



OPEN ACCESS

EDITED BY
Bifeng Hu,
Jiangxi University of Finance and
Economics, China

REVIEWED BY
Yuan Wang,
Chinese Academy of Sciences (CAS),
China
Ke Zhang,
Nanjing Agricultural University, China

*CORRESPONDENCE
Rutian Bi
✉ btrt@sxau.edu.cn

RECEIVED 19 January 2026
REVISED 16 March 2026
ACCEPTED 18 March 2026
PUBLISHED 13 April 2026

CITATION

Li P, Li X, He P, Wang J, Yang C, Yao Z,
Bi R, Xu L, Zhu H and Yang F (2026) Soil
total nitrogen prediction using sentinel-2
simulated bands and machine learning:
a laboratory spectroscopy study in
Hemerocallis citrina Baroni fields.
Front. Soil Sci. 6:1780422.
doi: 10.3389/fsoil.2026.1780422

COPYRIGHT

© 2026 Li, Li, He, Wang, Yang, Yao, Bi, Xu,
Zhu and Yang. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/)
(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.

Soil total nitrogen prediction using sentinel-2 simulated bands and machine learning: a laboratory spectroscopy study in *Hemerocallis citrina* Baroni fields

Ping Li¹, Xuran Li², Peng He¹, Jingshu Wang¹, Chenbo Yang^{1,3},
Zelong Yao¹, Rutian Bi^{1*}, Lishuai Xu¹, Hongfen Zhu¹
and Fan Yang⁴

¹College of Resources and Environment, Shanxi Agricultural University, Jinzhong, China, ²Rural Energy
and Environment Agency, Ministry of Agriculture and Rural Affairs, Beijing, China, ³College of
Agriculture, Shanxi Agricultural University, Jinzhong, China, ⁴China Meteorological Administration/
National Observation and Research Station of Desert Meteorology, Taklimakan Desert of Xinjiang/
Taklimakan Desert Meteorology Field Experiment Station of China Meteorological Administration/
Xinjiang Key Laboratory of Desert Meteorology and Sandstorm/Key Laboratory of Tree-ring Physical
and Chemical Research, China Meteorological Administration, Institute of Desert Meteorology,
Urumqi, China

Soil total nitrogen (STN) is a crucial indicator of crop productivity and soil health. Accurate monitoring of STN is essential for optimizing nitrogen management and achieving sustainable agricultural development. An adequate and timely STN supply serves as a key physiological basis for promoting effective tillering, flower stalk development, and continuous multibatch bud formation in *Hemerocallis citrina* Baroni. To address the challenges posed by the high-dimensionality of hyperspectral data and the dynamic spectral response of STN across different growth stages, this study employed spectral resampling to select feature bands based on Sentinel-2 sensor data (Simulation of Sentinel-2 Bands, SSB method). Specifically, hyperspectral data were collected under laboratory controlled conditions (constant temperature darkroom, standard light source, air-dried ground soil), simulated Sentinel-2 sensor bands through spectral resampling (SSB method), and constructed an STN prediction framework based on 8 machine learning algorithms (random forest, extreme gradient boosting, back propagation neural network (BPNN), genetic algorithm-optimized BPNN (GA-BPNN), convolutional neural networks (CNN), and a hybrid CNN-bidirectional long short-term memory-attention model). The model performance was comprehensively evaluated using the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and mean bias error (MBE). This study aims to establish laboratory-scale soil-spectral chemical relationship baselines, providing band selection and algorithm validation references for subsequent field remote sensing applications, rather than directly developing field operational systems. The results showed that: (1) the three-band spectral index TBI3 exhibited the highest correlation with STN across the full growth period ($R=0.7354$). The optimal indices for specific growth stages were TBI4, TBI3, and TBI5 for the spring seedling/leaf expansion, bolting/flowering, and bud emergence stages, respectively, with TBI-series indices exhibiting significantly superior performance compared to two-dimensional indices; (2) the GA-BPNN model achieved the highest accuracy for the full growth period,

