Article

Performance of Machine Learning Models in Predicting Common Bean (*Phaseolus vulgaris* L.) Crop Nitrogen Using NIR Spectroscopy

Marcos Silva Tavares [1,*], Carlos Augusto Alves Cardoso Silva [1], Jamile Raquel Regazzo [1], Edson José de Souza Sardinha [2], Thiago Lima da Silva [1], Peterson Ricardo Fiorio [1] and Murilo Mesquita Baesso [2]

[1] Luiz de Queiroz Higher School of Agriculture, University of São Paulo—USP, Piracicaba 13635-900, SP, Brazil; jamile.regazzo@usp.br (J.R.R.); thiagolim@usp.br (T.L.d.S.); fiorio@usp.br (P.R.F.)
[2] Faculty of Animal Science and Food Engineering, University of São Paulo—USP, Pirassununga 13418-900, SP, Brazil
* Correspondence: ms.tavares@usp.br

Abstract: Beans are the main direct source of protein consumed by humans in the world and their productivity is directly linked to nitrogen. The short crop cycle imposes the need for fast methodologies for N quantification. In this work, we evaluated the performance of four machine learning algorithms in nitrogen estimation using NIR spectroscopy, comparing predictions between complete spectral data and only intervals obtained with the variable importance in projection (VIP). Doses of 0, 50, 100, and 150 kg ha^{-1} of N were applied and leaf reflectance was collected. Weka software was used to test the algorithms. The selection of the most effective spectral zones was made with the variable importance in projection (VIP). The intervals of 700–740 nm and 983–995 nm were considered the most important for the study of nitrogen. More efficient predictions were verified for RF and KNN models (R^2 = 0.89, RMSE = 2.23 g kg^{-1}; and R^2 = 0.80, RMSE = 2.89 g kg^{-1}, respectively) when only the most important spectral regions were included. The efficiency of nitrogen prediction based on NIR reflectance combined with machine learning was verified, which can serve as an important tool in precision agriculture.

Keywords: algorithms; reflectance; precision agriculture; NIR spectroscopy; bean crop

1. Introduction

Beans (*Phaseolus vulgaris* L.) are an essential staple food and one of the most produced agricultural crops worldwide, especially in Central and South America, being the main source of vegetable protein and the most important legume for direct human consumption [1–3]. In addition, beans are used for other purposes, as a biological nitrogen-fixing agent in soil and animal feed [4]. Brazil stands out in the world for being one of the largest producers and consumers of common beans (*Phaseolus vulgaris* L.), with a total area of 2,613,086 hectares harvested in 2021 [5,6].

Nitrogen (N) is considered an essential element to obtain high yields, with a significant influence on common bean growth, due to its strong link with plant photosynthesis, vegetative growth, and grain yield [7]. Within the context of precision agriculture, it is extremely important to quantify the leaf N content to establish the amount to be applied to the soil, considering the phenological stage of the crop. Meanwhile, the excessive supply of N can result in plants that are more susceptible to attack by pests and diseases, promote exaggerated vegetative development, delay the reproductive phase of plants, favor the indiscriminate growth of weeds in the cultivation area, cause contamination in soil and water, and significantly increase the cost of production.

Conventional chemical analyses for the quantification of leaf N in plants are destructive, laborious, time-consuming, and require specific equipment and inputs [8,9]. Due to the fast cycle (60 to 80 days) of the bean crop, the development of fast and in situ methodologies for N prediction are of great importance and highly anticipated in the smart agriculture scenario. Management to match the supply of N with the needs of the crop is a viable alternative to achieve high yield and reduce environmental impact, making the application more efficient [10].

In this context, spectroscopy has shown potential as a fast and non-invasive approach to observe plant characteristics, making it possible to monitor and optimize nitrogen fertilization through reflectance data, for example [11,12]. This method is based on the detection of biochemical changes that affect the photosynthetic activity, structure, and stability of chemical bonds, which promotes changes in reflectance [13]. Recently, near-infrared spectroscopy (NIRS) has been extensively studied as an innovative and economically viable technique, being successfully employed in the estimation of nutrients in plants [14,15]. Near-infrared (NIR) reflectance has been shown to be positively correlated with chlorophyll content in leaves and is useful for studying N [8].

Among the most widely used techniques to deal with hyperspectral data, partial least squares regression (PLSR) stands out as the most popular model [16,17]. However, this model presents some difficulty in capturing nonlinear connections in spectroscopic data [18]. This context increases the demand for other alternative methods capable of dealing with nonlinear connections present in datasets obtained by spectroscopy. Because hyperspectral measurements produce high and complex amounts of data, one type of approach that could ideally handle this is machine learning [19]. Machine learning algorithms such as trees, rules, support vector machines, and artificial neural networks are advantageous when compared to linear methods in treating nonlinear problems [20]. In recent years, machine learning has advanced towards building predictive models focused on quantifying plant characteristics using spectra [21]. Despite this, the use of other machine learning models for leaf N estimation is still incipient; to the best of our knowledge, this study is the first to evaluate and compare the potential of four machine learning algorithms in the prediction of leaf N in the bean crop. The use of machine learning algorithms can promote a rapid estimation of N in plants, contributing to the management of nitrogen fertilization as an important tool in the context of precision agriculture.

The main objective of this work was to evaluate and compare the performance of four machine learning algorithms (Random Forest—RF, K-nearest neighbors—KNN, Artificial Neural Network—ANN, and M5Rules—M5) in the prediction of leaf nitrogen in the common bean crop by means of hyperspectral data in the NIR range (700 to 1300 nm). The specific objectives of this research also include the following: (1) to establish which NIR spectral zones are most appropriate for leaf nitrogen estimation; (2) to compare the predictive performance of the algorithms in the spectral range of 700 to 1300 nm and considering only the specific intervals selected by the variable importance in projection (VIP).

2. Materials and Methods

2.1. Study Place and Experimental Design

The experiment was carried out in a greenhouse at the School of Animal Science and Food Engineering of the University of São Paulo (FZEA-USP), Campus Pirassununga/SP, Brazil. According to the Köppen classification, the climate in the study area is humid subtropical (Cw) with an average annual temperature of 20.6 °C and an average annual rainfall of 1238 mm [22].

The substrate used for sowing was composed of three parts of soil (Quartzarenic Neosol), 2 parts of tanned cattle manure, and 2 parts of crushed sugarcane straw, aimed at maintaining the soil uncompacted and aerated. The physicochemical properties of each substrate component are described in detail in Table 1.

Table 1. Physicochemical properties of the materials used for substrate composition.

