BMC Plant Biology

RESEARCH Open Access

Integrating Taguchi design and machine learning models for trait stability and predicting forage quality in naturally occurring grass pea (*Lathyrus* spp.)

Muhammed İkbal Çatal[1], Cengiz Sancak[2*], Seyid Amjad Ali[3] and Muhammad Aasim[4*]

Abstract

This study investigated the forage quality of four naturally occurring Lathyrus species (*Lathyrus czeczottianus*, *L. pratensis*, *L. roseus*, and *L. rotundifolius subsp.* miniatus), including one endemic taxon, collected from the Rize province of Turkey. Forage traits of two years were analyzed using the Taguchi design (TD) of experiment to determine trait stability and species-based differences and robustness. Machine learning models (Random forest – RF and Light gradient boosting – LightGBM) were applied to predict forage quality parameters. Results revealed *L. pratensis* as the most robust species for critical traits such as crude ash ratio, crude protein ratio, and K/(Ca + Mg) based on signal-to-noise ratio. The ML models showed high predictive accuracy for mineral and digestibility-related traits ($R^2 > 0.97$). The RF model exhibited superior predictive capacity for fiber traits compared to the LightGBM model. The combined use of TD and ML-based prediction demonstrates a potential framework for evaluating species-level differences, supporting data-driven breeding, optimizing forage quality, and sustainable livestock production in diverse environments. While ML models demonstrated promising predictive performance, future studies incorporating larger datasets and different cross-validation techniques are needed to enhance model robustness and improve the general applicability of these findings.

Keywords Taguchi design, Machine learning, Lathyrus, Forage traits prediction, Mineral composition

Introduction

Grass pea (*Lathyrus sativus* L.) belongs to the genus Lathyrus and family Fabaceae. It is a self-pollinating annual vine having long, grass-like leaflets (Lambein). Lathyrus species are plants that grow naturally in various regions of Turkey, commonly in Eastern and Southeastern Anatolia. However, recent studies indicate that Lathyrus species exhibit a remarkable distribution and diversity in the central, inland, and coastal regions of the Black Sea [1]. It is documented that a total of 61 species of Lathyrus are found in the flora of Turkey, and approximately 20 species were identified in the vegetation of the Black Sea Region. Whereas it was documented that 8 *Lathyrus* species, with 2 endemics, are grown naturally in Rize province of Turkey [2].

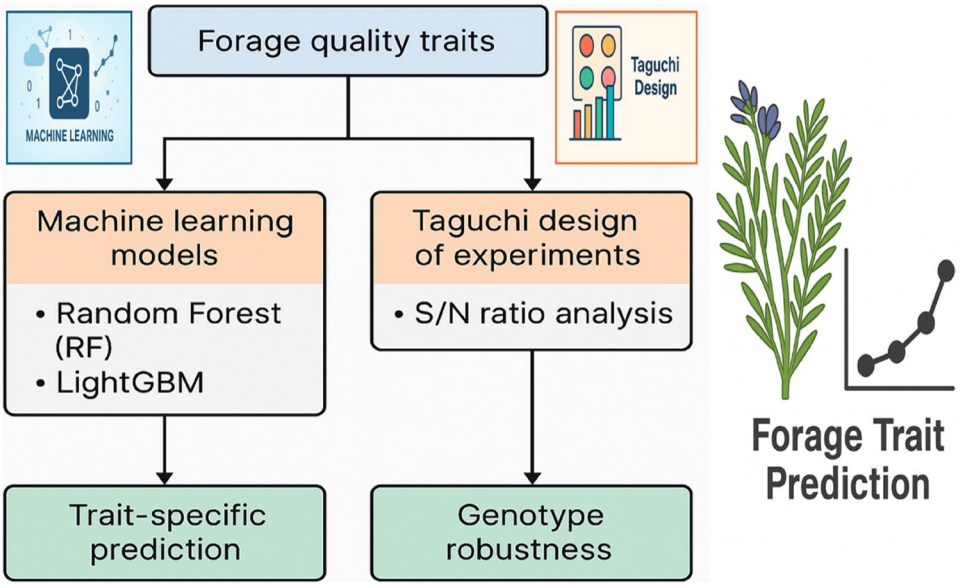
Grass pea is a traditional crop used both for animal and human consumption as food. Its drought tolerance and adaptability to a wide range of soil types, including marginal areas, make it an excellent forage legume [3], due

*Correspondence:
Cengiz Sancak
sancak@ankara.edu.tr
Muhammad Aasim
maasim@sivas.edu.tr

Full list of author information is available at the end of the article

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Graphical Abstract



to high protein content. However, the utilization of grass pea grain is limited by the presence of a water-soluble non-protein amino acid β-N-Oxalyl Diamino propionic acid (β-ODAP). The grains of the grass pea plant act as a neurotoxin that cripples the lower limbs when consumed in large quantities over a long period of time, and can cause neurolathyrism disease in monogastric animals and humans [4]. Therefore, the selection of suitable L*athyrus* species with low β-ODAP is critical for the safe use of grass pea as a sustainable and high-protein forage and food crop.

In modern livestock nutrition management, selecting a superior species based on its forage quality traits [5], and optimization is done by using classical statistical techniques. In this study, the Taguchi Design (TD) was incorporated to quantify forage traits based on signal-to-noise (S/N) ratios. The TD methodology is used for process design optimization by offering a systematic and efficient method for conducting experiments under varying operational environments [6], and the impact of variation is reduced during processing [7, 8]. It allows to analyze noise factors (variations in the input data) to enhance robustness and reliability. The signal-to-noise (S/N) ratio measures the quality in relation to target values [9] and is highly advantageous in plant sciences due to optimizing design parameters, leading to significant improvements in crop yield and resource management [10, 11].

The integration of artificial intelligence (AI), especially machine learning (ML) algorithms, is transforming the predictive modeling landscape in agriculture by extracting patterns from complex, nonlinear, and high-dimensional biological and agronomic data [12]. Among the widely adopted models, Random Forest (RF) and Light Gradient Boosting Machine (LightGBM) have shown remarkable performance for diverse agricultural applications ranging from yield forecasting to nutritional trait prediction, and stress diagnosis [13, 14, 15]. The main hypothesis of using ensemble-based ML is to predict multiple forage quality parameters more accurately by using routinely available input data. Furthermore, it is hypothesized that performance will vary by trait due to underlying biological complexity and data structure. In recent years, applications of ML models for forage and grassland crops have been reported to predict, validate, or optimize forage-related traits [16, 17]. From a practical perspective, AI-driven forage quality has the potential to support nutrient balancing and feed planning [18].

In this study, forage quality characteristics of 4 Lathyrus species, 1 of which is endemic (*Lathyrus czeczottianus Bassler-Endemic*, *Lathyrus pratensis L.*, *Lathyrus roseus Steven*, *Lathyrus rotundifolius* Willd. subsp. *miniatus* (M.Bieb. ex-Steven) P.H. Davis), growing naturally in Sıraköy locality of Çamlıhemşin district of Rize province, were determined. It was hypothesized that traits would differ among species due to biological variation. Whereas TD and ML models would capture these differences based on species and year as a source of variation.

Materials and methods

Forage quality analysis

This study was carried out in 2019 and 2020 by collecting spontaneously growing Lathyrus species in Sıraköy, Çamlıhemşin, Rize province (40° 47’ 55.66” N, 40° 55’ 59.73” E), at an average altitude of 1960 m above sea level

and 38 km away from the district of Çamlıhemşin. The climatic data of Rize province obtained from the General Directorate of Meteorology from 1964 to 2020 were obtained. Whereas soil analysis of the research area was also determined. Plant materials were collected in accordance with national regulations, and sampling was carried out in compliance with local conservation guidelines to ensure that the sustainability of natural populations was not adversely affected.

Four *Lathyrus* species, including one endemic (*Lathyrus czeczottianus-Endemic*, *Lathyrus pratensis*, *Lathyrus roseus*, *Lathyrus rotundifolius* subsp. *miniatus*), were used. All four species are grown naturally in the area of study. The formal identification of *Lathyrus* species used in this study was performed by Asst. Prof. Hüseyin BAYKAL, Recep Tayyip Erdoğan University, Rize, Turkey. Samples of each taxon collected are publicly available at the herbarium of Recep Tayyip Erdoğan University, Faculty of Agriculture. A total of 15 plants were collected in 3 replicates from the specified locations at the full maturity stage (BBCH scale, stage 79: nearly all pods have reached final length) for quality analysis. The developmental stages were identified according to the extended BBCH scale for grain legumes [19].

Crude ash content was determined by burning the hay samples at 550 °C for 4 h in a porcelain crucible in a muffle furnace. Other forage quality traits (Acid Detergent Fiber (ADF), Neutral Detergent Fiber (NDF), Acid Detergent Protein (ADP), Phosphorus (P), Potassium (K), Calcium (Ca), and Magnesium (Mg)) were determined using a Near-Infrared Spectroscopy (NIRS) system (Foss NIR Systems, Model 6500). Before analysis, all samples were standardized by drying at 60 °C and milling to a 1 mm particle size. The system was operated using WinISI II (v1.5) software with the IC-0904FE calibration package. To ensure the accuracy of the predictions, the calibration was validated against reference values obtained from laboratory wet chemistry. The reliability of the results was assessed through the Standard Error of Prediction (SEP) and the Ratio of Performance to Deviation (RPD). For the primary traits (CP, ADF, and NDF), the RPD values are > 2.0, indicating high predictive power and confidence in the reported data. Hemicellulose ratio, Digestible Dry Matter (DDM), Dry Matter Intake (DMI), Relative Feed Value (RFV), Digestible Energy (DE), and Metabolic Energy (ME) values were calculated according to ADF and NDF ratios, and Ca/P and K/(Ca + Mg) ratios were determined according to macro element contents.

Hemicellulose = NDF - ADF

Digestible Dry Matter (DDM) = 88.9 - (0.779 × ADF%) [20].

Dry Matter Intake (DMI) = 120 / (%NDF) [21].

Relative Feed Value (RFV) = (DDM × DMI) / 1.29 [21].

Digestible Energy (DE) = 0.27 + 0.0428 × (%DDM) [22].