with a test R^2 of 0.6284, along with the lowest MAE ($0.0693 \text{ g}\cdot\text{kg}^{-1}$) and RMSE ($0.0879 \text{ g}\cdot\text{kg}^{-1}$), demonstrating outstanding generalization capability; and (3) the GA-BPNN model outperformed the other models in comparative analyses across different growth stages, and the growth stage-specific integrated modeling method showed higher prediction accuracy and enhanced resistance to overfitting (both training and test R^2 exceeded 0.6, with the gap reduced to 0.0064). Based on these findings, we propose a technical framework termed “SSB-SPXY-GA-BPNN-growth stage adaptation”, which provides theoretical and methodological support for precise STN monitoring and variable-rate fertilization.

KEYWORDS

genetic algorithm-optimized BPNN, growth stage-specific modeling, laboratory spectroscopy, soil total nitrogen, spectral index

Highlights

What are the main findings?

- The three-band spectral index TBI3, constructed using Sentinel-2 bands, demonstrates the highest correlation with soil total nitrogen content, significantly outperforming traditional two-dimensional indices. GA-BPNN model achieves the highest prediction accuracy among multiple machine learning algorithms while exhibiting outstanding generalization capability.

What is the implication of the main finding?

- The construction of three-band spectral indices provides a superior spectral indicator for high-accuracy, non-destructive monitoring of soil nitrogen, enhancing the reliability of remote sensing inversion. GA-BPNN model establishes a core methodological foundation for developing a “growth stage-adaptive” precision fertilization technology system, holding significant practical value for advancing precision agricultural management and green development.

1 Introduction

Soil total nitrogen (STN) is a critical reservoir in terrestrial nitrogen cycling (1) and directly regulates soil fertility evolution, crop nutrient uptake, and agricultural productivity levels, thereby constituting a cornerstone of global food security (2). However, nitrogen management faces a well-defined ecological threshold: excessive application not only diminishes yield benefits but also escalates environmental costs sharply (3–5). This highlights the imperative for a paradigm shift towards “with STN dynamics as the core, nitrogen input is optimized according to local conditions”. Therefore, to ensure crops nitrogen demand and maintain STN utilization efficiency, it is necessary to balance the yield target and ecological sustainability by accurately monitoring the change in STN content to realize the green transformation of agriculture with low input, high efficiency, and high yield. Accurate monitoring of STN in specialty crop systems remains challenging due to the

dynamic spectral response of soil-vegetation systems across growth stages (6–8). Recent advances in spectral resampling techniques enable the simulation of satellite sensor bands (e.g., Sentinel-2) from laboratory hyperspectral data, providing a cost-effective approach for sensitive band selection prior to field-scale application (9–11). Specifically, Sentinel-2’s red-edge bands (B5–B8A) have demonstrated unique potential for vegetation nitrogen assessment, yet their applicability for soil nitrogen monitoring in high-value horticultural crops remains underexplored (8, 12–15).

Hyperspectral remote-sensing technology, with the capability to acquire continuous surface spectral information, has emerged as a vital tool for STN content monitoring. However, its high-dimensional data features (encompassing hundreds to thousands of spectral bands) introduce numerous irrelevant wavelengths, leading to information redundancy and exponentially increasing computational complexity (9), thereby constraining further improvements in STN prediction efficiency. Spectral resampling techniques adjust the spectral resolution of raw hyperspectral data to simulate satellite sensor bands (e.g., Sentinel-2) (10, 11), enabling the construction of multi-dimensional spectral indices tailored to specific sensor characteristics (12). For soil nitrogen monitoring, this approach has achieved promising accuracy. Gao et al. (13) integrated spectral resampling with machine learning models (Support Vector Machines (SVM), Back Propagation Neural Network (BPNN), and Random Forest (RF) to simulate the GF-1, Landsat-8, WorldView-2, and Sentinel-2 bands to achieve high-precision prediction of soil microaggregate organic carbon (Sentinel-2 simulated bands + BPNN model, coefficient of determination $[R^2]=0.773$, Geng et al. (14) combined fractional-order derivative processing with ZY1-02D satellite data to improve organic carbon estimation accuracy ($R^2=0.66$) in Northeast China’s black soil region using 0.6-order three-dimensional spectral indices and RF modeling. This strategy has also been proven effective in monitoring vegetation parameters (15). Collectively, these studies validate that the tripartite technical paradigm of “spectral resampling-multidimensional index construction-machine learning modeling” can overcome the dimensionality limitations of traditional hyperspectral analysis. This method provides a high-accuracy and low-cost solution for regional-scale STN dynamic monitoring and crop nutrient management and offers significant