Soil (Quartzarenic Neosol): 0–20 cm								
pH (CaCl$_2$)	P (res) mg. dm^{-3}	S (PPM)	K (res) mmolc. dm^{-3}	Ca mmolc. dm^{-3}	Mg mmolc. dm^{-3}	Al mmolc. dm^{-3}	H + Al mmolc. dm^{-3}	C.T. (g/kg)
4.4	8	4	0.3	16	4	4.1	30	9.3
Tanned Cattle Manure								
pH	MS total (%)	M.O (%)	Gray (%)	C Total (%)	C org (%)	N (g/kg)	P$_2$O$_5$ (g/kg)	Relation C/N
6.6	63.5	17.6	80	10.2	6.8	10.2	5.9	10:1
Crushed Sugarcane Straw								
pH	MS total (%)	M.O (%)	Gray (%)	C Total (%)	C org (%)	N (g/kg)	P$_2$O$_5$ (g/kg)	Relation C/N
4.9	7.6	50.6	28.4	29.4	12.8	9.1	4.8	32:1

A completely randomized design was used with four treatments (0, 50, 100, and 150 kg N ha^{-1}) and 12 replications (Figure 1). The seeds of the BRS FC104 cultivar were donated by the Brazilian Agricultural Research Corporation (EMBRAPA), the unit responsible for the improvement of the common bean plant, headquartered in the city of Santo Antônio de Goiás, Goiás/Brazil. Start-up fertilization (40, 140, and 140 kg ha^{-1} of N, P$_2$O$_5$, and k$_2$O, respectively) and nitrogen fertilization were based on the technical recommendations of the fertilization manual of the Agronomic Institute of Campinas (IAC) for an expected yield of more than 5000 kg ha^{-1}, using Urea (45% of N) as nitrogen fertilizer [23]. The nitrogen doses stipulated for each treatment were applied when the plants were in the V4 phenological stage, characterized by the complete development of the first trifoliate leaf.

Figure 1. Experimental scheme completely randomized.

A total of 48, 5 L plastic containers, 18 cm wide and 21 cm in diameter, were used for sowing beans, fully filled with the substrate to the brim, in which 4 seeds were deposited for germination in each of them on 1 May 2023. At the time of 15 days after sowing (DAS), only the two plants with the best vegetative vigor per pot were maintained. The spacing between plants was 10 cm and between rows of pots 40 cm, equivalent to a population density of 250,000 plants per hectare.

The water supply of the plants was carried out manually with the use of a graduated beaker, and a daily depth of 5 mm of water was applied. Temperature and humidity data (average, maximum, and minimum daily) were collected daily using a digital thermo-hygrometer, as shown in Figure 2.
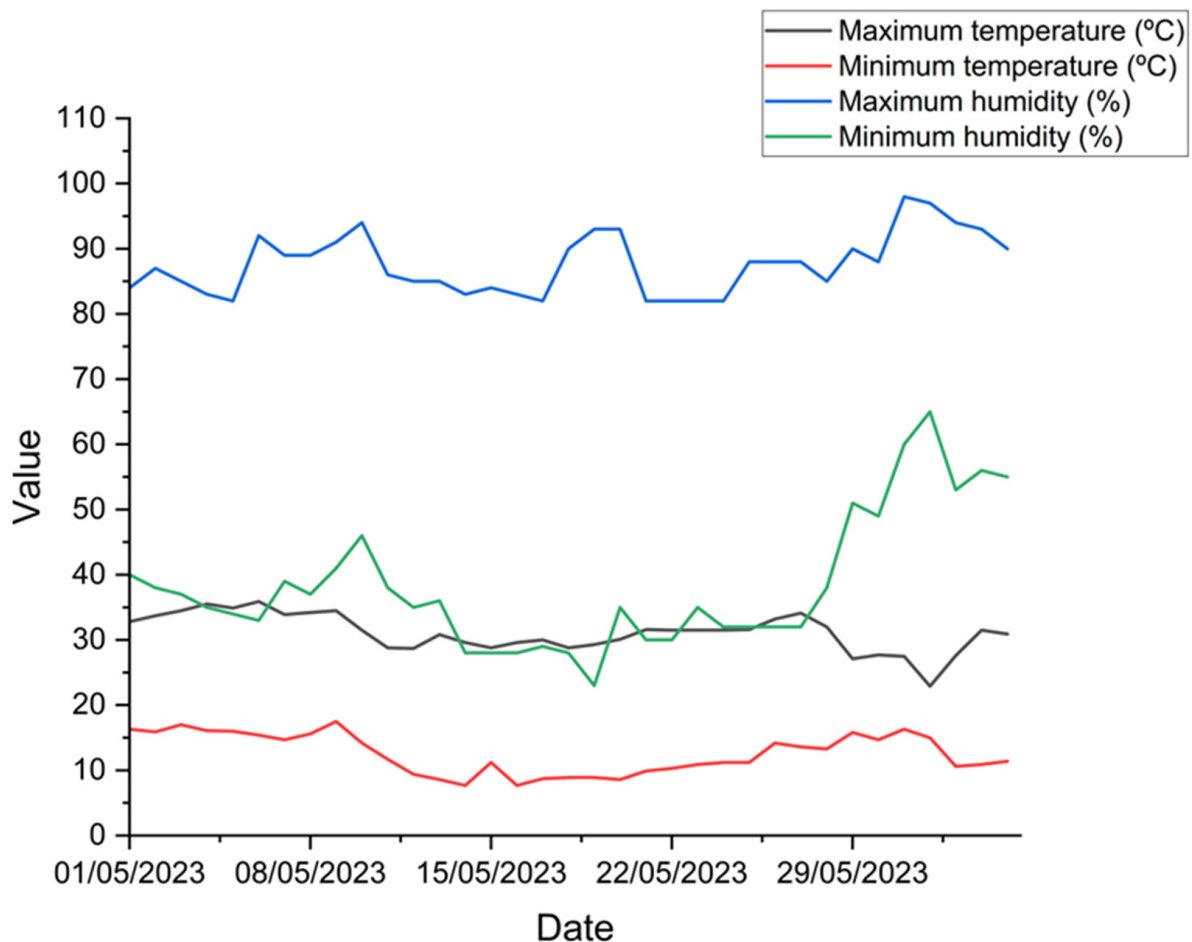

Figure 2. Temperature and humidity data collected during the experimental period.

2.2. Foliar Collection, Hyperespectral Data Acquisition, and Quantification of Leaf Nitrogen

In the V4 phenological phase at 32 days after sowing (DAS), when the third trifoliate leaf was completely open and flat, all 96 plants with uniform size (Appendix A) were sampled, and 3 leaflets were removed from each, totaling 288 samples. Immediately after collection, the leaflets were packed in sealed plastic bags that were inserted into a 20 L thermal box containing ice to maintain the turgidity of the leaf material and transported to the geoprocessing laboratory for spectral reading. For this purpose, a proximal spectro-radiometer FieldSpec3 (ASD—Analytical Spectral Devices Inc., Boulder, CO, USA) was used, using a computer with the software RS3. The equipment was used in the spectral range of 350–2500 nm with a spectral resolution of 1.4 nm from 350 to 1050 nm and 2 nm from 1050 to 2500 nm with an interval of 1 nm. To read the leaves, the Leaf Clip® probe (ASD—Analytical Spectral Devices Inc., Boulder, CO, USA) was attached to the device. The Leaf Clip® is able to maintain the same light intensity and orthogonal incidence in all

readings, thus acting as a fully controlled method. The measurements were carried out in the laboratory to keep the environment as controlled as possible. In this process, we tried to correct inconsistencies in the readings caused by external factors, such as noise, environmental variations (humidity and temperature), or even the scattering of light [24]. The sensor was calibrated every 5 min, using a barium sulfate plate for the white pattern (100% reflectance) and a black surface 0.004 in thick x 0.935 OD painted black vinyl for the black pattern (0% reflectance). In this study, spectral data were obtained in the range of 350 to 2500 nm, but only the range of 700–1300 nm, corresponding to the NIR, was stored and used for predictions of leaf N content. Generally, the range of NIR spectroscopy is situated from 700 to 1300 nm [25–27]. The quantification of the spectral response of the leaves was measured by the reflectance factor, which consists of the ratio between the radiance of the target (leaf) and the radiance of a plate with 100% reflectance, providing as a percentage how much of the energy incident on the leaves is reflected. After reading, the leaflets were placed in paper bags and dried in an oven with forced ventilation at a constant temperature of 65 °C up to constant weight. The dry leaf material was crushed to form a fine powder and sent for quantification of the nitrogen content. The chemical quantification of the N content was based on the Kjeldahl method [28].