Metabolic Energy (ME) = 0.821 × DE (Mcal/kg) [23].

Statistical analysis

The data were subjected to analysis of variance in the JMP 13 statistical program [24] for two-way ANOVA analysis. The difference between mean values of the treatments was determined by the F test, grouped at 5% ($P < 0.05$) level according to Tukey's significance test. Results were also analyzed by constructing a Pearson correlation matrix to identify the positive and negative correlations. A heatmap was used for the visual presentation of the correlation rather than inferential testing. All pairwise correlations were computed using the average trait values for each species. To ensure the robustness of the statistical analyses and machine learning models, the data were structured as follows: For each of the four *Lathyrus* species, three independent biological replicates were collected annually over two consecutive years (2019 and 2020). This resulted in a total of 24 unique observations (4 species × 3 replicates × 2 years) for each forage quality trait. Rather than averaging the yearly data, each year's results were treated as separate data points to allow the machine learning algorithms to account for interannual environmental variations. These 24 observations formed the primary dataset for Taguchi optimization and subsequent predictive modeling.

Taguchi model analysis

Taguchi experimental design is suitable for biological systems with complex interactions and variability, and is used to evaluate the effects of multiple factors on forage quality traits [7, 9]. In this study, two parameters (species and year) were used to investigate Signal-to-Noise (S/N) ratios of all characteristics tested to assess robustness and optimize performance parameters. In this study, the year was treated as a noise factor and species as the control factor. The robustness of the model was interpreted biologically as the ability of the species to maintain stable nutritional traits under varying environmental conditions. For this purpose, the Minitab 21.0 program was used, and S/N ratios of all measured forage-quality traits (CAR, DMR, CPR, Hmcl, DDM, DMI, RFV, DE, ME) were analyzed using the "Larger-the better". Fibre fractions, including ADF, NDF, and K(Ca + Mg), were evaluated using the smaller-the-better criterion because lower values improve forage quality and animal health. Mineral concentrations (P, K, Ca, Mg) and the Ca/P ratio were automated at the nominal-the-best criterion due to their optimal physiological ranges in ruminant nutrition. Because Taguchi S/N ratios are computed using logarithmic transformations, the resulting values may be either positive or negative and represent the relative signal strength compared with experimental noise rather than raw trait values.

Machine learning analysis

Two different ML models were employed to investigate the forage quality parameters of four different *Lathyrus* species. The objective of the ML analysis was to predict individual forage quality traits of naturally occurring Lathyrus species and explore the predictive capability of two different ML models for these traits [25]. For this purpose, RF and LightGBM models were selected, and model's performance was tabulated based on four different performance metrics listed in Eqs. 1–4.

Coefficient of determination (R^2)

The R^2 is a useful statistical metric used for assessing the predictive strength of regression models. A high R^2 score implies a high prediction accuracy of the model and a strong association between the independent and dependent variables.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2}{\sum_{i=1}^{n}\left(Y_i - \widetilde{Y}\right)^2} \tag{1}$$

Mean Squared Error (MSE)

Is used to measure the average squared difference between predicted values and actual values in regression models. It implies how close a regression line is to the measured data points,

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2 \tag{2}$$

Mean Absolute Error (MAE)

The MAE is a crucial metric used for evaluating predictive models by measuring the average magnitude of errors between predicted and observed values.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \widehat{Y}_i\right| \tag{3}$$

Mean Absolute Percentage Error (MAPE)

The MAPE is a valuable metric used for assessing and forecasting accuracy, expressed as a percentage.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \widehat{Y}_i}{Y_i}\right| \times 100 \tag{4}$$

Machine learning models

RF and LightGBM are tree-based ensemble methods that construct decision rules through threshold-based partitioning of the feature space and are therefore insensitive to the scale of input variables. The RF is the most extensively used advanced decision tree model that trains several trees simultaneously using bagging, or bootstrap aggregation [26]. It is generally employed for understanding complex nonlinear interactions, outperforming linear statistical models. In this model, all trained trees contribute to the outcome, fundamental principle of the entire operational system is presented in Eq. 5 [27].

$$y = \sum_{i-1}^{n}(\alpha_i - \alpha_i^*)k(x, x_i) + b \tag{5}$$

LightGBM is a Microsoft-developed model based on the gradient boosting algorithm [25], and its operational principle involves decision trees for improved forecast accuracy, making it a valuable tool in agricultural science and plant modeling. Although its application is limited to plant sciences, despite its robustness [28]. In recent years, the use of both techniques for solving problems by classification or regression models in plant and agricultural sciences has been increasing [29].

The dataset followed a 4 × 2 factorial design (species × year) with three replicates per treatment combination, yielding a total of 24 observations. Species and year were used as predictor variables, and each replicate was treated as an independent observation during model development and validation. Both predictors were handled as categorical variables and encoded numerically before model fitting. No feature scaling was applied. Multicollinearity was not considered a concern in this study because only two independent categorical predictors (species and year) were used in the models. These variables represent distinct experimental factors and do not exhibit linear dependence.

For each trait, a separate ML model was trained using the leave-one-out cross-validation (LOO-CV) method [30], which is appropriate for small datasets as it maximizes the use of available training data. In each iteration, a single observation was held out for testing, while the remaining observations were used to train the model.

Hyperparameter optimization was conducted via grid search. The open-source Python programming language and the Sklearn package were utilized for the coding process. For the Random Forest model, the search space included n_estimators (10–400), max_depth (10–30), max_features ('sqrt', 'log2'), min_samples_split (2–10), and min_samples_leaf (1–4). For LightGBM, the tuning parameters comprised num_leaves (4–32), learning_rate (0.01–0.1), min_data_in_leaf (5–50), and

colsample_bytree (0.05–1.0). The final model configuration was selected based on the minimum cross-validated mean squared error (MSE).

Results

ANOVA analysis of forage quality traits of lathyrus species

The species × year interaction was found statistically significant for most of the traits; results were interpreted in a year-specific manner rather than as main effects. The year-specific differences were presented when needed. The analysis of variance (ANOVA) results for forage quality traits exhibited statistically significant effects for all traits except DMI by species ($P<0.01$), indicating substantial interspecific variability among the species for most parameters (Table 1). Conversely, the insignificant impact of the year was recorded only for CAR and [K/(Ca + Mg)], suggesting the limited impact of environmental variation on most traits. The interaction between year and species was statistically significant for CAR, CPR, ADF, NDF, Hmcl, DDM, DE, and ME. Comparison of species revealed better performance of *L. pratensis*, and exhibited the highest values for key nutritional traits, including CAR (12.98%), CPR (19.96%), DDM (61.03), DMI (1.96), Relative Feed Value (RFV, 93.06), DE (2.88 Mcal/kg), and ME (2.36 Mcal/kg). It also exhibited the highest mineral values for phosphorus (P, 0.29%), potassium (K, 1.95%), and the K/(Ca + Mg) ratio (1.02). Whereas *L. rotundifolius* subsp. Miniatus exhibited the highest DMR (95.97%), ADF (48.08%), Ca (1.86%), Mg (0.41%), and Ca/P ratio (8.95). The highest values for ADP (1.99%) and Hmcl (27.85%) were recorded from *L. czeczottianus* species. Conversely, the highest NDF (70.09%) and the lowest CPR (13.66%), RFV (68.93), DE (2.49 Mcal/kg), and ME (2.04 Mcal/kg) were recorded for *L. roseus*. Results also showed the significant interactions' impact for CAR, CPR, ADF, NDF, Hmcl, DDM, DE, and ME, suggesting that the influence of species on these traits varied depending on the year (Table 1).

Taguchi-based response to noise (S/N) analysis

The S/N ratio values of years and species highlight the temporal stability or variation of forage quality traits. The delta (Δ) value refers to the difference between maximum and minimum values, and it reflects how consistent each parameter is under similar environmental conditions in response to time (Figs. 1a-i and 2a-i).

The species-based S/N ratio analysis provides insights into biological or varietal sensitivity. The delta of CAR

Table 1 Significance levels and forage quality values of *Lathyrus* species for the studied traits

Variation	CAR	DMR	CPR	ADF	NDF	ADP	Hmcl	DDM	DMI
Year	*	ns	ns	ns	ns	ns	ns	ns	ns
Species	**	**	**	**	**	**	**	**	ns
Year * Species	*	ns	**	**	**	ns	**	*	ns
Species	**CAR****	**DMR****	**CPR****	**ADF****	**NDF****	**ADP****	**Hmcl****	**DDM****	**DMI**
L. czeczottianus	6.73b	92.72b	15.59b	40.30b	68.15a	1.99a	27.85a	57.50a	1.76
L. pratensis	12.98a	91.55b	19.96a	35.77c	61.01c	1.21d	25.23ab	61.03a	1.96
L. roseus	3.80c	93.90ab	13.66b	47.44a	70.09a	1.40c	22.65b	51.94b	1.71
L. rotundifolius	7.32b	95.97a	10.90c	48.08a	65.02b	1.45b	16.94c	51.44b	1.85
Mean	7.71	93.54	15.03	42.90	66.07	1.51	23.17	55.48	1.82
SD	3.84	1.89	3.81	5.92	3.97	0.33	4.66	4.61	0.11
Minimum	3.80	91.55	10.90	35.77	61.01	1.21	16.94	51.44	1.71
Maximum	12.98	95.97	19.96	48.08	70.09	1.99	27.85	61.03	1.96
Variation	**RFV**	**DE**	**ME**	***P***	**K**	**Ca**	**Mg**	**Ca/*P***	**K/(Ca + Mg)**
Year	ns	ns	ns	ns	ns	ns	ns	ns	*
Species	**	**	**	**	**	**	**	**	**
Year * Species	ns	*	*	ns	ns	ns	ns	ns	*
Species	**RFV****	**DE****	**ME****	***P*****	**K****	**Ca****	**Mg****	**Ca/*P*****	**K/(Ca + Mg)****
L. czeczottianus	78.49b	2.73a	2.24a	0.20b	1.34b	1.67b	0.30b	8.47a	0.67b
L. pratensis	93.06a	2.88a	2.36a	0.29a	1.95a	1.65b	0.25c	5.74b	1.02a
L. roseus	68.93d	2.49b	2.04b	0.20b	0.45c	1.48c	0.32b	7.35ab	0.24c
L. rotundifolius	73.52c	2.47b	2.02b	0.21b	0.41c	1.86a	0.41a	8.95a	0.18d
Mean	78.50	2.64	2.17	0.23	1.04	1.67	0.32	7.63	0.53
SD	10.46	0.20	0.16	0.04	0.74	0.16	0.07	1.43	0.39
Minimum	68.93	2.47	2.02	0.20	0.41	1.48	0.25	5.74	0.18
Maximum	93.06	2.88	2.36	0.29	1.95	1.86	0.41	8.95	1.02

ns not significant, *CAR* Crude Ash Ratio, *DMR* Dry Matter Ratio, *CPR* Crude Protein Ratio, *ADF* Acid Detergent Fiber, *NDF* Neutral Detergent Fiber, *ADP* Acid Detergent Protein, *Hmcl* Hemicellulose, *DDM* Digestible Dry Matter, *DMI* Dry Matter Intake, *RFV* Relative Feed Value, *DE* Digestible Energy, *ME* Metabolic Energy