practical value for advancing precision agriculture and sustainable development. yet its application for STN estimation in perennial horticultural crops across multiple growth stages requires further validation.

While current STN inversion studies have widely adopted machine learning models to overcome the limitations of traditional linear methods such as Partial Least Squares Regression (PLSR) (16, 17), significant challenges remain in model selection and optimization. On the one hand, contradictory performance has been observed for identical algorithms across different scenarios: For the RF model, with Xiao et al. (18) achieved only $R^2=0.570$ for total nitrogen (TN) inversion using Moderate-Resolution Imaging Spectroradiometer (MODIS) data in Shandong Province, while Mashaba-Munghemezulu et al. (19) attained significantly higher accuracy ($R^2=0.90$) for STN prediction in smallholder maize farms by integrating Sentinel-2 raw bands (including red and red-edge bands), spectral indices, and environmental variables. The Extreme Gradient Boosting (XGBoost) algorithm demonstrated optimal performance ($R^2=0.627$) when driven by multi-source remote sensing data in Hebei coastal wetlands (20), though systematic cross-regional comparisons with other machine learning models such as RF and BPNN remain notably lacking in existing research. On the other hand, deep learning has demonstrated significant potential, as evidenced by convolutional neural networks (CNN) achieving high-precision STN mapping ($R^2=0.702$) in the southern hilly regions through the integration of topographic and hyperspectral data (17). However, the applicability and performance boundaries between deep learning and shallow machine learning methods remain vaguely defined, as evidenced by comparative studies showing that (1) BPNN outperforms PLSR by 6.5% in STN content inversion for black soil regions (16), and (2) SVM achieves superior STN prediction accuracy ($R^2=0.676$) over both BPNN and PLSR in Datong County, Qinghai Province (21). These findings demonstrate the strong dependence of model performance on data structure and environmental context (22), while highlighting the lack of synergistic optimization among sensitive feature identification, multi-dimensional index construction, and algorithm selection. In particular, a research gap exists in STN monitoring for *Hemerocallis citrina* Baroni, an economically significant crop, where spectral nitrogen response mechanisms during critical growth stages (from spring seedling to flowering) remain poorly explored, requiring urgent development of both multi-model adaptability frameworks and fertilization decision systems. Consequently, establishing an integrated modeling system featuring “dynamic sensitive-band selection-algorithm transfer adaptation-growth-stage-specific coupling” is critical to minimize nitrogen regulation errors through systematic analysis of *Hemerocallis citrina* Baroni soil spectral features, thereby addressing critical theoretical gaps in precision management of specialty crops.

The Yunzhou District of Datong City, a representative area of China’s northern farming-pastoral ecotone, was selected as the study area, where the distinctive semi-arid climate and intense wind erosion have significantly aggravated topsoil structural degradation and nutrient loss (23), with STN content becoming a critical limiting

factor in the development of local specialty agricultural industries. As a perennial specialty crop with high economic value in the region, the core patterns of spectral response mechanisms of soil total nitrogen across different growth stages of *Hemerocallis citrina* Baroni are closely linked to the physiological priorities and growth development needs of the plant (24). The main objectives and innovations of this study are threefold: (1) To analyze the correlation between spectral indices and STN across the full growth period and specific growth stages of *Hemerocallis citrina* Baroni, while identifying optimal band combinations for each phenological phase; (2) To comprehensively compare the predictive accuracy of eight multivariate machine learning models for STN estimation; and (3) To propose and validate a novel technical framework termed “SSB-SPXY-GA-BPNN-growth stage adaptation”.

