,\ \text{WH} - 2.62) + 2.51 \times \text{DL} \\ & \times \text{Lines Elazig Pop} + 6.51 \times \max(0,\ \text{WH} - 2.62) \\ & \times \text{Lines L563}\end{aligned}$$

The estimation equation discovered by MARS can also be written as follows, considering the basis functions in Table 5.

$$\begin{aligned}\text{PL} = {} & 43.3 + 4.26 \times \text{B}_4(\text{x}) - 7.1 \times \text{B}_2(\text{x}) \\ & - 0.116 \times \text{B}_1(\text{x}) + 2.51 \times \text{B}_5(\text{x}) + 6.51 \times \text{B}_3(\text{x}).\end{aligned}$$

As a result of the MARS model, there will be a decrease in plant length when WH ≤ 2.62 cm.

When L563 line and WH > 2.62 cm, plant length will increase 6.51 cm with each increase in WH value. "Karadag" line plants are 4.26 times more tall. In "Elazig Pop" line plants, the plant length will increase by 2.51 cm with each increase of DL. Briefly, WH and DL variables from plant characteristics and L563, Karadag and Elazig Pop plant lines are important factors in plant length growth.

Characteristics that most affect plant height in grass pea plants, as shown in Table 6, are WH-wet stalk (100%) lines 563 (57.1%) and DL-dry leaf (45%) lines Karadag (37.1%) and lines Elazig Pop (29.6%).

The importance of the independent variables in Table 6 is shown in MARS and Bagging MARS algorithms in Fig. 1 show the change of PL according to WH, Lines 563 and DL. The most important variable always has a value of 100 (Gocheva-Ilieva and Iliev 2010). To a degree this distribution matches the relationship between factors, which is WH: Lines 563: DL: Lines Karadag: Lines Elazig Pop = 100:57.1:45:37.1:29.6 (for MARS model). Likewise, for the Bagging MARS model to a degree this distribution matches the relationship between factors, which is WH: Lines 563: DL: Lines Karadag: Lines Elazig Pop = 100:59.3:44:38.7:26.4, respectively.

Various values were given to the plant characteristics and lines, i.e., independent variables, to estimate the plant height (PL). For example:

PL = 42.978 kg for WL = 10, WH = 5, DL = 2, DH = 0.90– Lines = "Coloratus";

PL = 47.996 for WL = 10, WH = 5, DL = 2, DH = 0.90– Lines = "ElazigPop";

PL = 47.235 for WL = 10, WH = 5, DL = 2, DH = 0.90–Lines = "Karadag".

The Bagging MARS algorithm has also been applied to estimate plant height. The prediction equation of the Bagging MARS algorithm as below:

Table 2 Descriptive statistics of the studied variables

Lines	*n*	PL			WL		
		$\overline{X} \mp s_{\overline{x}}$	Min	Max	$\overline{X} \mp s_{\overline{x}}$	Min	Max
Leo	3	42.05 ± 2.29	37.50	44.83	16.82 ± 3.84	9.15	20.99
Ela	3	41.01 ± 2.51	36.00	43.83	6.85 ± 1.34	4.18	8.42
504	3	39.97 ± 2.11	36.20	43.50	8.22 ± 1.64	5.00	10.34
Line-17	3	42.03 ± 0.50	41.30	43.00	7.70 ± 1.07	5.87	9.59
481	3	41.87 ± 2.78	36.50	45.80	10.34 ± 0.88	8.87	11.90
563	3	39.3 ± 7.25	29.00	53.30	8.64 ± 2.67	3.77	12.97
528	3	43.1 ± 2.86	39.30	48.70	8.61 ± 0.50	7.62	9.13
Coloratus	3	40.78 ± 2.74	36.00	45.50	10.73 ± 0.91	9.20	12.34
Karadağ	3	45.28 ± 3.22	38.83	48.50	10.57 ± 2.40	6.56	14.87
İptaş	3	43.53 ± 2.04	40.30	47.30	8.96 ± 1.13	6.79	10.59
Elazığ Prop	3	48.2 ± 2.31	44.60	52.50	10.73 ± 3.45	4.56	16.48
Gürbüz	3	41.23 ± 1.16	39.20	43.20	8.92 ± 0.49	7.94	9.51
General	36	42.36 ± 0.83	29.00	53.30	9.76 ± 0.63	3.77	20.99
		WH			DL		
Leo	3	5.64 ± 1.59	2.46	7.23	2.93 ± 0.65	1.65	3.72
Ela	3	3.04 ± 1.28	1.20	5.51	1.44 ± 0.28	0.88	1.76
504	3	2.77 ± 0.85	1.48	4.36	1.70 ± 0.32	1.07	2.04
Line-17	3	3.04 ± 0.82	2.22	4.67	1.71 ± 0.23	1.29	2.08
481	3	3.95 ± 1.30	1.70	6.22	2.02 ± 0.12	1.83	2.23
563	3	2.31 ± 0.99	0.83	4.19	1.89 ± 0.53	0.90	2.69
528	3	2.59 ± 0.09	2.42	2.73	1.73 ± 0.10	1.57	1.91
Coloratus	3	4.46 ± 1.33	2.88	7.10	1.91 ± 0.19	1.60	2.27
Karadağ	3	4.05 ± 1.40	1.76	6.59	2.04 ± 0.40	1.44	2.80
İptaş	3	3.28 ± 0.58	2.35	4.35	1.72 ± 0.25	1.26	2.12
Elazığ Prop	3	5.38 ± 1.63	2.61	8.25	2.26 ± 0.69	1.04	3.43
Gürbüz	3	2.35 ± 0.24	1.87	2.61	1.95 ± 0.12	1.71	2.08
General	36	3.57 ± 0.32	0.83	8.25	1.94 ± 0.11	0.88	3.72
			DH				
Leo		3	0.94 ± 0.25		0.44		1.25
Ela		3	0.70 ± 0.30		0.29		1.28
504		3	0.61 ± 0.16		0.34		0.90
Line-17		3	0.71 ± 0.19		0.50		1.08
481		3	0.75 ± 0.20		0.43		1.13
563		3	0.53 ± 0.19		0.21		0.86
528		3	0.53 ± 0.03		0.48		0.59
Coloratus		3	0.76 ± 0.21		0.43		1.16
Karadağ		3	0.76 ± 0.20		0.42		1.12
İptaş		3	0.61 ± 0.13		0.40		0.85
Elazığ Prop		3	1.20 ± 0.35		0.62		1.84
Gürbüz		3	0.56 ± 0.03		0.51		0.60
General		36	0.72 ± 0.06		0.21		1.84

$\overline{X}$ mean, *s*: std. deviation, $s_{\overline{x}}$ std. error. *PL* plant length, *WL* Wet leaf, *WH* wet stalk, DL: dry leaf, *DH* Dry stalk

$$
\begin{aligned}
PL = {} & (45.04219 + 4.047408 \times LinesColoratus - 5.156891 \times LinesElazigPop - 1.990009 \times Linesiptas \\
& + 10.88588 \times LinesKaradag + 6.179116 \times LinesL481 + 3.82726 \times LinesLine17 + 1.561443 \times h(WL - 9.15) \\
& - 12.18191 \times h(WL - 10.24) - 7.89232 \times h(3.4 - WH) - 2.248094 \times h(WH - 3.4) + 18.69571 \times h(1.87 - DL) \\
& + 53.65689 \times h(DL - 1.87) - 41.92593 \times h(0.54 - DH) - 56.97137 \times h(DH - 0.54) + 72.3931 \times h(DH - 0.7) \\
& + 41.41957 - 8.813093 \times LinesColoratus - 3.493608 \times LinesL563 - 15.62442 \times LinesLeo + 4.487601 \times LinesLine17 \\
& - 1.194401 \times h(7.64 - WL) + 9.12968 \times h(WL - 7.64) - 5.225173 \times h(WL - 9.2) - 1.820791 \times h(2.46 - WH) \\
& - 4.293672 \times h(WH - 2.46) - 37.44489 \times h(DL - 1.6) + 25.2655 \times h(DL - 1.99) + 20.59166 \times h(DH - 0.5) \\
& + 47.60403 - 9.83256 \times LinesColoratus - 2.24397 \times LinesL481 + 6.528843 \times LinesL563 - 13.19993 \times LinesLeo \\
& + 8.591584 \times h(WL - 7.96) - 22.97611 \times h(WL - 9.15) + 16.93404 \times h(WL - 9.51) + 16.93404 \times h(WL - 9.51) \\
& - 47.12881 \times h(DL - 1.67) + 38.07226 \times h(DL - 1.79) - 30.12386 \times h(0.7 - DH) - 5.583796 \times h(DH - 0.7) \\
& + 43.97622 - 6.461374 \times LinesLeo - 8.838332 \times h(2.62 - WH) + 4.845721 \times h(DL - 1.99) + 44.12306 \\
& - 7.061523 \times LinesColoratus - 11.72334 \times LinesKaradag - 4.042661 \times LinesL563 - 36.45193 \times LinesLeo \\
& - 6.216233 \times h(2.61 - WH) - 8.846126 \times h(DL - 1.75) + 31.7359 \times h(DL - 2))/5
\end{aligned}
$$

