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.

Hyperspectral inversion of SPAD in Camellia oleifera based on integrated Taylor-CC and ANN with optimal wavelength selection

Xingzi Kou
Anhui Agricultural University
Xinyue Jiang
Nanjing Normal University
Yue Zhang
Anhui Agricultural University
Chunyu Ge
Anhui Agricultural University
Zhibin Sun
Nanjing Normal University
Xuehai Tang
tangxuehai@ahau.edu.cn
Anhui Agricultural University

Research Article

Keywords:

Posted Date: November 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11500533/v1

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

Additional Declarations: No competing interests reported.

Hyperspectral inversion of SPAD in *Camellia oleifera* based on integrated Taylor-CC and ANN with optimal wavelength selection

Xingzi Kou[a,b], Xinyue Jiang[c], Yue Zhang[a,b], Chunyu Ge[a,b], Zhibin Sun[c,d], Xuehai Tang[a,b,*]

[a] *School of Forestry and Landscape Architecture, Anhui Agricultural University, Hefei 230036 China*

[b] *Anhui Provincial Key Laboratory of Forest Resources and Silviculture, Anhui Agricultural University, Hefei 230036, China*

[c] *School of Mathematical Sciences, Nanjing Normal University, Nanjing 210023, China*

[d] *Jiangsu Key Laboratory for Numerical Simulation of Large Scale Complex Systems, Nanjing Normal University, Nanjing 210023, China*

Corresponding author:

Xuehai Tang;

School of Forestry and Landscape Architecture, Anhui Agricultural University, Hefei, 230036, China;

phone: +8618956022306;

E-mail: tangxuehai@ahau.edu.cn

Hyperspectral inversion of SPAD in *Camellia oleifera* based on integrated Taylor-CC and ANN with optimal wavelength selection

Xingzi Kou • Xinyue Jiang • Yue Zhang • Chunyu Ge • Zhibin Sun • Xuehai Tang

Abstract

Background Reliable assessment of leaf chlorophyll status can support the monitoring of *Camellia oleifera*, but point-based SPAD measurements provide limited spatial coverage. Hyperspectral data offer a non-destructive alternative, although the transferability of spectral-SPAD models across phenological stages remains uncertain. This study evaluated a wavelength-selection and modeling framework for canopy SPAD estimation in *Camellia oleifera*.

Methods Canopy hyperspectral reflectance and corresponding SPAD measurements were collected during the spring-shoot (April), fruit-expanding (August), and full-blooming (November) stages. Taylor expansion-based correlation coefficient (Taylor-CC) analysis was used to select core wavelengths across a range of correlation thresholds. Artificial neural network (ANN) models were developed and compared with random forest and support vector regression approaches. Stage-specific and cross-stage predictions were assessed using R2, root mean square error (RMSE), and the ratio of performance to deviation (RPD). Forward-difference analysis was used to characterize the local sensitivity of model predictions to reflectance perturbations.

Results Stage-matched models performed better on their corresponding datasets than in cross-stage predictions. The ANN models achieved R^2 values of 0.8350, 0.7984, and 0.7280 for April, August, and November, respectively, with stage-matched RPD values above 1.9. In cross-stage predictions, R^2 remained below 0.5 and RPD below 1.5. Taylor-CC-based wavelength configurations performed favorably relative to full-spectrum inputs in the reported comparisons, although the magnitude of improvement varied across model and stage combinations. The April model showed the lowest forward-difference sensitivity among the evaluated models; this indicates a weaker local response under the specified perturbation procedure, rather than

establishing physiological causation or external robustness.

Conclusions The results indicate limited transferability of canopy spectral-SPAD relationships among the three sampled datasets and support stage-aware calibration under the conditions studied. April is a candidate monitoring window for the sampled plantation and acquisition protocol, but validation across years, sites, and cultivars is required before recommending a generally optimal monitoring time. Taylor-CC wavelength selection combined with ANN modeling warrants further evaluation using independent field data.

Keywords *Camellia oleifera* · SPAD · Taylor-CC · Hyperspectral · Remote sensing · Artificial neural network

Introduction

Camellia oleifera Abel. is a unique and economically important woody oilseed tree in China, yielding tea oil with a well-balanced fatty acid profile and high nutritional value — often regarded as the "Oriental olive oil" [1,2]. The *Camellia oleifera* industry plays a strategic role in national edible oil security, rural revitalization in mountainous areas, and ecological conservation. In recent years, supportive policies have driven continuous expansion in planting area and yield, leading to rapid industrial scaling. However, widespread low-yield plantations and extensive management practices remain major challenges, severely limiting profitability and underscoring the urgent need to shift toward precision, intelligence-driven cultivation [3,4]. Critical to this transition is the rapid, non-destructive, and accurate monitoring of key physiological traits such as leaf chlorophyll content, which reflects photosynthetic capacity and nitrogen status. Such capability is essential for implementing science-based irrigation and fertilization, early pest/disease detection, and yield prediction—key to overcoming current management bottlenecks and unlocking yield potential [5-8]. Hence, developing efficient and reliable remote sensing inversion methods for chlorophyll content is of great practical significance for precision management in *Camellia oleifera* plantation.

Chlorophyll is the core pigment in plant photosynthesis, and its content serves as

a key physiological indicator reflecting photosynthetic capacity, nitrogen status, and stress responses [9-11]. Accurate, rapid estimation of leaf chlorophyll content is therefore vital for assessing plant nutrition and forest ecosystem health. Traditional laboratory methods, such as acetone-ethanol extraction followed by spectrophotometric, offer high accuracy but are destructive, labor-intensive, and unsuitable for large-scale dynamic monitoring [12,13]. Portable chlorophyll meters (e.g., SPAD-502 Plus) allow rapid field diagnosis but are limited in spatial coverage, hindering landscape-level applications [14,15]. Although SPAD values correlate well with chlorophyll content in many crops, including *Camellia oleifera* [16,17], the method remains point-based and contact-dependent, sampling only a few leaf spots that may not represent canopy-scale conditions, thus constraining operational efficiency in broad-area monitoring [18,19].

Hyperspectral remote sensing provides a promising alternative. By capturing reflectance across hundreds of narrow, contiguous bands, it detects subtle spectral variations linked to biochemical constituents such as chlorophyll [20-22]. Unlike few-band SPAD meters, hyperspectral data provide rich, continuous spectral information that reflects chlorophyll content and its interactions with leaf structure and water status. This enables the inversion of SPAD values—a reliable chlorophyll proxy—from spectral data, supporting non-destructive, spatially extensive monitoring [23,24]. In recent years, significant progress has been made in research on inverting crop SPAD values using hyperspectral data [25,26]. The research approach typically follows two core steps: (1) selecting informative wavelengths to reduce dimensionality and enhance model robustness, and (2) establishing quantitative models linking spectral features to SPAD values. Common wavelength-selection methods include Successive Projections Algorithm (SPA), Competitive Adaptive Reweighted Sampling (CARS), and Pearson Correlation Coefficient (PCC) [27-29]. More recently, the Taylor expansion-based Correlation Coefficient (Taylor-CC) has shown strength in identifying key wavelengths related to plant physiological parameters by leveraging functional relationships within continuous spectral curves. For modeling, machine learning algorithms such as Partial Least Squares Regression

(PLSR), Support Vector Machine (SVM), and Artificial Neural Networks (ANN) often outperform traditional linear models in capturing complex, nonlinear relationship between hyperspectral data and SPAD values [30-32].

For *Camellia oleifera*, existing studies have explored hyperspectral estimation of chlorophyll/SPAD, often using PLSR on full-spectrum or pre-selected wavelengths [33]. While these works provide a foundation, room for improvement remains. Most wavelength-selection approaches rely on discrete statistical correlations and may not fully exploit the continuous, functional nature of spectral curves, potentially missing shape- and derivative-related information [34]. Moreover, although ANN has demonstrated superior nonlinear fitting capability over PLSR in chlorophyll estimation for crops such as apple and winter wheat, studies integrating advanced wavelength selection like Taylor-CC with high-performance ANN models remain scarce for *Camellia oleifera* [35,36].

To address these gaps, this study proposes a novel hyperspectral inversion framework that integrates Taylor-CC wavelength selection with ANN modeling for SPAD estimating in *Camellia oleifera*. The Taylor-CC method treats spectra as continuous differentiable functions, approximating reflectance at one wavelength via Taylor expansion about another, thereby better capturing inter-wavelength functional dependence and selecting a compact, informative wavelength subset. The specific objectives are: (1) to identify an optimal wavelength set using Taylor-CC; (2) to develop an ANN inversion model based on these wavelengths and compare its performance against full-spectrum ANN, RF, and SVR; (3) to evaluate the accuracy and robustness of the integrated framework in estimating SPAD, aiming to provide an efficient, reliable technical solution for precise nutrient management and growth monitoring in *Camellia oleifera* cultivation.

Materials and methods

Study area

This study was conducted at the *Camellia oleifera* plantation of Anhui Dechang Seedlings Co., Ltd., located in Hepeng Town, Shucheng County, Lu'an City, Anhui

Province, China. The plantation center is at 116°46′56.377″ E, 31°11′25.040″ N (CGCS2000; Figure 1). The region has a humid subtropical monsoon climate, with a mean annual temperature of 15°C, annual precipitation of 1000 – 1700 mm, and an average frost-free period of 205 days. The area experiences distinct seasonal variation, characterized by hot, humid summers and relatively cold winters, with ample sunshine and rainfall evenly distributed across the growing season. The canopy spectra and SPAD data used in this article were all obtained through actual measurements conducted by the research team in the study area.