2.3. Model Description and Performance Analysis

The models were implemented in the Environment for Knowledge Analysis (WEKA) software, version 3.8.6. WEKA is an innovative, open-source tool for all research communities working on supervised and unsupervised learning methodologies and was developed at the University of Waikato, New Zealand [29].

Random Forest (RF) was proposed by Breiman (2001) and is a nonlinear ensemble decision tree-based algorithm, which deals with high-dimensional input datasets by constructing and averaging some random decision trees for regression or classification [30]. Among the important tuning parameters for this machine learning algorithm, 2 stand out: (1) the number of trees to grow; and (2) the number of variables per node [14]. RF was implemented with cross-validation of 10 folds, a batchsize of 100, 100 interactions, 1 seed, 0 maxdepth, and 0 execution for slots.

KNN is an algorithm that can be used in regression and classification problems and is considered one of the simplest [31]. This model adopts the concept of nearest neighbors with an initial value of 'K' to find the similarity between the data points and then forwards the new data point to the category with the closest similarity [32]. Similarity is measured by calculating the Euclidean distance $ED(x, y)$, as shown in Equation (1). In this work, the algorithm was implemented with 5k-nearest neighbors, no distance weighting, meansquare = false, 2 number decimal places, and cross-validation with 10 folds.

$$ED(x, y) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{1}$$

M5Rules is a powerful implementation of Quinlan's M5 algorithm, having as a rule to recursively partition the data space and fit a prediction model within each partition [33,34]. M5Rules was configured according to Weka's standard, having a batchsize of 100, as false debug, false unpruned, with the minimum number of instances equaling 4.0, the decimal number of seats equaling 4, and cross-validation with 10 folds.

Artificial neural networks have gained significant attention due to their ability to mimic the functioning of the human brain and their effectiveness in solving complex problems [35]. The multi-layer perceptron-based ANN model is a supervised learning algorithm and consists of multiple layers of interconnected nodes, with each node performing a simple calculation using a weighted sum of its inputs and an activation function [36]. The standard Weka parameters were maintained with a learning rate of 0.3, a batchsize of 100, 0.2 momentum, the decimal number of places equaling 2.16 neurons in the hidden layer, the validation set size equal to 0, validation threshold equal to 20, and cross-validation with 10 folds.

The performance of the machine learning models was calculated with all reflectance data (700–1300 nm) and with the selection of the spectral zones most correlated with the actual nitrogen content obtained in the laboratory, adopting as a metric the importance index of the variable in the projection (VIPi), mathematically described by Equation (2). Variable importance in projection (VIP) was calculated using partial least squares regression with 15 components and a threshold equal to 1 in Matlab R2022b software. The graphs with the scores for each wavelength were extracted and processed in OriginPro 2024 software. The variable importance in projection (VIP) classifies the independent variables according to their explanatory power, making possible a more efficient evaluation of the problem of multiple collinearity between variables [37]. In our study, we were rigorous, adopting the 1.0 value as the cut-off value to select the most responsive wavelengths in the prediction of N.

$$VIP_i = \sqrt{\frac{\sum_{K=1}^{L} W_{ik}^2 \times SSY_k \times I}{SSY_{total} \times L}} \tag{2}$$

where *Wik* is the value of the weight for the variable *i*, *SSYk* is the sum of the squares of the variance for the *k*-th (*k-ésimo*), *I* is the number of independent variables, *SSYTotal* is the sum total of the squares given by the variables, and *L* is the total number of components.

The coefficient of determination (R^2) and the root mean squared error (*RMSE*) were used as the main metrics to measure the amount of explained variance and, consequently, the performance of the models used in this study in the estimation of N. The stability of the models was evaluated by comparing the difference in the values of R^2 and *RMSE*. The R^2 expresses the fit between the values estimated by the model and the actual nitrogen value, corresponding to that measured in the laboratory. The R^2 can range from 0 to 1; the closer the value is to 1, the more accurate the model is in predicting. The *RMSE* represents the deviation (error) between the estimated value and the measured value. In this case, the predictive accuracy of the model is improved proportionately as *RMSE* values are reduced. Equations (3) and (4) were used to calculate the metrics.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - P_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} \tag{3}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - P_i)^2}{n}} \tag{4}$$

where *Pi* represents the predicted value of the regression model, *y* represents the mean value of the measured value, *yi* represents the measured value, and *n* represents the sample size.

3. Results

3.1. Foliar Nitrogen Content

The descriptive analysis of the nitrogen content present in common bean leaves submitted to four levels of nitrogen fertilization is shown in Figure 3. In general, the arithmetic mean of the leaf N content increased as a function of the applied doses, reaching values of 37.02, 46.68, 51.72, and 58.9 g kg^{-1} for T1 (0 kg ha^{-1} N), T2 (50kg ha^{-1} N), T3 (100 kg ha^{-1} N), and T4 (150 kg ha^{-1} N), respectively.

The leaf N content obtained in this study is in agreement with those obtained by Nunes et al. (2021) who studied 16 varieties of common bean submitted to two doses of N (0–100 kg ha^{-1}) and found a mean leaf N content of 37.6–51.7 g kg^{-1} in BRS amethyst and 36.1–51.7 g kg^{-1} in Dama TAA in the absence of N [38]. Regarding the dose of 100 kg ha^{-1}, our mean result, 51.7 g kg^{-1}, was slightly higher than the value observed by the authors, 41.8 g kg^{-1}, for cultivar IAC millennium.
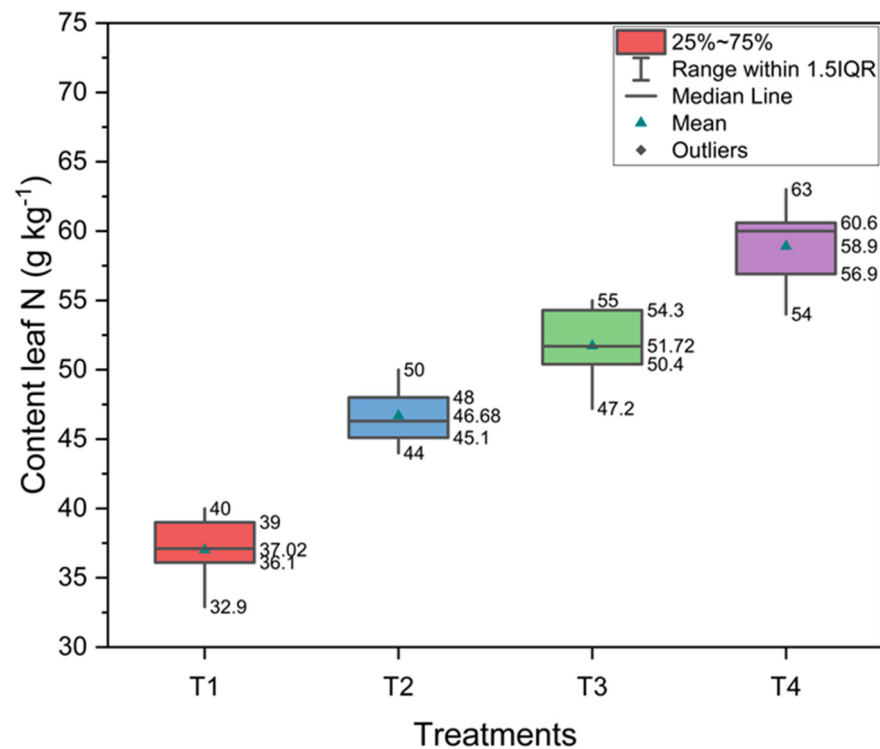
Figure 3. Average values of nitrogen content in common bean leaves submitted to nitrogen application. T1: 0 kg ha^{-1} N, T2: 50 kg ha^{-1}, T3: 100 kg ha^{-1}, T4: 150 kg ha^{-1}.

