Check for updates

OPEN ACCESS

EDITED BY Yu Fenghua, Shenyang Agricultural University, China

REVIEWED BY Renato Herrig Furlanetto, University of Florida, United States Haijun Qi, Guangdong Academy of Agricultural Sciences (GDAAS), China

*CORRESPONDENCE Wen He Nw@gxib.cn

RECEIVED 30 April 2025 ACCEPTED 17 June 2025 PUBLISHED 07 July 2025

CITATION

Song Z, He W, Yao Y, Yu L, Huang J, Xu Y and Wang H (2025) The PLSR-ML fusion strategy for high-accuracy leaf potassium inversion in karst region of Southwest China. *Front. Plant Sci.* 16:1620971. doi: 10.3389/fpls.2025.1620971

COPYRIGHT

© 2025 Song, He, Yao, Yu, Huang, Xu and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

The PLSR-ML fusion strategy for high-accuracy leaf potassium inversion in karst region of Southwest China

Zhihao Song^{1,2}, Wen He^{2*}, Yuefeng Yao², Ling Yu³, Jinjun Huang², Yong Xu¹ and Haoyu Wang¹

¹College of Geomatics and Geoinformation, Guilin University of Technology, Guilin, China, ²Guangxi Key Laboratory of Plant Conservation and Restoration Ecology in Karst Terrain, Guangxi Institute of Botany, Guangxi Zhuang Autonomous Region and Chinese Academy of Sciences, Guilin, China, ³School of Computer Science and Engineering, Guilin University of Aerospace Technology, Guilin, China

Potassium is a critical macronutrient for plant growth, yet accurately and rapidly estimating its content in karst regions remains challenging due to complex terrestrial conditions. To address this, we collected leaf potassium content and reflectance data from 301 plant samples across nine karst regions in Guangxi Province. Our results showed that hybrid models combining Partial Least Squares Regression (PLSR) with three machine learning algorithms-Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Multi-Layer Perceptron (MLP)namely PLSR-RF, PLSR-XGBoost, and PLSR-MLP, demonstrated exceptional accuracy in estimating leaf potassium content. Validation coefficient of determination (R²) values reached 0.89, 0.94, and 0.96, respectivelyrepresenting improvements of 206%, 147%, and 108% over standalone algorithms. This performance gain was attributed to rigorous overfitting control: PLSR's dimensionality reduction synergized with ensemble machine learning (RF, XGBoost, MLP) to eliminate redundant spectral features while retaining predictive signals. Furthermore, fractional differentiation preprocessing significantly improved the correlation between spectral reflectance and potassium content, enhancing model robustness. Two spectral regions (700-1100 nm, 1400-1800 nm) were identified as key predictors, aligning with known potassium-related biochemical absorption features. Collectively, the integration of these strategies offers a robust framework for nutrient monitoring in ecologically fragile karst ecosystems.

KEYWORDS

karst region, leaf potassium content, machine learning, fractional differentiation, spectral reflectance

1 Introduction

The karst landscapes of southwestern China constitute a globally significant geomorphological system (D'Ettorre et al., 2024). Characterized by distinctive lithological structures and heterogeneous vegetation assemblages, this ecologically fragile region serves as a vital reservoir of terrestrial biodiversity. Nevertheless, intensive anthropogenic activities-particularly shifting slash-and-burn agriculture and unsustainable slope farming practices-have induced substantial degradation of surface vegetation cover (Jiang et al., 2014). This degradation may adversely affect the availability of essential nutrient such as potassium, which plays a critical role in plant growth by regulating water balance, facilitating nutrient transport (Wang et al., 2013; Hasanuzzaman et al., 2018), and enhances plant resilience against biotic and abiotic stressors (Anschütz et al., 2014). Foliar potassium concentration serves as a robust phytochemical indicator strongly correlated with plant physiological status, providing critical insights into vegetation health assessment. Consequently, precise quantification of foliar potassium levels emerges as a methodological imperative for elucidating plant adaptive strategies in degraded karst ecosystems.

Conventional laboratory methods for leaf potassium analysis, relying on destructive wet chemistry techniques, face inherent limitations in operational efficiency and scalability. In contrast, hyperspectral reflectance technology has emerged as a transformative non-destructive solution, enabling rapid in situ nutrient assessment through advanced spectroscopic platforms. The integration of hyperspectral remote sensing systems has particularly enhanced real-time potassium monitoring capabilities, demonstrating remarkable success in precision agriculture applications (Lin et al., 2024; Azadnia et al., 2023). However, species-specific variations result in differing spectral band sensitivities to potassium content across plant taxa. Current research has yet to fully resolve uncertainties in characterizing potassium-related spectral responses, necessitating further investigation into their underlying mechanisms. For example, Lu et al. (2020) found that the spectral reflectance of rice leaves in the shortwave infrared region (1300-2000 nm) is particularly sensitive to potassium content. Similarly, Lyu et al. (2023) identified potassium-sensitive bands in grape leaves at 410 nm, 490-500 nm, and 1242 nm. These discrepancies between rice and grape studies underscore the variability in potassium-sensitive spectral regions across species, highlighting both the challenges in universal band selection and the critical need for taxa-specific calibration. This variability becomes particularly relevant in ecologically unique regions such as karst landscapes, which host specialized plant communities. Karst-adapted species exhibit distinct spectral signatures compared to non-karst flora due to their divergent evolutionary adaptations and environmental stressors (Yue et al., 2010). Consequently, region-specific studies are imperative to map the spectral sensitivity patterns of leaf potassium in karst ecosystems, enabling accurate nutrient monitoring and supporting ecological conservation in these biodiverse yet fragile habitats.

While spectral information enables precise characterization of potassium signatures in leaf spectral response curves, noise interference remains a significant concern (Xie et al., 2020). Hyperspectral data acquisition is inherently susceptible to artifacts introduced by sample properties (e.g., particle size and surface texture) and environmental variability (Kong et al., 2023). Spectral differentiation transformations serve as a robust preprocessing technique to mitigate background noise and unwanted spectral reflectance variations. These transformations enhance spectral sensitivity, amplify diagnostic features, and optimize predictive model performance (Yang C, et al., 2021). First- and second-order derivatives are widely employed to improve spectral signal-to-noise ratios. However, integer-order differentiation exhibits limitations in resolving subtle spectral features when curvature variations are gradual, often leading to feature loss (Li et al., 2024). In contrast, fractional differentiation operates at finer computational intervals, enabling enhanced spectral information extraction from in situ leaf measurements (Benkhettou et al., 2015). In addition, using fractional differentiation can further sharpen peak shapes and perform better in detecting subtle signal changes in positive and negative spectral peaks (Tan et al., 2024). This methodological refinement directly translates to improved precision in estimating critical biochemical parameters, such as foliar potassium levels, which will be rigorously evaluated in our experimental framework.

In the field of nutrient content inversion, mainstream empirical approaches can be broadly categorized into four types (Berger et al., 2020): empirical parameter regression (Jay et al., 2017), linear nonparametric regression (Furlanetto et al., 2024), physically based methods, and nonlinear nonparametric regression (i.e., machine learning) (Furlanetto et al., 2023; Flynn et al., 2023). Empirical parameter regression typically employs narrowband vegetation indices (e.g., NIR/SWIR combinations) for rapid estimation of nutrients. However, due to the lack of distinct absorption features for certain elements and the influence of spectral signal coupling, these methods often suffer from limited generalizability (Li et al., 2021). Linear nonparametric regression techniques such as Partial Least Squares Regression (PLSR) and Principal Component Regression (PCR) utilize full-spectrum information and avoid manual feature selection. Still, their reliance on linear assumptions makes it difficult to capture the complex nonlinear relationships between spectral responses and plant biochemical properties (Atzberger et al., 2010). Physically based radiative transfer models (e.g., PROSAIL) aim to simulate the nutrient-spectrum relationship from a mechanistic perspective. Nonetheless, the weak absorption features of nutrients can be easily confounded with canopy water content and structural parameters, leading to ill-posed inversion problems (Féret et al., 2019). In contrast, machine learning methods are well-suited for nutrient estimation due to their strong capabilities in modeling complex nonlinear relationships and handling large-scale datasets (He et al., 2021).