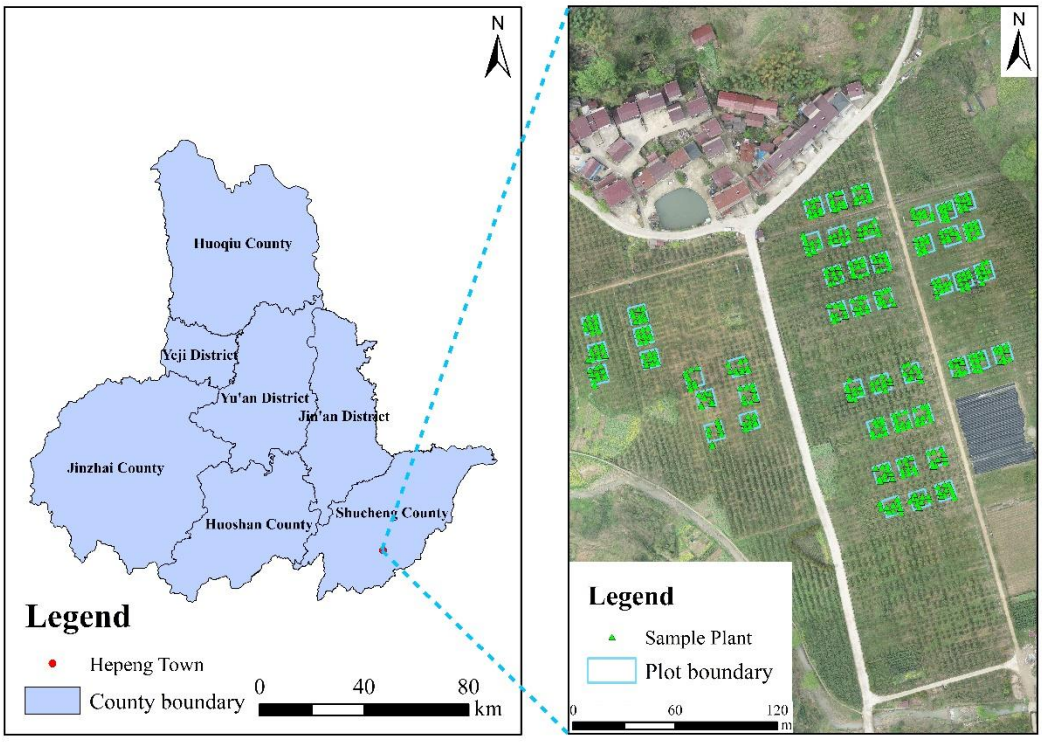


Fig. 1 Location of the study area

Sample selection

This study selected all *Camellia oleifera* varieties from the ‘Changlin’ series, including ‘Changlin’ No. 4, ‘Changlin’ No. 27, ‘Changlin’ No. 40, and "Changlin" No. 53. In 2010, 1-year-old *Camellia oleifera* seedlings were planted, and by the time this study began, all *Camellia oleifera* plants had entered a stable fruiting period. This study selected the spring shoot stage, fruit expanding stage, and full blooming stage as the research subjects.

Data collection

Hyperspectral reflectance

Canopy spectral data were collected using an ASD FieldSpec4 Wide-Res spectroradiometer (Analytical Spectrum Devices Inc., USA), covering the spectral range of 350~2500 nm. The instrument's default output provides resampled data at 1 nm intervals, resulting in 2151 spectral bands. Given that the average height of the sampled *Camellia oleifera* trees ranged from 2.0 to 2.5 m, and considering the 1.5 m length and 25° field of view of the spectrometer's standard fiber optic cable, a 3 m fiber optic extension was connected to the spectrometer via an adapter. The fiber optic probe was mounted in a handheld fixture and secured to the top of a telescopic pole. During measurements, the probe was positioned 2.0 m above the center of the tree crown, maintaining a consistent vertical height for all sampled trees. In this configuration, the probe—oriented perpendicular to the canopy plane—covered a circular field of view approximately 0.9 m in diameter, which encompassed most of the upper and mid-crown portions of an individual tree.

SPAD

Leaf chlorophyll content, expressed as SPAD values, was measured with a SPAD-502 portable chlorophyll meter (Minolta Camera Co., Osaka, Japan). The device operates based on the differential light absorption of chlorophyll at two specific wavelengths: 650 nm (red, corresponding to the chlorophyll absorption peak) and 940 nm (near-infrared, serving as a reference). The difference in transmittance at these two wavelengths is used to calculate a unitless SPAD index proportional to relative chlorophyll content, following the principle of the Beer-Lambert law.

Prior to fieldwork, the instrument was calibrated using the provided standard calibration plate. For each sampled plant, three to five fully expanded, healthy leaves were selected from different canopy positions (upper, middle, and lower layers) and orientations. On each leaf, measurements were taken at three pre-defined positions (tip, middle, and base), avoiding major veins and visibly damaged areas. At each position, three replicate readings were recorded, and the mean was calculated to represent the SPAD value for that leaf. The probe was held perpendicular to the leaf

surface during measurement to minimize the influence of leaf angle, and readings were taken under stable diffuse light conditions (e.g., on overcast days or during early morning hours) to reduce interference from direct sunlight. The average of all sampled leaves per plant was used as the representative SPAD value for that individual. All data were stored in the device memory and subsequently transferred to a computer for further processing and analysis.

Introduction to data usage

Field data and hyperspectral measurements were collected during three key phenological stages of *Camellia oleifera*: the spring shoot stage (April 2024), the fruit expanding stage (August 2023), and the full blooming stage (November 2023). Correspondingly, three time-series datasets were established and designated as Stage April, Stage August, and Stage November. Each dataset comprises the canopy hyperspectral reflectance (350–2500 nm, 240 samples per stage) and synchronously measured SPAD values, denoted as Data April/August/November and SPAD April/August/November, respectively. Using the Taylor-CC method, a set of core wavelengths was selected from each stage, recorded as Wavelength April, Wavelength August, and Wavelength November. Based on these three hyperspectral datasets, three separate ANN models were developed, referred to as Model April, Model August, and Model November.

Key wavelength selection of hyperspectrum in response to canopy SPAD of Camellia oleifera

In this study, dimensionality reduction was applied to the April, August, and November datasets to mitigate the challenges associated with high-dimensional hyperspectral data, thereby enhancing the robustness and generalizability of the subsequent predictive models.

Taylor Correlation Coefficient (Taylor-CC)

This study employs a novel method based on the Taylor expansion of a smooth function to estimate the correlation between hyperspectral signals at two distinct wavelengths [37]. This method assumes that the continuous reflectance function f is

194 at least twice differentiable at two adjacent wavelengths, x and y. Consequently, the
195 reflectance y measured at wavelength can be expressed as:

196 $$f(y) = f(x) + f'(x)(y-x) + \frac{1}{2}f''(x)(y-x)^2 + o((y-x)^2). \quad (1)$$

197 Thus, the estimated reflectance at y, using the derived information up to the
198 second-order derivative at x, becomes：

199 $$\hat{f}(y) = f(x) + f'(x)(y-x) + \frac{1}{2}f''(x)(y-x)^2. \quad (2)$$

200 Similarly, $f(x)$ and $\hat{f}(x)$ can be obtained by performing Taylor expansion at
201 wavelength y.

202 $$c(x,y) = \begin{cases} 1 - \frac{1}{2}(\left|\hat{f}(x) - f(x)\right| + \left|\hat{f}(y) - f(y)\right|), \left|\hat{f}(x) - f(x)\right| + \left|\hat{f}(y) - f(y)\right| \le 2 \\ 0, else \end{cases}. \quad (3)$$

203 Based on Equation (3), a correlation matrix $C(f,x) = [c(x_i, x_j)] \in R^{K \times K}$ for K
204 wavelengths can be constructed. It should be emphasized that $C(f,x)$ is a symmetric
205 matrix. Since a wavelength variable is perfectly correlated with itself, the diagonal
206 elements satisfy $C(x_i, x_i) = 1$ for $1 \le i \le K$. The off-diagonal elements
207 $0 \le c(x_i, x_j) \le 1, (i \ne j, 1 \le i, j \le K)$ quantify relationship between the $i-th$ and $j-th$
208 wavelengths.

209 ***Core wavelength screening***

210 The objective of core wavelengths screening is to reduce data dimensionality by
211 eliminating wavelengths that exhibit high mutual correlation. While common
212 dimensionality reduction techniques include Principal Component Analysis,
213 Equidistant Mapping and Local Linear Embedding. This study adopts the Taylor-CC
214 method to identify a subset of highly informative, correlated core wavelengths,
215 thereby effectively reducing spectral dimensionality [38].

216 Figure 2a–2c present the correlation matrices obtained by the Taylor-CC method
217 for Data April, Data August, and Data November, respectively, using a threshold of
218 0.8. Similar correlation patterns are observed under other threshold settings. Because
219 the spectral reflectance can be treated as a continuous smooth function, the correlation

matrices display a pronounced diagonal structure: correlation coefficients are significantly elevated only within segmented intervals along the main diagonal, showing a gradient from blue to yellow that reflects high local self-similarity. In contrast, correlation values are nearly zero in off-diagonal regions, indicating negligible cross-correlation between distant wavelengths and underscoring the global independence of the spectral signals. This diagonal pattern clearly reveals the local correlation and global independence of the hyperspectral data at different scales, providing a sound theoretical basis and a visual reference for subsequent feature selection, segment-specific modeling, and noise separation.