3.2. Leaf Spectral Analysis

The mean reflectance curves of the cultivar BRS FC104 comprised in the near-infrared range (NIR (700 to 1300 nm)) and resulting from the application of the four nitrogen treatments (Figure 3) are shown in Figure 4. The spectral behavior of leaves with different concentrations of N showed little variation in reflectance, being between 0.45 and 0.55 on the *y*-axis, where the two smallest reflectance curves (green and blue lines) represent the concentrations of 60.6 and 51.7 g kg^{-1} of N, respectively.

Near-infrared wavelengths (between 720 and 1300 nm) refer to the scattering of light along the mesophyll under the influence of internal leaf structures such as cell wall width, intercellular air spaces, and the amount of mesophyll per unit leaf area within the mesophyll [39,40]. Reflectances around 0.5 obtained in healthy bean plants were measured by Machado et al., (2015) in the NIR region, showing satisfactory vegetative vigor of the crop at 25 DAS [41]. When fertility is adequate, plants are more photosynthetically active, which characterizes greater absorption of electromagnetic energy in the visible region and greater reflectance in the red border and near-infrared regions [42]. Other plant species also express similar spectral behavior with respect to reflectance. Assessing the hyperspectral response of species *Megathyrsus maximus*, *Pennisetum purpureum* Schumach, *Philodendron* sp. *Tradescantia pallida* cv. *purpurea*, *Cordyline fruticosa* (L.), and *Cordyline fruticosa* (L.), in the near-infrared region (700–1300 nm), reflectance of approximately 50% with progressive decreases up to 1058 nm were observed by [40].

Two spectral zones most correlated with leaf nitrogen content were identified using the variable importance in projection (VIP) across the NIR spectrum, as shown in Figure 5. The first is located in the range of 700 to 740 nm, with the highest value (4.1) observed at the 708 nm wavelength (blue dotted line). In the second, variable importance in projection (VIP) higher than 1 occurred only in the spectral range from 983 to 995 nm, with the maximum value of 1.29 observed at the 988 nm wavelength.
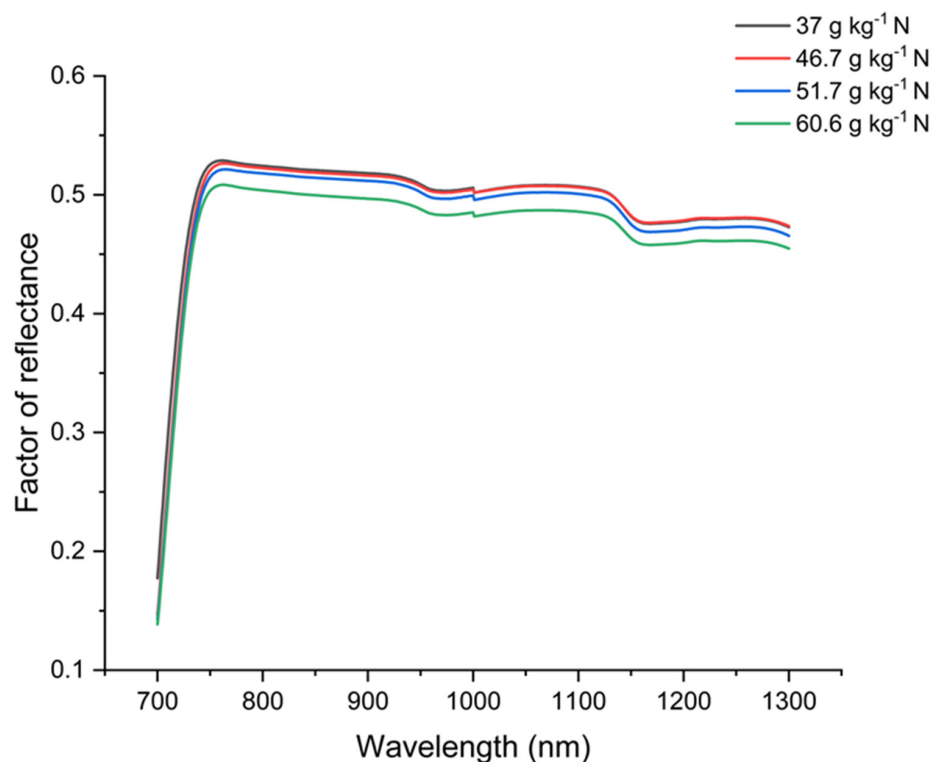
Figure 4. Spectral curve of medium reflectance (700 to 1300 nm) in BRS FC104 bean leaves grown in a greenhouse and submitted to 4 levels of nitrogen fertilization.
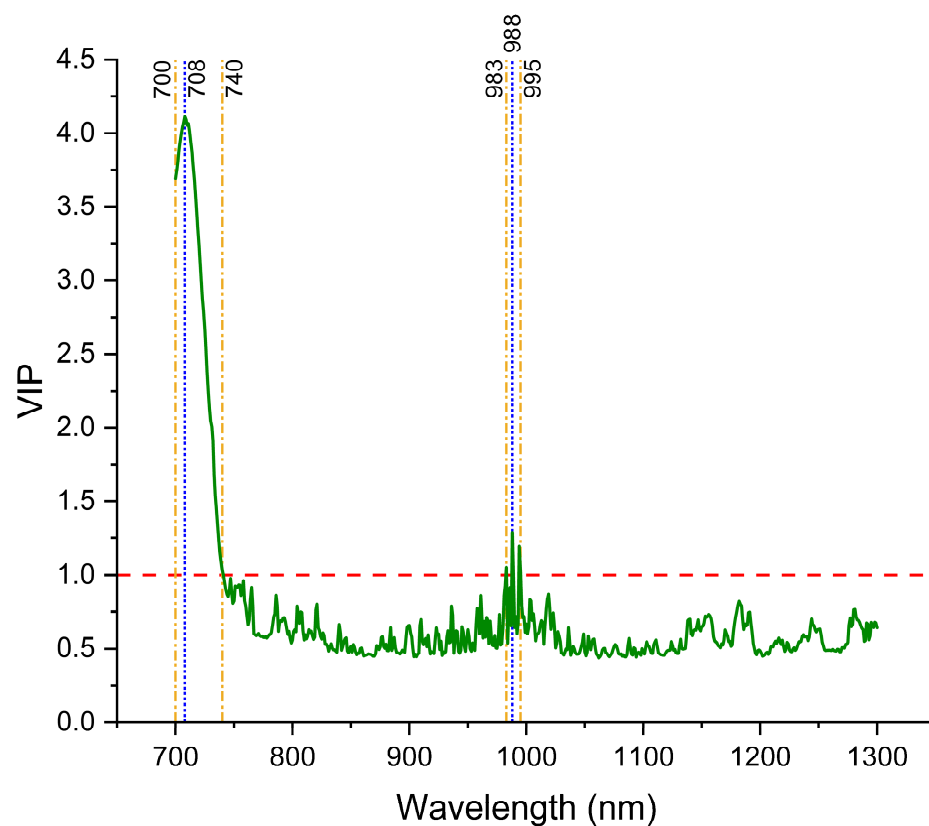

Figure 5. Selection of the most effective wavelengths for N prediction based on variable importance in projection (VIP) scores, showing two spectral zones highlighted in yellow.

