Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Feasibility of applying hyperspectral imaging as a rapid method for estimating potassium content across different plant species datasets

Piotr Baranowski (p.baranowski@ipan.lublin.pl)
Institute of Agrophysics PAS: Instytut Agrofizyki im Bohdana Dobrzanskiego Polskiej Akademii Nauk
https://orcid.org/0000-0003-0979-177X
Anna Siedliska
Jaromir Krzyszczak
Artur Banach
Marcin Siłuch
Piotr Bartmiński
Agnieszka Wolińska
Przemysław Tkaczyk

Research Article

Keywords: potassium, hyperspectral imaging, precision farming, machine learning, regression models

Posted Date: November 9th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2190182/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Feasibility of applying hyperspectral imaging as a rapid method for estimating potassium content across different plant species datasets

Anna Siedliska[1], Piotr Baranowski[1*], Jaromir Krzyszczak[1], Artur Banach[2], Marcin Siłuch[3], Piotr Bartmiński[3], Agnieszka Wolińska[2], Przemysław Tkaczyk[4]

[1]Institute of Agrophysics, Polish Academy of Sciences, ul. Doświadczalna 4, 20-290 Lublin, Poland

[2]The John Paul II Catholic University of Lublin, Department of Biology and Biotechnology of Microorganisms, ul. Konstantynów. 1 I, 20-708 Lublin, Poland

[3]Institute of Earth and Environmental Sciences, Maria Curie-Skłodowska University in Lublin, al. Kraśnicka 2d, 20-718 Lublin, Poland

[4] Department of Agricultural and Environmental Chemistry, University of Life Sciences in Lublin, Akademicka 15, 20-950 Lublin, Poland.

*corresponding author: Piotr Baranowski, p.baranowski@ipan.lublin.pl

Abstract

Purpose The accurate and frequent estimation of the leaf plant potassium concentration enabled by hyperspectral imaging techniques has allowed growers to optimize fertilizer applications and reduce the negative impact on the environment. In this study, we examined the feasibility of using leaf spectral data to accurately estimate the potassium content in sugar beet and celery plants.

Methods Leaf images in the visible and near infrared region (VNIR: 400–1000 nm) and short-wavelength infrared region (SWIR: 1000-2500 nm) were captured by a hyperspectral camera. The potassium content was measured by ordinary destructive laboratory methods. The correlation-based feature selection (CFS) algorithm was implemented to select important wavelengths that carried the most useful information for predicting the potassium content in plant leaves. Four multivariate regression methods were tested to find a model with strong predictive performance.

Results The experimental results showed that the Random Forest (RF) model using 12 bands (425, 443, 479, 599, 631, 662, 798, 863, 897, 921, 1978 and 2053 nm) had the highest accuracy for predicting potassium content in sugar beet, celery and both plant datasets ($R_p^2 = 0.85$, $R_p^2 = 0.79$, and $R_p^2 = 0.81$, respectively).

Conclusion The results confirm the universality of the described method. Although further validation studies involving other plant species are needed, it appears that the spectral reflectance technique could be a promising

tool for the rapid, noninvasive and cost-effective estimation of K content in plant leaves, contributing to a significant step forward in precision fertilization management.

Keywords: potassium; hyperspectral imaging; precision farming; machine learning; regression models

Introduction

Potassium (K) is a macronutrient that is crucial for plant growth and development. It plays an important role in many physiological and biochemical processes, such as photosynthesis, osmoregulation, the transportation of metabolites, the synthesis of enzymes, proteins, starches, and cellulose, and plant stress alleviation (Ahmad and Maathuis 2014; Chen et al. 2021a; Lu et al. 2021). Potassium deficiency causes restrained development of the rooting system, slow growth, low resistance to stress and diseases, delayed maturity, and reduced yields (Hafsi et al. 2014; Hasanuzzaman et al. 2018). Although soil has rich K reserves, only 1-2% is available to plants (Zhang and Kong 2014). Therefore, the appropriate supplementation of potassium to plants during their crucial stages of growth is necessary to promote plant growth and increase crop yield. Excess doses and ineffective utilization of K fertilizers may lead to pollution of the environment and underground water. Therefore, knowledge of spatial and temporal variability is necessary to improve crop yield production while reducing operation costs and environmental pollution. The current methods available for the diagnosis of crop K deficiency are mainly based on visual observation followed by chemical diagnosis (de Bang et al. 2021), which is destructive, tedious and neither economically viable on a large scale nor environmentally acceptable. Precision farming, which combines sensors, robots, GPS, mapping tools and data analytics software to improve crop quality and farm profitability while reducing the negative impact on the environment, needs rapid, accurate and timely determination of crop K status for large fields. Hyperspectral imaging (HSI) offers rapid and nondestructive monitoring of plant growth conditions (Asaari et al. 2019), soil parameters (Shen et al. 2020; Lin et al. 2020) and agricultural product quality (Siedliska et al. 2018), so it can be successfully used for precise agriculture.

Studies have shown that the HSI technique can not only be successfully used to estimate nutrient element contents, especially nitrogen, in plants (Li et al. 2021; Pancorbo et al. 2021; Song et al. 2021; Guo et al. 2021; Wan et al. 2022) but also effectively overcome the limitations of traditional monitoring methods. Compared with the prediction of N content, there are a limited number of reports that have attempted to estimate the potassium content in plants.

Highly accurate results were obtained for monitoring K levels in rice (Lu et al. 2019, 2021), wheat (Pimstein et al. 2011; Mahajan et al. 2014; Weksler et al. 2021), apple tree leaves (Guo et al. 2017; Chen et al.

2021b), olive (Gómez-Casero et al. 2007) and tea plants (Wang et al. 2020). Peng et al. (2020) used wavelengths from 300 to 1000 nm to quantify the K contents of tree, shrub and grass species. Hyperspectral data within the spectral range of 908–1735 nm were used to estimate the K content in tea leaves (Wang et al. 2020). These studies have focused on the use of signals from the VNIR to estimate K nutritional status and have not taken advantage of numerous spectral features related to leaf biochemistry, which can be observed in the SWIR spectrum. A determination coefficient of 0.89 has been reached for the prediction of K concentrations in barley leaves based on hyperspectral data in the range of 1,000 to 2,500 nm (Grieco et al. 2022). The same algorithm was used to quantify the K content in maize and soybean leaves with an accuracy of 83% (Pandey et al. 2017)

Although many studies have attempted to estimate K content with reflectance spectra, no studies have explored the quantitative models of K nutrition in vegetable crops and their transferability across different plant species. The objectives of this paper were (1) to analyze the relationship between leaf spectral reflectance and K content in plants and (2) to develop a model based on hyperspectral data for estimating plant K nutritional status across different plant species.

Materials and methods

Plant materials

Here, 108 celery plants (*Apium graveolens* L., cv. Neon) and 108 sugar beet plants (*Beta vulgaris* L., cv. Tapir) were used for the experiment. The celery (produced by SEMO) and sugar beet (produced by SESVanderHave) seeds were purchased commercially. The celery and sugar beet seeds were sown in plastic pots containing peat moss. After 7 weeks, seedlings of similar sizes were transplanted to 20 cm-diameter pots (one seedling per pot) containing sand as the growth media. The plants were grown in the greenhouse under natural sunlight supplemented with LED light using a day/night photoperiod of 12/12 h, with temperatures ranging from 20 °C to 22 °C during March – June and September and from 24 °C to 26 °C during July – August. Since the ultimate goal of regression analysis was to predict the amount of K in plant leaves across different plant species, we wanted to record a wide range of measured K content values. Therefore, in this experiment, the plants from each species were divided equally into four groups with 27 plants each and were subjected to four different K rates, where K was applied to the pots to stimulate different nutrient levels in plant leaves to test the hyperspectral imaging system. Plants were irrigated with a treatment solution containing various quantities of potassium sulfate (K_2SO_4) to obtain four different fertilization doses, K33, K67, K100, and K133. The K concentration of the control treatment solution (K100) was 164.8 mg L^{-1} (Inthichack et al., 2012). Other treatments included 1/3, 2/3, and 4/3 of this value. To provide a source of nitrogen and phosphorus, the treatment solution contained 33.3 mg/l of N and 13.3 mg/l of P applied as NH_4NO_3 and $Ca(H_2PO_4)_2$, respectively. The concentrations of the micronutrients B, Fe, Zn and Mo were 0.28, 2.4, 1.0, 0.35 and 0.05 mg/l, respectively. They were applied as a commercial fertilizer, Micro Plus (produced by Intermag, Olkusz, Poland). Each plant was irrigated with treatment solution every two days for 60 days. Data collection was performed three times (21, 31 and 51 days after transplanting) at differing stages of plant development. The developmental stages of sugar beet correspond to leaf development (BBCH 14) and rosette growth (BBCH 28 and 35). For celery, we recorded data at BBCH stages 14, 42 and 45. In these stages, potassium supplementation is expected to have the greatest effect on plant growth and yield (Prośba-Białczyk., 2017). At each stage, the leaf gas exchange of nine samples from each experiment was measured. After that, the plants were scanned using hyperspectral imaging and transferred to the laboratory for chemical analysis.