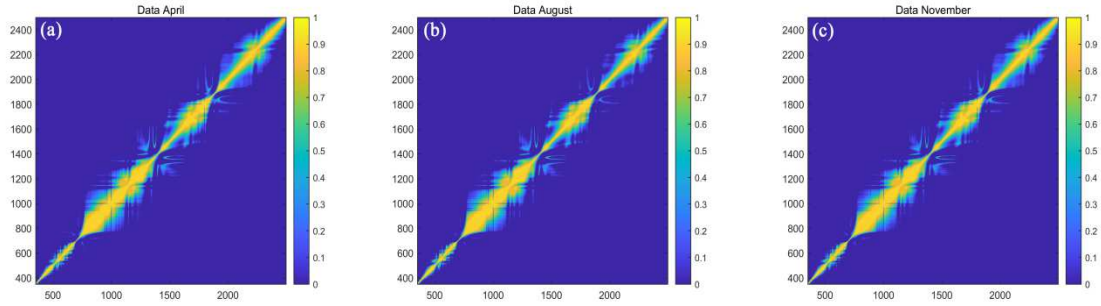


Fig. 2 (a) The correlation of Wavelength April from the Taylor-CC method; (b) the correlation of Wavelength August from the Taylor-CC method; (c) the correlation of Wavelength November from the Taylor-CC method

Hyperspectral estimation model of Camellia oleifera canopy SPAD

According to the correlation between the wavelengths selected by the method of Taylor-CC, the core wavelengths under different thresholds (from 0.5 to 0.99, step size 0.01) can be selected to input the models.

Artificial neural network (ANN)

Since its inception in the 1940s, the Artificial Neural Network (ANN) has evolved into a foundational machine learning model, particularly with the recent advances in computational power and the availability of large-scale datasets. ANN is a mathematical model inspired by the structure and functional principles of biological neural networks [39-41]. After understanding and abstracting the structure of human brain and the response mechanism of external stimuli, it simulates the processing mechanism of complex information by the neural system of human brain on the basis of network topology knowledge.

ANN is a machine learning model that consists of many nodes (also known as neurons) connected to each other. Each node represents a specific output function called the activation function. The direction of the connection between each two nodes represents the direction of the signal transmission with weights. It mainly consists of three parts: input layer, hidden layer and output layer. The input layer receives the external input signal and passes it to the hidden layer. Then the hidden layer processes and weights the input signal with an activation function to generate the output signal and pass it to the next layer or final output layer. The output layer processes and weights the output of the hidden layer with the activation function to generate the final output result.

Sensitivity of a neural network

Calculating the sensitivity of a neural network is a method for evaluating the degree to which input features affect the output results. It provides information about the relative importance of each input feature, helping with tasks such as variable selection, interpretation of model predictions [42], and optimization of models [43].

In this study, Forward Difference Method in Numerical Methods was used to calculate the sensitivity of a neural network [44].

The specific computational steps are as folloWangws:

(1) Input data preparation: Define the input feature matrix $X \in RN \times p$, where N denotes the number of samples and P represents the dimension of input feature. Each row of the matrix corresponds to an independent sample.

(2) Determination of target neuron: Identify the target output nodes to be analyzed. In this study, the single regression output of the network (e.g., the predicted SPAD value) is taken as the target variable for sensitivity calculation.

(3) Minor disturbance of input features: Apply a positive disturbance to the j-th input feature by increasing all sample values corresponding to the j-th feature in the input matrix by a preset disturbance value δ, yielding a perturbed input matrix:

$$X_{perturb}^{(j)} = X, X_{perturb}^{(j)}[j,:] = X[j,:] + \delta \quad (4)$$

The disturbance term δ must be sufficiently small to ensure that a local linear

approximation between the input feature's original value and the perturbed value holds. In this study, the disturbance amount is set as $\delta = 0.001$.

(4) Calculation of network output before and after perturbation: Input the original input matrix X and the perturbed input matrix $X_{perturb}^{(j)}$ into the trained network separately, and obtain the corresponding output vectors.

$$\hat{Y}_{\text{base}} = f(X),\ \hat{Y}_{perturb}^{(j)} = f\left(X_{perturb}^{(j)}\right) \tag{5}$$

where $f(.)$ denotes the mapping function of the trained neural network.

(5) Sensitivity calculation: The partial derivative of the network output with respect to the *j-th* input feature is estimated using the forward difference approximation. The mean value of the absolute values across all samples is taken as the sensitivity index of the feature:

$$S_j = \frac{1}{N}\sum_{n=1}^{N}\left|\frac{f\left(x_n + \delta e_j\right) - f\left(x_n\right)}{\delta}\right| \tag{6}$$

where e_j denotes the unit vector in the j-th coordinate direction, and x_j represents the input vector of the *n-th* sample.

(6) Iteration over all input features: Repeat steps (3) to (5) sequentially for *j=1,2,...,p*. Finally, the sensitivity vector for all input features is obtained.

$$S = \left[S_1, S_2, ..., S_p\right] \tag{7}$$

Model building

According to the correlation between the wavelengths selected by the method of Taylor-CC, the core wavelengths under different thresholds (from 0.5 to 0.99, step 0.01) can be selected to input the models. Using Data April, Data August, Data November and selected core wavelengths, ANN is used to compare performances of Model April, Model August, Model November. In this study, all input/output data (e.g., reflectance and SPAD) were scaled onto [-1,1], so as to eliminate bias due to differences in data scales.

A single-hidden-layer Feedforward Neural Network was employed as the

modeling framework. Two key hyperparameters—the number of hidden nodes and the training function—were determined via systematic grid search coupled with 5-fold cross-validation on the modeling dataset. The candidate set for the number of hidden nodes was set as H = {5,10,20,30,50}, while the candidate set for the training function was defined as T = {trainlm, trainbr, trainscg}, corresponding to the Levenberg-Marquardt, Bayesian regularization, and conjugate gradient algorithms, respectively. All combinations in the above candidate sets were traversed, resulting in a total of $|H| \times |T| = 15$ hyperparameter combinations. For each hyperparameter combination $(h,\tau) \in H \times T$, the modeling data were split into five folds. The network was trained on four folds and validated on the remaining fold. The mean coefficient of determination $\overline{R^2}$ over the five folds was taken as the performance metric. The optimal combination is defined as:

$$(h^*, \tau^*) = \text{argmax}_{(h,\tau)\in H\times T}\overline{R^2}(h,\tau) \quad (8)$$

Once (h^*, τ^*) was determined, the network was retrained on the full modeling dataset using (h^*, τ^*), and the resulting model is used for subsequent prediction on the independent test set and sensitivity analysis. The unified settings for network training are as follows: the maximum number of training epochs is 1000, the training target error is 10 e^{-6}, and maximum consecutive validation failures is 10. The data were split 8:2 into training and validation subsets.

The entire workflow of network training, prediction evaluation, and sensitivity analysis were conducted over 50 independent repeated trials. In each trial, hyperparameter search and model training were performed independently, and the following results were recorded: prediction performance metrics (R^2, RMSE, RPD), the optimal hyperparameter combination (h^*, τ^*), and the sensitivity vector of each input feature. Finally, the model predictive performance and feature importance were comprehensively evaluated using the mean and standard deviation of the 50 experimental results, in order to mitigate the impact of random initialization on the stability of the conclusions.

$$S^{(i)} \in R^{1\times p}(i = 1,2,...,50) \quad (9)$$

To systematically compare the generalization ability of the models constructed for the three phenological periods, each established model was adopted not only to estimate SPAD for its corresponding stage but also to perform cross-period SPADI prediction across the other two periods. By applying Model April, Model August and Model November to the prediction tasks of all three growth stages respectively, and comparing the predicted values of each model with the measured values, the prediction accuracy of the three models was quantitatively obtained. Accordingly, a comprehensive evaluation of their cross-period generalization performance was realized.

This study further to evaluate the SPAD prediction contribution of characteristic wavelengths at different growth stages—Wavelength April, Wavelength August, and Wavelength November. It is assumed that the core wavelengths of each growth stage have been screened using the Taylor-CC method under given threshold conditions, and the corresponding wavelength serial numbers will be recorded. Since the SPAD prediction process relies on the synergistic effect of core wavelengths and the prediction model, differences in input wavelengths will directly affect model construction and prediction outcomes. This study designs two wavelength input strategies to quantitatively evaluate the prediction efficiency of characteristic wavelengths at different stages. Taking the prediction of April-stage SPAD by Model August as an example: if the wavelength sequential numbers of Wavelength August are adopted to extract the corresponding wavelengths from the reflectance data of the April stage for model construction, this strategy is denoted as Method I; if the wavelength sequential numbers of Wavelength April itself are adopted to extract the corresponding wavelengths from the reflectance data of the April stage for model construction, this strategy is denoted as Method II. It should be noted that due to significant differences in spectral characteristics among different growth stages, the number and indices of core wavelengths screened by the Taylor-CC method for each stage under the same threshold conditions are usually inconsistent.

Evaluation metrics

Cross-validation can be used to evaluate the performances of the three models, Model April, Model August, Model November, and the performance of the two methods, Method I and Method II. In this study, root mean square error (RMSE), determination coefficient (R^2), residual prediction difference (RPD) are used to evaluate the performance of the above models and methods. They are defined as:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - x_i)^2} \quad (10)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(x_i - y_i)^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \quad (11)$$

$$RPD = \frac{SD}{RMSE} \quad (12)$$

where x_i is the i-th observed value, y_i is the i-th model predicted value, N is the size of x and y, $\bar{x}$ is the average of x, and $\bar{y}$ is the average of y, SD is the standard deviation of the observed values. Model performs better when RMSE is smaller, R^2 or RPD is higher.