However, significant challenges in model fitting persist when applying machine learning algorithms to vegetation parameter estimation (Doktor et al., 2014). The performance of machine learning models critically depends on feature selection - excessively large feature sets or overcomplicated architectures frequently lead to overfitting, compromising both training accuracy and model generalizability. Conversely, insufficient feature quantities and oversimplified models may result in underfitting. This issue is particularly pronounced in field spectroscopy data characterized by high dimensionality and multicollinearity. To address these challenges, researchers have implemented multiple mitigation strategies: (1) expanding training datasets to improve statistical representation; (2) employing dimensionality reduction techniques; (3) adopting robust cross-validation protocols; (4) applying regularization methods (Zhang et al., 2021); and (5) developing ensemble learning frameworks (Wang R, et al., 2020).

Hyperspectral data is characterized by high dimensionality and multivariate features, and the issue of feature redundancy has yet to be effectively resolved (Liu et al., 2021). This necessitates systematic dimensionality reduction of hyperspectral data to ensure model robustness. Notable implementations include Cao et al. (2021), who successfully mitigated overfitting in maize leaf nitrogen estimation through optimized spectral compression, and Ni et al. (2024) achieving superior predictive performance (R²=0.98) in sucrose quantification models via principal component analysis (PCA). This empirical evidence collectively substantiates that dimensionality reduction techniques, particularly PCA, significantly enhance both model accuracy (p<0.01) and algorithmic stability compared to untreated hyperspectral inputs. Building upon these methodological advancements, our study innovatively integrates partial least squares (PLS)-optimized PCA with ensemble machine learning frameworks to establish a robust estimation model for leaf potassium content in karst ecosystems, specifically designed to improve generalizability across heterogeneous geological environments.

Based on field spectrometer data, this study used a fractional differential spectroscopy method combined with multiple models to estimate the potassium content in the leaves of mixed forests in the Guangxi karst region. The main objectives of this research are as follows: (1) To assess the distribution of wavelengths sensitive to potassium content in plant leaves in the karst region; (2) To explore the role of fractional differentiation in estimating potassium content in karst plant leaves based on spectroradiometer data; and (3) To investigate whether combined models can overcome the overfitting issues encountered in machine learning models when estimating potassium content in karst plant leaves.

2 Materials and methods

2.1 Study area

The investigation was conducted in the karst-dominated terrain of Guangxi Zhuang Autonomous Region, Southwest China (20°54'-26°24'N, 104°28'-112°04'E; Figure 1). This geomorphologically complex area exhibits altitudinal gradients ranging from coastal plains (0 m) to montane systems (2141 m ASL), bisected by the Tropic of Cancer and bounded by tropical marine systems to the south. These latitudinal and topographic configurations engender a monsoonal climate regime with pronounced seasonality, manifesting in mean annual temperatures of 17.5-23.5°C and precipitation gradients from 841.2 mm (leeward basins) to 3387.5 mm (windward slopes). Nine standardized plots (200 m² each) were established across karst terrains, covering three vegetation succession stages: primary forests, secondary forests, and shrublands. This stratified design effectively captures karst ecosystem heterogeneity.

2.2 Data collection

Longitudinal foliar sampling spanned July 2018 to September 2020 across all study plots. Within each plot, phyllosphere specimens were systematically collected from 8–15 dominant species, establishing a comprehensive karst flora spectral database comprising 301 samples representing 37 families, 59 genera, and 70 species. To ensure spatial representativeness, sampling followed triaxial orientation protocols (0°[N], 120°, and 240°) within the horizontal plane.

Spectral acquisition employed a high-resolution field spectroradiometer (Fieldspec4, ASD Inc., USA) with 3 nm VNIR (350–1000 nm) and 8 nm SWIR (1001–2500 nm) spectral resolution (Shah et al., 2019). Three photometric replicates per tree were obtained through standardized protocol: 1) periodic radiometric calibration (10-minute intervals) using integrated reference panels; 2) constrained by field operation limitations (4hour battery endurance), two mature leaves per branch underwent non-destructive scanning; 3) branch-level spectral signatures were averaged to derive tree-specific reflectance profiles.

Post-spectral analysis, target leaves were immediately preserved in sterile bags (Whirl-Pak[®]) under controlled conditions (ICERSICE940 incubator, 4°C). Samples underwent laboratory processing within 24 h: 1) oven-drying at 75°C to constant mass; 2) mechanical homogenization to 100-mesh particle size; 3) quantitative potassium determination via flame photometric analysis (Sherwood 410, \pm 0.01 ppm detection limit) following standard digestion protocols (Reddy and Veeranki, 2013).

2.3 Methodology

2.3.1 Fractional differentiation

The fractional differentiation extends the concept of traditional integer-order differentiation to any arbitrary order, enabling continuous interpolation between integer orders (Hong et al., 2019). This method emphasizes subtle changes in spectral information (Wang Z, et al., 2020). Currently, the classic definitions of fractional differentiation include Riemann-Liouville (R-L), Grünwald-Letnikov (G-L), and Caputo (Pu et al., 2008; Wang et al., 2018). In this study, the Grünwald-Letnikov (G-L) definition was mainly adopted to derive the differentiation to the n-th order, as shown in Equation 1.



$$d^{\nu}f(x) = \lim_{h \to \infty} \frac{1}{h^{\nu}} \sum_{m=0}^{\frac{t-a}{h}} (-1)^m \frac{\Gamma(\nu+1)}{m! \Gamma(\nu-m+1)} f(x-mh)$$
(1)

where v is the order of differentiation, h is the step size, t and a are the upper and lower bounds of differentiation, respectively, and $\Gamma(\cdot)$ is the Gamma function, as defined in Equation 2.

$$\Gamma(\beta) = \int_0^\infty e^{-t} t^{\beta - 1} dt = (\beta - 1)!$$
(2)

where β is an arbitrary variable. In this study, the leaf spectra were differentiated within the range of 0 to 3 orders (at intervals of 0.1 order).

2.3.2 Partial least squares regression

Partial Least Squares Regression (PLSR) is a multivariate data analysis technique that combines the features of Principal Component Analysis (PCA) and Multiple Linear Regression (MLR). It is used to predict a set of dependent variables from a large number of collinear independent variables. This method assumes that the datasets of independent and dependent variables are $Z = [z_1, z_2, \dots z_k]_{n \times k}$ and $Q = [q]_{n \times 1}$, respectively. First, the first latent variable f_1 is extracted from Z, which is a linear combination of $z_1, z_2, \dots z_k$, and maximizes the correlation with the dependent variable Q. Then, a regression model is established for Q using f_1 . If the regression equation achieves the desired accuracy, component extraction is stopped; otherwise, the next component is extracted until the model reaches a satisfactory level of accuracy. The regression model is given by Equation 3, and each latent variable is defined as shown in Equation 4:

$$q = f_1 a_1 + f_2 a_2 + \dots + f_k a_k \tag{3}$$

$$f_m = w_{m1}z_1 + w_{m2}z_2 + \dots + w_{mk}z_k \tag{4}$$

where *m* is the number of principal components, *k* is the number of independent variables, *a* is the regression coefficient of *y* with respect to *f*, and *w* is the linear coefficient of *f* with respect to *z*.