Chemical analysis of macronutrient content

The potassium content of plant leaves was measured using a standard chemical procedure. Fresh leaves were cut and oven-dried (105 °C for 0.5 h followed by 80 °C) until a constant weight was attained. Dried leaves (ca. 0.3 g) were digested in triplicate in a mixture of 65% HNO_3 and 30% H_2O_2 (6:2) using ultra-pure reagents. Digestion was conducted using a microwave Ethos One mineralizer (Milestone Srl., Italy). The obtained solutions were diluted to 100 ml and analyzed for K using a Hitachi Z-8200 spectrophotometer (Hitachi, Japan). The results were expressed as mg/g dry weight. The total leaf phosphorous content was determined colorimetrically (AutoAnalyser AA3, BranLuebbe) according to the manufacturer's instructions: Method No. G-189-97 Rev.3. Before analysis, the samples were digested in 8 ml of concentrated sulfuric acid (VI) using an Ethos One (Milestone) mineralizer. The total leaf nitrogen content was determined by a LECO TruSpec analyzer using a dry combustion method at a temperature of 950 °C.

Leaf gas exchange

Leaf gas exchange, such as the photosynthetic rate and transpiration rate, was measured using an LCi-SD Portable Photosynthesis System (ADC Bio Scientific Ltd., UK). The measurements were carried out between 10:00 and 12:00 on the first fully expanded leaves from nine plants for each treatment.

Hyperspectral imaging system

To extract spectral information from each leaf, two cameras were used to capture both VNIR and SWIR images. The cameras were mounted 40 cm above a belt conveyor (Reall, Lublin, Poland) with the belt speed regulated for each camera separately (to conduct line scanning of the plant leaves). The illumination system Brilum (Piaseczno, Poland) model LAVADO416 has 4 lighting modules, each of which contain four 20-W halogen lamps (Philips, Netherlands) fixed in two opposite frames positioned at an angle of 45° toward the conveyor belt surface. The VNIR images covering 269 spectral bands in the range of 400 nm to 1000 nm were obtained with a spectral sampling interval of 2.8 nm using a VNIR camera with an ImSpector V10E imaging spectrograph. The SWIR camera with an N25E 2/3" imaging spectrometer (Specim, Oulu, Finland) captured a total of 224 images from each plant leaf, with each value corresponding to a specific wavelength inside the 1000–2500 nm range.

Spectral data preprocessing

The spectral data obtained from hyperspectral instruments usually contain a considerable amount of background information and noise in addition to the sample information. Spectral preprocessing enhances the absorbance features of spectra by reducing abnormalities in spectral measurements, such as noise, uncertainties, variabilities, interactions, and unrecognized features (Mishra et al. 2020). Data at both ends of the spectrum obtained from VNIR (<400 nm and >970 nm) and SWIR (<1100 nm and >2400 nm) were excluded due to increased noise most likely caused by ambient light or a low signal-to-noise ratio (Duranovich et al. 2020). First, raw spectra obtained from plant leaves were transformed using baseline correction, in which the lowest value was subtracted from all the remaining values in the spectrum. Second, the second derivative Savitzky–Golay filter (second-order polynomial fitting and nine-data-points window size) was applied. The data processing methods were implemented using Unscramble X10.3 software (CAMO Software, Inc., Oslo, 133 Norway).

Optimal wavelength selection

Selection of the wavelengths carrying the most information is necessary for reducing multicollinearity and redundancy between adjacent wavelengths, as well as for the improvement of the predictive performance of models. If the redundant wavelengths are eliminated and the optimal variables are selected, the modeling process can be simplified and the model performance improved while also facilitating online industrial applications or simple cost-effective multispectral systems. In the current study, the CFS algorithm with the greedy stepwise selection method was applied to the second derivative transformation of the original spectra. The usefulness of the CFS method for selecting the optimal feature subset was confirmed in previous studies (Siedliska et al. 2017; Guo et al. 2020).

Modeling framework

In this study, three different datasets (two datasets for each plant species and one for both plant species together) were used for the determination of the K content in plant leaves. These datasets contained varying types of input information: measured leaf K content and leaf reflectance at 400–2400 nm. To evaluate the model performance, leaf samples in each dataset were randomly divided into calibration and prediction datasets at a ratio of 75:25 using a standard procedure within an open-source computer program called the Waikato Environment for Knowledge Analysis (WEKA). This subsampling proportioning for calibration of the prediction sets was recommended and commonly used for plant samples (Li et al. 2020; Siedliska et al. 2021) To evaluate the effectiveness of machine learning approaches, four regression algorithms were selected for comparison and further understanding of the relationship between dependent and independent variables: the k-nearest neighbors algorithm (kNN), support vector regression (SVR), random forest (RF) and linear regression (LR). The chosen algorithms represent various approaches to machine learning, covering the classes of functions, decision trees and lazy regressors (used only to prepare their inputs until classification time). These algorithms were successfully used in previous studies for regression problems associated with hyperspectral data analysis (Siedliska et al. 2018; Garg et al. 2019; Wu et al. 2020; Gargade and Khandekar 2021). The features of all the algorithms used, their differences, and their advantages and disadvantages are presented in Table. 1. All the algorithm tuning steps were performed on the calibration data, and evaluations of model performance were performed on the prediction dataset. The accuracy and performance of each model were evaluated using the coefficient of determination (R^2), root mean square error of calibration (RMSEC) and prediction (RMSEP), and mean square error of calibration (MSEC) and prediction (MSEP).

Table 1 The description and main parameters of the model used for prediction of potassium content in sugar beet and celery plants

Name of classifier's library	Description of algorithm	Acronym	Chosen parameters of classifier
kNN	K-nearest neighbours classifier (also known as "Instance based" learning) is a simple classification and regression algorithm based on similarity (distance) calculation between the instances. It implements rote learning which is based on a local average calculation (Altman, 1992). **Advantages:** Simplicity and intuitiveness of use. Highly flexible decision boundary adjusting the value of K. No training time for classification. When adding new data to the dataset, the prediction is adjusted without having to retrain a new model. It's very easy to tune hyperparameter K. **Disadvantages:** Problems with large datasets (prediction complexity, long processing time). The algorithm assumes equal importance to all features. Sensitive to outliers. A single mislabeled example can change the class boundaries.	kNN	Num of neighbours: 3 Batch Size: 100; Debug: false; Distance Function: Euclidean distance;
Linear Regression	A class for using linear regression for prediction (Adichien, 1967). **Advantages:** Performs exceptionally well for linearly separable data. Easy to implement, interpret and efficient to train. Possible extrapolation beyond a specific data set. Resistant to overfitting problems. **Disadvantages:** Sensitive to outliers. Quite prone to underfitting. In many cases the assumption of the linearity between dependent and independent variables is too big simplification (multicollinearity must be removed before applying the algorithm).	LR	Attribute selection method: M5 method; Batch size:100; Debug: false; Num of Decimal Places:4
Support Vector Regression	A supervised learning algorithm that is used to predict discrete values. It uses the same principle as the Support Vector Machines (SVM). The algorithm tries to find the best fit line which is the hyperplane that has the maximum number of points (Smola and Schölkopf , 2004). **Advantages:** Excellent generalization capability. Usually, high prediction accuracy. Robust to outliers. Easy to be implemented. Decision model can be easily updated. **Disadvantages:** Not always effective with large datasets. Does not perform very well when the data set has more noise. Underperforms, if the number of features for each data point exceeds the number of training data samples.	SVR	Debug: false; Kernel function Type: Polynominal Kernel; C: 1; gamma: 0.1; Epsilon: 0.01; Normalize: true;
Random Forests	Classifier for constructing a forest of random trees (Breiman, 2001). **Advantages**: Suitable for complicated tasks because of handlling variables very fast. Very effective when working with big data with numerous variables. Contains algorithms for filling in missing values in datasets. **Disadvantages:** The computations may go far more complex compared to other algorithms (time and resources consuming).	RF	Debug: false; MaxDepth: 0; Num of Features : 0; Num of Trees: 10; Seed : 1