2 Materials and methods

2.1 Study site

Yunzhou District ($39^{\circ}43'N-40^{\circ}16'N$, $113^{\circ}20'E-113^{\circ}55'E$), Datong City, Shanxi Province, is situated in China’s northern agro-pastoral ecotone, with a total cultivated area of 486.67 km² (25). Characterized by an average elevation of 1,157 m and a northwest-to-southeast topographic gradient (Figure 1A), the region experiences a temperate continental monsoon climate. This climate features an annual mean temperature of 6.4 °C, precipitation of 392 mm (primarily concentrated in summer), 3,000 sunshine hours annually, and significant diurnal temperature variations (daily range: 12.6–14.4 °C) (26). *Hemerocallis citrina* Baroni is the primary cash crop in the region and is cultivated using large-scale intensive farming systems covering approximately 110 km² (27). Its growth cycle comprises of five distinct phenological stages: spring seedling, leaf expansion, bolting, bud emergence, and flowering (28). Current fertilization practices face three critical challenges: (1) blind fertilizer application owing to inadequate soil testing, (2) excessive nitrogen use causing plant lodging, and (3) low nutrient-use efficiency from extensive fertilization methods.

2.2 Soil collection and pretreatment

2.2.1 Soil sample collection and spectral data acquisition

Densely cultivated *Hemerocallis citrina* Baroni areas were selected as sampling sites, and 60 sampling points were established in the central zones, as illustrated in Figure 1A. For each point, composite topsoil samples (0–20 cm depth) were collected using a five-point sampling method across the five key growth stages (The specific sampling schedule is shown in Figure 1C, with a 15-day interval between sampling points for different growth stages to ensure repeated sampling across all developmental phases. The sampling location accuracy was maintained within ± 3 meters, and no fertilizers were applied throughout the entire growth cycle.). A total of 5 growth stages $\times$ 60 sampling points were obtained, and all 300 samples were