Several studies have proven the efficiency of the interval between 700–740 nm in the study of leaf N in several plant species. Evaluating the importance of spectral bands in the prediction of N in sugarcane crops, Silva et al. (2023) observed variable importance in projection (VIP) ranging from 1 to 1.5 for sugarcane [43]. The accuracy of reflectance spectroscopy is further refined by selecting the most responsive wavelengths for analysis, as different plant features are more discernible at specific wavelengths [44,45]. This selection process is critical for generating reliable and actionable data [21].

RF stood out as the most accurate model in the prediction of N when the entire NIR spectrum (700–1300 nm) was used, expressing an excellent coefficient of determination ($R^2 = 0.84$) and lower error (RMSE = 2.69) between observed and predicted values (Figure 6A). The scatter plots showing the performance of the KNN ($R^2 = 0.77$ and RMSE = 3.86) and M5Rules ($R^2 = 0.70$ and RMSE = 3.76) models reveal reduced capacity when compared to RF. ANN was the least appropriate model to deal with estimation of N from NIR reflectance, expressing the largest error.
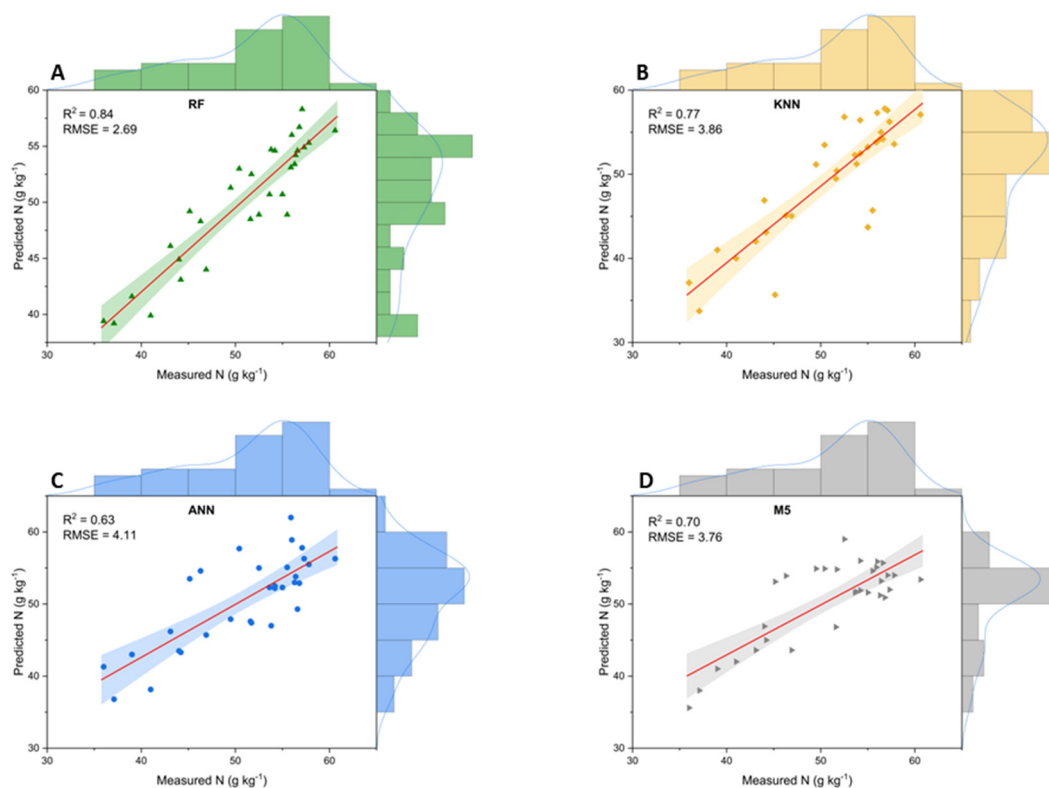

Figure 6. Performance of machine learning models in predicting N using raw NIR reflectance data in the spectral range of 700 to 1300 nm. (**A**)—Random Forest (RF); (**B**)—K-nearest neighbors (KNN); (**C**)—Artificial Neural Network (ANN); (**D**)—M5Rules (M5).

These results reflect the robustness of the RF model applied to the prediction of leaf N in the bean crop. One explanation for this is the fact that this algorithm is composed of multiple trees trained through bagging and a random variable selection process, having excellent capability against noise and outliers in the database [9,21]. RF demonstrates aptitude for data with nonlinearity inherent in the relationship between spectral variables and biophysical or biochemical parameters. Applied to predictions of nitrogen and leaf chlorophyll in maize crops, Random Forest outperformed the ANN, M5Rules, decision trees (REPT), support vector machine (SVM), and ZeroR learning algorithms (ZR), using hyperspectral data as input [9]. Evaluating models based on in situ hyperspectral data to predict nitrogen concentration in three legumes (soybean, teparian bean, moth bean) with

four machine learning algorithms, Flynn et al. (2023) found the superiority of RF ($R^2 = 0.72$) compared to KNN, PLS, and SVM [46].

On the other hand, the ANN model has a high capacity for nonlinear approximation and excellent generalization [47]. However, in this study, ANN did not obtain satisfactory performance when compared to the other models. This fact may have occurred due to the need for successive modifications in the hyperparameters of the network for optimization in the prediction [48].