2.3.3 Random forest

Random Forest (RF) is a machine learning algorithm based on decision trees (Breiman, 2001). RF resamples multiple samples from the training dataset and constructs a decision tree for each sample. Finally, the output value is calculated as the average of the predictions from all decision trees (Yang T, et al., 2021). RF has two important parameters: the number of trees and the number of features considered for splitting at each node. Initially, the number of decision trees was set to 50, and was then gradually increased in steps of 50 until it reached 200. The feature parameters for each

node split were set as sqrt, log2, and 10. The optimal parameters were determined through grid search (Yang and Shami, 2020a).

2.3.4 Extreme gradient boosting

XGBoost is an improved algorithm based on Gradient Boosted Decision Trees (GBDT), proposed by Chen and Guestrin (2016), which efficiently constructs boosted trees and supports parallel computation. Compared with traditional GBDT, which only utilizes first-order differential information, XGBoost performs a second-order Taylor expansion on the loss function, thereby improving the efficiency of finding the optimal solution. Finally, XGBoost obtains the overall prediction by summing the predictions of multiple decision trees, as shown in Equation 5.

$$\hat{y}_n = \sum_n^M f_m x_n f_m \in F \tag{5}$$

where \hat{y}_n represents the final predicted value of the model, M denotes the number of combined decision trees, which is the number of trees to be tuned, f_m is the *m*-th tree, x_n represents the *n*-th input sample, and F is the set of all tree models.

2.3.5 Multilayer perceptron

A Multilayer Perceptron (MLP) is a feedforward neural network composed of multiple neurons or nodes, which learns complex nonlinear mappings through connections between input and output vectors. It utilizes a parallel hierarchical structure consisting of an input layer, hidden layers, and an output layer, with information being transmitted through connection weights among these layers to predict target variables (Ehteram et al., 2020). In an MLP, the sum of the input signals received by a node is transformed through a nonlinear activation function to generate the output signal (Gardner and Dorling, 1998).

$$s_j = \sum_{i=1}^{n_0} w_{ij} x_i + a_j \tag{6}$$

$$z_j = f(s_j) = (1 + e^{-c_j})^{-1}$$
(7)

In Equations 6, 7, s_j represents the input to the *j*-th neuron in the hidden layer, a_j is the bias for the *j*-th neuron in the hidden layer, w_{ij} is the weight between the *i*-th input neuron and the *j*-th neuron in the hidden layer, $f(b_j)$ is the activation function, and z_j is the output of the *j*-th neuron. The final output of the MLP is obtained by computing a weighted sum of the hidden layer outputs, as shown in Equation 8:

$$O_k = \sum_{j=1}^{n_1} w_{jk} z_j + a_k \tag{8}$$

where O_k is the output of the *k*-th neuron in the output layer, w_{jk} is the weight between the *j*-th neuron in the hidden layer and the *k*-th neuron in the output layer, and n_1 represents the number of neurons in the hidden layer.

2.3.6 Combined models, sample segmentation, and accuracy assessment

The partial least squares regression (PLSR)-derived latent variables served as input variables for three machine learning

architectures: RF, XGBoost, and MLP. Subsequently, the integrated models PLSR-RF, PLSR-XGBoost, and PLSR-MLP were established. This hybrid dimensionality reduction approach effectively mitigated high-dimensionality challenges inherent in spectral data while controlling algorithmic complexity. During latent variable extraction from fractionally differentiated spectra, we implemented a variance retention threshold, where the process was terminated once the cumulative explained variance reached 75%, to preserve critical spectral features.

To effectively split the data into training and validation sets, the train_test_split function from the scikit-learn library in Python 3.10 was used. This function allows for random splitting of the dataset into different subsets, ensuring the independence of model training and validation. The training set accounted for 4/5 of the total samples, while the validation set accounted for 1/5. The model accuracy was evaluated using the coefficient of determination (R²), mean squared error (MSE), and mean absolute error (MAE).

2.3.7 Model parameter optimization

To ensure optimal predictive performance, the key hyperparameters of each model were systematically optimized. For the Partial Least Squares Regression (PLSR) model, the optimal number of components (n_components) was determined through exhaustive manual search over a predefined range (1 to 20) with model performance evaluated via 10-fold cross-validation. For the three machine learning models integrated with PLSR-Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Multi-Layer Perceptron (MLP)—hyperparameter tuning was performed using grid search with 10-fold cross-validation (Yang and Shami, 2020a).

In the RF model, the primary parameters optimized included the number of trees (n_estimators, e.g., 100, 200, 300) and the maximum tree depth (max_depth, e.g., 5, 10, 15). For the XGBoost model, key parameters such as the learning rate (learning_rate, e.g., 0.01, 0.05, 0.1), maximum depth (max_depth), and the number of estimators (n_estimators) were adjusted. In the MLP model, optimization focused on the architecture of hidden layers (hidden_layer_sizes, e.g., (100), or (100, 50)), activation function (activation, e.g., ReLU), solver algorithm (solver, e.g., Adam), and the L2 regularization term (alpha).

3 Results

3.1 Descriptive statistics of the samples

A total of 301 leaf samples were collected and analyzed for their total potassium content (expressed in units of 10 g/kg). The results showed that the potassium content ranged from 0.06 to 5.87, with a mean value of 0.81 (Figure 2). The coefficient of variation was calculated to be 1.30, indicating a high degree of variability among the samples. This substantial variation provides a solid foundation for model development and accuracy evaluation in subsequent analysis.



3.2 Fractional differentiation of reflectance and its correlation

correlation between spectral features and the target variable in practical applications.

Figure 3 illustrates the variations in spectral reflectance with different fractional differentiations. Compared to integer-order differentiations (0th, 1st, 2nd, and 3rd), fractional differentiation exhibits smaller amplitudes and smoother transitions. This gradual transformation maintains the detailed features of the spectral curves and prevents the abrupt fluctuations typically observed in integer-order differentiations. These results suggest that fractional differentiation, demonstrates greater advantages in the analyzing of complex experimental designs.

Figure 4 illustrates the distribution of absolute correlation coefficients between fractional differentiation spectra and leaf potassium content across fractional differentiation orders ranging from FD (0.0) to FD (3.0), with wavelengths spanning from 400 to 2500 nm. Before fractional differentiation (FD (0.0)), the spectral bands between 400-505 nm and 640-680 nm show significant correlation with leaf potassium content, though the correlation coefficients are relatively low. As the order of fractional differential (FD) increases-particularly between FD (1.5) and FD (3.0)-the spectral information in the ranges of 700-1100 nm and 1400-1800 nm shows stronger correlations with leaf potassium content, with most correlation coefficients exceeding 0.2. The maximum absolute correlation coefficient generally increases from FD (0.0) to FD (2.2), reaching a peak value of 0.46, before declining at higher orders. These findings highlight that selecting an appropriate fractional differentiation order, such as FD (2.2), can effectively improve the

3.3 Performance evaluation of individual models

The performance of the Partial Least Squares Regression (PLSR) model under fractional differentiation is shown in Figure 5a. Across the FD range from 0.0 to 3.0, the R^2 values for the training set consistently exceed those of the validation set by approximately 0.2 to 0.3, suggesting the presence of a certain level of overfitting in the PLSR model. The validation set achieves its highest R^2 value of 0.51 when the fractional differentiation is set to 0.8. Although the model's fitting accuracy is relatively low, it demonstrates stable performance without significant overfitting.

As shown in Figures 5b–d, the RF, XGBoost, and MLP models all exhibit a marked discrepancy in R^2 values between the training and validation sets, reflecting a clear tendency toward overfitting. In comparison to RF and XGBoost, the MLP model demonstrates marginally superior validation performance, with a maximum R^2 of 0.46, outperforming RF (0.29) and XGBoost (0.38).