According to this equation, in the first bootstrap an increase in plant height grass pea is expected for those with Lines Coloratus, Lines Karadag, Lines 481 and Lines Line-17 with WL > 9.15, 1.87 ≤ DL, DL > 1.87 and DH > 0.7. The effects of these increases on the model were positive and were 4.047, 10.886, 6.179, 3.827, 1.561, 18.696, 53.657 and 72.393, respectively. In the second bootstrap, plants Line-17 with WL > 7.64, DL > 1.99 and DH > 0.5 has been positive effects on plant height. These effects are 4.487, 9.130, 25.266 and 20.592, respectively. Other effects were negative. In the third bootstrap, Lines 563 with WL > 7.96, WL > 9.51 and DL > 1.79 are predicted to increase in plant height. The effects of these increases were 6.529, 8.592, 16.934 and 38.072, respectively. In other cases, a decrease in plant height is expected. In the fourth bootstrap, only plants with DL > 1.99 will increase plant height by 4.846 units. In the fifth bootstrap, only grass pea with DL > 2 will have an increase of 31.754 units in plant height.

Table 3 Correlation matrix of plant features

	PL	WL	WH	DL	DH
PL					
WL	0.517**				
WH	0.502**	0.755***			
DL	0.540**	0.971***	0.706***		
DH	0.514**	0.649***	0.952***	0.657***	

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

Table 4 Model performance for MARS and bagging MARS algorithms

	MARS	Bagging MARS
Pearson's correlation coefficients (r)	0.867	0.901
Coefficient of determination (R^2)	0.752	0.811
Adjusted coefficient of determination (Adj. R^2)	0.701	0.811
Standard deviation ratio (SD ratio)	0.498	0.435
Root mean square error (RMSE)	2.459	2.148
Relative root mean square error (RRMSE)	5.937	5.212
Performance index (PI)	3.180	2.878
Mean error (ME)	0	− 0.056
Relative approximation error (RAE)	0.058	0.050
Mean absolute percentage error (MAPE)	4.733	4.101
Mean absolute deviation (MAD)	1.941	1.735

As in the MARS model, in the Bagging MARS model, various values are given to the plant characteristics and lines for the estimation of the grass pea plant height (PL). For example,

PL = 42.978 kg for WL = 10, WH = 5, DL = 2, DH = 0.90 and Lines = "Coloratus";

PL = 43.036 for WL = 9.5, WH = 4.5, DL = 1.85, DH = 0.95 and Lines = "İptas";

PL = 47.678 for WL = 9.5, WH = 4.5, DL = 1.85, DH = 0.95 and Lines = "ElazigPop";

PL = 55.284 for WL = 9.5, WH = 4.5, DL = 1.85, DH = 0.95 and Lines = "L563".

When the previous values for plant characteristics values were not changed, different estimates were obtained for different lines. From this, it was understood that the lines in plant height are an important determining factor.

Estimation of PL values in MARS and Bagging MARS models were close to each other. Considering the goodness of fit criteria, it was seen that the Bagging MARS algorithm gave slightly better results than the MARS algorithm. The graph showing MARS, Bagging MARS and actual values (PL) together is presented in Fig. 2.