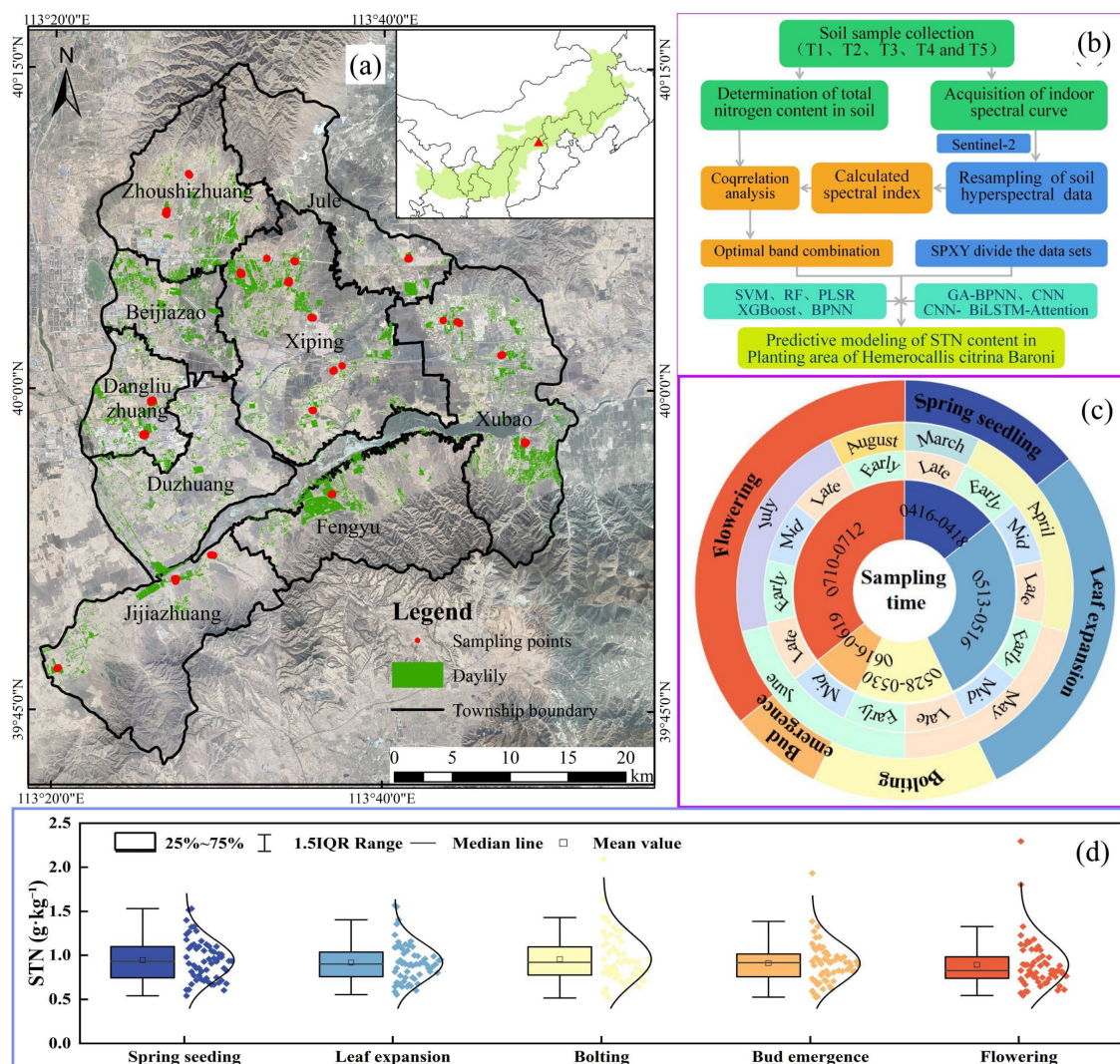


FIGURE 1

(A) Geographical location of the study area and distribution of sampling sites, (B) Technical workflow diagram (where T1–T5 correspond to different growth stages), (C) Division of different growth periods and sampling schedule, and (D) Variations in soil total nitrogen (STN) content across different growth stages. The coefficient of variation (CV) for STN during each growth stage ranged from 23.96% to 31.86%, indicating moderate variation intensity. Among these, the CV was highest during the flowering stage (31.86%) and the heading stage (29.75%), significantly higher than that during the leaf expansion stage (23.96%) and the spring seedling stage (24.62%).

integrated to form the complete growth dataset (full growth period soil data: range: 0.52–2.30 g·kg⁻¹, mean $\pm$ SD: 0.92 $\pm$ 0.25 g·kg⁻¹, CV = 27.27%). After removing plant roots, weeds, and gravel impurities, the samples were homogenized in sterile plastic bags, air-dried, ground, and sieved through a 0.154 mm mesh. STN content was determined via the Kjeldahl method (unit: g·kg⁻¹) (9). The average content of STN in different growth stages was 0.94 $\pm$ 0.23 g·kg⁻¹ (spring seedling), 0.92 $\pm$ 0.22 g·kg⁻¹ (leaf expansion), 0.96 $\pm$ 0.28 g·kg⁻¹ (bolting), 0.91 $\pm$ 0.23 g·kg⁻¹ (bud emergence) and 0.89 $\pm$ 0.28 g·kg⁻¹ (flowering), which showed some dynamic changes and distribution differences. Spectral reflectance curves were acquired using a FieldSpec4 portable spectroradiometer (Analytical Spectral Devices (ASD), Inc., Longmont, CO, USA), covering wavelengths of 350–2500 nm. Measurements were conducted in a darkroom with a 50 W halogen light source, maintaining a 10° field-of-view optical probe 15 cm above the

samples. Before each test, white reference calibration was performed using five spectral curves measured for each soil sample, with their average value serving as the representative reflectance spectrum for that specimen (9).