Result

Threshold-dependent selection of core SPAD-sensitive spectral wavelengths based on Taylor-CC method

To reduce the arbitrariness associated with evaluating model and wavelength performance under a single correlation threshold, 50 thresholds levels (0.5 to 0.99, step 0.01) were tested. Using the Taylor-CC method, wavelengths exhibiting high mutual correlation were iteratively removed; the remaining subset, characterized by low or no inter-correlation, was retained as the core wavelengths set for each period (denoted as Wavelength April, August, and November). The core wavelengths selected for the three datasets under the 50 thresholds are displayed in Fig. 3a-c.

The results indicated a clear threshold-dependent pattern in the selected wavelengths across all three periods. As the threshold increased, the number of selected wavelengths rose consistently, with the August dataset yielding the largest selected set, followed by November and then April. This suggested that the fruit expanding stage exhibited richer spectral information sensitive to SPAD variation.

High-frequently selected wavelengths were predominantly located in the visible blue region, the red- edge to near-infrared transition zone, and portions of the short-wave infrared, underscoring the stability and consistency of these spectral regions for SPAD retrieval across seasons. In contrast, wavelengths in the 900-1100 nm, 1500-1700 nm, and >2000 nm ranges showed more scattered selection patterns and greater sensitivity to threshold changes, indicating that their contribution to SPAD estimation is more context-dependent. Overall, a subset of key sensitive wavelengths remained relatively stable across seasons, while the number and distribution of auxiliary wavelengths varied notably with seasonal conditions.

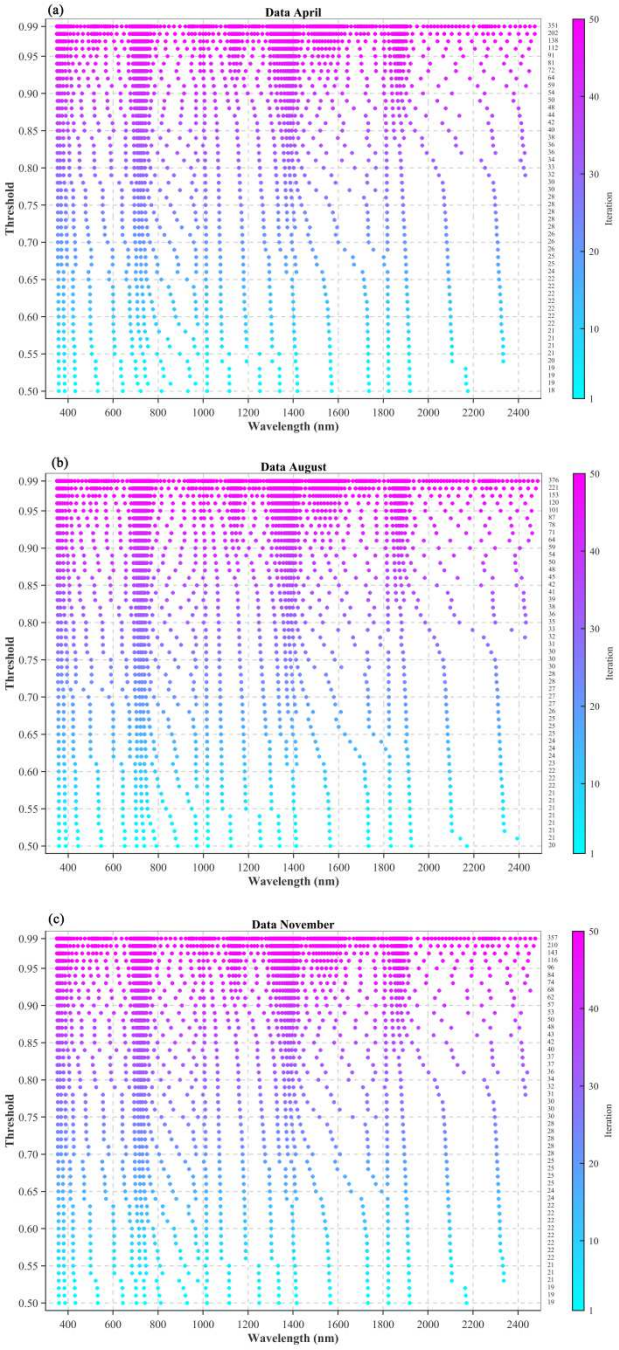


Fig. 3. (a) The core wavelengths of Data April identified via Taylor-CC method at 50 different thresholds; (b) the core wavelengths of Data August identified via Taylor-CC method at 50 different thresholds; (c) the core wavelengths of Data November identified via Taylor-CC method at 50 different thresholds. The specific number of wavelengths selected by the Taylor-CC method is marked on the right side.

Evaluation of model inversion performance

The SPAD prediction performance of Model April, Model August, and Model November under cross-stage scenarios was systematically evaluated using R^2, RMSE and RPD, the results are summarized in Tables 1-3. All three models performed best

406 on their corresponding phenological stages. Model April achieved the highest
407 accuracy for April SPAD (R^2 = 0.8350 ± 0.0256, RPD = 2.4882 ± 0.1944, and RMSE
408 = 2.0963 ± 0.1605). Model August and Model November also obtained satisfactory
409 stage-specific results, with R^2 values of 0.7984 ± 0.0192 and 0.7280±0.0363,
410 respectively, and RPD ＞ 1.9, confirming the advantage of using stage-dedicated
411 models.

412 In contrast, cross-stage prediction led to a pronounced decline in performance.
413 For example, Model August yielded an R^2 of only 0.1936±0.5380 for April SPAD,
414 with RPD ＜ 1.3. Model April attained an R^2 of merely 0.4152±0.1661 for August
415 SPAD. Across all cross-stage trials, R^2 remained below 0.5 and RPD below 1.5,
416 indicating that the models lack the generalizability required for reliable SPAD
417 estimation outside their original phenological context. Notably, Model April exhibited
418 the most consistent performance in cross-stage predictions, whereas Model August
419 presented larger fluctuations and lower stability.

420 **Table 1** Performance assessments of SPAD April predicted by Model April, Model August, Model
421 November when using Method I. All values are derived from the test set and expressed as mean ±
422 standard deviation (SD) across 50 repeated training iterations.

	Model April	Model August	Model November
R^2	0.8350 ± 0.0256	0.1936 ± 0.5380	0.4111 ± 0.2240
RMSE	2.0963 ± 0.1605	4.4561 ± 1.3330	3.9085 ± 0.7126
RPD	2.4882 ± 0.1944	1.2504 ± 0.3111	1.3690 ± 0.2440

423 **Table 2** Performance assessments of SPAD August predicted by Model April, Model August,
424 Model November when using Method I.

	Model April	Model August	Model November
R2	0.4152 ± 0.1661	0.7984 ± 0.0192	0.3666 ± 0.2843
RMSE	4.1375 ± 0.5293	2.4462 ± 0.1175	4.2569 ± 0.8560
RPD	1.3390 ± 0.1483	2.2393 ± 0.1105	1.3259 ± 0.2191

425 **Table 3** Performance assessments of SPAD November predicted by Model April, Model August,
426 Model November when using Method I.

	Model April	Model August	Model November
R2	0.4111 ± 0.2240	0.3666 ± 0.2843	0.7280 ± 0.0363
RMSE	3.9085 ± 0.7126	4.2569 ± 0.8560	3.1713 ± 0.2101
RPD	1.3690 ± 0.2440	1.3259 ± 0.2191	1.9337 ± 0.1274

427 **Performance assessment using metrics**

In *the* cross-stage prediction of SPAD April, Method II demonstrated clear advantages. When predicting April SPAD with Model November, Method II attained R^2 = 0.4541 ± 0.1925, an increase of 0.043 over Method I (R^2 = 0.4111 ± 0.2240), while RMSE decreased from 3.9085 ± 0.7126 to 3.7675 ± 0.6601. The smaller standard deviations of both *metrics* under Method II indicate improved prediction stability. A similar, though less pronounced, trend was observed for Model August predicting April SPAD: R^2 rose from 0.1936 to 0.2191, RMSE decreased from 4.4561 to 4.2955, and RPD increased slightly from 1.2504 to 1.3257.

In the cross-stage prediction of SPAD August, Method II provided the most notable stability gain in the cross-stage scenario of Model April. Here, R^2 improved from 0.4152 ± 0.1661 (Method I) to 0.4356 ± 0.1202 (Method II) , and the standard deviation of RMSE and R^2 were reduced from 0.5293 to 0.4202, and from 0.1661 to 0.1202. This was the combination where Method II achieved the greatest improvement in stability among all cross-stage scenarios. For Model November predicting August SPAD, the two methods yielded comparable results.

In the cross-stage prediction of SPAD November, the difference between the two strategies were marginal. Model August produced nearly identical outputs under both methods. For Model April, Method I yielded slightly higher R^2 (0.4111 ± 0.2240) and RPD (1.3690 ± 0.2440) than Method II (R^2 = 0.4023 ± 0.1275, RPD = 1.3140 ± 0.1180), though the latter exhibited smaller standard deviations and thus greater robustness.

Table 4 Performance assessments of SPAD April predicted by Model April, Model August, Model November when using Method II.