Table 5 MARS model parameter estimation

Coefficients	*B*	Estimate	Std. error	*t* value	Pr(>|t|)
(Intercept)		43.2539	0.721	59.918	<2e− 16***
bx[, − 1]h(WH-2.62)	$B_1(x)$	− 0.1158	0.3111	− 0.372	0.712290
bx[, − 1]h(2.62-WH)	$B_2(x)$	− 7.0990	1.1137	− 6.374	4.93e− 07***
bx[, − 1]h(WH-2.62) × linesl563	$B_3(x)$	6.5146	1.7550	3.712	0.000837***
bx[, − 1]LinesKaradag	$B_4(x)$	4.2571	1.6483	2.583	0.014927*
bx[− 1]DL × LinesL481	$B_5(x)$	2.5091	0.7026	3.571	0.001221**

B basis function

Table 6 Relative importance of model independent variables

Variables	Relative importance (%)	
	MARS	Bagging MARS
WH	100	100
Lines 563	57.1	59.3
DL	45	44
Line Karadag	37.1	38.7
Lines Elazig pop	29.6	26.4

Discussion

In a study examining the morphological characteristics of different cultivars of the grass pea plant, the average plant height was between 19.23 and 52.53 cm. The investigated characteristics are shown in the dendongram of plant varieties (Kosev and Vasileva 2019). In terms of plant height, it is in agreement with the values found in this study. The plant height was found to be between 56.2 and 111.2 cm in the grass pea plant with different accessions, which was found to be considerably higher than the values obtained in this study (Grela et al. 2012). In another study, plant height was