In summary, although the PLSR model has limited fitting accuracy in predicting leaf potassium content, it demonstrates good stability. The training set R^2 remains between 0.6 and 0.7, while the validation set R^2 stays between 0.3 and 0.5. In contrast, the RF, XGBoost, and MLP models perform well on the training set but poorly on the validation set, indicating potential overfitting.



Therefore, among these four individual models, the PLSR model is the most suitable for estimating leaf potassium content.

3.4 Performance evaluation and analysis of combined models

The PLSR-RF model (Figure 6a) demonstrates strong fitting and generalization capabilities, as evidenced by its stable performance across most FD settings. The training set achieves consistently high R^2 values around 0.9, while the validation set maintains moderately high R^2 values ranging from approximately 0.75 to 0.89. Notably, within the FD range of 0.5 to 1.3, the validation performance improves sharply, with the R^2 value increasing from 0.01 to 0.77. The model achieves optimal performance at a fractional differentiation of FD (2.7), where the training set R^2 is 0.98, with MSE and MAE of 0.01 and 0.07, respectively. For the validation set, the R^2 value is 0.89, with MSE and MAE of 0.21 and 0.29, respectively. The PLSR-XGBoost model shows significant fluctuations across different FD settings, particularly for the training set. Despite these fluctuations, the difference in R^2 values between the training and validation sets decreases significantly when the fractional differentiation exceeds 1.2 (Figure 6b). This indicates that the combined model effectively mitigates overfitting. When the fractional differentiation is set to FD (2.7), the model performance reaches its peak, with R^2 , MSE, and MAE values are 0.99, $1.8*10^{-5}$, and 0.003 for the training set, and 0.94, 0.1, and 0.22 for the validation set, respectively. These findings indicate that PLSR combined with XGBoost provides more stable predictions under higher fractional differentiation levels.

The PLSR-MLP model performs poorly at low fractional differentiation values (FD < 0.8), with validation R^2 remaining below 0.4 between FD (0.2) and FD (0.6). Notably, at FD (0.3), the model exhibits signs of underfitting, as indicated by similarly low performance on both the training and validation sets. This suggests that the MLP has limited adaptability to raw data or data processed with low-order fractional differentiation (Figure 6c).



Absolute value distribution of correlation coefficients between fractional differentiation spectra and leaf potassium content, and the maximum absolute value of correlation coefficients for different fractional differentiations.



FIGURE 5

Determination Coefficients (R²) of different single models ((a) PLSR, (b) RF, (c) XGBoost, (d) MLP) for leaf potassium content estimation using fractional differentiation spectra: comparison of training and validation sets across different fractional differentiation orders (FD (0.0) to FD (3.0)).

However, as FD increases, the model's performance improves significantly. At FD (2.8), the R^2 values for both the training and validation sets reach 0.99 and 0.96, respectively, with MSE and MAE values of 0.01 and 0.05 for the training set, and 0.07 and 0.16 for the validation set, indicating excellent model performance at this optimal order.

Overall, the three combined models exhibit distinct responses to fractional differentiation. PLSR-RF improves with increasing FD but shows signs of overfitting. PLSR-XGBoost generalizes well when FD > 1.0, despite early instability. While PLSR-MLP achieves the highest accuracy in this study (Figure 7), PLSR-XGBoost involves fewer hyperparameter adjustments, demonstrates high computational efficiency, and facilitates easy deployment Therefore, although PLSR-MLP is the optimal model in terms of predictive performance, PLSR-XGBoost may offer a more practical solution for real-world potassium prediction tasks, especially in scenarios with limited computational resources or where rapid deployment is required.

3.5 Model comparison and selection of the optimal model

In this study, seven models, namely PLSR, RF, XGBoost, MLP, PLSR-RF, PLSR-XGBoost, and PLSR-MLP, were applied to predict the plant leaf potassium content using spectral differentiation transformation techniques in the karst region of Guangxi Province. The optimal fractional differentiation prediction results for each model are shown in Figure 7. Based on the coefficient of determination (R^2) on the validation sets, the top three models are PLSR-MLP (R^2 =0.96), PLSR-XGBoost (R^2 =0.94), and PLSR-RF (R^2 =0.89), respectively. In comparison, the RF model alone showed the worst performance, with an R^2 of only 0.29 on the validation sets.

Among these seven models, the PLSR-RF, PLSR-XGBoost, and PLSR-MLP models all effectively predict potassium content in plant leaves in the southwestern karst region. Relative to individual models, the three combined models exhibit improvements of 206%, 147%, and 108% in R^2 on the validation set, respectively. These substantial gains suggest that the combined modeling approach effectively mitigates overfitting and enhances generalization capability.

3.6 Advantages of fractional differentiation

The fractional differentiation is determined to be the optimal spectral transformation approach for all seven models (Table 1). The application of fractional differentiation significantly enhances the models' performance in estimating leaf potassium content. For the PLSR model, the optimal fractional differentiation is FD (0.8), resulting in a validation R² of 0.51, a marked improvement over the 0th order (R² = 0.26), 1st order (R² = 0.39), 2nd order (R² = 0.33), and 3rd order (R² = 0.35). The PLSR-RF model achieves its best performance at





FD (2.7), with a validation R^2 of 0.89, significantly outperforming the 0th order ($R^2 = 0.005$), 1st order ($R^2 = 0.58$), 2nd order ($R^2 = 0.82$), and 3rd order ($R^2 = 0.86$). The PLSR-XGBoost model performs optimally at FD (2.7), with a validation R^2 of 0.94, significantly outperforming the 0th order ($R^2 = 0.08$), 1st order ($R^2 = 0.58$), 2nd order ($R^2 = 0.83$), and 3rd order ($R^2 = 0.89$). Finally, the PLSR-MLP model achieves its

highest validation R^2 of 0.96 at FD (2.8), outperforming all integer orders from 0.0 to 3.0.

The results show that the optimal differentiation orders in all seven models are fractional rather than integer. This highlights the advantage of fractional differentiation in improving the accuracy and robustness of leaf potassium content estimation.

Model	Orders	Training Sets R2	Training Sets MSE	Training Sets MAE	Validation Sets R2	Validation Sets MSE	Validation Sets MAE
PLSR	FD (0.0)	0.44	0.49	0.49	0.26	1.41	0.77
	FD (1.0)	0.60	0.35	0.40	0.39	1.17	0.60
	FD (2.0)	0.55	0.39	0.40	0.33	1.28	0.60
	FD (3.0)	0.65	0.31	0.35	0.35	1.25	0.58
	FD (0.8)	0.76	0.21	0.33	0.51	0.94	0.62
RF	FD (0.0)	0.75	0.22	0.27	0.07	1.78	0.75
	FD (1.0)	0.77	0.2	0.25	0.15	1.63	0.71
	FD (2.0)	0.87	0.11	0.19	0.29	1.31	0.59
	FD (3.0)	0.76	0.21	0.25	0.3	1.43	0.64
	FD (1.9)	0.87	0.11	0.19	0.29	1.34	0.58
XGBoost	FD (0.0)	0.33	0.58	0.47	0.001	1.9	0.76
	FD (1.0)	0.72	0.23	0.31	0.17	1.59	0.69
	FD (2.0)	0.98	0.01	0.09	0.3	1.33	0.61
	FD (3.0)	0.84	0.14	0.21	0.31	1.32	0.61
	FD (2.2)	0.99	8*10 ⁻⁸	2.2*10 ⁻⁴	0.38	1.19	0.61
MLP	FD (0.0)	0.01	0.85	0.65	0.19	1.55	0.75
	FD (1.0)	0.42	0.51	0.42	0.45	1.05	0.7
	FD (2.0)	0.73	0.23	0.32	0.17	1.59	0.97
	FD (3.0)	0.91	0.08	0.17	0.12	1.67	0.98
	FD (0.9)	0.73	0.23	0.37	0.46	1.03	0.7
PLSR-RF	FD (0.0)	0.51	0.41	0.39	0.005	1.91	0.77
	FD (1.0)	0.93	0.06	0.15	0.58	0.79	0.49
	FD (2.0)	0.99	0.007	0.06	0.82	0.34	0.36
	FD (3.0)	0.99	0.004	0.03	0.86	0.26	0.32
	FD (2.7)	0.98	0.01	0.07	0.89	0.21	0.29
PLSR-XGBoost	FD (0.0)	0.52	0.41	0.38	0.05	2.0	0.77
	FD (1.0)	0.99	1.4*10 ⁻⁶	8.4*10 ⁻⁴	0.58	0.81	0.51
	FD (2.0)	0.99	3.1*10 ⁻⁵	0.004	0.84	0.3	0.33
	FD (3.0)	0.99	1.9*10 ⁻⁵	0.003	0.93	0.13	0.24
	FD (2.7)	0.99	1.8 ^{*10-5}	0.003	0.94	0.1	0.22
PLSR-MLP	FD (0.0)	0.38	0.55	0.45	0.07	1.77	0.71
	FD (1.0)	0.86	0.12	0.21	0.79	0.39	0.35
	FD (2.0)	0.99	0.007	0.06	0.91	0.06	0.16
	FD (3.0)	0.99	0.001	0.02	0.91	0.04	0.19
	FD (2.8)	0.99	0.01	0.07	0.96	0.05	0.16