	Model April	Model August	Model November
R2	0.8350 ± 0.0256	0.2191 ± 0.7239	0.4541 ± 0.1925
RMSE	2.0963 ± 0.1605	4.2955 ± 1.5859	3.7675 ± 0.6601
RPD	2.4882 ± 0.1944	1.3257 ± 0.3567	1.1480 ± 0.2457

Table 5 Performance assessments of SPAD August predicted by Model April, Model August, Model November when using Method II.

	Model April	Model August	Model November
R2	0.4356 ± 0.1202	0.7984 ± 0.0192	0.3715 ± 0.3159
RMSE	4.0761 ± 0.4202	2.4462 ± 0.1175	4.2291 ± 0.9084

RPD	1.3540 ± 0.1316	2.2393 ± 0.1105	1.3371 ± 0.2198

Table 6 Performance assessments of SPAD November predicted by Model April, Model August, Model November when using Method II.

	Model April	Model August	Model November
R2	0.4023 ± 0.1275	0.3715 ± 0.3159	0.7280 ± 0.0363
RMSE	4.6879 ± 0.4694	4.2291 ± 0.9084	3.1713 ± 0.2101
RPD	1.3140 ± 0.1180	1.3371 ± 0.2198	1.9337 ± 0.1274

Performance assessment using sensitivity

According to the sensitivity analysis (Tables 7-12), the computed sensitivity of the models was influenced jointly by the target growth stage, the month of model training, and the wavelength-input method. Overall, sensitivity tended to increase the training period shifted from April to August and November. Moreover, Method II generally exhibited lower sensitivity values and smaller dispersion compared with Method I.

Among the three models, Model April maintained the lowest sensitivity across all prediction targets, with relatively small fluctuations, indicating a weaker response to phenological shifts and greater stability in cross-stage application. In contrast, Model August and Model November typically exhibited higher sensitivity. Specifically, Model November showing the highest values in most scenarios along with lager dispersion. This pattern suggests Model December is more reactive to changes in the target variable and more susceptible to inter-stage differences—a trait that was especially pronounced with Method I.

Although the late-season models displayed higher overall responsiveness, this elevated sensitivity did not translate into better cross-stage adaptability. On the contrary, it was frequently accompanied by greater outcome instability. By comparison, Model April consistently showed low sensitivity and better agreement across cross-stage predictions, reflecting stronger robustness. Furthermore, while preserving the relative ranking of model sensitivities, Method II further reduced both the magnitude of sensitivity and its dispersion, thereby enhancing the stability of cross-stage predictions.

In summary, the sensitivity analysis confirms that Model April is the least affected by growth-stage variations and demonstrates the most stable cross-stage

performance, offering clear advantages in terms of model robustness and overall applicability.

Table 7 Sensitivity of SPAD April predicted by Model April, Model August, Model November when using Method I.

	Model April	Model August	Model November
Sensitivity	67.6119 ± 30.6093	140.8332 ± 50.4041	191.5266 ± 280.2925

Table 8 Sensitivity of SPAD April predicted by Model April, Model August, Model November when using Method II.

	Model April	Model August	Model November
Sensitivity	67.6119 ± 30.6093	140.0873 ± 54.5715	159.6048 ± 83.4655

Table 9 Sensitivity of SPAD August predicted by Model April, Model August, Model November when using Method I.

	Model April	Model August	Model November
Sensitivity	55.2748 ± 22.9767	156.9412 ± 93.3075	191.5266 ± 280.2925

Table 10 Sensitivity of SPAD August predicted by Model April, Model August, Model November when using Method II.

	Model April	Model August	Model November
Sensitivity	48.5263 ± 23.5530	140.0873 ± 54.5715	163.698 ± 116.3468

Table 11 Sensitivity of SPAD Novermber predicted by Model April, Model August, Model November when using Method I.

	Model April	Model August	Model November
Sensitivity	65.6105± 31.3554	143.4885 ± 53.8420	149.6603 ± 78.8458

Table 12
Sensitivity of SPAD November predicted by Model April, Model August, Model November when using Method II.

	Model April	Model August	Model November
Sensitivity	61.8219 ± 30.4770	136.9420 ± 55.0060	149.6603 ± 78.8458

Discussion

Rationale for selecting spectral feature contribution evaluation methods

Accurately quantifying the contribution of individual hyperspectral wavelength to SPAD estimation is crucial for interpreting model decisions, identifying key spectral features, and enhancing model transparency. In this study, the sensitivity of each input wavelength in the ANN was assessed using the forward-difference method. This approach approximates the partial derivative of the network output (predicted SPAD) with respect to a given wavelength by applying a small, uniform perturbation (δ =

0.001) to its reflectance and computing the mean absolute change in the output across all samples. The resulting sensitivity index provides a continuous, comparable measure of the marginal influence of each wavelength on the final prediction. The sensitivity analysis revealed pronounced differences among the three stage-specific models. Model April consistently exhibited the lowest sensitivity (e.g., 67.6119 ± 30.6093 for April SPAD prediction) and the smallest fluctuation across all prediction scenarios, suggesting that its predictions rely on a broad synergy of wavelengths rather than on a few highly sensitive ones. This characteristic likely contributes to its greater robustness when applied to inputs from other phenological stages. In contrast, Model November showed the highest sensitivity values and the greatest fluctuation (e.g., 191.5266 ± 280.2925 for April SPAD prediction), indicating that its decisions are more dependent on a narrow set of spectral features that are prominent only under specific phenological conditions. This pattern is consistent with previous findings that over-reliance on a small number of high-contribution features can reduce a model's adaptability to spatiotemporal variations.

For the fully-connected ANN architecture used here, the forward-difference method offers a practical balance between computational efficiency and interpretability, justifying its selection over more computationally intensive alternatives such as permutation importance [45]. Although gradient-based attribution methods widely used in deep convolutional networks, for the fully connected ANN employed in this study, the forward difference method had proven effective in identifying key wavelengths with physiological significance. For instance, the highly sensitive wavelengths identified by this method were consistently concentrated in the red-edge region (720-750 nm), which aligned well with the physical mechanisms of chlorophyll absorption and scattering in leaves. This finding was also consistent with the conclusion that red-edge parameters are the core features for chlorophyll inversion, as determined using continuous wavelet analysis [46]. Notably, although red-edge wavelengths were identified as important in all models, their absolute sensitivity values were substantially lower in Model April than in the later-season models. This finding implies that the relative influence of the underlying physical drivers of the

spectral signal—such as chlorophyll content, leaf structure, and background moisture—shifts across phenological stages. Such stage-dependent weighting of spectral features provides a principled basis for developing phenology-aware inversion schemes, rather than relying on a single universal model.

Model selection and advantages for SPAD inversion

In hyperspectral remote sensing of vegetation chlorophyll, conventional machine learning algorithms such as RF and SVM are frequently employed due to their computational efficiency and interpretability, and have demonstrated utility in crop monitoring [47]. However, the relationship between spectral data and leaf chlorophyll content (represented by SPAD) is inherently high-dimensional, nonlinear, and modulated by phenology, stress, and canopy structure. Traditional models often struggle to capture these complex interactions fully. In contrast, ANNs excel at learning high-order, nonlinear mappings from high-dimensional spectral inputs to target biochemical parameters, owing to their multi-layer, distributed representation [48]. This study conducted modeling based on hyperspectral data from three key phenological stages of *Camellia oleifera*. The results indicated that the ANN model achieved the highest overall accuracy and stability in SPAD estimation, corroborating similar findings in other crops [49].

A key contribution of this work is the proposed synergistic framework that couples functional wavelength selection via Taylor-CC with ANN-based nonlinear modeling. Unlike conventional filter-based methods (e.g., CARS, SPA) that rely on point-wise statistical correlations, the Taylor-CC method treats the reflectance spectrum as a smooth, differentiable function. By computing first- and second-order Taylor expansions between adjacent wavelengths, it quantifies local functional dependence, thereby more reliably identifying a compact set of informative, low-redundancy features in critical spectral regions such as the red-edge and water-absorption bands [50]. This approach not only reduces input dimensionality but also provides a physiologically interpretable basis for feature selection.

When the resulting core wavelengths were used as input to the ANN, the model

attained superior predictive performance on an independent test set, with significantly better stability than models trained on full-spectrum data or features selected by traditional methods. The framework effectively filters out non-sensitive bands and noise, allowing the ANN to focus on robust chlorophyll-spectrum relationships. Consequently, the model maintained relatively high accuracy even in cross-phenological prediction tasks, underscoring the value of a functionally grounded feature-selection step [51].

Moreover, the Taylor-CC method enables a transparent, multi-metric evaluation of different models across varying data subsets (e.g., growth stages, homogeneous zones) by simultaneously considering standard deviation, correlation coefficient, and root-mean-square error—offering a more comprehensive assessment than single performance metrics [52,53]. In summary, the integration of Taylor-CC wavelength selection with ANN modeling addresses a common bottleneck in hyperspectral vegetation parameter retrieval: the synergy between physically informed feature reduction and powerful nonlinear fitting. This framework not only enhances the interpretability and generalization ability of the inversion model but also provides a transferrable methodological foundation for precision monitoring of *Camellia oleifera* and similar woody crops.

Optimal inversion strategy and application of integrated approach

The ANN model demonstrated consistent predictive capability across 50 independent runs. The averaged performance metrics were $R^2 = 0.8313$ (SD = 0.0331), RMSE = 2.1178 (SD = 0.4819), and RPD = 2.4631 (SD = 0.1751), indicating robust and stable SPAD estimation. Analysis of the performance distribution across runs revealed that a subset of iterations substantially exceeded the average. Setting a threshold at the top 25% ($R^2 \geq 0.8460$) identified 13 high-performance runs, with the best iteration achieving $R^2 = 0.8777$, RMSE = 1.8101, and RPD = 2.8651. Several runs approached an RPD of 3.0, which is commonly regarded as the threshold for excellent predictive ability. These results suggest that the upper performance bound of the model remains considerable, contingent on favorable hyperparameter and weight-initialization

conditions.