*: $P<0.05$, **: $P<0.01$

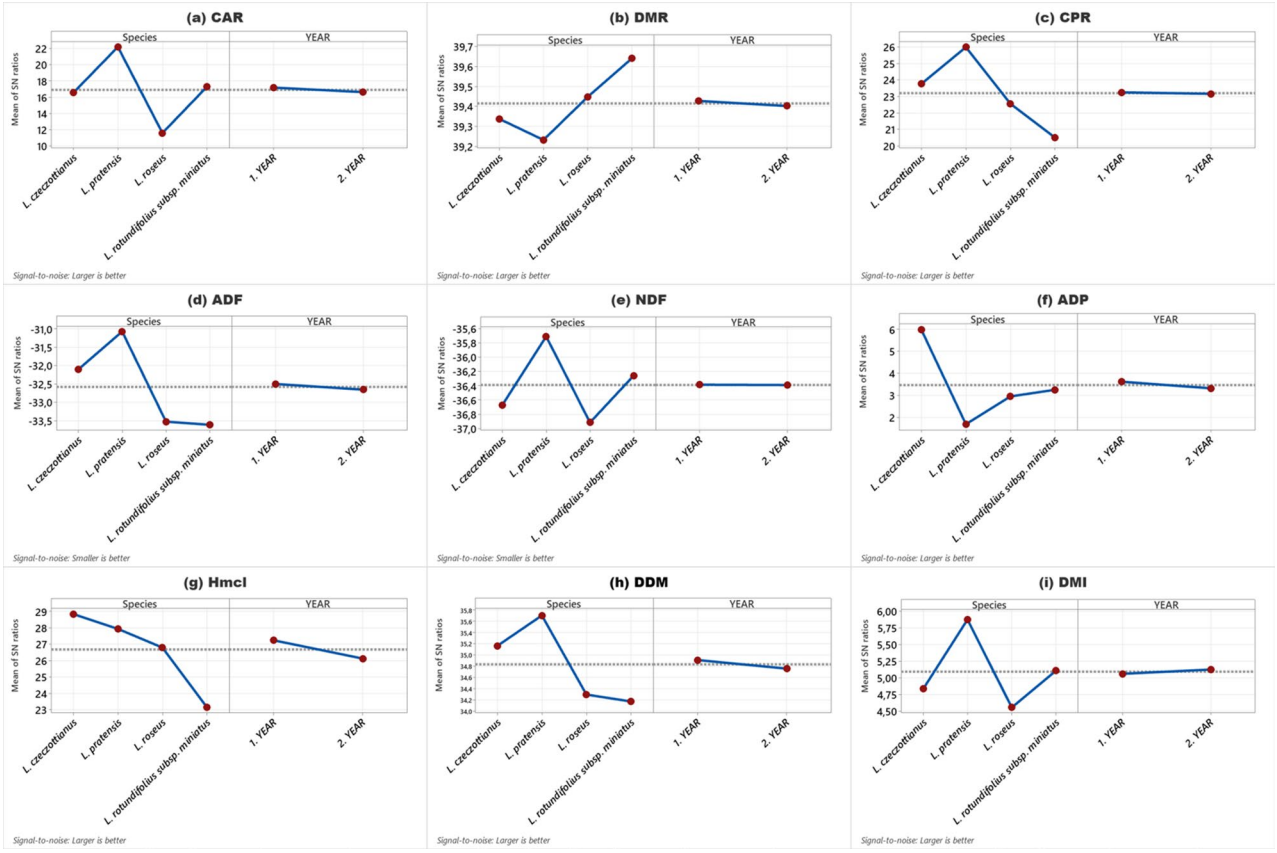


Fig. 1 Signal-to-noise ratio of forage traits of Lathyrus species (**a**) CAR, (**b**) DMR, (**c**) CPR, (**d**) ADF, (**e**) NDF, (**f**) ADP, (**g**) Hmcl, (**h**) DDM, and (**i**) DMI

(Δ = 10.62) and CPR (Δ = 5.48) showed very high variability for species, respectively. Whereas a considerable variation was recorded for Hmcl (Δ = 5.67) and DDM (Δ = 1.53), indicating the significance of species. Relatively low deltas for DMR (Δ = 0.41) and ME (Δ = 1.37) suggest relatively robust performance regardless of species. The ADF (Δ = 2.53) and NDF (Δ = 1.20) varied moderately for species, showing that fiber quality is partly genetically determined, and low-ADF/NDF species could be preferred for breeding or recommendation. The most extreme trial-based variation was seen in the K/(Ca + Mg) ratio, with a delta of 14.95, indicating big differences among species in mineral balance. Likewise, K also showed a very high delta (Δ = 13.44), highlighting how species-specific traits or their nutrient uptake efficiency can drastically affect this parameter.

On the other hand, P, Ca, and Mg showed moderate to high variability (Δ = 4.97, 4.71, and 4.40, respectively), suggesting that genetic traits or root nutrient uptake characteristics of different species are critical in determining mineral content. Whereas Ca/P ratio also showed trial-based variation (Δ = 3.60), which could influence mineral bioavailability and livestock performance. In general, most of the parameters remained stable and low for years, showing less impact of environmental changes on biological or varietal differences. Whereas relatively high variability was observed for species, showing the impact of biological differences or management practices. Based on Taguchi results, *L. pratensis* was the most ideal species, followed by *L. czeczottianus* and *L. rotundifolius*, respectively. Whereas *L. roseus* was the least desired species among the tested species (Table 2).

Forage quality trait correlation analysis

To understand the correlation between key agronomic and nutritional parameters on forage quality, a correlation matrix was constructed using average values of quality traits of four Lathyrus species, illustrating indicative patterns rather than definitive relationships. Results indicate potential correlations for CPR and DE, suggesting that forage samples with higher protein content also tend to exhibit higher digestible energy levels. Whereas CAR showed a positive association with DMR and DMI, which may reflect a potential relationship between carbohydrate-related traits and forage utilization. Similarly, DDM exhibited positive associations with RFV and DE, suggesting possible links between digestibility and overall feed quality.

Results also exhibit negative correlations for ADF and DDM, possibly due to indigestible plant cell wall

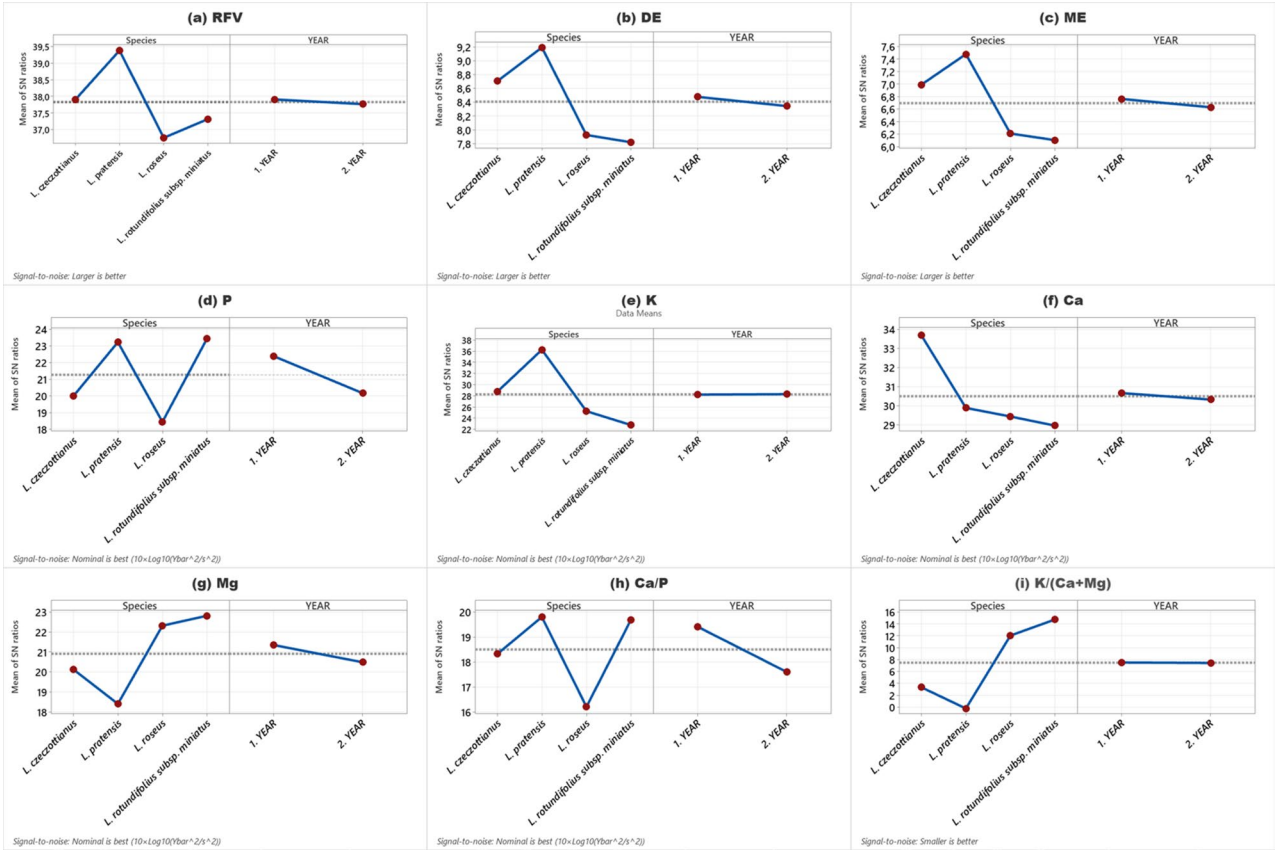


Fig. 2 Signal to noise ratio of nutrient traits of Lathyrus species (**a**) RFV, (**b**) DE, (**c**) ME, (**d**) P, (**e**) K, (**f**) Ca, (**g**) Mg, (h) Ca/P, and (**i**) K/(Ca + Mg)