TABLE 1 Performance comparison of seven models at different fractional differentiation orders (0.0, 1.0, 2.0, 3.0) and the optimal fractional differentiation order, based on evaluation metrics including R², MSE, and MAE for both individual and combined models.

This table presents the R², MSE, and MAE for the training and validation sets under the optimal fractional order, with the best-performing results shown in bold.

4 Discussion

4.1 Distribution of sensitive wavelengths

This study demonstrates that the spectral ranges of 700-1100 nm and 1400-1800 nm are critical for accurately estimating potassium content in plant leaves. Previous studies have identified the 964-1024 nm range as important for detecting potassium status in mature rubber tree leaves (Hu et al., 2024). In addition, specific wavelengths such as 720 nm and 1027 nm have been shown to play essential roles in predicting potassium content in rapeseed leaves (Zhang et al., 2013). The sensitive band in the 1400-1800 nm range identified in this study also aligns closely with the findings of Pimstein et al. (2011), further validating the relevance of this region for potassium estimation. Potassium is an essential ion in plant cells, involved in regulating osmotic pressure, activating enzymatic processes, and controlling stomatal dynamics (Nieves-Cordones et al., 2014; Yu et al., 2023). These physiological activities influence leaf cellular structure and water status, thereby indirectly affecting spectral reflectance. In the 700-1100 nm range, particularly within the near-infrared region (700-900 nm), spectral responses are strongly associated with internal leaf structure, which is sensitive to variations in tissue density and cellular arrangement. Since potassium plays a key role in water transport, cell turgor, and tissue development, changes in potassium levels can induce structural modifications that alter reflectance in this region (Lyu et al., 2023). Moreover, the short-wave nearinfrared region (900-1100 nm) captures spectral signals related to leaf water content and biochemical composition, both of which are closely linked to potassium-mediated regulation (Dos Santos et al., 2023).

The presence of sensitive bands in the 1400-1800 nm range is closely linked to the various physiological roles of potassium in plant growth. Potassium influences leaf water transpiration by regulating stomatal opening, which in turn affects spectral reflectance (Lin et al., 2024). Consequently, potassium-sensitive bands are often found near the peak wavelengths of water absorption, such as 1450 nm and 1950 nm (Yu et al., 2023). However, some wavelengths farther from these water absorption peaks also show high sensitivity, likely due to changes in plant chemical composition and physiology under the unique environmental conditions of the karst regions. Previous studies have demonstrated significant differences in stoichiometric characteristics between plants in karst and non-karst regions (Zhang et al., 2019). Potassium is crucial for activating enzymes involved in starch, protein, and fat synthesis, as well as promoting the synthesis of plant hormones that regulate meristem growth (Amirruddin et al., 2020). These functions may contribute to the sensitive bands distanced from water absorption peaks. Therefore, the presence of such bands in the 1400-1800 nm range likely reflects potassium's regulatory effects on physiological traits linked to long-term adaptation of plants to the karst environment.

4.2 The capabilities of fractional differentiation

Spectral data are often affected by instrument noise, environmental conditions, sample surface scattering, and background signals (Liu et al., 2023). Preprocessing techniques help mitigate these interferences, yielding a purer spectral signal that prevents the model from being affected by irrelevant signals and reduces errors (Li et al., 2025). Among these techniques, differentiation—particularly fractional differentiation—has emerged as a powerful method for capturing subtle spectral details and improving the accuracy of spectral-based estimations.

While traditional preprocessing techniques such as SNV and MSC effectively reduce scattering effects and smooth spectra, they are limited in handling high-noise spectral data (Oliveri et al., 2019). Differentiation processing of near-infrared spectra effectively removes noise while extracting subtle inflection points and spectral changes (Wang et al., 2018). Yang et al. (2022) demonstrated that applying differentiation to crop spectra significantly improves model prediction accuracy. Similarly, Shen et al. (2020) found that fractional differentiation significantly improves the accuracy of soil organic matter (SOM) content estimation. These studies highlight the significant advantages of differentiation in spectral preprocessing. Our findings similarly show that differentiation enhances the correlation between leaf potassium content and spectral reflectance, thereby improving estimation accuracy.

Differentiation includes both integer-order and fractional differentiation (Jin and Wang, 2022). Integer-order differentiation typically involves the first and second differentiations. However, the large intervals between these first and second differentiations result in significant differences between the nth and (n+1)th differentiation curves. This limitation causes integer-order differentiation to overlook finer spectral details (Anon, 2020). In contrast, fractional differentiation can extract detailed spectral information over smaller intervals while minimizing the introduction of excessive high-frequency noise (Zununjan et al., 2024; Song et al., 2023). The advantages of fractional differentiation stem from its unique mathematical structure, which, through the Grünwald-Letnikov definition, achieves a generalized difference structure, smooth attenuation, and long memory effects (Scherer et al., 2011). This enables fractional differentiation to more accurately capture spectral detail variations in data with complex background noise. Ge et al. (2022) demonstrated that fractional differentiation is highly effective for processing hyperspectral data in soil salinization risk assessment, with models using fractional differentiation proving more stable than those using integer-order differentiation. This conclusion from Ge et al. (2022) aligns with our findings, where fractional differentiation outperformed integerorder differentiation in estimating potassium content in plant leaves in the karst region.

However, the application of fractional differentiation also presents challenges. Low-order differentiation transformations

provide limited improvement in correlation, while higher-order differentiation does not significantly enhance correlation coefficients between spectral reflectance and potassium content. Additionally, the optimal fractional differentiation varies across models, and similar studies on nutrient inversion in plant leaves suggest that the best fractional differentiation should be chosen based on the specific model being used.

4.3 Control overfitting

The results indicate that the RF, XGBoost, and MLP models generally exhibit overfitting (Figure 5). Due to their strong nonlinear fitting abilities (Bentéjac et al., 2021), these models tend to capture noise and irrelevant features when handling highdimensional data, resulting in overfitting (Ying, 2019).

Common methods to control overfitting include dimensionality reduction, regularization, cross-validation, feature selection (Barbosa et al., 2024), and ensemble models. Several studies have explored the application of these methods in controlling overfitting. For example, Teresa et al. (2022) showed that dimensionality reduction effectively addresses over-parameterization in deep learning. Du et al. (2024) estimated rapeseed growth parameters using an ensemble learning algorithm, achieving better performance than individual machine learning models. For dimensionality reduction, we employed a PLSbased PCA method to extract latent variables that are highly correlated with the target variable. These latent variables were used as input features for the RF, XGBoost, and MLP models, effectively reducing the risk of overfitting in complex datasets.