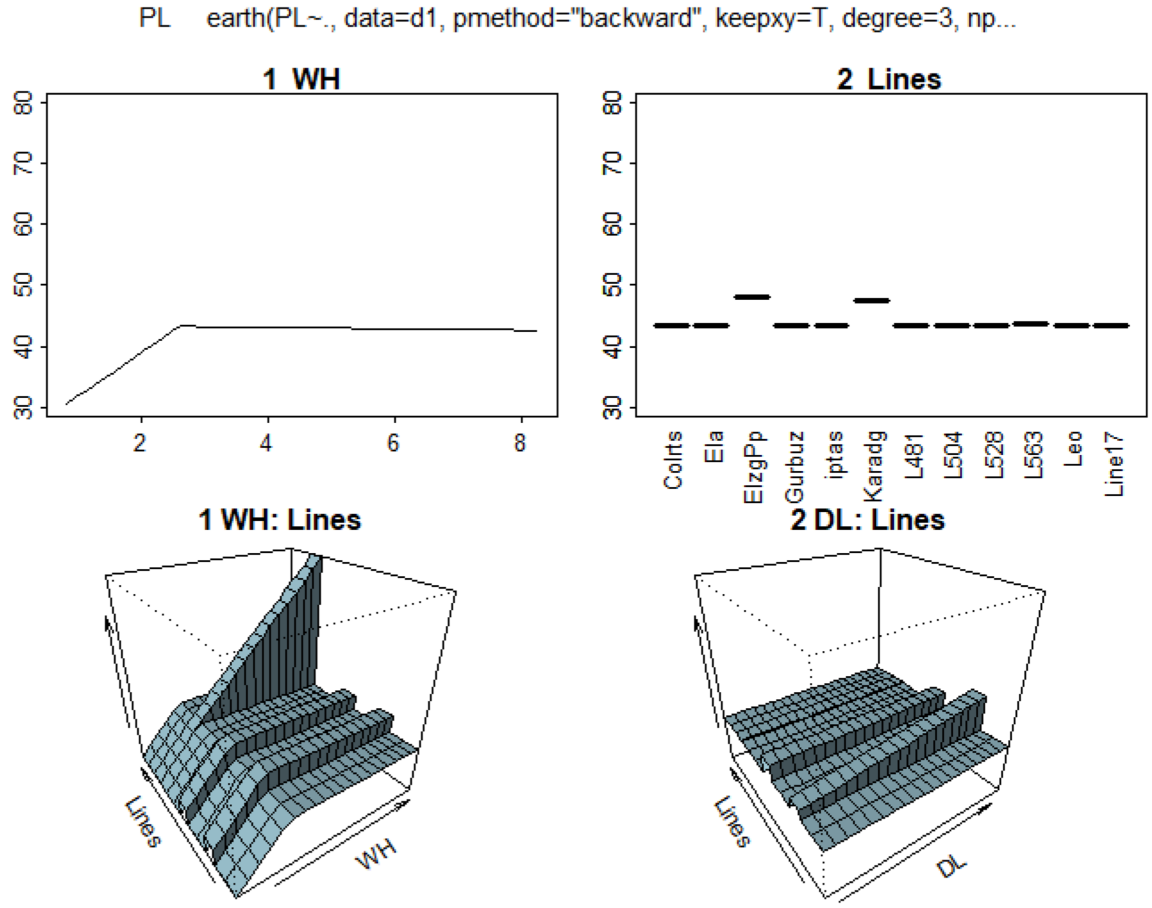


Fig. 1 Variation of PL according to WH, DL and lines

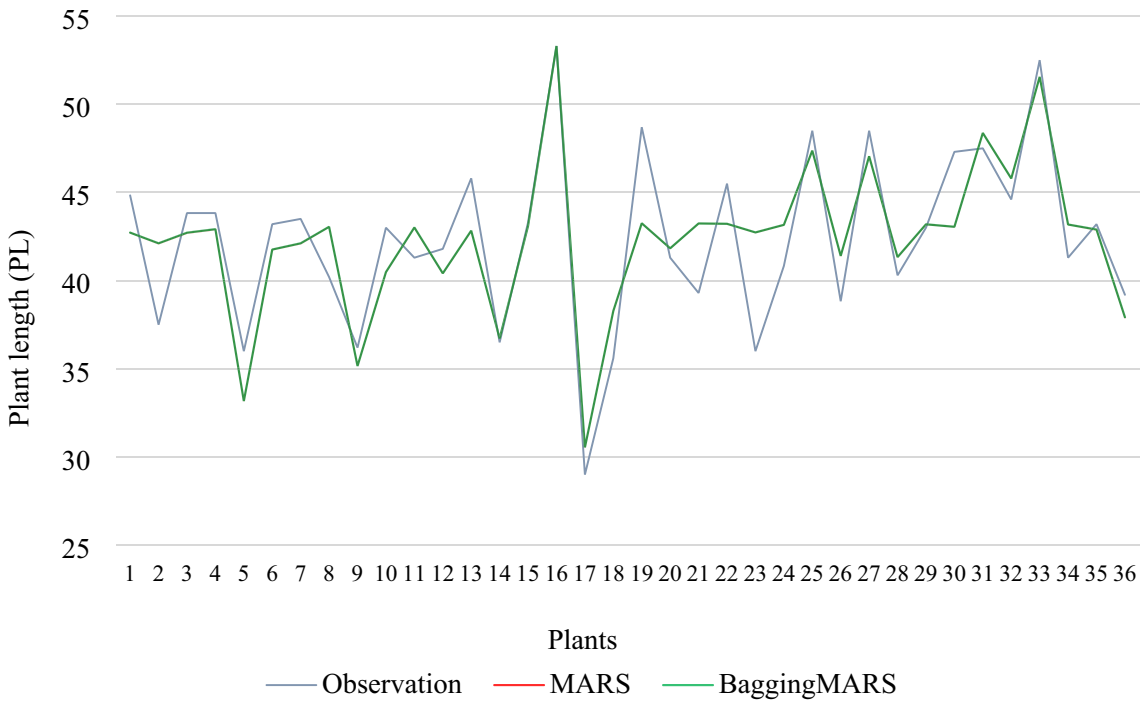


Fig. 2 Observed versus predicted PL values

measured between 32 and 87 cm in grass pea grown according to different clipping methods (Księżak and Bojarszczuk 2020). According to of grass pea grown cultivars and mutants the plant height of average was 104.2–111.5 for KRAB cultivar and 110.1–119.1 for DEREK cultivar (Rybiński et al. 2006). It is higher than the values found in this study. In a study Kökten and Bakoğlu (2011) conducted by in a randomized block trial arrangement with 3 replications in Elazig province of the Turkey, the average plant height was calculated 50.12 cm, the number of main branches 4.73 per plant, the number of side branches 11.54 per plant, and the height of the lower pod 11.81 cm, the number of pods per plant 18.13 and the number of seeds per pod as 2.06. While according to row spacing of 20, 30 and 40 cm plant height, number of main branches, lower pod height, number of pods per plant and number of seeds per pod did not show a statistically significant difference; the number of side branches showed a statistically significant difference according to row spacing. Oten et al. (2017a, b) investigated the relationship between grain and hay yield with some characteristics in grass pea by correlation and path analysis. The highest correlations with hay yield were between green forage yield (0.970), cut yield (0.904) and biological yield (0.805), respectively. In the path analysis of the direct and indirect relationships of different characters on hay yield in the grass pea, the highest direct effects were, respectively, on wet grass yield (51.32%), cut yield (28.81%) and grain yield (21.90%).