The selection of the most significant wavelengths by variable importance in projection (VIP) for N prediction, located in the spectral ranges between 700–740 nm and 983–995 nm (Figure 5), increased the coefficient of determination of the RF, KNN, and M5Rules algorithms to 0.89, 0.80, and 0.76, respectively, improving the predictive capacity (Figure 7). On the other hand, RNA's performance was reduced by 3% when the two spectral intervals obtained by variable importance in projection (VIP) were used.
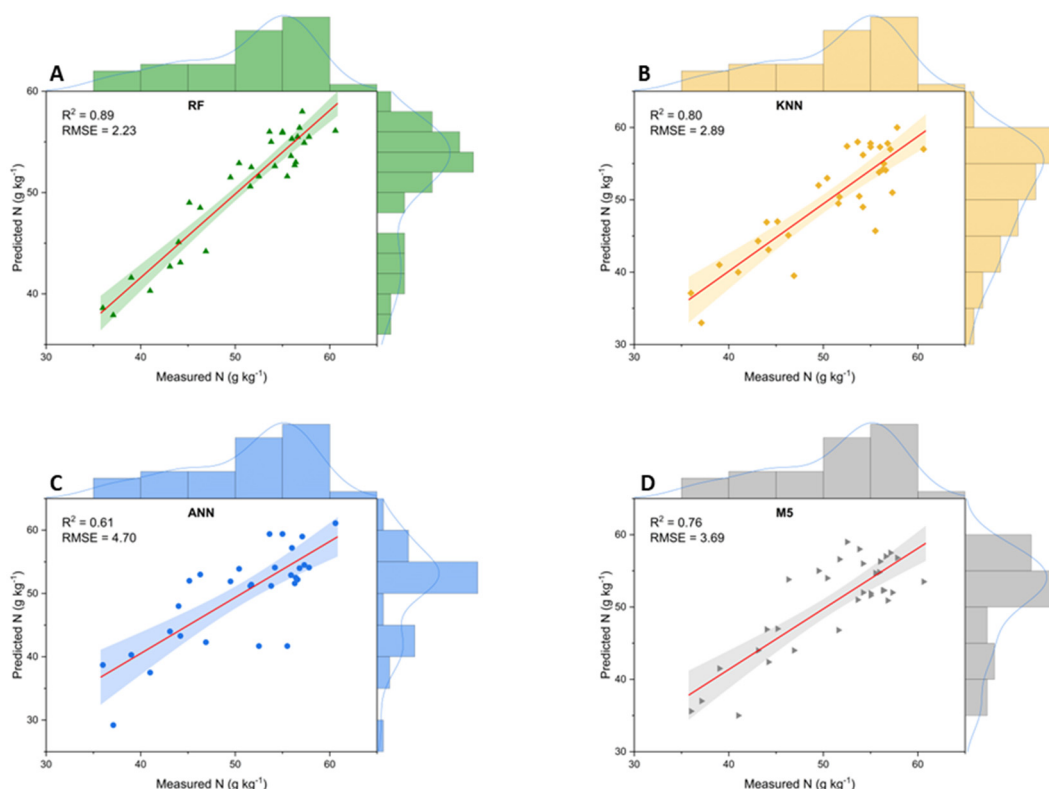

Figure 7. Performance of machine learning models in estimating N using only the two most effective spectral regions (700–740 nm and 983–995 nm) for N prediction selected by (**A**)—Random Forest (RF); (**B**)—K-nearest neighbors (KNN); (**C**)—Artificial Neural Network (ANN); (**D**)—M5Rules (M5).

These results are similar to those obtained in the literature. Fiorio et al. (2024) obtained more efficient predictions of leaf nitrogen content in sugarcane from hyperspectral reflectance data using PLSR, considering only spectral ranges obtained by variable importance in projection (VIP) [12]. The research conducted by Azadnia et al. (2023) studied the prediction of N, phosphorus (P), and potassium (K) in apple trees with spectroscopy data and also proved the performance gain of machine learning algorithms as a function of the choice of the most effective intervals using variable importance in projection (VIP) [14].

Comparing the performance results of KNN dealing with raw spectral data and only with the wavelengths selected by variable importance in projection (VIP), a 25% reduction in error is observed, from 3.86 g kg^{-1} to 2.89 g kg^{-1}. This considerable difference may have occurred because the KNN algorithm is more sensitive to data quality than

other algorithms, and the choice of more relevant variables increases its capacity for generalization, interpretability, and computational efficiency [49,50].

4. Conclusions

Random Forest was the algorithm with the highest performance in the prediction of leaf nitrogen in the common bean crop when compared to KNN, M5Rules, and ANN, considering the range from 700 to 1300 nm. On the other hand, ANN had the worst performance compared to the other algorithms. Two spectral bands were more effective for prediction of leaf nitrogen from the reflectance obtained in the NIR; the first is 700–740 nm and the second is 983–995 nm. The use of the two spectral bands selected by variable importance in the projection (VIP) resulted in more efficient predictions, increasing the performance of the RF, KNN, and M5Rules models. The results of this work prove the efficiency of the prediction of the N content in the leaves of the bean crop based on NIR reflectance data combined with machine learning algorithms. This approach can optimize nitrogen fertilization management, serving as an important tool in the context of precision agriculture. Finally, additional studies under field conditions, including with other agricultural crops, are recommended to consolidate the use of machine learning in the prediction of N considering spectroscopy.

Author Contributions: Conceptualization, M.S.T. and C.A.A.C.S.; methodology, M.S.T., E.J.d.S.S. and T.L.d.S.; software, J.R.R., T.L.d.S., M.S.T. and E.J.d.S.S.; validation, M.S.T., J.R.R., T.L.d.S. and M.M.B.; formal analysis, P.R.F., M.S.T. and C.A.A.C.S.; resources, J.R.R.; data curation, T.L.d.S., M.S.T. and E.J.d.S.S.; writing—original draft preparation., J.R.R. and M.S.T. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The raw data supporting the conclusions of this article may be made available by the authors upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Figure A1. Note. Bean plants with size uniformity prior to collection for spectral reading.