In addition, hyperparameter optimization is a crucial strategy for mitigating overfitting and improving model generalization (Bischl et al., 2023). By tuning parameters such as the number of estimators, learning rate, and maximum tree depth (for RF and XGBoost), or the number of hidden layers and neurons (for MLP), models can better balance bias and variance. In this study, we employed grid search combined with cross-validation to optimize the key hyperparameters of each model, thereby reducing overfitting and enhancing predictive robustness. These findings are consistent with previous studies, which have demonstrated that well-tuned models generally outperform those using default configurations, particularly in high-dimensional datasets (Quan, 2024).

Combining dimensionality reduction with machine learning shows great potential for predicting nutrient content in plant leaves. For instance, Mahajan et al. (2024) used a PLSR-based machine learning model to predict potassium content in cashew leaves, achieving an R^2 of 0.66. Zhou et al. (2024) combined PCA with machine learning to predict cadmium content in lettuce leaves, obtaining an R^2 of 0.92 for the validation set. In our study, potassium content estimation in karst plants achieved an R^2 of 0.96 in the prediction set. This result confirms the effectiveness of PLS-based dimensionality reduction for retrieving leaf nutrient content across multiple species. This approach provides a valuable reference for future research.

In summary, combined machine learning models effectively control overfitting and enhance prediction performance. However, our research is limited to the leaf scale, and further validation is needed for their effectiveness in controlling overfitting when applied to UAV or satellite platforms. Future studies should explore the applicability of these models at larger scales and with higher-resolution data to comprehensively assess their generalization and practical value. Moreover, selecting the best model should not rely solely on prediction accuracy; factors such as model complexity, training time, and computational cost must also be taken into account to ensure the model's feasibility and efficiency in real-world applications.

5 Conclusions

This study identifies key spectral bands (700–1100 nm, and 1400–1800 nm) that are critical for estimating potassium content in plant leaves. These bands correspond to important physiological processes, including photosynthesis, pigment concentration, and water regulation, which are influenced by potassium. Fractional differentiation effectively reduces noise and captures subtle spectral features, significantly improving the accuracy of potassium estimation compared to traditional integer-order differentiation.

Furthermore, the study addresses overfitting in machine learning models by combining dimensionality reduction, and advanced algorithms such as Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Multilayer Perceptron (MLP). This integrated approach resulted in a high prediction accuracy ($R^2 = 0.96$) for potassium content in karst region plants.

In summary, this research advances potassium estimation through hyperspectral data by optimizing data preprocessing, and enhancing model performance. These findings provide valuable insights for plant nutrient monitoring, particularly in complex ecological environments, and offer a foundation for future research on large-scale remote sensing applications.

Data availability statement

The datasets presented in this article are not readily available because The data that has been used is confidential. Requests to access the datasets should be directed to Wen He, hw@gxib.cn.

Author contributions

ZS: Methodology, Validation, Writing – original draft. WH: Conceptualization, Investigation, Methodology, Validation, Writing – review & editing. YY: Supervision, Writing – review & editing. LY: Investigation, Writing – review & editing. JH: Data curation, Investigation, Writing – review & editing. YX: Writing – review & editing. HW: Investigation, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported

by the Natural Science Foundation of Guangxi, China (2024GXNSFAA010318), the Key Research and Development Program of Guangxi, China (GuikeAB22035060), the Basic Research Fund of Guangxi Institute of Botany (GUI ZHIYE 23005) and the Fund of Guangxi Key Laboratory of Plant Conservation and Restoration Ecology in Karst Terrain (No.22-035-26).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Amirruddin, A. D., Muharam, F. M., Ismail, M. H., Tan, N. P., and Ismail, M. F. (2020). Hyperspectral spectroscopy and imbalance data approaches for classification of oil palm's macronutrients observed from frond 9 and 17. *Comput. Electron. Agric.* 178, 105768. doi: 10.1016/j.compag.2020.105768

Anon (2020). A possible fractional order derivative and optimized spectral indices for assessing total nitrogen content in cotton. *Comput. Electron. Agric.* 171, 105275. doi: 10.1016/j.compag.2020.105275

Anschütz, U., Becker, D., and Shabala, S. (2014). Going beyond nutrition: Regulation of potassium homoeostasis as a common denominator of plant adaptive responses to environment. *J. Plant Physiol.* 171, 670–687. doi: 10.1016/j.jplph.2014.01.009

Atzberger, C., Guérif, M., Baret, F., and Werner, W. (2010). Comparative analysis of three chemometric techniques for the spectroradiometric assessment of canopy chlorophyll content in winter wheat. *Comput. Electron. Agric.* 73, 165–173. doi: 10.1016/j.compag.2010.05.006

Azadnia, R., Rajabipour, A., Jamshidi, B., and Omid, M. (2023). New approach for rapid estimation of leaf nitrogen, phosphorus, and potassium contents in appletrees using Vis/NIR spectroscopy based on wavelength selection coupled with machine learning. *Comput. Electron. Agric.* 207, 107746. doi: 10.1016/ j.compag.2023.107746

Barbosa, G. N. N., Andreoni, M., and Mattos, D. M. F. (2024). Optimizing feature selection in intrusion detection systems: Pareto dominance set approaches with mutual information and linear correlation. *Ad Hoc Networks* 159, 103485. doi: 10.1016/j.adhoc.2024.103485

Benkhettou, N., Brito da Cruz, A. M. C., and Torres, D. F. M. (2015). A fractional calculus on arbitrary time scales: Fractional differentiation and fractional integration. *Signal Process.* 107, 230–237. doi: 10.1016/j.sigpro.2014.05.026

Bentéjac, C., Csörgő, A., and Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artif. Intell. Rev.* 54, 1937–1967. doi: 10.1007/s10462-020-09896-5

Berger, K., Verrelst, J., Féret, J. B., Wang, Z., Wocher, M., Strathmann, M., et al. (2020). Crop nitrogen monitoring: Recent progress and principal developments in the context of imaging spectroscopy missions. *Remote Sens. Environ.* 242, 111758. doi: 10.1016/j.rse.2020.111758

Bischl, B., Binder, M., Lang, M., Pielok, T., Richter, J., Coors, S., et al. (2023). Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. *WIREs Data Min. Knowledge Discov.* 13, e1484. doi: 10.1002/widm.v13.2

Breiman, L. (2001). Random forests. Mach. Learn. 45, 5-32. doi: 10.1023/ A:1010933404324

Cao, C., Wang, T., Gao, M., Li, Y., Li, D., and Zhang, H. (2021). Hyperspectral inversion of nitrogen content in maize leaves based on different dimensionality reduction algorithms. *Comput. Electron. Agric.* 190, 106461. doi: 10.1016/j.compag.2021.106461

Chen, T., and Guestrin, C. (2016). "XGBoost: A scalable tree boosting system," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco California USA: Association for Computing Machinery. 785–794. doi: 10.1145/2939672.2939785

D'Ettorre, U. S., Liso, I. S., and Parise, M. (2024). Desertification in karst areas: A review. *Earth-Science Rev.* 253, 104786. doi: 10.1016/j.earscirev.2024.104786

Doktor, D., Lausch, A., Spengler, D., and Thurner, M. (2014). Extraction of plant physiological status from hyperspectral signatures using machine learning methods. *Remote Sens.* 6, 12247–12274. doi: 10.3390/rs61212247

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Dos Santos, G. L. A. A., Reis, A. S., Besen, M. R., Furlanetto, R. H., Rodrigues, M., Crusiol, L. G. T., et al. (2023). Spectral method for macro and micronutrient prediction in soybean leaves using interval partial least squares regression. *Eur. J. Agron.* 143, 126717. doi: 10.1016/j.eja.2022.126717