In a study conducted by Balabanlı and Kara (2003) in the province of Isparta in Turkey, the plant height variation of damson grown in different lines was investigated in a randomized block design with 3 replications. As a result of the analysis of variance, statistically significant differences were determined between the lines in the average of 2 years. The highest plant height was obtained with 61.5 cm from the line numbered 555, and the lowest plant height with 51.7 and 52.5 cm was obtained from the lines 567 and 559.

In the study of Gül et al. (2004), plant height in different daisy lines in Diyarbakır province in Turkey varied between 51.33 and 57 cm according to the combined results of 3 years. The plant height in this study was higher than the length. This difference is due to different variety, line and environment.

Morphological observation data of some grass pea lines collected from the natural flora of Antalya in Turkey were evaluated. As a result of the dendogram examination made according to the UPMGA (Unweighted Pair-Group Method with Arithmetic Mean) method; the studied lines showed similarity between 86% and 99% and they were divided into 3 main groups. Lines 1, 8, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 27, 29, 30 and 37 used in the study were in group 1, line 7 in group 3 and the control type used other lines were included in the 2nd group (Oten et al. 2017a, b).

When literature studies were searched with MARS and Bagging MARS algorithms, no study was found on pea plants. However, some studies have been reached on

livestock and other plants in the field of agriculture. In one of the a for mentioned studies, the factors affecting the honey yield per hive of beekeeping enterprises in Turkey's Elazig province were examined and Multivariate Adaptive Regression Splines (MARS) and Bagging MARS algorithms were applied, which showed excellent performance as an appropriate algorithm that did not show overfitting problems. Bagging MARS algorithm successfully received the interaction of educational level, working status in other actions except for beekeeping, with number of full beehives compared to MARS algorithm (Kulekçi et al. 2022). The model performance of the methods used is consistent with the results of this study.

Conclusion

The predictive performances of the models created using MARS and Bagging MARS algorithms were obtained very close to each other for estimated plant length. However, Bagging MARS algorithm has been found to be a more suitable method. Based on the empirical results using Bagging MARS approach, the best model of live weight prediction is the model with the value of R^2 is 0.811. The most significant influencing predictor of the plant length is DL.

According to the results of the MARS algorithm, Line 563, Line Karadag and Line 481 with WH > 2.62 g, respectively, can be considered as important selection criteria in breeding studies to reach higher plant height in grass pea. As a result of the Bagging MARS algorithm, the most important criteria in increasing the plant height were DH > 0.7 g, DL > 1.87 g, DL > 1.99 g, DH > 0.5 g, DL > 1.79 g, WL > 9.51 g and DL > 2 g in different bootstraps, respectively, can be considered as. Therefore, it was concluded that MARS and Bagging MARS modeling was a good statistical tool to reveal plant morphological features of grass pea in the application.

Author contribution statement MİÇ: he compiled the necessary data and obtained material. Language revision in manuscript and response letter. SC: he made statistical analyzes suitable for the study. He wrote the method for statistical analysis and wrote the statistical comments, the full findings and part of the discussion. He made the necessary revision in the article and checked the statistical analysis He wrote response letter. AB: he obtained the data for the study. He wrote the introduction part of the study. He wrote part of the discussion and checked the general. Checked the revision and made additions to the article. He checked response letter.

Data availability The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

References

Akin M, Eyduran SP, Eyduran E, Reed BM (2020) Analysis of macro nutrient related growth responses using multivariate adaptive regression splines. Pl. Cell. Tissue Organ Culture (PCTOC) 140:661–670. https://doi.org/10.1007/s11240-019-01763-8

Akyol M (2011) Yaşam Çözümlemesine Yeni Bir Yaklaşım: MARS. Ankara Üniversitesi Sağlık Bilimleri Enstitüsü, Doktora Tezi, Ankara

Anonim (2019) T.C. Başbakanlık Devlet Meteoroloji İşleri Genel Müdürlüğü, Rize İl Müdürlüğü Kayıtları.