Results

Leaf nutrient concentrations and changes in photosynthesis parameters at different potassium fertilization doses

The average K concentrations, transpiration rates, photosynthetic rates (bars) and standard deviations (whiskers) across three stages of plant growth and for various doses of soil K supply for sugar beet and celery are presented in Fig. 1 and Fig. 2. It was observed that independent of the studied species, the highest concentrations of K in the plant leaves occurred in the first stage of plant development for all K supply doses (Fig. 1a and 2a). The K concentrations in the first development stage for all levels of K supply were up to 2 times higher than those for the two other stages of plant development. This result confirms that at the plant development stage, K accumulates mainly in the leaves, while at the leaf development stage, when root development prevails, the accumulation of K in the leaves diminishes as a result of the higher K needs for root growth. Figs. 1 (b-c) and 2 (b-c) also show that both transpiration rates and photosynthetic rates reach the highest values for the recommended K supply dose (K100) and the second stage of plant development. A shortage (K33) or excess (K133) of K caused a considerable decrease in transpiration and photosynthetic rates for both studied species at each developmental stage. These results indicated that an inappropriate K supply could limit photosynthesis through stomatal restrictions, which is consistent with previous studies (Cakmak 2005; Xu et al. 2020).

The K, P and N contents in the plant leaves under various doses of K supplied to the soil, combined with data from the entire experiment (3 stages of plant development), are shown in Fig. 3 and Fig. 4 in the form of boxplots. These plots display the distribution of the data based on a five-number summary, including minimum and maximum values (ends of the whiskers), the first and third quartiles (2 horizontal lines closing boxes), the median (the bold lines inside the boxes) and the outliers (individual points in the plots). As expected, the medians of the leaf potassium content increased gradually with increasing doses of potassium for both studied species, with the K133 level in sugar beet being the only exception (Fig. 3a and Fig. 4a). The lowest variations in leaf K content, expressed as the lowest range between the minimum and maximum values, were observed for K33, both in celery and sugar beet, and the highest variation was observed for K133. The range of K contents between the first and third quartiles is the lowest in both studied species for K33, which confirms that at this level of K concentration in the soil, the accumulation in the plant was limited and did not change across plant growth stages. This suggests that during plant growth, the content of K in the leaves changes considerably depending on the availability of this macroelement in the soil.

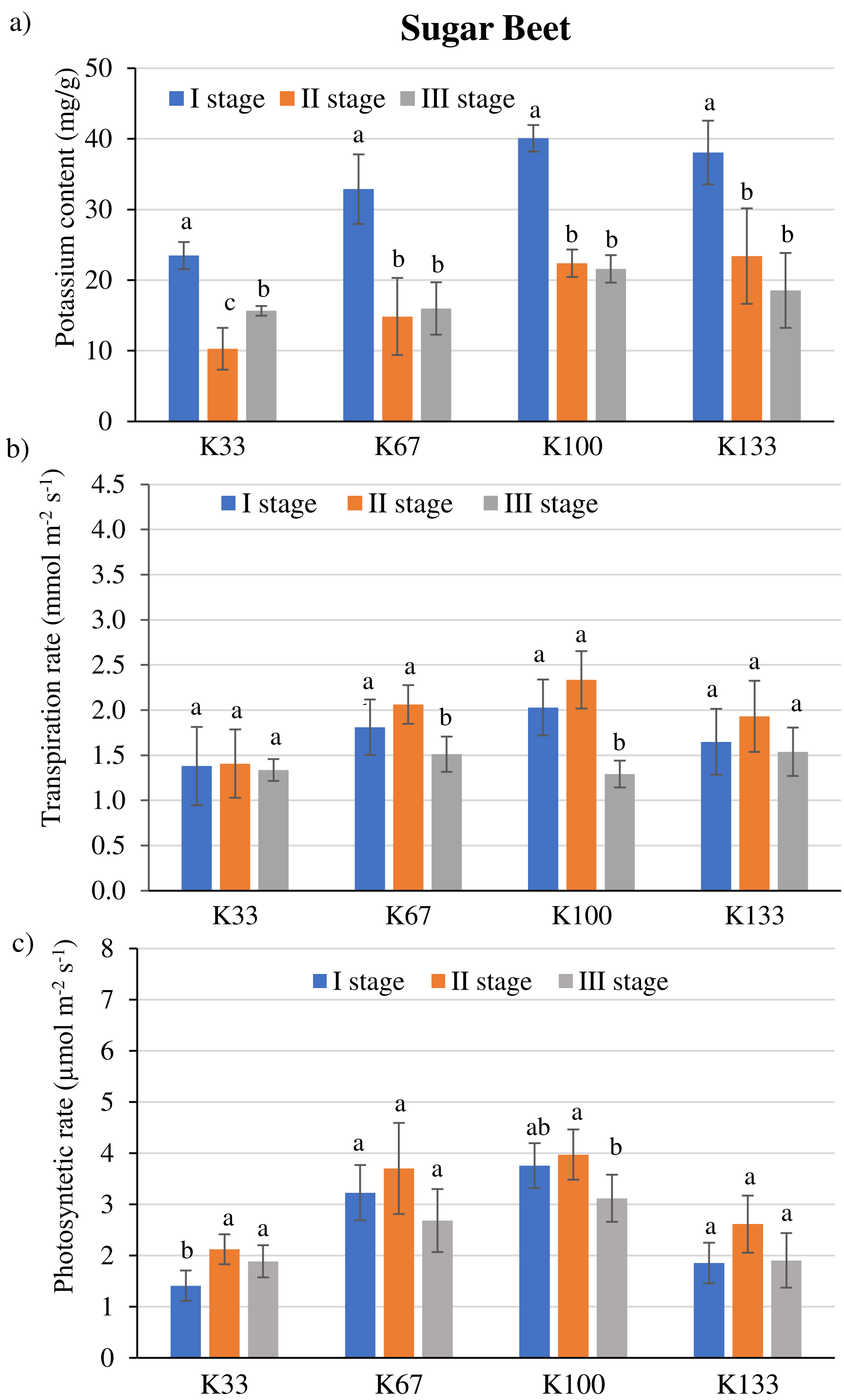

Fig. 1 Bar charts of leaf potassium content (a), transpiration rate (b) and photosynthetic rate (c) in response to different potassium fertilization doses on sugar beet. Values followed by different letters in the same lines are significantly different according to Duncan's multiple range test ($p < 0.05$)