Du, R., Lu, J., Xiang, Y., Zhang, F., Chen, J., Tang, Z., et al. (2024). Estimation of winter canola growth parameter from UAV multi-angular spectral-texture information using stacking-based ensemble learning model. *Comput. Electron. Agric.* 222, 109074. doi: 10.1016/j.compag.2024.109074

Ehteram, M., Ahmed, A. N., Ling, L., Fai, C. M., Latif, S. D., Afan, H. A., et al. (2020). Pipeline scour rates prediction-based model utilizing a multilayer perceptron-colliding body algorithm. *Water* 12, 902. doi: 10.3390/w12030902

Féret, J. B., Le Maire, G., Jay, S., Berveiller, D., Bendoula, R., Hmimina, G., et al. (2019). Estimating leaf mass per area and equivalent water thickness based on leaf optical properties: Potential and limitations of physical modeling and machine learning. *Remote Sens. Environ.* 231, 110959. doi: 10.1016/j.rse.2018.11.002

Flynn, K. C., Baath, G., Lee, T. O., Gowda, P., and Northup, B. (2023). Hyperspectral reflectance and machine learning to monitor legume biomass and nitrogen accumulation. *Comput. Electron. Agric.* 211, 107991. doi: 10.1016/j.compag.2023.107991

Furlanetto, R. H., Crusiol, L. G. T., Gonçalves, J. V. F., Nanni, M. R., de Oliveira Junior, A., de Oliveira, F. A., et al. (2023). Machine learning as a tool to predict potassium concentration in soybean leaf using hyperspectral data. *Precis. Agric.* 24, 2264–2292. doi: 10.1007/s11119-023-10040-w

Furlanetto, R. H., Crusiol, L. G. T., Nanni, M. R., de Oliveira Junior, A., and Sibaldelli, R. N. R. (2024). Hyperspectral data for early identification and classification of potassium deficiency in soybean plants (Glycine max (L.) merrill). *Remote Sens.* 16, 1900. doi: 10.3390/rs16111900

Gardner, M. W., and Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environ.* 32, 2627–2636. doi: 10.1016/S1352-2310(97)00447-0

Ge, X., Ding, J., Teng, D., Xie, B., Zhang, X., Wang, J., et al. (2022). Exploring the capability of Gaofen-5 hyperspectral data for assessing soil salinity risks. *Int. J. Appl. Earth Observation Geoinformation* 112, 102969. doi: 10.1016/j.jag.2022.102969

Hasanuzzaman, M., Bhuyan, M. H. M. B., Nahar, K., Hossain, M. S., Mahmud, J. A., Hossen, M. S., et al. (2018). Potassium: A vital regulator of plant responses and tolerance to abiotic stresses. *Agronomy* 8, 31. doi: 10.3390/agronomy8030031

He, W., Li, Y., Wang, J., Yao, Y., Yu, L., Gu, D., et al. (2021). Using field spectroradiometer to estimate the leaf N/P ratio of mixed forest in a karst area of southern China: A combined model to overcome overfitting. *Remote Sens.* 13, 3368. doi: 10.3390/rs13173368

Hong, Y., Liu, Y., Chen, Y., Liu, Y., Yu, L., Liu, Y., et al. (2019). Application of fractional-order derivative in the quantitative estimation of soil organic matter content through visible and near-infrared spectroscopy. *Geoderma* 337, 758–769. doi: 10.1016/j.geoderma.2018.10.025

Hu, W., Tang, W., Li, C., Wu, J., Liu, H., Wang, C., et al. (2024). Handling the challenges of small-scale labeled data and class imbalances in classifying the N and K statuses of rubber leaves using hyperspectroscopy techniques. *Plant Phenomics* 6, 0154. doi: 10.34133/plantphenomics.0154

Jay, S., Maupas, F., Bendoula, R., and Gorretta, N. (2017). Retrieving LAI, chlorophyll and nitrogen contents in sugar beet crops from multi-angular optical remote sensing: Comparison of vegetation indices and PROSAIL inversion for field phenotyping. *Field Crops Res.* 210, 33–46. doi: 10.1016/j.fcr.2017.05.005

Jiang, Z., Lian, Y., and Qin, X. (2014). Rocky desertification in Southwest China: Impacts, causes, and restoration. *Earth-Science Rev.* 132, 1–12. doi: 10.1016/j.earscirev.2014.01.005

Jin, J., and Wang, Q. (2022). Hyperspectral indices developed from the low order fractional derivative spectra can capture leaf dry matter content across a variety of species better. *Agric. For. Meteorology* 322, 109007. doi: 10.1016/j.agrformet.2022.109007

Kong, Y., Liu, Y., Geng, J., and Huang, Z. (2023). Pixel-level assessment model of contamination conditions of composite insulators based on hyperspectral imaging technology and a semi-supervised ladder network. *IEEE Trans. Dielectrics Electrical Insulation* 30, 326–335. doi: 10.1109/TDEL2022.3226164

Li, J., Ma, Y., Zhang, J., and Kong, D. (2025). Rapid detection of fertilizer information based on Raman spectroscopy and machine learning. *Spectrochimica Acta Part A: Mol. Biomolecular Spectrosc.* 324, 124985.

Li, W., Xiang, Y., Liu, X., Tang, Z., Wang, X., Huang, X., et al. (2024). Estimation of soil moisture content based on fractional differential and optimal spectral index. *Agronomy* 14, 184. doi: 10.3390/agronomy14010184

Li, H., Yang, W., Lei, J., She, J., and Zhou, X. (2021). Estimation of leaf water content from hyperspectral data of different plant species by using three new spectral absorption indices. *PloS One* 16, e0249351. doi: 10.1371/journal.pone.0249351

Lin, D., Chen, Y., Qiao, Y., Qin, D., Miao, Y., Sheng, K., et al. (2024). A study on an accurate modeling for distinguishing nitrogen, phosphorous and potassium status in summer maize using in *situ* canopy hyperspectral data. *Comput. Electron. Agric.* 221, 108989. doi: 10.1016/j.compag.2024.108989

Liu, R., Ning, X., Cai, W., and Li, G. (2021). Multiscale dense cross-attention mechanism with covariance pooling for hyperspectral image scene classification. *Mobile Inf. Syst.* 2021, 9962057. doi: 10.1155/2021/9962057

Liu, L., Zareef, M., Wang, Z., Li, H., Chen, Q., and Ouyang, Q. (2023). Monitoring chlorophyll changes during Tencha processing using portable near-infrared spectroscopy. *Food Chem.* 412, 135505. doi: 10.1016/j.foodchem.2023.135505

Lu, J., Yang, T., Su, X., Qi, H., Yao, X., Cheng, T., et al. (2020). Monitoring leaf potassium content using hyperspectral vegetation indices in rice leaves. *Precis. Agric.* 21, 324–348. doi: 10.1007/s11119-019-09670-w

Lyu, H., Grafton, M., Ramilan, T., Irwin, M., and Sandoval, E. (2023). Assessing the leaf blade nutrient status of pinot noir using hyperspectral reflectance and machine learning models. *Remote Sens.* 15, 1497. doi: 10.3390/rs15061497

Mahajan, G. R., Das, B., Kumar, P., Murgaokar, D., Patel, K., Desai, A., et al. (2024). Spectroscopy-based chemometrics combined machine learning modeling predicts cashew foliar macro- and micronutrients. *Spectrochimica Acta Part A: Mol. Biomolecular Spectrosc.* 320, 124639. doi: 10.1016/j.saa.2024.124639