components. Whereas NDF exhibited a negative correlation with DMI and RFV, confirming that higher fiber content can limit intake and overall feed value. A negative relationship was observed between ADF/NDF and traits like CAR, DE, and CPR, emphasizing that high fiber reduces energy and nutrient efficiency. To further highlight interdependence, a correlation network graph was generated, focusing on strong relationships ($|r| \geq 0.75$). This visualized how key traits like CPR, DDM, RFV, and DE form a positively reinforcing cluster, while ADF and NDF act as suppressors in the system. These patterns are essential in targeting simultaneous improvement of multiple traits (Fig. 3).

Machine learning analysis

The performance of both RF and LightGBM models demonstrates promising predictive accuracy for most of the forage quality parameters within the given dataset. Results revealed high R^2 values for both RF (0.995) and LightGBM (0.996) models, along with very low or negligible MSE and MAE values. The consistent high accuracy of both models to K might be due to its strong correlation with other forage quality markers. The K/(Ca + Mg) ratio exhibited a similar trend to K, and high R^2 scores were recorded for both models, underscoring the synergy between the nutrient dynamics in predicting plant mineral balances. For ADP, both models showed excellent predictive performance, with the LightGBM model (0.984) slightly outperforming the RF model (R^2 of 0.977). Results were further validated by negligible MSE and MAE, possibly owing to its relationships with fiber components and dry matter digestibility.

Both models also predicted moderate R^2 scores for Ca, P, CPR, NDF, DDM, ME, Hmcl, and ADF. Results revealed R^2 scores of 0.807 (RF) and 0.761 (LightGBM), with relatively very low errors, possibly due to interactions with other macronutrients like P and Mg. The results of P were recorded as 0.733 for the RF and 0.666 for the LightGBM model. Results of CPR were almost identical and recorded as R^2 of 0.842 (RF) and 0.828 (LightGBM). However, relatively high MAPE (9.0%) scores indicated some underlying variability, perhaps due to physiological or environmental factors.

The NDF results were an R^2 score of 0.749 for the RF and 0.718 for the LightGBM model. However, a moderate MAPE score of 2.7–2.8 suggests that fiber complexity or environmental modulation could add noise. The DDM revealed relatively lower R^2 scores of 0.656 for RF and 0.537 for LightGBM, suggesting that both models struggle to capture the variability. The ME was also

Table 2 Response table of the signal-to-noise ratio for forage quality parameters of different *Lathyrus* species

Data	CAR (%)	DMR (%)	CPR (%)	ADF (%)	NDF (%)	ADP (%)	Hmcl (%)	DDM	DMI
Target	LTB	LTB	LTB	STB	STB	LTB	LTB	LTB	LTB
Year 1	17.19	39.43	23.24	-32.50	-36.38	3.62	27.24	34.91	5.061
Year 2	16.64	39.40	23.16	-32.65	-36.39	3.32	26.12	34.76	5.125
Delta	0.55	0.03	0.09	0.15	0.001	0.30	1.11	0.15	0.064
L. czeczottianus	16.56	39.34	23.76	-32.10	-36.67	5.99	28.83	35.16	4.834
L. pratensis	22.21	39.23	25.98	-31.08	-35.71	1.68	27.93	35.71	5.874
L. roseus	11.59	39.45	22.55	-33.52	-36.91	2.95	26.79	34.30	4.555
L. rotundifolius	17.30	39.64	20.50	-33.61	-36.26	3.25	23.16	34.17	5.108
Delta	10.62	0.41	5.48	2.53	1.20	4.32	5.67	1.53	1.319
Opt. Sp.	2	4	2	2	2	4	4	2	2

Data	RFV	DE (Mcal/kg)	ME (Mcal/kg)	*P* (%)	K (%)	Ca (%)	Mg (%)	Ca/*P*	K/(Ca + Mg)
Target	LTB	LTB	LTB	NTB	NTB	NTB	NTB	NTB	STB
Year 1	37.90	8.475	6.76	22.38	28.24	30.68	21.35	19.41	7.52
Year 2	37.76	8.342	6.63	20.16	28.32	30.33	20.49	17.60	7.45
Delta	0.14	0.133	0.13	2.22	0.08	0.34	0.86	1.81	0.08
L. czeczottianus	37.89	8.703	6.99	19.99	28.77	33.69	20.13	18.32	3.36
L. pratensis	39.37	9.190	7.48	23.22	36.26	29.90	18.41	19.80	-0.21
L. roseus	36.75	7.924	6.21	18.45	25.28	29.45	22.32	16.21	12.05
L. rotundifolius	37.32	7.818	6.11	23.42	22.81	28.98	22.81	19.69	14.74
Delta	2.62	1.372	1.37	4.97	13.44	4.71	4.40	3.60	14.95
Opt. Sp.	2	2	2	1	1	2	1	1	4

LTB larger-the-better, *STB* smaller-the-better, *NTB* nominal-the-better, *CAR* Crude Ash Ratio, *DMR* Dry Matter Ratio, *CPR* Crude Protein Ratio, *ADF* Acid Detergent Fiber, *NDF* Neutral Detergent Fiber, *ADP* Acid Detergent Protein, *Hmcl* Hemicellulose, *DDM* Digestible Dry Matter, *DMI* Dry Matter Intake, *RFV* Relative Feed Value, *DE* Digestible Energy, *ME* Metabolic Energy

predicted moderately, and R^2 of 0.657 for RF and 0.537 for LightGBM was observed. Although Hmcl exhibited high R^2 scores (0.794 for RF and 0.706 for LightGBM), relatively high MAPE scores (11.5–14.8%) were recorded, indicating the complex plant physiology and structure of Hmcl that may require deeper feature engineering. The RF model predicted ADF better (0.863) compared to the LightGBM (0.747), with relatively higher MSE in LightGBM (8.912). Whereas low R^2 (~0.435–0.471) Ca/P with significantly high MAPE (12–13%) implies inconsistent performance, reflecting the possible impact of external factors such as mineral antagonism. The DMI was the least predicted factor, and both models failed to predict accurately, resulting in negative R^2 scores. Detailed error metrics of the DMI trait were not documented in Table 3 because both produced negative R2 values, indicating poor predictive performance.

The results of Model-dependent performance variations revealed better performance of the RF model compared to the LightGBM model for fiber-related (ADF and DDM) metrics, possibly due to high errors attributed to the LightGBM model. Both models were excellent for RFV prediction (R^2 ~0.911 to 0.918). LightGBM was better due to lower MAPE (2.589 vs. 3.12). Conversely, slightly higher MSE and MAE were recorded for the LightGBM model despite showing a higher R^2 score. Variables such as K, K/(Ca + Mg), ADP, and RFV with relatively $R^2 > 0.9$ and low MAPE < 5% make them traits with comparatively stable predictive behavior in the present dataset. The graphical presentation of actual and predicted scores of all traits is presented in Figs. 4a-h and 5a-i.

A comprehensive evaluation of both models demonstrates strong capabilities in predicting a wide range of forage quality parameters, based on the complexity of the biological trait. Both models predicted moderately complex traits involving intake dynamics and fiber fractions. Whereas consistent and accurate prediction was attributed to parameters such as K, digestible proteins, and certain mineral ratios.

Discussion

The present study highlights the potential of integrating multiple statistical and ML models to evaluate forage quality traits in Lathyrus species. Results of ANOVA analysis illustrated the substantial genetic variability in forage quality traits among the four Lathyrus species. The results underscore the importance of species selection in optimizing forage production, the significant impact of species on almost all factors, showing strong genetic control over nutrient composition and fiber characteristics. Results align with findings from recent studies