Balabanlı C, Kara B (2003) Determination of some agronomic characteristics and yield potential common chickling (*Lathyrus Sativus* L.) Lines under Isparta conditions. J Field Crops Central Res Inst 12(1):57–63

Barpete S (2015) Genetic associations, variability and diversity in biochemical and morphological seed characters in Indian grass pea (*Lathyrus sativus L.*) accessions. Fresenius Environ Bull 24:1–6

Basaran U, Acar Z, Önal-Ascı Ö, Mut H, Ayan İ (2007) Mürdümük (*Lathyrus* sp.) Türlerinin Önemi, Tarimda Kullanim Olanaklari Ve Zararli Madde İçerikleri. J of Fac of Agric 22(1):139–148

Breiman L (1996) Bagging predictors. Mach Learn 24:123–140

Catal Mİ, Bakoglu A (2018) In vitro regeneration techniques in the grass pea (*Lathyrus sativus* L.) plant. Eur J Forest Sci 6(2):56–62

Eyduran E (2020) ehaGoF: Calculates Goodness of Fit Statistics. R package version 0.1.0. URL: https://CRAN.Rproject.org/package=ehaGoF. (Access date: Sep 30, 2022).

Eyduran E, Akin M, Eyduran SP (2019) Application of Multivariate Adaptive Regression Splines in Agricultural Sciences through R Software. Nobel Akademik Yayıncılık, p112, ISBN:978–605–2149–81–2.

Frank IE (1995) Tutorial: modern nonlinear regression methods. Chemom Intell Lab Syst 27:1–19

Friedman JH (1991) Multivariate adaptive regression spline. Annals Statist 19:1–141

Gandomi AH, Roke DA (2013) Intelligent formulation of structural engineering systems. In: Seventh MIT Conference on Computational Fluid and Solid Mechanics-Focus: Multiphysics and Multiscale, 12–14 June, Cambridge, USA

Genç H, Şahin A (2001) Batı Akdeniz ve Güney Ege Bölgesinde yetişen *Bazı Lathyrus* L. türleri üzerinde sitotaksonomik araştırmalar. III S.d.u. Fen Bilimleri Dergisi 5(1):98–112

Gocheva-Ilieva SG, Iliev IP (2010) Parametric and nonparametric empirical regression models: case study of copper bromide laser generation. Hindawi Publishing Corparat Mathemat Prob Eng. https://doi.org/10.1155/2010/697687

Grela ER, Rybin´ski W, Matras J, Sobolewska S, (2012) Variability of phenotypic and morphological characteristics of some *Lathyrus sativus L.* and *Lathyrus cicera L.* accessions and nutritional traits of their seeds. Genetic Res Crop Evolut 59:1687–1703

Grzesiak W, Zaborski D (2012) Examples of the use of data mining methods in animal breeding. (Book) ISBN 978–953–51–0720–0.

Gül İ, Sümerli M, Yılmaz Y (2004) Determination of yield and yield components of some Grasspea (*Lathyrus sativus* L.) lines under diyarbakir conditions. Tarım Bilimleri Dergisi 10(4):416–421

Hastie T, Tibshirani R, Friedman J (2001) The elements of statistical learning: data mining. Inference and Prediction Springer-Verlag, New York

Hasyim M, Rahayu DS, Muliwati NE, Hayuhantika D, Puspasari R, Anggreini D, Hastari RC, Hartanto S, Utomo H (2018) Bootstrap aggregating multivariate adaptive regression splines bagging MARS to analyse the lecturer research performance in private university wma-mathcomtech 2018. IOP Conf Series J Phys Conf Ser 1114:012117. https://doi.org/10.1088/1742-6596/1114/1/012117

Kenicer GJ, Kajita T, Pennington RT, Murata J (2005) Systematics and biogeography of Lathyrus (*Leguminosae*) based on internal transcribed spacer and cpDNA sequence data. Am J Botany 92:1199–1209