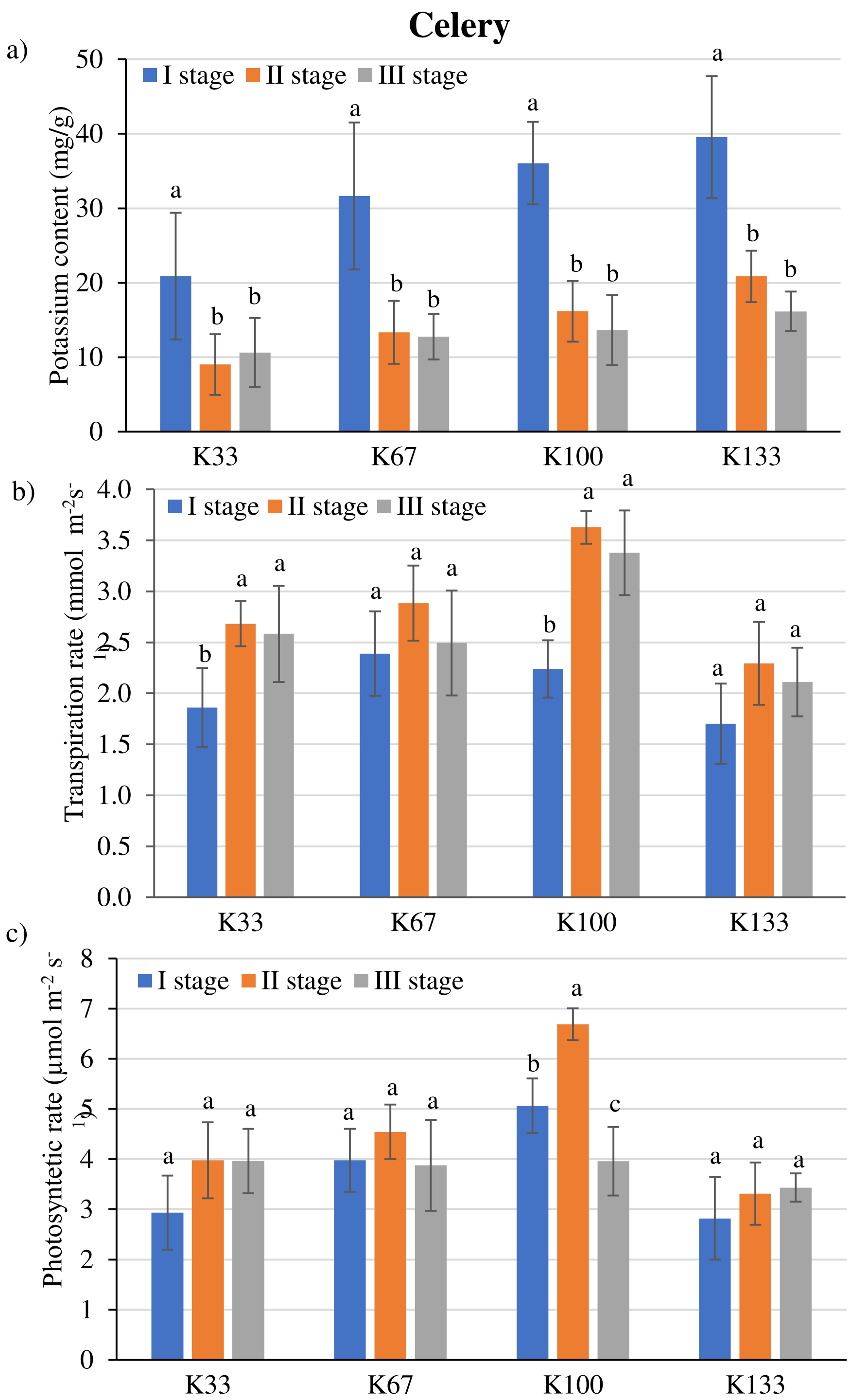

Fig. 2 Bar charts of leaf potassium content (a), transpiration rate (b) and photosynthetic rate (c) in response to different potassium fertilization doses on celery. Values followed by different letters in the same lines are significantly different according to Duncan's multiple range test ($p < 0.05$)

K fertilization rates had no significant effect on the P content in sugar beet foliage (Fig. 3b). This result is in agreement with a study by Abdel-Motagally et al. (2009). On the other hand, the P concentration in K33, K67 and K100 for celery plants slightly increased with an increase in K supply (Fig. 4b). However, a further increase in the K dose (K133) caused a decrease in the P concentration in celery leaves. The data presented in Fig. 3c and Fig. 4c revealed that the leaf nitrogen content in both plant species was not significantly affected by the K application dose.

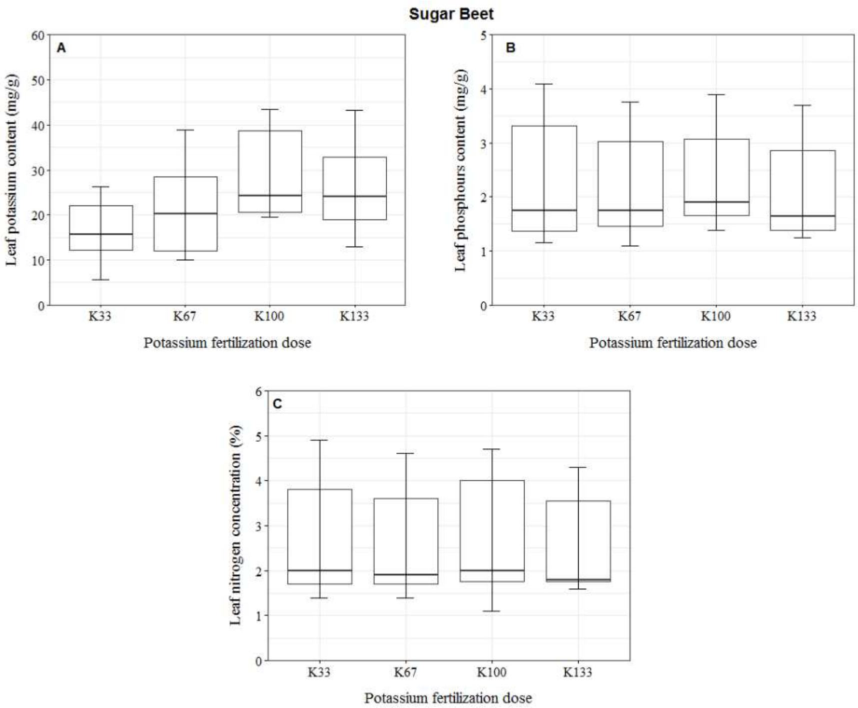

Fig. 3 Boxplots of leaf potassium (A), phosphorus (B) and nitrogen (C) contents in response to different potassium fertilization doses on sugar beet. Boxes represent interquartile range, whiskers represent minimum and maximum values, dots represent outliers, and bold lines represent median values

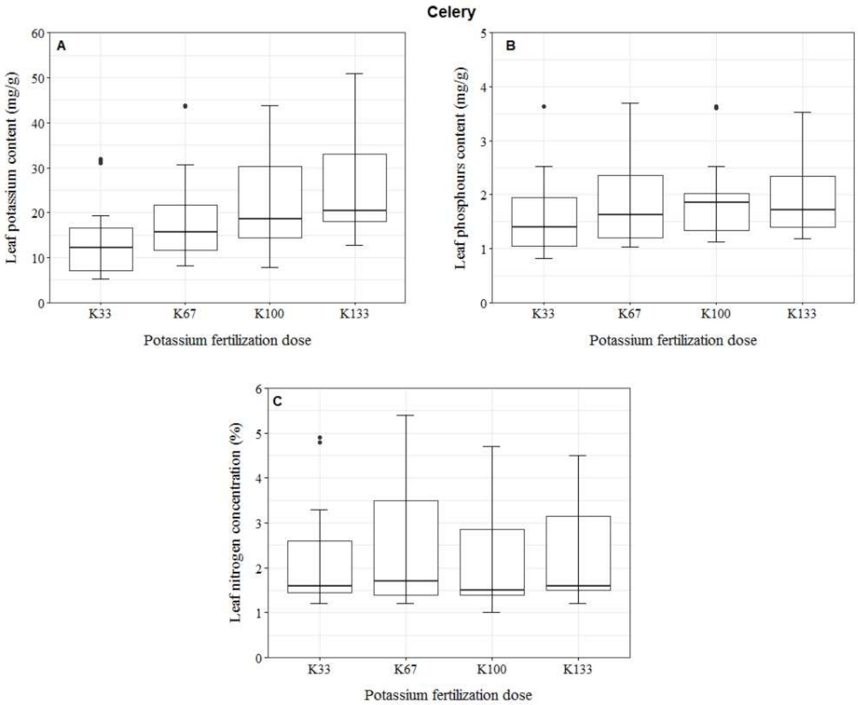

Fig. 4 Boxplots of leaf potassium (A), phosphorus (B) and nitrogen (C) contents in response to different potassium fertilization doses on celery. Boxes represent interquartile range, whiskers represent minimum and maximum values, dots represent outliers, and bold lines represent median values

Spectral features of leaves

Example raw reflectance spectra and spectra transformed with the second derivative of chosen leaves from different experiments (various K supply doses) are presented in Fig. 5. The leaf spectra showed similar patterns across all sugar beet and celery treatments. However, the K100 spectra differed significantly from those for the other treatments, especially in the NIR and SWIR ranges. In the second derivative spectra, this difference was also visible, especially for the peaks at 700, 1400, 1850, 1920, 2300 and 2350 nm. Because the main objective of the study was to estimate the K content in the plant leaves using the hyperspectral characteristics, the spectra transformed with the second derivative were used as inputs into the chosen machine learning algorithms.