A distinct performance outlier was observed near the 13th run, where all metrics degraded simultaneously. This anomaly likely stemmed from an unfortunate random split that created an unrepresentative training-validation distribution or from weight initialization trapping the network in a poor local minimum. Importantly, the vast majority of runs exhibited stable metrics within the expected range, confirming that the outlier represents an occasional convergence failure rather than a structural flaw in the model architecture. The adopted strategy of reporting the mean and standard deviation over 50 independent runs is designed to dampen the influence of such stochastic variations, thereby yielding a more representative and reliable estimate of the model's true performance.

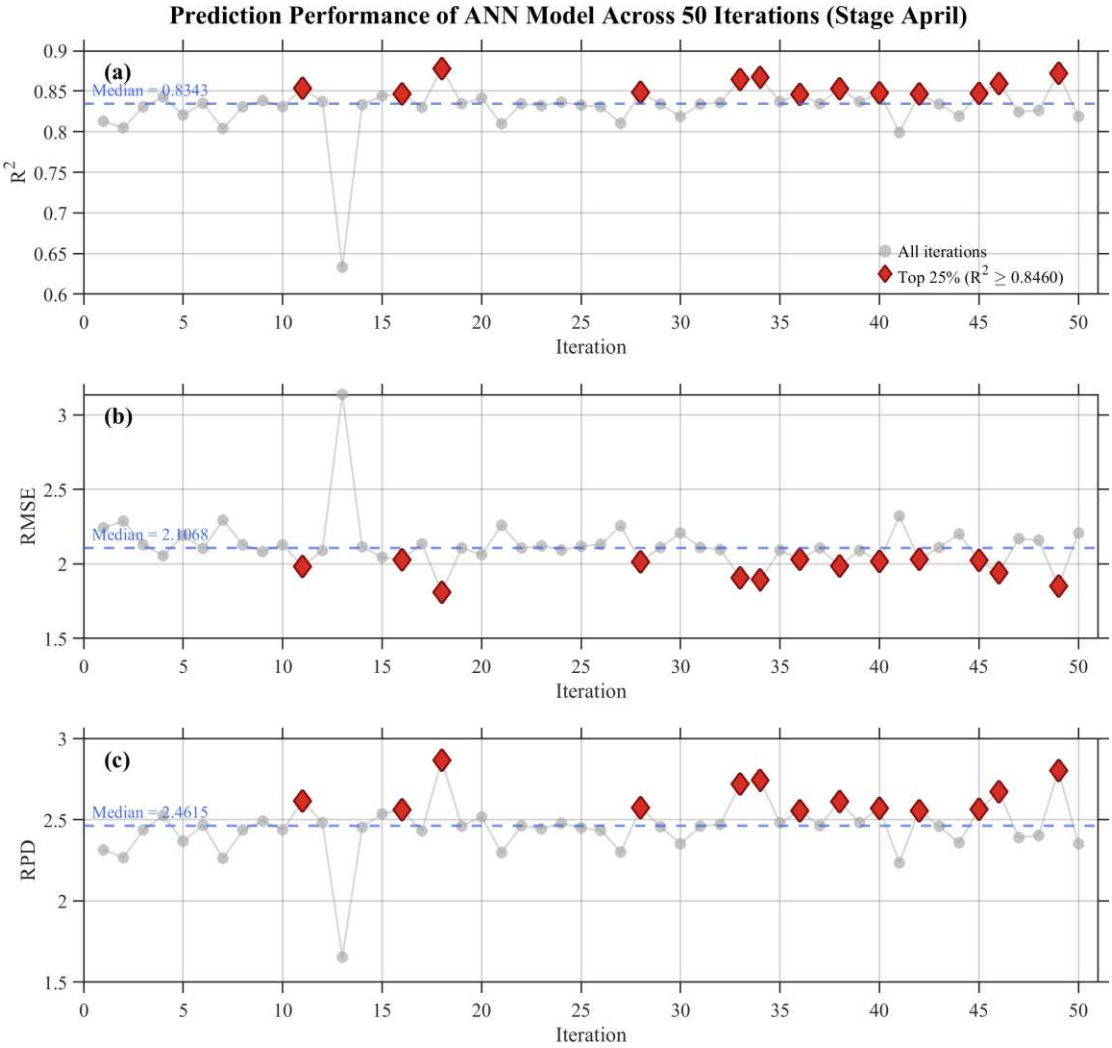


Fig. 4 Prediction performance metrics (R^2, RMSE, and RPD) of Stage April ANN model across 50 repeated training iterations for leaf SPAD inversion. Red diamonds indicate the top 25% iterations ($R^2 \geq 0.8460$); dashed blue lines denote the median of each metric

This study investigated SPAD prediction using hyperspectral data from three key phenological stages: March 2024 (i.e. one-stage), August 2023 (i.e. two-stage), and November 2023 (i.e. three-stage). A total of nine phenological-stage combination experiments for SPAD prediction were conducted, and the accuracy of three modeling approaches was compared. For both RF and Support Vector Regression (SVR), the SCabs algorithm was employed to screen sensitive feature wavelength ranges from the response spectra through 50 repeated experiments. In each repetition, RF utilized five-fold cross-validation with grid search to optimize the number of decision trees and the minimum leaf node size, while SVR simultaneously optimized the kernel function (e.g., linear and RBF), penalty coefficient C, and the γ parameter for the RBF kernel. For the full-spectrum data, feature pre-selection was performed directly, and an ANN served as the baseline model, with five-fold cross-validation used to select the optimal number of hidden layer nodes and training function. The generalization capability of all models was evaluated using R^2, RMSE, and RPD, with the results presented in Fig. 5.

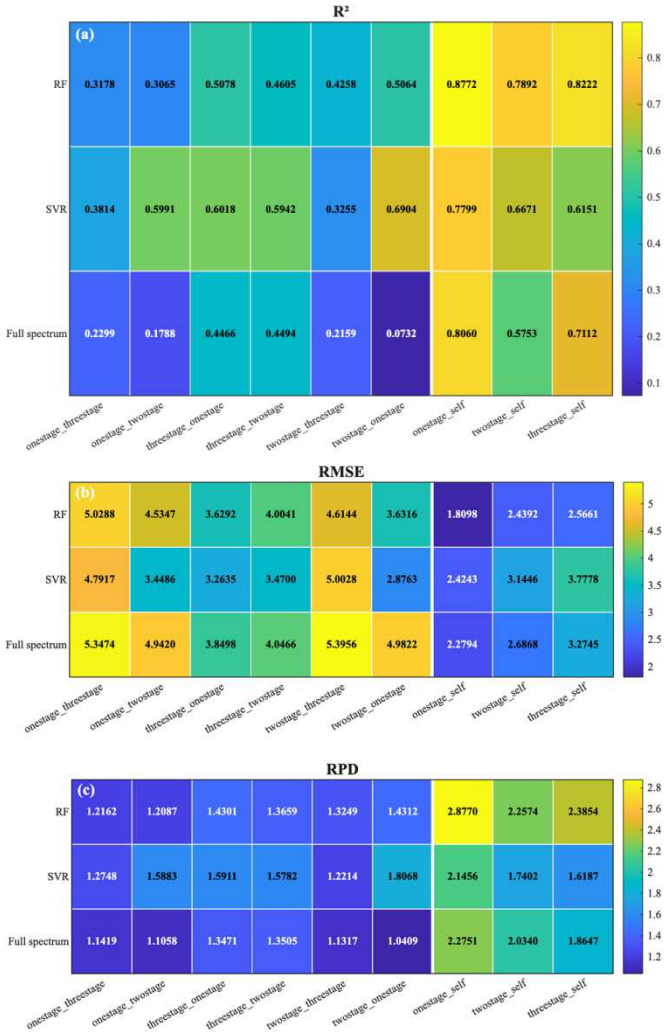


Fig. 5.A comparison of SPAD estimation accuracy among three modeling methods for different phenological stage combinations.

In the inversion of leaf SPAD values, among all phenological groups screened by Taylor-CC, the performance of RF and SVR was comparable, with RF exhibiting slightly better optimal performance. In the single-phenology self-prediction (onestage_self) group, RF achieved the best accuracy with R^2 = 0.8772, RMSE = 1.8098, and RPD = 2.8770. In contrast, the ANN model employed in this study demonstrated significantly higher accuracy under the exact same phenological grouping, indicating its superior ability to fit the complex nonlinear relationship between spectral data and SPAD values. Under any given modeling method, all Taylor-CC-screened phenological groups consistently outperformed the full-spectrum models. Taking RF as an example, the optimal R^2 for the full-spectrum model was only 0.7112, whereas the Taylor-CC-screened onestage_self group showed a 23.3%

improvement in R^2, a 44.8% reduction in RMSE, and a 54.3% increase in RPD. In the extreme scenario of cross-phenology prediction, the full-spectrum model achieved an R^2 of only 0.0732 in the twostage_onestage group, nearly losing its predictive capability. In contrast, the corresponding Taylor-CC-screened group still provided effective predictions, demonstrating its ability to effectively eliminate redundant information and noise, thereby significantly enhancing model accuracy and robustness.