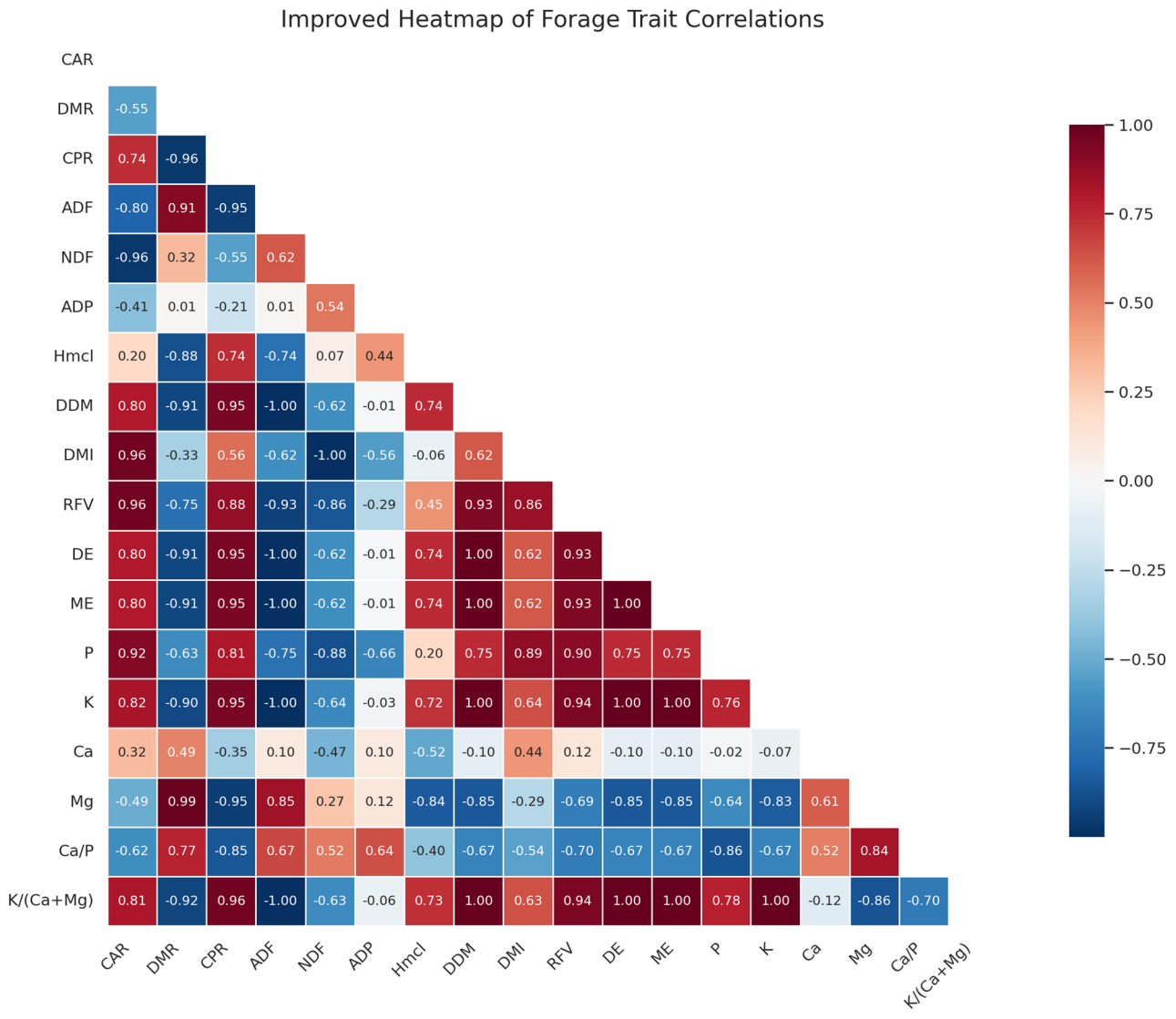


Fig. 3 Heat map analysis of forage and nutrient traits of Lathyrus species

emphasizing the role of genetic diversity in determining forage quality in legumes [31, 32]. Forage traits like CAR and [K/(Ca + Mg)] were regulated by year, indicating the impact of environmental variability on certain mineral and ash-related parameters [33]. Whereas CAR, ADF, NDF, and DDM parameters were regulated by the interaction of year × species, reflecting that climatic factors can modulate species-specific responses related to digestibility and feed value [34].

Results further revealed the superior nutritional quality of *L. pratensis*, characterized by high CPR, DDM, DE, ME, and RFV values. These traits collectively indicate the nutritional quality of forage and regulate animal intake and productivity. The superior CPR and DDM suggest efficient nutrient assimilation and lower fiber contents. Whereas high DE and ME indicate sustainable energy provision for ruminants. The elevated mineral content (K and P) further enhances the suitability of *L. pratensis* for livestock nutrition. The findings support prioritizing this species in regions that lack protein-rich, digestible feed for livestock. Results were further supported by considering other species, where high fiber and mineral content, including the highest ADF and Ca/P ratio, were recorded for *L. rotundifolius* subsp. *miniatus*. Although elevated Ca and Mg levels can be beneficial for forage crops, the high ADF content may limit digestibility, potentially reducing animal intake. Whereas elevated hemicellulose and ADP were recorded for *L. czeczottianus*, *which* revealed moderate quality and can be used as supplemental forage. Among the tested species, *L. roseus* was the least preferred species due to lower CPR, DE, ME, and RFV values. These findings emphasize the critical role of species selection in supporting preliminary data-driven evaluation of forage quality [35].

Table 3 Performance metrics of ML models for different forage traits of Lathyrus species

Model	**RF**				**Model**	**LightGBM**			
Metrics	R^2	MSE	MAE	MAPE	Metrics	R^2	MSE	MAE	MAPE
CAR	0.973	0.302	0.319	3.602	CAR	0.974	0.300	0.341	3.911
DMR	0.405	3.344	1.381	1.473	DMR	0.358	3.609	1.530	1.636
CPR	0.842	2.209	1.241	8.966	CPR	0.828	2.399	1.330	9.657
ADF	0.863	4.836	1.893	4.521	ADF	0.747	8.912	2.422	5.587
NDF	0.749	4.253	1.785	2.706	NDF	0.718	4.775	1.871	2.837
ADP	0.977	0.002	0.035	2.424	ADP	0.984	0.001	0.032	2.22
Hmcl	0.794	7.887	1.782	11.528	Hmcl	0.706	11.285	2.434	14.806
DDM	0.656	8.335	2.201	4.049	DDM	0.537	11.210	2.611	4.859
DMI	-0.005	*	*	*	DMI	-0.0889	*	*	*
RFV	0.911	7.764	2.360	3.120	RFV	0.918	7.144	1.963	2.589
DE	0.682	0.014	0.093	3.624	DE	0.537	0.021	0.112	4.344
ME	0.657	0.010	0.079	3.717	ME	0.537	0.014	0.092	4.344
P	0.733	0.000	0.018	8.363	P	0.666	0.001	0.021	9.351
K	0.995	0.002	0.034	4.539	K	0.996	0.002	0.033	4.312
Ca	0.807	0.004	0.048	2.915	Ca	0.761	0.005	0.056	3.386
Mg	0.767	0.001	0.028	9.047	Mg	0.736	0.001	0.029	9.509
Ca/P	0.471	1.264	0.927	12.089	Ca/P	0.435	1.350	0.973	12.996
K/(Ca+Mg)	0.995	0.001	0.019	5.529	K/(Ca+Mg)	0.995	0.001	0.019	5.501

*Error metrics for DMI are not reported due to negative R^2 values by ML models; CAR (Crude Ash Ratio), DMR (Dry Matter Ratio), CPR (Crude Protein Ratio), ADF (Acid Detergent Fiber), NDF (Neutral Detergent Fiber), ADP (Acid Detergent Protein), Hmcl (Hemicellulose), DDM (Digestible Dry Matter), DMI (Dry Matter Intake), RFV (Relative Feed Value), DE (Digestible Energy), ME (Metabolic Energy)

Although ANOVA identifies statistically significant mean differences for species, TD analysis provides additional insights by quantifying robustness using S/N ratios. The results of the S/N ratio-based analysis revealed the significant variability among treatments and years. The results of S/N for years showed relatively low variation for most parameters, suggesting the minimum impact of annual climatic variability or field management on nutrient composition compared to biological or treatment-related factors. The results of DMR, DMI, and ME exhibited stability over time due to very low delta for these parameters [36, 37]. On the contrary, K/(Ca + Mg), K, CAR, CPR, and Hmcl were the most sensitive and exhibited variability, illustrating the influence of species and climatic conditions on mineral balance and energy/protein components in forage [38, 39].

Likewise, variable variation among species due to genetic diversity or management treatments was also observed. Relatively high variation for the K/(Ca + Mg) ratio and K levels highlights the critical role of species selection. It is well established that excessive K uptake can reduce Ca and Mg absorption, leading to disorders [40, 41]. Likewise, the variation in CPR and Hmcl among species suggests a strong influence on digestibility and protein content [42]. Whereas moderate variation in ADF and NDF among the species revealed the potential for genotype-based improvement in fiber digestibility [43]. Lower ADF and NDF are desirable for better forage intake and energy utilization [44]. High variability in mineral parameters such as Ca, Mg, P, and particularly the Ca/P and K/(Ca + Mg) ratios, suggests the need for targeted management of soil fertility and plant nutrient uptake. Maintaining these ratios within recommended limits is essential for ensuring optimal animal health and performance [45], and variation in forage quality can be attributed to species differences and environmental adaptation.

The correlation analysis highlights distinct groups of positively and negatively associated forage traits. Correlation matrix is a powerful tool, and critical for breeding, species selection, and nutrient management in forage systems based on agronomic and nutritional parameters [46]. Understanding these relationships has profound implications for forage science due to balanced digestibility, nutrient density, and fiber structure to meet the requirements of different livestock systems. Several forage quality traits like CPR, DDM, and RFV serve as critical indicators of forage quality and animal performance potential, and their strong positive correlations highlight an opportunity for simultaneous improvement through breeding programs [46, 47]. It was observed that high protein content (CPR) also tends to exhibit higher digestible energy levels (DE), which confirms the role of protein as a contributor to energy density in ruminant diets [43]. Likewise, higher carbohydrate concentrations improve forage recovery and palatability [48], as CAR was correlated positively with DMR and DMI. The central role of determining feed quality was also confirmed by promoting positive associations between DDM with RFV and DE [49].