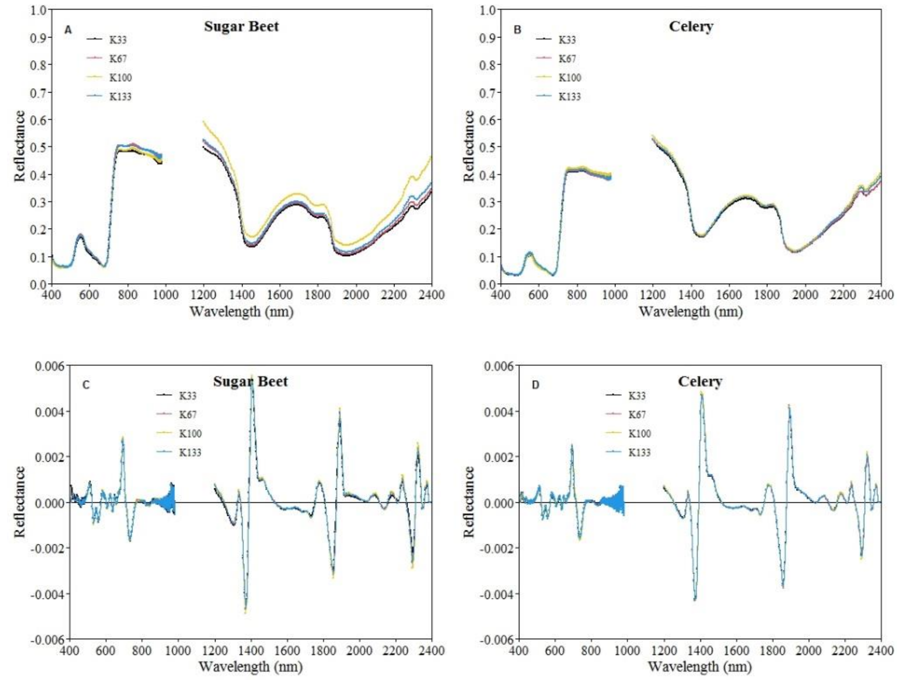

Fig. 5 Mean reflectance spectra (a, b) and their second derivatives (c, d) for sugar beet and celery leaves treated with four different potassium fertilization doses

K prediction models based on spectral characteristics

After the evaluation of the reflectance spectra of the plant leaves under different treatments, it was expected that in the band selection process for artificial intelligence processing, the spectral wavelengths from NIR/SWIR ranges would be dominant, which was partially confirmed for the spectra of the two studied species. However, after performing the automatic CFS algorithm for wavelength selection, among the twelve characteristic wavelengths selected for supervised regression analysis, there were six VNIR bands. All the bands selected for the prediction of K content were as follows: 425, 443, 479, 599, 631, 662, 798, 863, 897, 921, 1978 and 2053 nm.

The subsampling of the data for the calibration and prediction datasets yielded similar statistical properties and a wide variation in the K content in these datasets, which was especially important from the point of view of proper model creation and testing. The statistical properties of the calibration and prediction datasets are summarized in Table 2. The distributions of outcome and predictor variables on the calibration and prediction sets were similar. The effect of leaf potassium content (mg/g) in sugar beet on the calibration dataset had a median of 21.06 and a range of 5.52 to 43.45, and on the prediction dataset, its median was 23.61 with a range of 7.08-41.91. For celery leaves, the effect of the potassium leaf content on the calibration dataset had a median of 16.81 and a range of 5.24-50.69, and on the prediction dataset, its median was 19.37 with a range of 5.76 to 50.88 mg/g. This suggests that the K contents in the calibration and prediction datasets were within very similar ranges, indicating that the sample divisions were reasonable and that the prediction set could have been employed to assess the performance of the developed models.

Table 2 Summary statistics of measured K content (mg/g) in sugar beet and celery leaves in training and test data sets

Plant	Sample set	N	Range	Mean	Median	SD
Sugar beet	Training	81	5.52 - 43.45	22.25	21.06	9.39
	Test	27	7.08 - 41.91	25.65	23.61	10.62
Celery	Training	81	5.23 - 50.69	19.06	16.81	10.43
	Test	27	5.76 - 50.88	23.05	19.37	13.13

N, number of samples; SD, standard deviation

The results of the supervised regression modeling of the K content in sugar beet, celery and both species datasets, performed with the use of four selected algorithms from the calibration and prediction datasets, are presented in Table 3. This table contains the main statistical parameters of the created models, including MSE, RMSE and R^2 values. The results from this table show that all studied models yield good performance with R^2 values for the calibration dataset ranging from 0.83 to 0.98 and those for the prediction set ranging from 0.68 to 0.85 (considering all the algorithms used). The best-fitted model was selected based on the highest values of R^2 and the smallest values of RMSE. From all the selected model classes, the RF models based on 12 selected wavelengths showed the best effectiveness. The RF model for the sugar beet K content gave higher prediction accuracy (R_p^2 equal to 0.85, MSEP 4.93 and RMSEP 5.77) than that for celery (R_p^2 equal to 0.79, MSEP 6.57 and RMSEP 8.27). Comparably high prediction performance of the RF model for K content (R_p^2 equal to 0.81, MSEP 5.09 and RMSEP 6.33) was obtained based on the dataset incorporating both celery and sugar beet spectral characteristics. This result indicates the possibility for universal use of the created K content models, i.e., independent of the species and stage of plant growth. This result is especially encouraging as it was not found in previous studies. Of course, the authors are conscious that further studies incorporating more vegetable species differing in spectral features of their leaves and growth physiology are urgently needed to confirm the versatility of the created models.

Table 3 Performances of the models for predicting potassium content (mg/g) in the leaves of sugar beet, celery and both plants together

Model	Calibration set			Prediction set		
	MSEC	RMSEC	R_c^2	MSEP	RMSEP	R_p^2
			Sugar beet			
kNN	3.13	4.12	0.90	5.77	7.30	0.73
SVR	3.58	5.10	0.85	5.96	7.70	0.74
RF	1.80	2.30	0.98	4.93	5.77	0.85
LR	3.86	4.99	0.85	5.76	7.45	0.72
			Celery			
kNN	3.35	4.42	0.91	6.93	8.28	0.78
SVR	3.87	5.93	0.83	7.66	10.45	0.68
RF	1.68	2.13	0.98	6.57	8.26	0.79
LR	4.23	5.51	0.85	0.71	9.74	0.71
			Sugar beet and Celery			
kNN	3.52	4.79	0.89	5.48	6.61	0.81
SVR	4.44	6.12	0.83	5.28	6.99	0.78
RF	1.84	2.34	0.98	5.09	6.33	0.81
LR	4.61	6.03	0.83	5.62	7.10	0.77

The scatter plots of the regression analysis for predicted and measured values of K content in sugar beet, celery and the combined dataset for these two plant species are shown in Fig. 6, including the results of the RF models created based on the spectral data of 12 effective wavelengths. The plots contain the data points belonging to the calibration and prediction sets, marked as circles and triangles, respectively, and statistical parameters of model performance. In general, all three created models show a good fit to the 1:1 line (dashed line in the plot), indicating that the predicted values of the models were close to the value measured by the chemical method. All three RF models were prone to slightly overestimating low values and underestimating the highest values of K content in plant leaves.