Ni, J., Xue, Y., Zhou, Y., and Miao, M. (2024). Rapid identification of greenhouse tomato senescent leaves based on the sucrose-spectral quantitative prediction model. *Biosyst. Eng.* 238, 200–211. doi: 10.1016/j.biosystemseng.2024.01.013

Nieves-Cordones, M., Alemán, F., Martínez, V., and Rubio, F. (2014). K+ uptake in plant roots. The systems involved, their regulation and parallels in other organisms. *J. Plant Physiol.* 171, 688–695. doi: 10.1016/j.jplph.2013.09.021

Oliveri, P., Malegori, C., Simonetti, R., and Casale, M. (2019). The impact of signal pre-processing on the final interpretation of analytical outcomes – A tutorial. *Analytica Chimica Acta* 1058, 9–17. doi: 10.1016/j.aca.2018.10.055

Pimstein, A., Karnieli, A., Bansal, S. K., and Bonfil, D. J. (2011). Exploring remotely sensed technologies for monitoring wheat potassium and phosphorus using field spectroscopy. *Field Crops Res.* 121, 125–135. doi: 10.1016/j.fcr.2010.12.001

Pu, Y., Wang, W., Zhou, J., Wang, Y., and Jia, H. (2008). Fractional differential approach to detecting textural features of digital image and its fractional differential filter implementation. *Sci. China Ser. F: Inf. Sci.* 51, 1319–1339. doi: 10.1007/s11432-008-0098-x

Quan, S. J. (2024). Comparing hyperparameter tuning methods in machine learning based urban building energy modeling: A study in Chicago. *Energy Buildings* 317, 114353. doi: 10.1016/j.enbuild.2024.114353

Reddy, D. D., and Veeranki, K. (2013). Simple and inexpensive water extraction method for assaying potassium concentration in tobacco plant tissue. *Commun. Soil Sci. Plant Anal.* 44, 962–970. doi: 10.1080/00103624.2012.747603

Scherer, R., Kalla, S. L., Tang, Y., and Huang, J. (2011). The Grünwald–Letnikov method for fractional differential equations. *Comput. Mathematics Appl.* 62, 902–917. doi: 10.1016/j.camwa.2011.03.054

Shah, S. H., Angel, Y., Houborg, R., Ali, S., and McCabe, M. F. (2019). A random forest machine learning approach for the retrieval of leaf chlorophyll content in wheat. *Remote Sens.* 11, 920. doi: 10.3390/rs11080920

Shen, L., Gao, M., Yan, J., Li, Z. L., Leng, P., Yang, Q., et al. (2020). Hyperspectral estimation of soil organic matter content using different spectral preprocessing techniques and PLSR method. *Remote Sens.* 12, 1206. doi: 10.3390/rs12071206

Song, G., Wang, Q., and Jin, J. (2023). Fractional-order derivative spectral transformations improved partial least squares regression estimation of photosynthetic capacity from hyperspectral reflectance. *IEEE Trans. Geosci. Remote Sens.* 61, 1–10. doi: 10.1109/TGRS.2023.3270892

Tan, J., Ding, J., Wang, Z., Han, L., Wang, X., Li, Y., et al. (2024). Estimating soil salinity in mulched cotton fields using UAV-based hyperspectral remote sensing and a Seagull Optimization Algorithm-Enhanced Random Forest Model. *Comput. Electron. Agric.* 221, 109017. doi: 10.1016/j.compag.2024.109017

Teresa, N., Hogg, D. W., and Villar, S. (2022). Dimensionality reduction, regularization, and generalization in overparameterized regressions. *SIAM J. Mathematics Data Sci.* 4, 126–152. doi: 10.1137/20M1387821

Wang, R., Lu, S., and Feng, W. (2020). A novel improved model for building energy consumption prediction based on model integration. *Appl. Energy* 262, 114561. doi: 10.1016/j.apenergy.2020.114561

Wang, X., Zhang, F., Kung te, H., and Johnson, V. C. (2018). New methods for improving the remote sensing estimation of soil organic matter content (SOMC) in the Ebinur Lake Wetland National Nature Reserve (ELWNNR) in northwest China. *Remote Sens. Environ.* 218, 104–118. doi: 10.1016/j.rse.2018.09.020

Wang, Z., Zhang, X., Zhang, F., Chan, N. W., Kung te, H., Liu, S., et al. (2020). Estimation of soil salt content using machine learning techniques based on remotesensing fractional derivatives, a case study in the Ebinur Lake Wetland National Nature Reserve, Northwest China. *Ecol. Indic.* 119, 106869. doi: 10.1016/j.ecolind.2020.106869

Wang, M., Zheng, Q., Shen, Q., and Guo, S. (2013). The critical role of potassium in plant stress response. *Int. J. Mol. Sci.* 14, 7370–7390. doi: 10.3390/ijms14047370

Xie, T., Li, S., and Sun, B. (2020). Hyperspectral images denoising via nonconvex regularized low-rank and sparse matrix decomposition. *IEEE Trans. Image Process.* 29, 44–56. doi: 10.1109/TIP.83

Yang, C., Feng, M., Song, L., Wang, C., Yang, W., Xie, Y., et al. (2021). Study on hyperspectral estimation model of soil organic carbon content in the wheat field under different water treatments. *Sci. Rep.* 11, 18582. doi: 10.1038/s41598-021-98143-0

Yang, T., Lu, J., Liao, F., Qi, H., Yao, X., Cheng, T., et al. (2021). Retrieving potassium levels in wheat blades using normalised spectra. *Int. J. Appl. Earth Observation Geoinformation* 102, 102412. doi: 10.1016/j.jag.2021.102412

Yang, L., and Shami, A. (2020a). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing* 415, 295–316. doi: 10.1016/j.neucom.2020.07.061

Yang, W., Xiong, Y., Xu, Z., Li, L., and Du, Y. (2022). Piecewise preprocessing of near-infrared spectra for improving prediction ability of a PLS model. *Infrared Phys. Technol.* 126, 104359. doi: 10.1016/j.infrared.2022.104359

Ying, X. (2019). An overview of overfitting and its solutions. J. Physics: Conf. Ser. 1168, 022022. doi: 10.1088/1742-6596/1168/2/022022

Yu, Y., Yu, H., Li, X., Zhang, L., and Sui, Y. (2023). Prediction of potassium content in rice leaves based on spectral features and random forests. *Agronomy* 13, 2337. doi: 10.3390/agronomy13092337

Yue, Y., Wang, K., Zhang, B., Chen, Z., Jiao, Q., Liu, B., et al. (2010). Exploring the relationship between vegetation spectra and eco-geo-environmental conditions in karst region, Southwest China. *Environ. Monit. Assess.* 160, 157–168. doi: 10.1007/s10661-008-0665-z

Zhang, C., Bengio, S., Hardt, M., Recht, B., and Vinyals, O. (2021). Understanding deep learning (still) requires rethinking generalization. *Commun. ACM* 64, 107–115. doi: 10.1145/3446776

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

Zhang, Y., Zhou, C., Lv, W., Dai, L., Tang, J., Zhou, S., et al. (2019). Comparative study of the stoichiometric characteristics of karst and non-karst forests in Guizhou, China. J. Forestry Res. 30, 799–806. doi: 10.1007/s11676-018-0806-3

Zhou, L., Zhou, L., Wu, H., Jing, T., Li, T., Li, J., et al. (2024). Estimation of cadmium content in lactuca sativa L. Leaves using visible-near-infrared spectroscopy technology. *Agronomy* 14, 644. doi: 10.3390/agronomy14040644

Zununjan, Z., Turghan, M. A., Sattar, M., Kasim, N., Emin, B., and Abliz, A. (2024). Combining the fractional order derivative and machine learning for leaf water content estimation of spring wheat using hyper-spectral indices. *Plant Methods* 20, 97. doi: 10.1186/s13007-024-01224-0