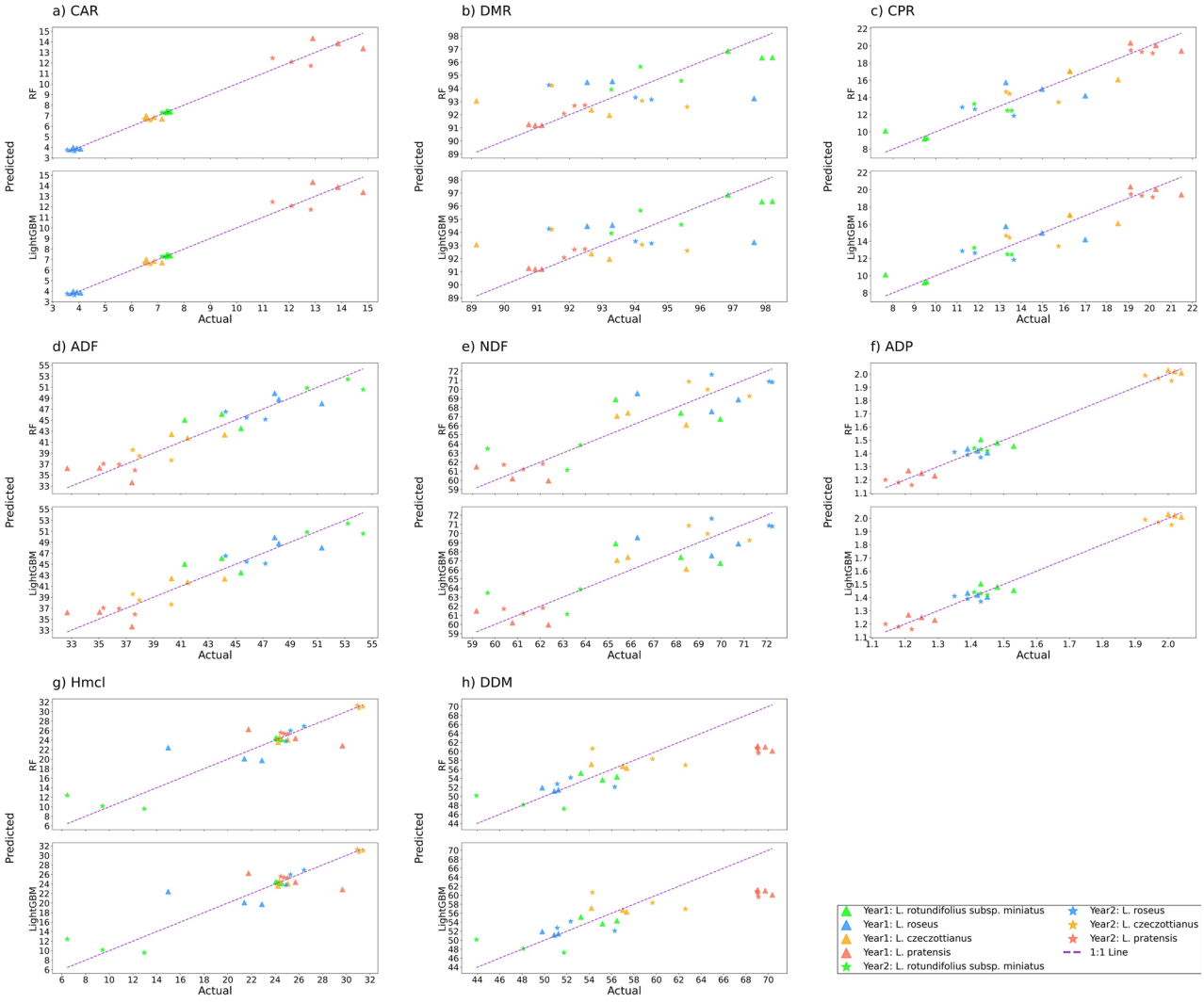


Fig. 4 Actual and predicted values of RF and LightGBM model of forage traits of Lathyrus species (**a**) CAR, (**b**) DMR, (**c**) CPR, (**d**) ADF, (**e**) NDF, (**f**) ADP, (**g**) Hmcl, and (**h**) DDM

A negative correlation between ADF and DDM exhibited the reduced digestibility due to indigestible plant cell wall components (mainly cellulose and lignin) contained in ADF [50]. Whereas a negative correlation of NDF with DMI and RFV indicates that higher fiber content can limit intake and overall feed value [51]. A negative relationship was observed between ADF/NDF and traits like CAR, DE, and CPR, emphasizing that high fiber reduces energy and nutrient efficiency. Conversely, the negative impact of ADF and NDF on digestibility and intake reinforces the necessity to select for low-fiber genotypes without compromising yield. These insights aid in developing multi-trait selection indices, supporting decisions on variety deployment and nutritional planning [46]. Moreover, improving correlations between energy and intake traits contributes to ruminant feed efficiency and sustainability of forage-based systems, especially in climate-stressed environments [43].

The RF and LightGBM models are highly efficient models used for agricultural traits prediction due to their ability to handle nonlinear interactions and their robustness against overfitting in high-dimensional data [52, 53]. Studies on nutrient contents, forage digestibility, yield, yield traits, or environmental responses have been documented [54]. Comparison of results based on metrics revealed the significance of agronomic decisions. A high R^2 score is generally desired for traits like mineral content and digestibility because these influence critical feeding decisions. Whereas MAE is crucial for nutrient threshold planning. MAPE metrics help to understand relative prediction accuracy. Traits like Hmcl and Ca/P showing MAPE > 10% suggest them for fine-grained recommendations. An analysis of the model's performance based on multiple metrics for robust model selection is recommended [55], especially for nutritional monitoring systems for legumes for precise prediction [54].

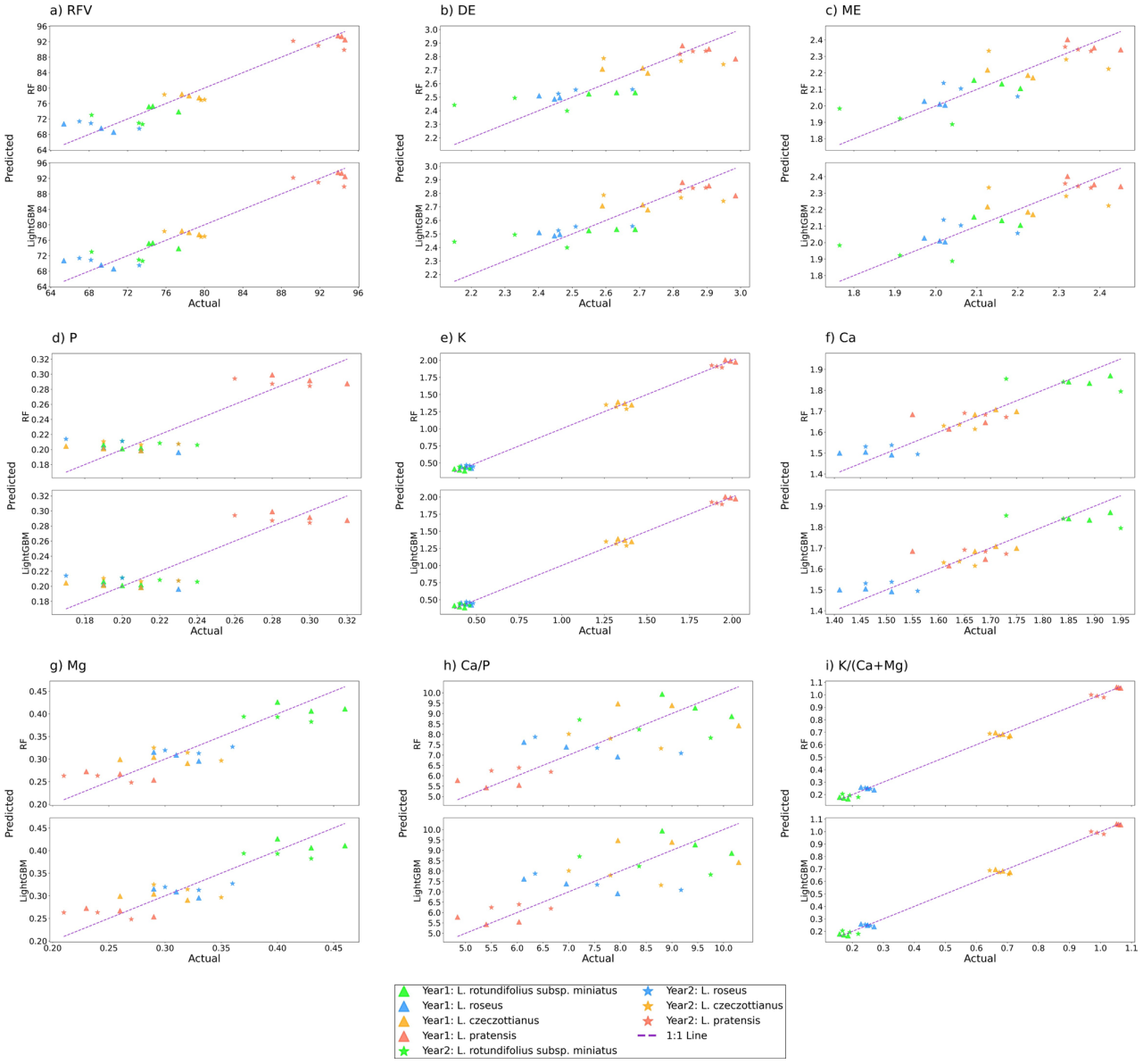


Fig. 5 Actual and predicted values of RF and LightGBM models of nutrient traits of Lathyrus species (**a**) RFV, (**b**) DE, (**c**) ME, (**d**) P, (**e**) K, (**f**) Ca, (**g**) Mg, (**h**) Ca/P, and (**i**) K/Ca + Mg)

In forage science, data regarding digestible matter, macro/micronutrients, and energy values interact in complex ways that require ensemble learning models to provide a valuable framework [56]. Results revealed high K and K/(Ca + Mg) ratio, which reflects their stable behavior in plant tissues and consistent soil-plant uptake patterns [57]. Protein digestibility depends largely on fiber (e.g., ADF, NDF) and N2 contents, and exhibits strong predictability in ML models [58]. Whereas moderate predictability of CPR and NDF suggests partial linearity and stable responses under various environmental conditions [59], possibly due to the low impact of genotype-environment interactions on protein and fiber-related traits. The DMI exhibited a negative or near-zero R^2 value, found extremely hard to predict due to the challenge in behavioral intake data using only physico-chemical data [60], weak correlation with input features, or failure of models to deal with non-linear physiological drivers. Hmcl and ADF exhibited high MAPE values despite being structural fiber components, due to environmental variability in lignification and plant maturity stages, which aren't always well captured in numerical features. Similar inconsistencies in predicting leaf mechanical traits using ML are attributed to the limited explanatory power of indirect features, such as spectral reflectance [17]. Conversely, Ca/P ratio exhibited high prediction error due to strong interdependence and antagonism between Ca and P uptake [57].