References

1. Antolin, L.A.S.; Heinemann, A.B.; Marin, F.R. Impact Assessment of Common Bean Availability in Brazil under Climate Change Scenarios. *Agric. Syst.* **2021**, *191*, 103174. [CrossRef]
2. Heinemann, A.B.; Costa-Neto, G.; Fritsche-Neto, R.; da Matta, D.H.; Fernandes, I.K. Enviromic Prediction Is Useful to Define the Limits of Climate Adaptation: A Case Study of Common Bean in Brazil. *Field Crops Res.* **2022**, *286*, 108628. [CrossRef]
3. FAO. *FAOSTAT Statistical Database*; Food and Agriculture Organisation of the United Nations: Rome, Italy, 2020.
4. Shumi, D. Response of Common Bean (*Phaseolus vulgaris* L.) Varieties to Rates of Blended NPS Fertilizer in Adola District, Southern Ethiopia. *Afr. J. Plant Sci.* **2018**, *12*, 164–179. [CrossRef]
5. Araujo Robusti, E.; Godoy Androcioli, H.; Ventura, M.U.; Hata, F.T.; Soares Júnior, D.; Menezes Júnior, A.d.O. Integrated Pest Management versus Conventional System in the Common Bean Crop in Brazil: Insecticide Reduction and Financial Maximization. *Int. J. Pest Manag.* **2023**, 1–11. [CrossRef]
6. da Silva Borges, M.P.; Trezzi, M.M.; Mendes, K.F.; Fuzinatto, E.; Pilatti, G.; da Silva, A.A. Tolerance of Brazilian Bean Cultivars to S-Metolachlor and Poaceae Weed Control in Two Agricultural Soils. *Agronomy* **2023**, *13*, 2919. [CrossRef]
7. Xie, K.; Ren, Y.; Chen, A.; Yang, C.; Zheng, Q.; Chen, J.; Wang, D.; Li, Y.; Hu, S.; Xu, G. Plant Nitrogen Nutrition: The Roles of Arbuscular Mycorrhizal Fungi. *J. Plant Physiol.* **2022**, *269*, 153591. [CrossRef] [PubMed]
8. Li, R.; Jia, X.; Hu, M.; Zhou, M.; Li, D.; Liu, W.; Wang, R.; Zhang, J.; Xie, C.; Liu, L.; et al. An Effective Data Augmentation Strategy for CNN-Based Pest Localization and Recognition in the Field. *IEEE Access* **2019**, *7*, 160274–160283. [CrossRef]
9. Silva, B.C.d.; Prado, R.d.M.; Baio, F.H.R.; Campos, C.N.S.; Teodoro, L.P.R.; Teodoro, P.E.; Santana, D.C.; Fernandes, T.F.S.; Silva Junior, C.A.d.; Loureiro, E.d.S. New Approach for Predicting Nitrogen and Pigments in Maize from Hyperspectral Data and Machine Learning Models. *Remote Sens. Appl.* **2024**, *33*, 101110. [CrossRef]
10. Fu, Y.; Yang, G.; Pu, R.; Li, Z.; Li, H.; Xu, X.; Song, X.; Yang, X.; Zhao, C. An Overview of Crop Nitrogen Status Assessment Using Hyperspectral Remote Sensing: Current Status and Perspectives. *Eur. J. Agron.* **2021**, *124*, 126241. [CrossRef]
11. Acosta, M.; Quiñones, A.; Munera, S.; de Paz, J.M.; Blasco, J. Rapid Prediction of Nutrient Concentration in Citrus Leaves Using Vis-NIR Spectroscopy. *Sensors* **2023**, *23*, 6530. [CrossRef]
12. Fiorio, P.R.; Silva, C.A.A.C.; Rizzo, R.; Demattê, J.A.M.; dos Santos Luciano, A.C.; da Silva, M.A. Prediction of Leaf Nitrogen in Sugarcane (*Saccharum* spp.) by Vis-NIR-SWIR Spectroradiometry. *Heliyon* **2024**, *10*, e26819. [CrossRef]
13. Sanaeifar, A.; Yang, C.; de la Guardia, M.; Zhang, W.; Li, X.; He, Y. Proximal Hyperspectral Sensing of Abiotic Stresses in Plants. *Sci. Total Environ.* **2023**, *861*, 160652. [CrossRef]
14. Azadnia, R.; Rajabipour, A.; Jamshidi, B.; Omid, M. New Approach for Rapid Estimation of Leaf Nitrogen, Phosphorus, and Potassium Contents in Apple-Trees Using Vis/NIR Spectroscopy Based on Wavelength Selection Coupled with Machine Learning. *Comput. Electron. Agric.* **2023**, *207*, 107746. [CrossRef]
15. Amaral, J.B.C.; Lopes, F.B.; Magalhães, A.C.M.d.; Kujawa, S.; Taniguchi, C.A.K.; Teixeira, A.d.S.; Lacerda, C.F.d.; Queiroz, T.R.G.; Andrade, E.M.d.; Araújo, I.C.d.S.; et al. Quantifying Nutrient Content in the Leaves of Cowpea Using Remote Sensing. *Appl. Sci.* **2022**, *12*, 458. [CrossRef]
16. Ji, F.; Li, F.; Hao, D.; Shiklomanov, A.N.; Yang, X.; Townsend, P.A.; Dashti, H.; Nakaji, T.; Kovach, K.R.; Liu, H.; et al. Unveiling the Transferability of PLSR Models for Leaf Trait Estimation: Lessons from a Comprehensive Analysis with a Novel Global Dataset. *New Phytol.* **2024**, *16*, 243. [CrossRef] [PubMed]
17. Chen, B.; Lu, X.; Yu, S.; Gu, S.; Huang, G.; Guo, X.; Zhao, C. The Application of Machine Learning Models Based on Leaf Spectral Reflectance for Estimating the Nitrogen Nutrient Index in Maize. *Agriculture* **2022**, *12*, 1839. [CrossRef]
18. Mustaqimah; Devianti; Munawar, A.A.; Sufardi, S. Capability of Short Vis-NIR Band Tandem with Machine Learning to Rapidly Predict NPK Content in Tropical Farmland: A Case Study of Aceh Province Agricultural Soil Dry Land, Indonesia. *Case Stud. Chem. Environ. Eng.* **2024**, *9*, 100711. [CrossRef]
19. Osco, L.P.; Ramos, A.P.M.; Faita Pinheiro, M.M.; Moriya, É.A.S.; Imai, N.N.; Estrabis, N.; Ianczyk, F.; Araújo, F.F.d.; Liesenberg, V.; Jorge, L.A.d.C.; et al. A Machine Learning Framework to Predict Nutrient Content in Valencia-Orange Leaf Hyperspectral Measurements. *Remote Sens.* **2020**, *12*, 906. [CrossRef]
20. Barcala, V.; Rozemeijer, J.; Ouwerkerk, K.; Gerner, L.; Osté, L. Value and Limitations of Machine Learning in High-Frequency Nutrient Data for Gap-Filling, Forecasting, and Transport Process Interpretation. *Environ. Monit. Assess* **2023**, *195*, 892. [CrossRef] [PubMed]
21. Li, D.; Hu, Q.; Ruan, S.; Liu, J.; Zhang, J.; Hu, C.; Liu, Y.; Dian, Y.; Zhou, J. Utilizing Hyperspectral Reflectance and Machine Learning Algorithms for Non-Destructive Estimation of Chlorophyll Content in Citrus Leaves. *Remote Sens.* **2023**, *15*, 4934. [CrossRef]
22. Alvares, C.A.; Stape, J.L.; Sentelhas, P.C.; de Moraes Gonçalves, J.L.; Sparovek, G. Köppen's Climate Classification Map for Brazil. *Meteorol. Z.* **2013**, *22*, 711–728. [CrossRef] [PubMed]
23. Cantarella, H.; Quaggio, J.A.; Júnior, D.M.; Boaretto, R.M.; Raij, B.v. *Boletim 100: Recomendações de Adubação e Calagem Para o Estado de São Paulo*; Elsiver: Singapore, 2022.
24. Patel, M.K.; Padarian, J.; Western, A.W.; Fitzgerald, G.J.; McBratney, A.B.; Perry, E.M.; Suter, H.; Ryu, D. Retrieving Canopy Nitrogen Concentration and Aboveground Biomass with Deep Learning for Ryegrass and Barley: Comparing Models and Determining Waveband Contribution. *Field Crops Res.* **2023**, *294*, 108859. [CrossRef]

25. Falcioni, R.; Oliveira, R.B.d.; Chicati, M.L.; Antunes, W.C.; Demattê, J.A.M.; Nanni, M.R. Estimation of Biochemical Compounds in Tradescantia Leaves Using VIS-NIR-SWIR Hyperspectral and Chlorophyll a Fluorescence Sensors. *Remote Sens.* **2024**, *16*, 1910. [CrossRef]

26. Miao, X.; Miao, Y.; Liu, Y.; Tao, S.; Zheng, H.; Wang, J.; Wang, W.; Tang, Q. Measurement of Nitrogen Content in Rice Plant Using near Infrared Spectroscopy Combined with Different PLS Algorithms. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2023**, *284*, 121733. [CrossRef] [PubMed]

27. Chen, Y.; Chen, M.; Zhang, S.; Ma, H.; Wang, J.; Lu, H.; Wu, Y. Rapid Determination of Geniposide in the Extraction and Concentration Processes of Lanqin Oral Solution by Near-Infrared Spectroscopy Coupled with Chemometric Algorithms. *Vib. Spectrosc.* **2020**, *107*, 103023. [CrossRef]

28. Lynch, J.M.; Barbano, D.M. Kjeldahl Nitrogen Analysis as a Reference Method for Protein Determination in Dairy Products. *J. AOAC Int.* **1999**, *82*, 1389–1398. [CrossRef] [PubMed]