Kökten K, Bakoğlu A (2011) Elazığ Koşullarında Mürdümük (*Lathyrus sativus* L.)'te Farklı Sıra Arasının Tohum Verimi ve Verim Öğeleri Üzerine Etkisi. Bingöl Üniversitesi Fen Bilimleri Dergisi 1(1):37–42

Kosev VI, Vasileva VM (2019) Morphological characterization of Grass pea (*Lathyrus sativus* L.) varieties. J Agricult Sci Sri Lanka 14(2):67–76

Księżak J, Bojarszczuk J (2020) The effect of sowing method on the yield of grass pea (*Lathyrus Sativus*) cultivated in an organic system. Appl Ecol Environ Res 18(4):5355–5366

Kulekçi M, Eyduran E, Altın AY, Tariq MM (2022) Usefulness of MARS and bagging MARS algorithms in prediction of honey production in beekeeping enterprises from Elazig Province of Turkey. Pakistan Journal of Zoology (PJZ) 54(3):1087–1093. https://doi.org/10.17582/journal.pjz/20200309160354

Kumar S, Bejiga G, Ahmed S, Nakkoul H, Sarker A (2011) Genetic improvement of grass pea for low neurotoxin (ODAP) content. Food Chem Toxicol 49:589–600

Kumar S, Gupta P, Barpete S, Sarker A, Amri A, Mathur PN, Baum M (2013) Grass pea. In: Singh M, Upadhyaya HD and Bisht IS (eds) Genetic and Genomic Resources for Grass pea Improvement. pp. 269–293.

Kunn M Johnson K (2013). Applied predictive modeling. NY. Doi: https://doi.org/10.1007/978-1-4614-6849-3.

Liddle AR (2007) Information criteria for astrophysical model selection. Monthly Not Royal Astron Soc Letters 377:L74–L78

Makoi JHJR, Ndakidemi PA (2011) Changes in plant growth, nutrient dynamics and accumulation of flavonoids and anthocyanins by manipulating the cropping systems involving legumes and cereals- a review. Aust J Agric Eng 2(3):56–65

Mendeş, M (2013) Uygulamalı Bilimler İçin İstatistik ve Araştırma Yöntemleri. Kriter Yayınevi, İstanbul, ISBN:978–605–5863–99–9.

Öten M, Kiremitçi S, Erdurmuş C (2017a) Determination of characters associated with seed yield and hay yield by path and correlation analysis in grass pea (*Lathyrus sativus* L.). Derim 34(1):72–78

Öten M, Kiremitçi S, Erdurmuş C (2017b) The determination of yield characteristics of some grass pea (*Lathyrus sativus* L.) lines collected from antalya natural flora. Ege Üniv Ziraat Fak Derg 54(1):17–26

Özyiğit Y, Bilgen M (2003) The Effect of Different Cutting Stages on The Yield, Quality and Agricultural Characteristics of Some Forage Legume - Bee Plants. Türkiye 5. Tarla Bitkileri Kongresi 13–17 Ekim 2003, Diyarbakir

R Core Team (2020) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. URL: https://www.R-project.org/ (access date: September 30, 2022)

RybiñskiBaszczakFornal WWJ (2006) Seed microstructure and genetic variation of characters in selected grass-pea mutants (*Lathyrus sativus L.*). International Agrophysics 20:317–326

Sutton CD (2005) Classification and Regression Trees, Bagging, and Boosting Handbook of Statistics Vol. 24: 303-329

Takma C, Atil H, Aksakal V (2012) Comparison of multiple linear regression and artificial neural network models goodness of fit to lactation milk yields. Kafkas Üniversitesi Veteriner Fakültesi Dergisi 18:941–944

Tamburino R, Guida V, Pacifico S, Parente Di Maro A (2012) Nutritional values and radical scavenging capacities of grass pea (*Lathyrus sativus* L.) seeds in valle agricola district. Italy. Aust Jour of Crop Sci 6(1):149–156

VazPatto MC, Skiba B, Pang ECK, Ochatt SJ, Lambein F, Rubiales D (2006) Lathyrus improvement for resistance against biotic and abiotic stresses: from classical breeding to marker assisted selection. Euphytica 147:133–147

Willmott C, Matsuura K (2005) Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim Res 30:79–82

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.