Model selection and interpretability revealed that the RF model offers better interpretability through feature importance rankings, better handling of missing values, and more robust performance in variables with nonlinear distributions, such as fiber fractions and energy traits [61]. The RF model performs better in biological yield forecasting in environments with a data set having more noise and high dimensionality. On the other hand, LightGBM tends to outperform larger datasets, but can be more sensitive to hyperparameter tuning and less robust in high-noise contexts, which may explain its higher error values for traits like DDM or Hmcl. Although negative R^2 scores were observed for DMI by both models, this trait showed poor predictive performance and was therefore excluded from the detailed error-metric reporting in Table 3, which reflects the high uncertainty and limited reliability of its prediction. Although LOOCV-based ML models demonstrated high predictive performance, small sample sizes may lead to overfitting and fail to generalize to new data, and results should therefore be interpreted as exploratory rather than definitive. It is recommended that integrating spectral and environmental data can be helpful to improve model robustness for traits that are highly sensitive to climate, soil, and management [57]. Future studies should include direct quantification of β-ODAP to evaluate potential toxicity. Likewise, application of neural network models, LSTM-based models [13], and feature selection optimization can be helpful to enhance prediction performance [62]. It is also recommended to follow the practical constraints on wild collections, and broader utilization should follow sustainable practices, prioritizing cultivated species rather than collecting from nature.

Conclusions

This study demonstrated that integrating Taguchi design with ML models provides a robust and efficient framework for evaluating forage quality traits in Lathyrus species. The results of TD and ML revealed *L. pratensis* as the most robust species for forage quality under variable conditions. These results support the targeted selection of Lathyrus species with a superior nutritional profile for future breeding. On the other hand, the predictive capability of ML models offers promising exploratory tools for decision support, allowing more precise farming of forage crops for diverse environments. However, their practical decision support applicability requires validation using a larger and more diverse dataset. The failure of ML models to predict the accuracy of DMI highlights the need for more comprehensive datasets, hyperparameter tuning, and trait-specific model optimization. Despite the robustness of the dual framework, data collection of species should be broadened to multiple locations and include both natural and cultivated species to increase genetic pools. Integrating multi-omics data with TD and ML may further enhance trait stability and accelerate precision breeding. Synergy between statistical design methods and artificial intelligence tools can significantly enhance precision breeding, trait stability, and sustainable forage production, and improve decision-making in legume improvement initiatives, especially in regions with limited agronomic resources.

Abbreviations

TD	Taguchi Design
S/N ratio	Signal-to-Noise ratio
AI	Artificial Intelligence
ML	Machine Learning
RF	Random Forest
LightGBM	Light Gradient Boosting Machine
R^2	Coefficient of Determination
MSE	Mean Square Error
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ADF	Acid Detergent Fiber
NDF	Neutral Detergent Fiber
ADP	Acid Detergent Protein
P	Phosphorus
K	Potassium
Ca	Calcium
Mg	Magnesium
DDM	Digestible Dry Matter
DMI	Dry Matter Intake
RFV	Relative Feed Value
DE	Digestible Energy
ME	Metabolic Energy
CAR	Crude Ash Ratio
DMR	Dry Matter Ratio
CPR	Crude Protein Ratio
Hmcl	Hemicellulose
LOO-CV	Leave-one-out cross-validation

Acknowledgements
Not applicable.

Authors' contributions
Conceptualization, C.S., and M.A.; methodology, M.I.Ç. C.S., M.A., and S.A.A.; software, C.S., M.A., and S.A.A.; validation, M.I.Ç., and S.A.A.; formal analysis, M.I.Ç., C.S., and M.A.; investigation, C.S., and M.I.Ç.; resources, M.I.Ç., and S.A.A.; data curation, M.I.Ç., and C.S; writing—original draft preparation, C.S., M.I.Ç. and M.A.; writing—review and editing, S.A.A.; visualization, M.A., and S.A.A.; supervision, C.S.; project administration, C.S., and M.A.; funding acquisition, M.I.Ç. All authors have read and agreed to the published version of the manuscript.

Funding
This study was supported by the Recep Tayyip Erdoğan University Development Foundation (Grant number: 02026002024142).

Data availability
The data and the code used in this article can be accessed under GPL-3.0 at Zenodo: https://doi.org/10.5281/zenodo.16538338.

Declarations

Ethics approval and consent to participate
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Field Crops, Faculty of Agriculture, University of Recep Tayyip Erdogan, Rize, Türkiye
[2]Department of Field Crops, Faculty of Agriculture, Ankara University, Ankara, Türkiye
[3]Department of Information Systems and Technologies, Bilkent University, Ankara, Türkiye
[4]Department of Precision Agriculture and Agricultural Robots, Faculty of Agricultural Sciences and Technologies, Sivas University of Science and Technology, Sivas, Türkiye

Received: 8 July 2025 / Accepted: 13 May 2026
Published online: 20 May 2026

References

1. Başaran U, Acar Z, Aşçı ÖÖ, Mut H, Ayan İ. Mürdümük (Lathyrus sp.) Türlerinin Önemi, Tarımda Kullanım Olanakları ve Zararlı Madde İçerikleri. Anadolu Tarım Bilim Derg. 2007;22:139–48.
2. GünerA, Vural M, Sorkun K, Rize, Florası. Vejetasyonu ve Yöre Ballarının Polen Analizi. TÜBİTAK Mat. Fiz. ve Biyol. Bilim Araş. Grubu, Proje No TBAG-650, Ankara, Türkiye. 1987. https://scholar.google.com/scholar?hl=tr&as_sdt=2005&sciodt=0%2C5&cites=15429481440133438243&scipsc=&q=G%C3%BCner+A%2C+Vural+M%2C+Sorkun+K%2C+Rize%2C+Floras%C4%B1.+Vejetasyonu+ve+Y%C3%B6re+Ballar%C4%B1n%C4%B1n+Polen+Analizi.+T%C3%9CB%C4%B0TAK+Mat.+Fiz.+ve+Biyol.+Bilim+Ara%C5%9F.+Grubu%2C+Proje+No+TBAG-650%2C+Ankara%2C+T%C3%BCrkiye.+1987.&btnG=.
3. Yan Z-Y, Spencer PS, Li Z-X, Liang Y-M, Wang Y-F, Wang C-Y, Li F-M. Lathyrus sativus (Grass Pea) and Its Neurotoxin ODAP. Phytochemistry. 2006;67:107–21.
4. Hanbury CD, White CL, Mullan BP, Siddique KHM. A review of the potential of lathyrus sativus L. and L. cicera L. Grain for Use as Animal Feed. Anim Feed Sci Technol. 2000;87:1–27.
5. Mia N, Sarker T, Halim MA, Alam A, Ali MS, Rahman MM, Hashem MA. Machine learning overview and its application in the livestock industry. Meat Res. 2025;5:109.
6. Freddi A, Salmon M, Freddi A, Salmon M. Introduction to the Taguchi Method. Des Princ Methodol Concept. 2019;First Prototyp:159–80.
7. Taguchi G, Wu Y. Introduction to Off-Line Quality Control. Central Japan Quality Control Assoc; 1985. https://scholar.google.com/scholar?hl=tr&as_sdt=0%2C5&q=Taguchi+G%2C+Wu+Y.+Introduction+to+Off-Line+Quality+Control.+Central+Japan+Quality+Control+Assoc%3B+1985.&btnG=.
8. Vakili R. The Application of Taguchi method to optimize pellet quality in broiler feeds. Acta Sci Anim Sci. 2023;45:e58931.
9. Farhangi H, Mozafari V, Roosta HR, Shirani H, Farhangi M. Optimizing growth conditions in vertical farming: enhancing lettuce and basil cultivation through the application of the Taguchi method. Sci Rep. 2023;13:6717.
10. Orłowska R, Pachota KA, Machczyńska J, Niedziela A, Makowska K, Zimny J, Bednarek PT. Improvement of anther cultures conditions using the Taguchi method in three cereal crops. Electron J Biotechnol. 2020;43:8–15.
11. Chen H-J, Lin H-C, Tang C-W. Application of the Taguchi method for optimizing the process parameters of producing controlled low-strength materials by using dimension stone sludge and lightweight aggregates. Sustainability. 2021;13:5576.
12. Coşkun ÖF, Aydın A, Toprak S, Başak H. Rootstock-mediated salinity resilience in cucumber (Cucumis sativus L.): integrating physiological traits, genomic stability and machine learning. BMC Plant Biol. 2025;26:121.
13. Cen H, Yu L, Pu Y, Li J, Liu Z, Cai Q, Liu S, Nie J, Ge J, Guo JA. Method to predict CO_2 mass concentration in sheep barns based on the RF-PSO-LSTM model. Animals. 2023;13:1322.
14. Morse-McNabb EM, Hasan MF, Karunaratne SA. Multi-variable sentinel-2 random forest machine learning model approach to predicting perennial ryegrass biomass in commercial dairy farms in Southeast Australia. Remote Sens. 2023;15:2915.
15. Jamshidi EJ, Yusup Y, Hooy CW, Kamaruddin MA, Hassan HM, Muhammad SA, Shafri HZM, Then KH, Norizan MS, Tan CC. Predicting oil palm yield using a comprehensive agronomy dataset and 17 machine learning and deep learning models. Ecol Inform. 2024;81:102595.
16. Wang H, Shao W, Hu Y, Cao W, Zhang Y. Assessment of six machine learning methods for predicting gross primary productivity in Grassland. Remote Sens. 2023;15:3475.
17. Ercan U, Kabas O, Moiceanu G. Prediction of leaf break resistance of green and dry alfalfa leaves by machine learning methods. Appl Sci. 2024;14:1638.
18. Li D, Yan G, Li F, Lin H, Jiao H, Han H, Liu W. Optimized machine learning models for predicting core body temperature in dairy cows: enhancing accuracy and interpretability for practical livestock management. Animals. 2024;14:2724.
19. Bleiholder H, Weber E, Lancashire P, Feller C, Buhr L, Hess M, Wicke H, Hack H, Meier U, Klose R. Growth stages of mono-and dicotyledonous plants. BBCH Monogr. 2001;158:20074619–180906.
20. Oddy VH, Robards G, Low SG. Prediction of in vivo dry matter digestibility from the fiber and nitrogen content of a feed. In: Robards GE, Packham RG, editors. Feed Information and Animal Production. Common wealth Agricultural Bureau. 1983. pp. 395-398. https://scholar.google.com/scholar?hl=tr&as_sdt=0%2C5&q=Oddy+VH%2C+Robards+G%2C+Low+SG.+Prediction+of+in+vivo+dry+matter+digestibility+from+the+fiber+and+nitrogen+content+of+a+feed.+NSW+Dep+Agric+Vet+Res+Stn+Glenfield.+1983.&btnG=.
21. Sheaffer CC, Peterson MA, Mccalin M, Volene JJ, Cherney JH, Johnson KD, Woodward WT, Viands DR. Acid detergent fiber, neutral detergent fiber concentration and relative feed value. In North American Alfalfa Improvement Conference, Minneapolis. 1995. https://www.naaic.org/stdtests/acidfiber.pdf.
22. Fonnesbeck PV, Clark DH, Garret WN, Speth CF. Predicting energy utilization from Alfalfa Hay from the Western Region. Proc Am Anim Sci (Western Sect). 1984;35:305–8.
23. Khalil J, Sawaya WN, Hyder SZ. Nutrient composition of atriplex leaves grown in Saudi Arabia. Rangel Ecol Manag. 1986;39:104–7.
24. Sall J, Stephens ML, Lehman A, Loring S. JMP start statistics: a guide to statistics and data analysis using JMP. Sas Inst. 2017. https://scholar.google.com/scholar_lookup?title=JMP+start+statistics%3A+A+guide+to+statistics+and+data+analysis+using+JMP&author=J+Sall&author=ML+Stephens&author=A+Lehman&author=S+Loring&publication_year=2017.
25. Aasim M, Katırcı R, Akgur O, Yildirim B, Mustafa Z, Nadeem MA, Baloch FS, Karakoy T, Yılmaz G. Machine Learning (ML) Algorithms and Artificial Neural Network for Optimizing In Vitro Germination and Growth Indices of Industrial Hemp. (Cannabis sativa L). Indus Crop Prod. 2022;181:114801.
26. Aggarwal CC. Neural Networks and Deep Learning. Springer; 2018. https://doi.org/10.1007/978-3-031-29642-0.
27. Pavlov YL. Random Forests. 2019. pp. 1–122. https://doi.org/10.1201/9780429469275-8.
28. Costa JM, Grant OM, Chaves MM. Thermography to explore plant–environment interactions. J Exp Bot. 2013;64:3937–49.
29. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T-Y. LightGBM: a highly efficient gradient boosting decision tree. Adv Neural Inf Process Syst. 2017;30.
30. Webb GI, Sammut C, Perlich C, Horváth T, Wrobel S, Korb KB, Noble WS, Leslie C, Lagoudakis MG, Quadrianto N, et al. Leave-One-Out Cross-Validation. Encyclopedia of Machine Learning. Boston, MA: Springer US; 2011. p. 600–1.
31. Jung HG, Allen MS. Characteristics of plant cell walls affecting intake and digestibility of forages by ruminants. J Anim Sci. 1995;73:2774–90.
32. Suttle N. Ruminant nutrition—digestion and absorption of minerals and vitamins. Ref. Modul. Food Sci. Elsevier; 2016. https://doi.org/10.1016/B978-0-08-100596-5.00964-1.
33. Wanchuk MR, McGranahan DA, Sedivec KK, Swanson KC, Hovick TJ. Prescribed fire increases forage mineral content in grazed rangeland. Int J Wildl Fire. 2024;33:7.
34. Oral HH. Forage yields and nutritive values of oat and triticale pastures for grazing sheep in early spring. PeerJ. 2024;12:e17840.
35. Tlahig S, Neji M, Atoui A, Seddik M, Dbara M, Yahia H, Nagaz K, Najari S, Khorchani T, Loumerem M. Genetic and seasonal variation in forage quality of lucerne (Medicago sativa L.) for resilience to climate change in arid environment. J Agric Food Res. 2024;15:100986.
36. Saeidnia F, Majidi MM, Dehghani MR, Mirlohi A, Araghi B. Multi environmental evaluation of persistence and drought tolerance in smooth bromegrass (Bromus inermis): genetic analysis for stability in combining ability. Crop Pasture Sci. 2021;72:7:565-574.
37. Mishra P, Hasham MM, Manjunatha GR, Ray S, Homa F, Pawariya V. Role of design of experiment in agriculture trials and its role in decision making. Int J Agric Stat Sci. 2023;19:389–99.
38. Moore KJ, Lenssen AW, Fales SL. Factors affecting forage quality. Forages Sci Grassl Agric. 2020;2:701–17.
39. Nielsen K. Agronomic analysis of the forage yield, nutritional value, species compatibility, and winter survival rates of sainfoin (Onobrychis viciifolia Scop.)