29. Bakthavatchalam, K.; Karthik, B.; Thiruvengadam, V.; Muthal, S.; Jose, D.; Kotecha, K.; Varadarajan, V. IoT Framework for Measurement and Precision Agriculture: Predicting the Crop Using Machine Learning Algorithms. *Technologies* **2022**, *10*, 13. [CrossRef]

30. Shi, M.; Hu, W.; Li, M.; Zhang, J.; Song, X.; Sun, W. Ensemble Regression Based on Polynomial Regression-Based Decision Tree and Its Application in the in-Situ Data of Tunnel Boring Machine. *Mech. Syst. Signal Process.* **2023**, *188*, 110022. [CrossRef]

31. Sha'abani, M.N.A.H.; Fuad, N.; Jamal, N.; Ismail, M.F. *KNN and SVM Classification for EEG: A Review*; Elisiver: Singapore, 2020; pp. 555–565. [CrossRef]

32. Roopashree, S.; Anitha, J.; Mahesh, T.R.; Vinoth Kumar, V.; Viriyasitavat, W.; Kaur, A. An IoT Based Authentication System for Therapeutic Herbs Measured by Local Descriptors Using Machine Learning Approach. *Measurement* **2022**, *200*, 111484. [CrossRef]

33. Wang, Y.; Ian, H. Induzindo Árvores Modelo Para Classes Contínuas. In Proceedings of the Induzindo Árvores Modelo para Classes Contínuas. In Proceedings of the 9th Conferência Europeia sobre Aprendizado de Máquina, Prague, Czech Republic, 23–25 April 1997.

34. Thai, T.H.; Omari, R.A.; Barkusky, D.; Bellingrath-Kimura, S.D. Statistical Analysis versus the M5P Machine Learning Algorithm to Analyze the Yield of Winter Wheat in a Long-Term Fertilizer Experiment. *Agronomy* **2020**, *10*, 1779. [CrossRef]

35. Afzal, S.; Ziapour, B.M.; Shokri, A.; Shakibi, H.; Sobhani, B. Building Energy Consumption Prediction Using Multilayer Perceptron Neural Network-Assisted Models; Comparison of Different Optimization Algorithms. *Energy* **2023**, *282*, 128446. [CrossRef]

36. Harsányi, E.; Bashir, B.; Arshad, S.; Ocwa, A.; Vad, A.; Alsalman, A.; Bácskai, I.; Rátonyi, T.; Hijazi, O.; Széles, A.; et al. Data Mining and Machine Learning Algorithms for Optimizing Maize Yield Forecasting in Central Europe. *Agronomy* **2023**, *13*, 1297. [CrossRef]

37. Zovko, M.; Žibrat, U.; Knapič, M.; Kovačić, M.B.; Romić, D. Hyperspectral Remote Sensing of Grapevine Drought Stress. *Precis. Agric.* **2019**, *20*, 335–347. [CrossRef]

38. Nunes, H.D.; Leal, F.T.; Mingotte, F.L.C.; Damião, V.D.; Junior, P.A.C.; Lemos, L.B. Agronomic Performance, Quality and Nitrogen Use Efficiency by Common Bean Cultivars. *J. Plant Nutr.* **2021**, *44*, 995–1009. [CrossRef]

39. Falcioni, R.; Moriwaki, T.; Pattaro, M.; Herrig Furlanetto, R.; Nanni, M.R.; Camargos Antunes, W. High Resolution Leaf Spectral Signature as a Tool for Foliar Pigment Estimation Displaying Potential for Species Differentiation. *J. Plant Physiol.* **2020**, *249*, 153161. [CrossRef]

40. Liu, L.; Zhang, S.; Zhang, B. Evaluation of Hyperspectral Indices for Retrieval of Canopy Equivalent Water Thickness and Gravimetric Water Content. *Int. J. Remote Sens.* **2016**, *37*, 3384–3399. [CrossRef]

41. Machado, M.L.; Pinto, F.d.A.C.; Paula Junior, T.J.d.; Queiroz, D.M.d.; Cerqueira, O.d.A.T. White Mold Detection in Common Beans through Leaf Reflectance Spectroscopy. *Eng. Agrícola* **2015**, *35*, 1117–1126. [CrossRef]

42. Liang, L.; Qin, Z.; Zhao, S.; Di, L.; Zhang, C.; Deng, M.; Lin, H.; Zhang, L.; Wang, L.; Liu, Z. Estimating Crop Chlorophyll Content with Hyperspectral Vegetation Indices and the Hybrid Inversion Method. *Int. J. Remote Sens.* **2016**, *37*, 2923–2949. [CrossRef]

43. Crusiol, L.G.T.; Sun, L.; Sun, Z.; Chen, R.; Wu, Y.; Ma, J.; Song, C. In-Season Monitoring of Maize Leaf Water Content Using Ground-Based and UAV-Based Hyperspectral Data. *Sustainability* **2022**, *14*, 9039. [CrossRef]

44. Falcioni, R.; Antunes, W.C.; Demattê, J.A.M.; Nanni, M.R. Biophysical, Biochemical, and Photochemical Analyses Using Reflectance Hyperspectroscopy and Chlorophyll a Fluorescence Kinetics in Variegated Leaves. *Biology* **2023**, *12*, 704. [CrossRef]

45. Maimaitijiang, M.; Sagan, V.; Sidike, P.; Daloye, A.M.; Erkbol, H.; Fritschi, F.B. Crop Monitoring Using Satellite/UAV Data Fusion and Machine Learning. *Remote Sens.* **2020**, *12*, 1357. [CrossRef]

46. Flynn, K.C.; Baath, G.; Lee, T.O.; Gowda, P.; Northup, B. Hyperspectral Reflectance and Machine Learning to Monitor Legume Biomass and Nitrogen Accumulation. *Comput. Electron. Agric.* **2023**, *211*, 107991. [CrossRef]

47. Khan, M.; Ullah, Z.; Mašek, O.; Raza Naqvi, S.; Nouman Aslam Khan, M. Artificial Neural Networks for the Prediction of Biochar Yield: A Comparative Study of Metaheuristic Algorithms. *Bioresour. Technol.* **2022**, *355*, 127215. [CrossRef] [PubMed]

48. Sreedhara, B.M.; Rao, M.; Mandal, S. Application of an Evolutionary Technique (PSO–SVM) and ANFIS in Clear-Water Scour Depth Prediction around Bridge Piers. *Neural Comput. Appl.* **2019**, *31*, 7335–7349. [CrossRef]

49. Narmilan, A.; Gonzalez, F.; Salgadoe, A.S.A.; Kumarasiri, U.W.L.M.; Weerasinghe, H.A.S.; Kulasekara, B.R. Predicting Canopy Chlorophyll Content in Sugarcane Crops Using Machine Learning Algorithms and Spectral Vegetation Indices Derived from UAV Multispectral Imagery. *Remote Sens.* **2022**, *14*, 1140. [CrossRef]
50. Yoosefzadeh-Najafabadi, M.; Earl, H.J.; Tulpan, D.; Sulik, J.; Eskandari, M. Application of Machine Learning Algorithms in Plant Breeding: Predicting Yield from Hyperspectral Reflectance in Soybean. *Front. Plant Sci.* **2021**, *11*, 624273. [CrossRef]