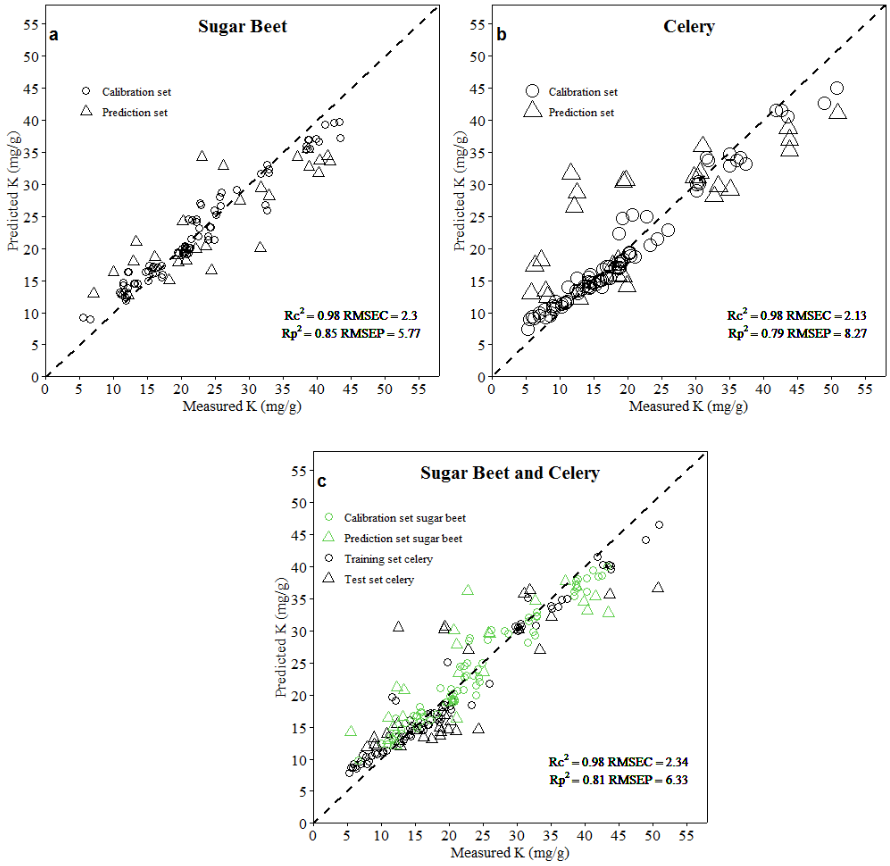

Fig. 6 The measured and predicted potassium contents in sugar beet (a), celery (b) and both plant datasets (c) obtained by the RF model based on spectral data from 12 selected wavelengths

Discussion

The presented results broaden the existing knowledge about the possibility of using hyperspectral reflectance data for evaluating the K content in the plants. The previous attempts to discover the relation between the reflectance spectra and the K content in the leaves was mainly focused on cereals (Pimstein et al. 2011; Mahajan et al. 2014; Pandey et al. 2017; Lu et al. 2021; Weksler et al. 2021; Grieco et al. 2022) and trees (Gómez-Casero et al. 2007; Guo et al. 2017; Chen et al. 2021b), while the root vegetables which are very sensitive for K deficiency (Prośba-Białczyk et al. 2017) were not considered. Our study is also an one of the few attempts to include in the supervised regression models the reflectance data including the VNIR/SWIR ranges. Among the wavelengths selected by the automatic CFS procedure, two were within the far SWIR range (1978 and 2053 nm) what is in agreement with the studies of Xu et al. (2020) and Rani et al. (2021), who emphasized the importance of these wavelengths in K content estimation. The absorption bands in this region are related to changes in starch (wavelengths near 1978 nm) and protein contents (2055 nm), which are highly influenced by K fertilization.

Although the automatic selection of the wavelengths for regression models was used in our study, the majority of VNIR wavelengths were close to these used by Peng et al. (2020) and Liu et al. (2020). In general the CFS algorithm is applied to remove irrelevant and redundant information from the spectra, based on the individual predictive ability of each feature and the degree of correlation between all the features (Hall 1998), however it was indicated in the previous studies that the bands selected using this algorithm strongly correlate with some physical features of the studied biological materials (Baranowski et al. 2013, Siedliska et al. 2017). K deficiency was suggested to produce an intensification of the reddish color as a consequence of increased anthocyanin content in the distal part of the plant (Rustioni et al. 2018), which explains the band selection results in the range between 520 and 650 nm. Wavelengths of approximately 443 nm and 471 nm are associated with carotenoid content, and those of approximately 662 nm are associated with chlorophyll b (Zhang et al. 2020), which has some correlation with plant potassium fertilization, as demonstrated in previous work (Xu et al. 2020).

To date, the possibility of applying the HSI technique for the determination of K status in plants has been poorly documented. Our results showed better performance of the machine learning regression model than that found in previous studies. For instance, Zhang et al. (2013) employed the hyperspectral image analysis technique and PLSR model to map the distribution of K in oilseed rape leaves and achieved the best prediction with a coefficient of determination of $R_p^2 = 0.746$. The potassium concentration of tea plant leaves has also been predicted with $R_p^2 = 0.84$ based on eight wavelengths selected from the spectral range of 908.15–1735.68 nm by using a successive projections algorithm (Wang et al. 2020). However, in this study, the effects of different fertilization conditions and growing seasons on the K concentration were not considered. The prediction of K content in the orange leaves was conducted by Osco et al. (2020) with an $R_p^2 = 0.76$ for one cultivar at a specific growth stage (vegetative stage). Another study of K content determination in rice leaves focused on the creation of three-band indices (Lu et al., 2019). In their study, prediction models with $R^2 = 0.74$ were obtained with selected bands that were similar to the ones used in our models. The red edge bands, from 660 to 800 nm, showed improved accuracies and stabilities for estimating K content. From all models studied in our research the best performance (R_p^2 0.85, 0.79, and 0.81 for sugar beet, celery and dataset incorporating both plants, respectively) was obtained for RF model. It may be because this model combines ensemble learning methods with the decision tree framework and averages the results to output a new result. This algorithm led to strong predictions in previous studies concerning problems due to complicated relationships between attributes (Breiman, 2001; Shuryak, 2022). Its practical use for K content estimation in plants is especially promising because of its ability to process datasets with many attributes related in a complicated way and capture linear and nonlinear dependencies and interactions. Moreover, this model can handle numerical as well as categorical variables, which is especially useful for analyzing data on macronutrient content in plant leaves across different plant species. The wide range of variability in the potassium content as a predictor in the elaborated models was a consequence of the experiment containing multiple treatment combinations (leaves of two species treated with four potassium fertilization doses measured at three development stages). Due to the broad range of K content values in the leaves of the studied plants, the main disadvantage of RF modeling, i.e., the inability of this approach to extrapolate data outside the calibration set, was minimized, which contributed to the best prediction accuracy of RF models. The efficiency of the three other regression algorithms used for K content estimation was slightly lower (in the prediction sets, R_p^2 changed from 0.68 to 0.81, RMSEP from 6.61 to 10.45 and MSEP from 5.28 to 7.66). From among these models, the kNN regressor gave the best prediction results for all the treatments during modeling and was not much worse than the RF algorithm. This good performance of the kNN algorithm was confirmed in previous studies (Osco et al. 2020) of biophysical materials, which emphasized the advantage of modeling based on local average calculations, which enables the description of nonlinear processes dominating the functioning of live organs. It was found in a previous study (Ragel et al. 2019) that the K^+ transition in the living organs of plants shows a steep curvilinear relationship between the concentration of K^+ in tissues and plant growth. Therefore, models based on linear correlations between variables, such as PLSR, which was used frequently in previous studies (Zhang et al., 2013), gave much worse correlations between the K content and spectral features of the leaves. Similar results were found for the LR model used in our study (its prediction accuracy was the worst among all the studied algorithms).