The experiments indicate that phenological matching significantly influences inversion performance: self-phenology prediction far outperforms cross-phenology prediction, with the onestage_self group achieving the best results among all combinations. In conclusion, this ablation study validates the necessity and superiority of the ANN modeling method combined with the Taylor-CC wavelength screening strategy in the hyperspectral inversion of leaf SPAD values.

Conclusions

This study developed and validated a novel framework that integrates Taylor-CC wavelength selection with ANN modeling for the efficient, non-destructive estimation of SPAD in *Camellia oleifera*. Based on a systematic analysis of hyperspectral data from three key phenological stages, the principal findings are as follows:

(1) The integrated Taylor-CC-ANN framework demonstrated clear superiority for SPAD inversion. By identifying a compact, highly informative subset of wavelengths through functional-correlation analysis, the approach effectively mitigates the high-dimensionality and redundancy of hyperspectral data. Models built on these core wavelengths achieved high, stage-specific accuracy (best model: $R^2 = 0.8350$, $RPD = 2.4882$), confirming the framework's ability to enhance both precision and robustness.

(2) SPAD inversion exhibited strong phenological specificity. While the April (spring shoot) model showed the highest accuracy and stability, all models declined significantly when applied cross-stage. This underscores that canopy spectral-chlorophyll relationships shift dynamically across the growing season,

necessitating stage-dedicated modeling for reliable, season-long monitoring.

(3) The spring shoot stage (April) is the optimal window for canopy physiological monitoring. The April model and its selected wavelengths displayed the greatest robustness and the least performance degradation in cross-stage predictions. Consequently, this period is recommended for building models with strong temporal extrapolation capability. For other stages where data are scarce, the April-based model and feature set provide a practical approximation.

In summary, this work delivers a reliable, high-precision technique for estimating chlorophyll in *Camellia oleifera* and establishes phenological stage as a critical driver of inversion performance, with the spring shoot stage being the most effective monitoring window. These findings support the move toward precision management in *Camellia oleifera* cultivation. Future efforts should focus on integrating the framework with UAV and satellite platforms to enable regional-scale dynamic monitoring.

Abbreviations

ANN: Artificial neural network

CARS: Competitive adaptive reweighted sampling

PCC: Pearson correlation coefficient

PLSR: Partial least squares regression

R^2 : Coefficient of determination

RBF: Radial basis function

RF: Random forest

RMSE: Root mean square error

RPD: Ratio of performance to deviation

SD: Standard deviation

SPA: Successive projections algorithm

SPAD: Soil and plant analyzer development

SVM: Support vector machine

SVR: Support vector regression

Taylor-CC: Taylor expansion-based correlation coefficient

UAV: Unmanned aerial vehicle

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and analysed during the present study are available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no conflict of interest.

Funding

This study was financially supported by National Natural Science Foundation of China (32171783, 32572046), and Based on UAV multi-sensor remote The Camellia oleifera varieties in the sensing data are accurate identify (24237018).

Authors' contributions

Xingzi Kou: Data curation, Methodology, Software, Writing - original draft, Writing -review & editing. Xinyue Jiang: Methodology, Software, Writing - original draft, Data curation, Validation. Yue Zhang: Data curation, Software, Writing -review & editing. Chunyu Ge: Investigation, Writing -review & editing, Visualization. Zhibin Sun: Conceptualization, Formal analysis, Methodology, Supervision. Xuehai Tang: Conceptualization, Funding acquisition, Methodology, Resources.

Acknowledgements

The authors are grateful to Anhui Dechang Seedlings Company Limited for providing the Camellia oleifera plantation field site and experimental materials for this study. We would like to thank all individuals who assisted with field hyperspectral measurements and leaf-sample data collection. We also appreciate the editors and

anonymous reviewers for their valuable comments and suggestions to improve the quality of this manuscript.

References

1. Xu P, Cai Y, Chen K, You R, Lu Y. Camellia oleifera oil: unveiling health benefits and exploring novel applications. Crit. Rev. Food Sci. Nutr. 2026;66:108–128. https://doi.org/10.1080/10408398.2025.2519688.
2. Qin P, Shen J, Wei J, Chen Y. A critical review of the bioactive ingredients and biological functions of Camellia oleifera oil. Curr. Res. Food Sci. 2024;8:100753. https://doi.org/10.1016/j.crfs.2024.100753.
3. Song W, Huang S, Li F, Tu P, Li Y, Zhang B, et al. Selection and preliminary evaluation of superior individual plant in Camellia oleifera. PeerJ 2025;13:e20283. https://doi.org/10.7717/peerj.20283.
4. Yang X, Jiang P, Luo Y, Shi Y. Non-destructive detection of fatty acid content of camellia seed based on hyperspectral. J. Oleo Sci. 2023;72:69–77. https://doi.org/10.5650/jos.ess22139.
5. Li X, Zhu B, Li S, Liu L, Song K, Liu J. A comprehensive review of crop chlorophyll mapping using remote sensing approaches: achievements, limitations, and future perspectives. Sensors 2025;25:2345. https://doi.org/10.3390/s25082345.
6. Leng M, Liu Y, Li B, Che Y, Feng H, Shu M, et al. Analysis of crop leaf area index, leaf chlorophyll content, and canopy chlorophyll content based on deep learning and hyperspectral remote sensing. Int. J. Remote Sens. 2025;46:6937–6955. https://doi.org/10.1080/01431161.2025.2549533.
7. Zhu G, Wang Q, Zhang S, Guo T, Liu S, Lu J. A meta-analysis of crop leaf nitrogen, phosphorus and potassium content estimation based on hyperspectral and multispectral remote sensing techniques. Field Crops Res. 2025;329:109961. https://doi.org/10.1016/j.fcr.2025.109961.
8. Duan D, Zhao C, Li Z, Yang G, Yang W. Estimating total leaf nitrogen concentration in winter wheat by canopy hyperspectral data and nitrogen vertical distribution. J. Integr. Agric. 2019;18:1562–1570. https://doi.org/10.1016/S2095-3119(19)62686-9.
9. Li Q, Li P, Yang H. Optimizing wavelet functions to improve the accuracy of hyperspectral estimation of chlorophyll in potato leaves. J. Plant Nutr. Fertil. 2025;31:2252–2265. https://doi.org/10.11674/zwyf.2025038.
10. Croft H, Chen JM, Luo X, Bartlett P, Chen B, Staebler RM. Leaf chlorophyll content as a proxy for leaf photosynthetic capacity. Glob. Change Biol. 2017;23:3513–3524. https://doi.org/10.1111/gcb.13599.
11. Chou S, Chen B, Chen J, Wang M, Wang S, Croft H, et al. Estimation of leaf photosynthetic capacity from the photochemical reflectance index and leaf pigments. Ecol. Indic. 2020;110:105867. https://doi.org/10.1016/j.ecolind.2019.105867.
12. Tsuchiya Y, Hirono Y, Sonobe R. Estimating chlorophyll content in tea leaves using spectral reflectance and deep learning methods. Ecol. Inform. 2025:103399. https://doi.org/10.1016/j.ecoinf.2025.103399.

13. Kandpal KC, Kumar A. Migrating from invasive to noninvasive techniques for enhanced leaf chlorophyll content estimations efficiency. Crit. Rev. Anal. Chem. 2024;54:2583–2598. https://doi.org/10.1080/10408347.2023.2188425.

14. Silva LDD, Bonfim RAAD, Barbosa MP, Sá MC, Oliveira LSD, Almeida MF, et al. Comparison of chlorophyll content estimation using chlorophyll meters and leaf chloroplast pigments extracted from Eucalyptus urophylla and Khaya senegalensis seedlings. Aust. J. Bot. 2024;72:BT24019. https://doi.org/10.1071/BT24019.

15. Shibaeva TG, Mamaev AV, Sherudilo EG. Evaluation of a SPAD-502 plus chlorophyll meter to estimate chlorophyll content in leaves with interveinal chlorosis. Russ. J. Plant Physiol. 2020;67:690–696. https://doi.org/10.1134/S1021443720040160.

16. Su T, Liu X, Lei B, Cao Z, Di J, Xu L, et al. Screening the characteristic hyperspectral wavelength variables of peanut leaves based on coupled algorithms to predict SPAD value in peanut. Int. J. Remote Sens. 2024;45:8334–8354. https://doi.org/10.1080/01431161.2024.2399332.

17. Zhu C, Ding J, Zhang Z, Wang J, Wang Z, Chen X, et al. SPAD monitoring of saline vegetation based on Gaussian mixture model and UAV hyperspectral image feature classification. Comput. Electron. Agric. 2022;200:107236. https://doi.org/10.1016/j.compag.2022.107236.

18. Wei S, Yin T, Yuan B, Ow GLF, Yusof MLM, Gastellu-Etchegorry JP, et al. Estimation of chlorophyll content for urban trees from UAV hyperspectral images. Int. J. Appl. Earth Obs. Geoinf. 2024;126:103617. https://doi.org/10.1016/j.jag.2023.103617.

19. Daughtry CS, Walthall CL, Kim MS, De Colstoun EB, McMurtrey III JE. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. Remote Sens. Environ. 2000;74:229–239. https://doi.org/10.1016/S0034-4257(00)00113-9.

20. Nagy A, Szabó A, Elbeltagi A, Nxumalo GS, Bódi EB, Tamás J. Hyperspectral indices data fusion-based machine learning enhanced by MRMR algorithm for estimating maize chlorophyll content. Front. Plant Sci. 2024;15:1419316. https://doi.org/10.3389/fpls.2024.1419316.