and grasses mixtures. PhD Thesis. University of Saskatchewan; 2021. https://harvest.usask.ca/handle/10388/13459.
40. McDowellL R. Minerals in animal and human nutrition. 2003. https://scholar.google.com/scholar_lookup?title=Minerals%20in%20Animal%20and%20human%20Nutrition&publication_year=2003&author=L.R.%20McDowell.
41. National Academies of Sciences and Medicine. Nutrient requirements of dairy cattle. 2021. https://www.ncbi.nlm.nih.gov/books/NBK600609/.
42. Brummer EC, Bouton JH, Casler MD, McCaslin MH, Waldron BL. Grasses and legumes: Genetics and plant breeding. Grassl Quietness Strength New Am Agric. 2009;155–71.
43. Duncan AJ, Poppi DP. Nutritional ecology of grazing and browsing ruminants. In The Ecology of Browsing and Grazing. Springer; 2008. pp. 89–116. 195©.
44. Buse KK, Carroll AL, Bradford BJ, Min D, Jagadish K, Kononoff PJ. The effect of replacing conventional alfalfa hay with lower-lignin alfalfa hay on feed intake, nutrient digestibility, and energy utilization in lactating Jersey cows. J Dairy Sci. 2024;107:9379–89.
45. Suttle N. Mineral nutrition of livestock. Cabi. 2022.
46. Chand S, Indu; Singhal RK, Govindasamy P. Agronomical and breeding approaches to improve the nutritional status of forage crops for better livestock productivity. Grass Forage Sci. 2022;77:11–32.
47. Sayar MS. Crude protein content and the other forage quality traits in some annual plant species from different families and assessments with biplot and correlation analysis. In: Kılı H, Gul I, editors. Alternative Protein Sources. Istanbul nobel Publishing; 2024. p. 17–49.
48. Stoycheva I, Vasileva V. Comparison of changes in structural carbohydrates and enzyme digestibility during vegetation in permanent and temporary grazing pastures. Bulg J Agric Sci. 2021;27.
49. Katoch R. Forage quality indices. In Techniques in Forage Quality Analysis; Springer: 2022. p. 11–16.
50. [Khudyakova KK, Kosolapova VG. Determining the structural carbohydrates and lignin levels in forage using the Van Soest and Kizel methods. Russ Agric Sci. 2022;48:400–4.
51. Oba M, Kammes-Main K. Symposium review: effects of carbohydrate digestion on feed intake and fuel supply. J Dairy Sci. 2023;106:2153–2160.
52. [Nazir N, Fatima S, Aasim M, Yaqoob F, Mahmood K, Ali SA, Awan SI. ul Haq, I. Zeugodacus fruit flies (Diptera: Tephritidae) host preference analysis by machine learning-based approaches. Com Elect Agric. 2024;222:109095.
53. [Parganiha V, Verma M. An effective soil analysis and crop yield prediction based on optimised light GBM in smart agriculture. J Agron Crop Sci. 2024;210:e12726.
54. Sun L, Lai M, Ghouri F, Nawaz MA, Ali F, Baloch FS, Nadeem MA, Aasim M, Shahid MQ. Modern plant breeding techniques in crop improvement and genetic diversity: From molecular markers and gene editing to artificial intelligence -. Crit Rev Plants. 2024;13(19):2676.
55. Aasim M, Ali SA, Altaf MT, Ali A, Nadeem MA, Baloch FS. Artificial neural network and decision tree facilitated prediction and validation of cytokinin-auxin induced in vitro organogenesis of sorghum (Sorghum bicolor L). Plant Cell Tiss Org Cult. 2023;153:611–24.
56. Smith C, Karunaratne S, Badenhorst P, Cogan N, Spangenberg G, Smith K. Machine learning algorithms to predict forage nutritive value of in situ perennial ryegrass plants using hyperspectral canopy reflectance data. Remote Sens. 2020;12:92.
57. Han P, Zhai Y, Liu W, Lin H, An Q, Zhang Q, Ding S, Zhang D, Pan Z, Nie X. Dissection of hyperspectral reflectance to estimate photosynthetic characteristics in upland cotton (Gossypium hirsutum L.) under different nitrogen fertilizer application based on machine learning algorithms. Plants. 2023;12(3):455.
58. Liu J, Feng Q, Liang T, Yin J, Gao J, Ge J, Hou M, Wu C, Li W. Estimating the forage neutral detergent fiber content of Alpine grassland in the Tibetan Plateau using hyperspectral data and machine learning algorithms. IEEE Trans Geosci Remote Sens. 2021;60:1–17.
59. Kruk M. Prediction of environmental factors responsible for chlorophyll a-induced hypereutrophy using explainable machine learning. Ecol Inf. 2023;75:102005.
60. Nguyen HD, Van CP, Nguyen TG, Dang DK, Pham TTN, Nguyen Q-H, Bui Q-T. Soil salinity prediction using hybrid machine learning and remote sensing in Ben Tre province on Vietnam's Mekong River Delta. Environ Sci Pollut Res. 2023;30(29):74340–57.
61. Alsaber A, Setiya P, Satpathi A, Aljamaan A, Pan J. Advancing pearl millet yield forecasting: Comparative analysis of individual and ensemble machine learning approaches over Rajasthan, India. PLoS ONE. 2025;20:e0317602.
62. Alsaeedi AH, Al-Mahmood HHR, Alnaseri ZF, Aziz MR, Al-Shammary D, Ibaida A, Ahmed K. Fractal feature selection model for enhancing high-dimensional biological problems. BMC Bioinformatics. 2024;25:12.

Publisher's note

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