The high accuracy of our models can be partially explained by the broad spectral range (400-2500 nm) used for the analysis, while previous studies mostly used the VNIR spectral range (Zhang et al. 2013; Li et al. 2018; Peng et al. 2020). Furthermore, our study is the first to use HSI for the prediction of the K content across two different plant species combined into one model.

Conclusions

The proposed approach uses reflectance spectral data in the visible and near-infrared regions of leaf surfaces to predict the K content in two vegetable plant species. It was found that leaf spectral reflectance, especially in visible and near-infrared regions, was very sensitive to plant K fertilization status. The advantage of the performed analyses and the created regression models is that a broad range of K concentrations and 3 stages of plant development were considered. An important outcome of this study involves the identification of the universal wavelengths for reflectance spectra (only partially coinciding with previously reported spectra), which contributes to the prediction of the K content in different plant species that is responsible for changes in anthocyanin, carotenoid and chlorophyll b contents as well as starch and protein concentrations. The current study showed the great potential of using hyperspectral imaging to predict K content with high accuracy across different plant species. The modeling approach based on testing several different regression procedures enabled the identification of an RF algorithm that combines ensemble learning methods with the decision tree framework as the best suited model for the task of K concentration analysis in plants. This model resulted in prediction accuracies reaching 85%, which seems very promising given that no earlier attempts to evaluate the K content in celery and sugar beet have been reported in the literature. The measurement and analytical methods used here are a good starting point for developing a multispectral imaging system that measures K content in real time. In the future, we plan to adapt the methods and results of this study to create a system that will be able to evaluate K status in a broad range of plant species. Such a multispectral set of sensors embedded in UAVs operating under field conditions is highly needed for precision farming and would reduce labor/time consumption and costly chemical K analyses.

REFERENCES

1. Abdel-Motagally FM, Attia K K (2009) Response of sugar beet plants to nitrogen and potassium fertilization in sandy calcareous soil. Int. J. Agric. Biol, 11(6), 695-700.
2. Ahmad I, Maathuis FJM (2014) Cellular and tissue distribution of potassium: Physiological relevance, mechanisms and regulation. J. Plant Physiol. 171:708–714. https://doi.org/10.1016/j.jplph.2013.10.016
3. Adichie J N (1967) Estimates of Regression Parameters Based on Rank Tests. Ann. Math. Statist. 38 (3) 894 – 904. https://doi.org/10.1214/aoms/1177698883
4. Altman NS (1992) An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression, The American Statistician, 46:3, 175-185. https://doi.org/10.1080/00031305.1992.10475879
5. Asaari MSM, Mertens S, Dhondt S, et al (2019) Analysis of hyperspectral images for detection of drought stress and recovery in maize plants in a high-throughput phenotyping platform. Computers and Electronics in Agriculture 162:749–758. https://doi.org/10.1016/j.compag.2019.05.018
6. Baranowski P., Mazurek W., Pastuszka-Woźniak J. (2013). Supervised classification of bruised apples with respect to the time after bruising on the basis of hyperspectral imaging data. Postharvest Biol. Tech., 86, 249-258.

7. Breiman L (2001) Random Forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324
8. Cakmak I (2005) The role of potassium in alleviating detrimental effects of abiotic stresses in plants. Journal of Plant Nutrition and Soil Science 168:521–530. https://doi.org/10.1002/jpln.200420485
9. Chen J, Guo Z, Chen H, et al (2021a) Effects of different potassium fertilizer types and dosages on cotton yield, soil available potassium and leaf photosynthesis. Arch. Agron. Soil Sci. 67:275–287. https://doi.org/10.1080/03650340.2020.1723005
10. Chen S, Hu T, Luo L, et al (2021b) Prediction of Nitrogen, Phosphorus, and Potassium Contents in Apple Tree Leaves Based on In-Situ Canopy Hyperspectral Reflectance Using Stacked Ensemble Extreme Learning Machine Model. J Soil Sci Plant Nutr. https://doi.org/10.1007/s42729-021-00629-3
11. de Bang TC, Husted S, Laursen KH, et al (2021) The molecular–physiological functions of mineral macronutrients and their consequences for deficiency symptoms in plants. New Phytologist 229:2446–2469. https://doi.org/10.1111/nph.17074
12. Duranovich FN, Yule IJ, Lopez-Villalobos N, et al (2020) Using Proximal Hyperspectral Sensing to Predict Herbage Nutritive Value for Dairy Farming. Agronomy 10:1826. https://doi.org/10.3390/agronomy10111826
13. Garg R, Aggarwal H, Centobelli P, Cerchione R (2019) Extracting Knowledge from Big Data for Sustainability: A Comparison of Machine Learning Techniques. Sustainability 11:6669. https://doi.org/10.3390/su11236669
14. Gargade A, Khandekar S (2021) Custard Apple Leaf Parameter Analysis, Leaf Diseases, and Nutritional Deficiencies Detection Using Machine Learning. In: Merchant SN, Warhade K, Adhikari D (eds) Advances in Signal and Data Processing. Springer, Singapore, pp 57–74
15. Gómez-Casero MT, López-Granados F, Peña-Barragán JM, et al (2007) Assessing Nitrogen and Potassium Deficiencies in Olive Orchards through Discriminant Analysis of Hyperspectral Data. J. Am. Soc. Hortic. Sci. 132:611–618. https://doi.org/10.21273/JASHS.132.5.611
16. Grieco M, Schmidt M, Warnemünde S, et al (2022) Dynamics and genetic regulation of leaf nutrient concentration in barley based on hyperspectral imaging and machine learning. Plant Sci. 315:111123. https://doi.org/10.1016/j.plantsci.2021.111123
17. Guo A, Huang W, Ye H, et al (2020) Identification of Wheat Yellow Rust Using Spectral and Texture Features of Hyperspectral Images. Remote Sens. 12:1419. https://doi.org/10.3390/rs12091419

18. Guo J, Zhang J, Xiong S, et al (2021) Hyperspectral assessment of leaf nitrogen accumulation for winter wheat using different regression modeling. Precision Agric 22:1634–1658. https://doi.org/10.1007/s11119-021-09804-z
19. Guo X, Zhu X, Li C, et al (2017) Hyperspectral Inversion of Potassium Content in Apple Leaves Based on Vegetation Index. Agric. Sci. 8:825–836. https://doi.org/10.4236/as.2017.88061
20. Hafsi C, Debez A, Abdelly C (2014) Potassium deficiency in plants: effects and signaling cascades. Acta Physiol Plant 36:1055–1070. https://doi.org/10.1007/s11738-014-1491-2
21. Hall, M.A., 1998. Correlation-based feature subset selection for machine learning. PhD thesis. University of Waikato, Hamilton, New Zealand.
22. Hasanuzzaman M, Bhuyan MHMB, Nahar K, et al (2018) Potassium: A Vital Regulator of Plant Responses and Tolerance to Abiotic Stresses. Agronomy 8:31. https://doi.org/10.3390/agronomy8030031
23. Inthichack P, Nishimura Y, Fukumoto Y (2012) Effect of potassium sources and rates on plant growth, mineral absorption, and the incidence of tip burn in cabbage, celery, and lettuce. Hortic Environ Biotechnol 53:135–142. https://doi.org/10.1007/s13580-012-0126-z
24. Li B, Xu X, Zhang L, et al (2020) Above-ground biomass estimation and yield prediction in potato by using UAV-based RGB and hyperspectral imaging. ISPRS J. Photogramm. Remote Sens. 162:161–172. https://doi.org/10.1016/j.isprsjprs.2020.02.013
25. Li L, Jákli B, Lu P, et al (2018) Assessing leaf nitrogen concentration of winter oilseed rape with canopy hyperspectral technique considering a non-uniform vertical nitrogen distribution. Ind. Crops Prod. 116:1–14. https://doi.org/10.1016/j.indcrop.2018.02.051
26. Li T, Zhu Z, Cui J, et al (2021) Monitoring of leaf nitrogen content of winter wheat using multi-angle hyperspectral data. Int. J. Remote Sens. 42:4672–4692. https://doi.org/10.1080/01431161.2021.1899333
27. Lin L, Gao Z, Liu X (2020) Estimation of soil total nitrogen using the synthetic color learning machine (SCLM) method and hyperspectral data. Geoderma 380:114664. https://doi.org/10.1016/j.geoderma.2020.114664
28. Liu S, Yang X, Guan Q, et al (2020) An Ensemble Modeling Framework for Distinguishing Nitrogen, Phosphorous and Potassium Deficiencies in Winter Oilseed Rape (Brassica napus L.) Using Hyperspectral Data. Remote Sensing 12:4060. https://doi.org/10.3390/rs12244060