2.2.2 Band resampling of sentinel-2 satellite sensor and spectral index construction

The spectral resampling method leverages the official spectral response functions of Sentinel-2 satellite sensors across various bands to perform weighted integration on raw high-resolution field spectral data (350–2500 nm), as detailed in the following formula. For each target band of the sensor, the original hyperspectral reflectance data undergoes convolution and normalization with the corresponding spectral response function within the wavelength range, yielding simulated discrete band reflectance

values. The specific formula is as follows:

$$R_{\text{band}} = \frac{\int_{\lambda_{\min}}^{\lambda_{\max}} R_{\text{hyper}}(\lambda) \cdot SRF_{\text{band}}(\lambda) d\lambda}{\int_{\lambda_{\min}}^{\lambda_{\max}} SRF_{\text{band}}(\lambda) d\lambda} \quad (1)$$

where, SRF_{λ} is the spectral response function, $R_{\text{hyper}}(\lambda)$ is the original hyperspectral reflectance.

The spectral coverage of Sentinel-2 spans the visible, red-edge, NIR, and SWIR regions (illustrated in Table 1), which exhibit direct and indirect correlations with STN content (20).

The spectral index method achieves nondestructive monitoring of surface parameters by constructing a multiband mathematical combination model that characterizes soil properties. The core mechanism involves selecting bands with significant spectral responses or building a mathematical index model to perform nonlinear combination operations on the reflectances of different bands, thereby extracting feature parameters that represent the target attributes. The indices constructed in this study are listed in the Table 2 (14).

To highlight the key methodology of combining laboratory hyperspectral data with satellite band simulations, this study employed the GA-BPNN modeling approach and conducted three modeling experiments: (1) using only the laboratory hyperspectral dataset, with the original 2151 bands (350–2500 nm) as input features; (2) using only the Sentinel simulation dataset, with 13 resampled Sentinel-2 bands; (3) the combined method (this study), utilizing Sentinel-2 simulation bands and artificially constructed spectral indices (2D and 3D). Table 3 summarizes the results, demonstrating that the combined method achieved the highest test R^2 (0.6284), surpassing both the laboratory hyperspectral dataset-only method ($R^2 = 0.2458$) – potentially due to the overdimensionality issue caused by $n=300$ and $p=2151$ – and the Sentinel simulation dataset-only method ($R^2 = 0.3731$), which was constrained by reduced spectral information. These findings

indicate that in small-sample STN modeling, spectral resampling combined with index construction can achieve the optimal balance between dimensionality reduction and information retention.

2.3 Feature selection method for the simulation of sentinel-2 bands

The core principle of the simulated sensor bands is to adapt hyperspectral data to the multispectral band structures of specific satellite sensors using spectral resampling techniques (Equation 1). Based on the 13 simulated spectral bands obtained above, this study employed Pearson correlation analysis to quantify the associations between various spectral indices (such as the Difference Index (DI), Ratio Index (RI), Normalized Difference Index (NDI), and Three-Band Indices (TBI1, TBI2, TBI3, TBI4, TBI5)) and soil total nitrogen (STN) content at different growth stages. The optimal band combinations for each index were then selected to serve as the characteristic inputs for STN retrieval during the respective growth periods. To assess potential multicollinearity among the selected band combinations, we calculated the Variance Inflation Factor (VIF) for the reflectance values of all bands involved. The results showed that the VIF values for the vast majority of band combinations were below 5, indicating that multicollinearity was at an acceptable level and would not severely interfere with model construction (15, 16).