21. Du L, Luo S. Spectral-frequency conversion derived from hyperspectral data combined with deep learning for estimating chlorophyll content in rice. Agriculture 2024;14:1186. https://doi.org/10.3390/agriculture14071186.

22. Cui B, Zhao Q, Huang W, Song X, Ye H, Zhou X. Leaf chlorophyll content retrieval of wheat by simulated RapidEye, Sentinel-2 and EnMAP data. J. Integr. Agric. 2019;18:1230–1245. https://doi.org/10.1016/S2095-3119(18)62093-3.

23. Wang Y, Wang X, Sun J, Wen Z, Wu H, Xiao L, et al. Estimation of winter wheat SPAD values by integrating spectral feature optimization and machine learning algorithms. Agronomy 2026;16:489. https://doi.org/10.3390/agronomy16040489.

24. Sudu B, Rong G, Guga S, Li K, Zhi F, Guo Y, et al. Retrieving SPAD values of summer maize using UAV hyperspectral data based on multiple machine learning algorithm. Remote Sens. 2022;14:5407. https://doi.org/10.3390/rs14215407.

25. Su T, Liu X, Lei B, Cao Z, Di J, Xu L, et al. Screening the characteristic hyperspectral wavelength variables of peanut leaves based on coupled algorithms to predict SPAD value in

peanut. Int. J. Remote Sens. 2024;45:8334–8354. https://doi.org/10.1080/01431161.2024.2399332.

26. Liu H, Lei X, Liang H, Wang X. Multi-model rice canopy chlorophyll content inversion based on UAV hyperspectral images. Sustainability 2023;15:7038. https://doi.org/10.3390/su15097038.

27. Yang C, Bai J, Sun H, Bi R, Song L, Muhammad A, et al. A new feature selection algorithm combining genetic algorithm, exponential decay function, and machine learning to realize hyperspectral estimation of winter wheat leaf area index. Comput. Electron. Agric. 2025;230:109851. https://doi.org/10.1016/j.compag.2024.109851.

28. Yuan W, Zhou H, Zhang C, Zhou Y, Wu Y, Jiang X, et al. Determination and visualization of moisture content in Camellia oleifera seeds rapidly based on hyperspectral imaging combined with deep learning. Spectrochim. Acta A Mol. Biomol. Spectrosc. 2025;330:125676. https://doi.org/10.1016/j.saa.2024.125676.

29. Chen B, Shi B, Gong J, Shi G, Jin H, Qin T, et al. Quality detection and variety classification of pecan seeds using hyperspectral imaging technology combined with machine learning. J. Food Compos. Anal. 2024;131:106248. https://doi.org/10.1016/j.jfca.2024.106248.

30. Wu X, Zhong L, Ding R, Wang C, Chen H, Zhong S, et al. Non-destructive estimation of SPAD and biomass in Lamiophlomis rotata using hyperspectral imaging and deep learning with DRSA-CARS feature selection. Front. Plant Sci. 2025;16:1640779. https://doi.org/10.3389/fpls.2025.1640779.

31. Zhang Y, Liang K, Zhu F, Zhong X, Lu Z, Chen Y, et al. Differential study on estimation models for indica rice leaf SPAD value and nitrogen concentration based on hyperspectral monitoring. Remote Sens. 2024;16:4604. https://doi.org/10.3390/rs16234604.

32. Ma C, Zhai L, Li C, Wang Y. Hyperspectral estimation of nitrogen content in different leaf positions of wheat using machine learning models. Appl. Sci. 2022;12:7427. https://doi.org/10.3390/app12157427.

33. Nian Y, Su X, Yue H, Anwar S, Li J, Wang W, et al. Winter wheat SPAD prediction based on multiple preprocessing, sequential module fusion, and feature mining methods. Agriculture 2024;14:2258. https://doi.org/10.3390/agriculture14122258.

34. Mucsi L, Litkey-Kovács D, Bonus K, Farmonov N, Elgendy A, Aji L, et al. Assessment of the effectiveness of spectral indices derived from EnMAP hyperspectral imageries using machine learning and deep learning models for winter wheat yield prediction. Remote Sens. 2025;17:3426. https://doi.org/10.3390/rs17203426.

35. Sonobe R, Sugimoto Y, Kondo R, Seki H, Sugiyama E, Kiriiwa Y, et al. Hyperspectral wavelength selection for estimating chlorophyll content of muskmelon leaves. Eur. J. Remote Sens. 2021;54:513–524. https://doi.org/10.1080/22797254.2021.1964383.

36. Feng H, Yang F, Yang G, Li Z, Ha P, Xing H. Estimation of chlorophyll content in apple leaves base on spectral feature parameters. Trans. Chin. Soc. Agric. Eng. 2018;34:182–188. https://doi.org/10.11975/j.issn.1002-6819.2018.06.023.

37. Sun Z, Jiang X, Tang X, Yan L, Kuang F, Li X, et al. Identifying core wavelengths of oil tree's hyperspectral data by Taylor expansion. Remote Sens. 2023;15:3137.

https://doi.org/10.3390/rs15123137.

38. Jiang X, Song Y, Sun Z, Kuang F, Tang X. Using a two-stage hybrid dimensionality reduction method on hyperspectral data to predict chlorophyll content of Camellia oleifera. Forests 2024;15:1937. https://doi.org/10.3390/f15111937.

39. Elsayed S, Gala H, Abd El-baki MS, Maher M, Elbeltagi A, Salem A, et al. Hyperspectral technology and machine learning models to estimate the fruit quality parameters of mango and strawberry crops. PLoS One 2025;20:e0313397. https://doi.org/10.1371/journal.pone.0313397.

40. Ebrahimi M, Sinegani AAS, Sarikhani MR, Mohammadi SA. Comparison of artificial neural network and multivariate regression models for prediction of Azotobacteria population in soil under different land uses. Comput. Electron. Agric. 2017;140:409–421. https://doi.org/10.1016/j.compag.2017.06.019.

41. Bogatenko T, Sergeev K, Slepnev A, Kurths J, Semenova N. Symbiosis of an artificial neural network and models of biological neurons: training and testing. Chaos 2023;33. https://doi.org/10.1063/5.0152703.

42. Hunter A. Feature selection using probabilistic neural networks. Neural Comput. Appl. 2000;9(2):124–132. https://doi.org/10.1007/s005210070023.

43. Oh MH, Kwon MW, Park K, Park BG. Sensitivity analysis based on neural network for optimizing device characteristics. IEEE Electron Device Lett. 2020;41:1548–1551. https://doi.org/10.1109/LED.2020.3016119.

44. Mandler H, Weigand B. A review and benchmark of feature importance methods for neural networks. ACM Comput. Surv. 2024;56:1–30. https://doi.org/10.1145/3679012.

45. Srivastava PK, Gupta M, Singh U, Prasad R, Pandey PC, Raghubanshi AS, et al. Sensitivity analysis of artificial neural network for chlorophyll prediction using hyperspectral data. Environ. Dev. Sustain. 2021;23:5504–5519. https://doi.org/10.1007/s10668-020-00827-6.

46. Guan C, Liu MY, Man WD, Zhang YB, Zhang QW, Fang H, et al. Estimation of chlorophyll content in Spartina alterniflora leaves based on continuous wavelet transformation and random forest algorithm. Spectrosc. Spectr. Anal. 2024;44:2993–3000. https://doi.org/10.3964/j.issn.1000-0593(2024)10-2993-08.

47. An G, Xing M, He B, Liao C, Huang X, Shang J, et al. Using machine learning for estimating rice chlorophyll content from in situ hyperspectral data. Remote Sens. 2020;12:3104. https://doi.org/10.3390/rs12183104.

48. Chungcharoen T, Donis-Gonzalez I, Phetpan K, Udompetaikul V, Sirisomboon P, Suwalak R. Machine learning-based prediction of nutritional status in oil palm leaves using proximal multispectral images. Comput. Electron. Agric. 2022;198:107019. https://doi.org/10.1016/j.compag.2022.107019.

49. Li Y, Chang Q, Liu X, Yan L, Luo D, Wang S. Estimation of maize leaf SPAD value based on hyperspectrum and BP neural network. Trans. Chin. Soc. Agric. Eng. 2016;32:135–142. https://doi.org/10.11975/j.issn.1002-6819.2016.16.019.

50. Zhang H, Yu L, Li D. Characteristic band selection of hyperspectral unmixing based on pixel simulation experiment. Spectrosc. Lett. 2025;58:794–806.

https://doi.org/10.1080/00387010.2025.2465699.

51. Chen Z, Wang X, Qiao S, Liu H, Shi M, Chen X, et al. A leaf chlorophyll content estimation method for Populus deltoides (Populus deltoides Marshall) using ensembled feature selection framework and unmanned aerial vehicle hyperspectral data. Forests 2024;15:1971. https://doi.org/10.3390/f15111971.

52. Chen R, Ren L, Yang G, Cheng Z, Zhao D, Zhang C, et al. Estimation of leaf chlorophyll content of maize from hyperspectral data using E2D-COS feature selection, deep neural network, and transfer learning. Agriculture 2025;15:1072. https://doi.org/10.3390/agriculture15101072.

53. Rasooli Sharabiani V, Soltani Nazarloo A, Taghinezhad E, Veza I, Szumny A, Figiel A. Prediction of winter wheat leaf chlorophyll content based on VIS/NIR spectroscopy using ANN and PLSR. Food Sci. Nutr. 2023;11:2166–2175. https://doi.org/10.1002/fsn3.3071.