29. Lu J, Li W, Yu M, et al (2021) Estimation of rice plant potassium accumulation based on non-negative matrix factorization using hyperspectral reflectance. Precision Agric. 22:51–74. https://doi.org/10.1007/s11119-020-09729-z

30. Lu J, Yang T, Su X, et al (2019) Monitoring leaf potassium content using hyperspectral vegetation indices in rice leaves. Precision Agric. https://doi.org/10.1007/s11119-019-09670-w

31. Mahajan GR, Sahoo RN, Pandey RN, et al (2014) Using hyperspectral remote sensing techniques to monitor nitrogen, phosphorus, sulphur and potassium in wheat (Triticum aestivum L.). Precision Agric. 15:499–522. https://doi.org/10.1007/s11119-014-9348-7

32. Mishra P, Roger JM, Rutledge DN, Woltering E (2020) SPORT pre-processing can improve near-infrared quality prediction models for fresh fruits and agro-materials. Postharvest Biol. Tech. 168:111271. https://doi.org/10.1016/j.postharvbio.2020.111271

33. Osco LP, Ramos APM, Faita Pinheiro MM, et al (2020) A Machine Learning Framework to Predict Nutrient Content in Valencia-Orange Leaf Hyperspectral Measurements. Remote Sens. 12:906. https://doi.org/10.3390/rs12060906

34. Pancorbo JL, Camino C, Alonso-Ayuso M, et al (2021) Simultaneous assessment of nitrogen and water status in winter wheat using hyperspectral and thermal sensors. Eur. J. Agron. 127:126287. https://doi.org/10.1016/j.eja.2021.126287

35. Pandey P, Ge Y, Stoerger V, Schnable JC (2017) High Throughput In vivo Analysis of Plant Leaf Chemical Properties Using Hyperspectral Imaging. Front Plant Sci 8:. https://doi.org/10.3389/fpls.2017.01348

36. Peng Y, Zhang M, Xu Z, et al (2020) Estimation of leaf nutrition status in degraded vegetation based on field survey and hyperspectral data. Sci Rep 10:4361. https://doi.org/10.1038/s41598-020-61294-7

37. Pimstein A, Karnieli A, Bansal SK, Bonfil DJ (2011) Exploring remotely sensed technologies for monitoring wheat potassium and phosphorus using field spectroscopy. Field Crops Res. 121:125–135. https://doi.org/10.1016/j.fcr.2010.12.001

38. Prośba-Białczyk U, Sacała E, Wilkosz M, Cieciura, M. (2017) Impact of seed stimulation and foliar fertilization with microelements on changes in the chemical composition and productivity of sugar beet. J. Elem, 22(4), 1525-1535.

39. Ragel P, Raddatz N, Leidi EO, et al (2019) Regulation of K+ Nutrition in Plants. Front. Plant Sci.10:

40. Rani P, Saini I, Singh N, et al (2021) Effect of potassium fertilizer on the growth, physiological parameters, and water status of Brassica juncea cultivars under different irrigation regimes. PLOS ONE 16:e0257023. https://doi.org/10.1371/journal.pone.0257023
41. Rustioni L, Grossi D, Brancadoro L, Failla O (2018) Iron, magnesium, nitrogen and potassium deficiency symptom discrimination by reflectance spectroscopy in grapevine leaves. Sci. Hortic.e 241:152–159. https://doi.org/10.1016/j.scienta.2018.06.097
42. Shen L, Gao M, Yan J, et al (2020) Hyperspectral Estimation of Soil Organic Matter Content using Different Spectral Preprocessing Techniques and PLSR Method. Remote Sens. 12:1206. https://doi.org/10.3390/rs12071206
43. Shuryak I (2022) Machine learning analysis of 137Cs contamination of terrestrial plants after the Fukushima accident using the random forest algorithm. J. Environ. Radioact. 241:106772. https://doi.org/10.1016/j.jenvrad.2021.106772
44. Siedliska A, Baranowski P, Pastuszka-Woźniak J, et al (2021) Identification of plant leaf phosphorus content at different growth stages based on hyperspectral reflectance. BMC Plant Biol 21:28. https://doi.org/10.1186/s12870-020-02807-4
45. Siedliska A, Baranowski P, Zubik M, et al (2018) Detection of fungal infections in strawberry fruit by VNIR/SWIR hyperspectral imaging. Postharvest Biol. Tech. 139:115–126. https://doi.org/10.1016/j.postharvbio.2018.01.018
46. Siedliska A, Baranowski P, Zubik M, Mazurek W (2017) Detection of pits in fresh and frozen cherries using a hyperspectral system in transmittance mode. J Food Eng. 215:61–71. https://doi.org/10.1016/j.jfoodeng.2017.07.028
47. Smola AJ, Schölkopf B (2004) A tutorial on support vector regression. Stat. and Comput. 14, 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88
48. Song X, Xu D, Huang C, et al (2021) Monitoring of nitrogen accumulation in wheat plants based on hyperspectral data. Remote Sensing Applications: Society and Environment 23:100598. https://doi.org/10.1016/j.rsase.2021.100598
49. Wan L, Zhou W, He Y, et al (2022) Combining transfer learning and hyperspectral reflectance analysis to assess leaf nitrogen concentration across different plant species datasets. Remote Sen. Environ.269:112826. https://doi.org/10.1016/j.rse.2021.112826

50. Wang Y-J, Jin G, Li L-Q, et al (2020) NIR hyperspectral imaging coupled with chemometrics for nondestructive assessment of phosphorus and potassium contents in tea leaves. Infrared Phys. Technol. 108:103365. https://doi.org/10.1016/j.infrared.2020.103365

51. Weksler S, Rozenstein O, Haish N, et al (2021) Detection of Potassium Deficiency and Momentary Transpiration Rate Estimation at Early Growth Stages Using Proximal Hyperspectral Imaging and Extreme Gradient Boosting. Sensors 21:958. https://doi.org/10.3390/s21030958

52. Wu C, Chen Y, Hong X, et al (2020) Evaluating soil nutrients of Dacrydium pectinatum in China using machine learning techniques. For Ecosyst 7:30. https://doi.org/10.1186/s40663-020-00232-5

53. Xu X, Du X, Wang F, et al (2020) Effects of Potassium Levels on Plant Growth, Accumulation and Distribution of Carbon, and Nitrate Metabolism in Apple Dwarf Rootstock Seedlings. Front. Plant Sci. 11:

54. Zhang C, Kong F (2014) Isolation and identification of potassium-solubilizing bacteria from tobacco rhizospheric soil and their effect on tobacco plants. Appl. Soil Ecol. 82:18–25. https://doi.org/10.1016/j.apsoil.2014.05.002

55. Zhang T, Fan S, Xiang Y, et al (2020) Non-destructive analysis of germination percentage, germination energy and simple vigour index on wheat seeds during storage by Vis/NIR and SWIR hyperspectral imaging. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 239:118488. https://doi.org/10.1016/j.saa.2020.118488

56. Zhang X, Liu F, He Y, Gong X (2013) Detecting macronutrients content and distribution in oilseed rape leaves based on hyperspectral imaging. Biosyst. Eng. 115:56–65. https://doi.org/10.1016/j.biosystemseng.2013.02.007

Statements and Declarations

Funding: This paper was partially financed by funds from the Polish National Centre for Research and Development in the frame of the project “MSINiN”, contract number: BIOSTRATEG3/343547/8/NCBR/2017. The funding source had no role in the design of this study, data collection analyses or interpretation, the decision to publish results or the preparation of the manuscript.

Competing interests: All the authors declare they have no financial interests.

Data availability: The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.