2.4 Construction and hyperparameter settings of the inversion model

The study used sample set partitioning based on joint X-Y distances (SPXY) to divide the training and test sets, with the core idea of considering the spatial distribution of both independent (X) and dependent (Y) variables, ensuring that the selected samples are representative in both the feature space and the response variable space. The data were divided in a 2:1 ratio between the training and

TABLE 1 Sentinel-2 satellite sensor information.

Sensor	Band number	Spectral region (nm)	Description	Country
Sentinel-2	13	B1:433–453	Aerosols	ESA
		B2:458–523	Blue	
		B3:543–578	Green	
		B4:650–680	Red	
		B5:698–713	Red Edge1	
		B6:733–748	Red Edge2	
		B7:773–793	Red Edge3	
		B8:785–900	NIR	
		B8A:855–875	Red Edge4	
		B9:935–955	Water vapor	
		B10:1360–1390	Cirrus	
		B11:1565–1655	SWIR1	
		B12:2100–2280	SWIR2	

ESA, European Space Agency; NIR, near-infrared; SWIR, shortwave infrared.

TABLE 2 Soil spectral indices.

Dimensional	Index	Formula references
Two-dimensional (2D)	Difference index (DI)	Bi-Bj
	Ratio index (RI)	Bi/Bj
	Normalized difference index (NDI)	(Bi-Bj)/(Bi+Bj)
Three-dimensional (3D)	Three-band index 1 (TBI1)	Bi/(Bj*Bz)
	Three-band index 2 (TBI2)	Bi/(Bj+Bz)
	Three-band index 3 (TBI3)	(Bi+Bj)/Bz
	Three-band index 4 (TBI4)	(Bi-Bj)/(Bi+Bz)
	Three-band index 5 (TBI5)	(Bi-Bj)/(Bi-2*Bj+Bz)

test sets. To evaluate model stability and generalization capability, a 5-fold cross-validation procedure was implemented during the training process for all algorithms. Furthermore, a paired t-test was performed on the prediction results to determine the statistical significance of performance differences between the optimal model and other models, with a significance level set at $p < 0.05$. Eight models-PLSR, RF, SVM, XGBoost, BPNN, GA-BPNN, CNN, and CNN-BiLSTM-Attention models-were applied and compared for statistical modeling (17, 21, 29, 30). This framework encompasses three paradigms: traditional machine learning (SVM, RF, PLSR, XGBoost), shallow neural networks (BPNN, GA-BPNN), and deep learning (CNN, CNN-BiLSTM-Attention). Specifically, SVM achieves nonlinear mapping through structural risk minimization and ϵ -insensitive loss; RF mitigates overfitting risks via Bagging integration of multiple de-correlated decision trees; PLSR resolves multicollinearity by synchronously decomposing spectral and target matrices; XGBoost enhances gradient boosting efficiency through second-order Taylor expansion and regularization optimization; BPNN employs a single hidden layer topology combined with the LM algorithm for rapid convergence; GA-BPNN overcomes local optima via global search with genetic algorithms; CNN automatically extracts spectral hierarchical features; while CNN-BiLSTM-Attention integrates spatial feature extraction, temporal dependency modeling, and SE channel attention mechanisms to achieve multi-scale representation. While deep learning typically requires large-scale datasets, this study primarily employs the CNN-BiLSTM-Attention model. This approach leverages convolutional layers to automatically learn hierarchical spectral features, significantly reducing the need for manual feature engineering. The bidirectional long short-term memory network captures sequential dependencies in spectral responses across different growth stages, while the attention mechanism adaptively identifies nitrogen-sensitive bands, thereby enhancing prediction accuracy and model interpretability. The model architecture is detailed in Figure 2. The hyperparameters of the models were selected by grid search and cross validation, Bayesian optimization, and genetic algorithm parameters. As shown in Table 4.

Given the limited sample size across different growth stages ($n=40$ per training set) and the complexity of models such as CNN and GA-BPNN, there is a potential risk of overfitting in growth stage-specific modeling. To mitigate this risk, an early stopping parameter was implemented for the neural network models (BPNN and GA-BPNN: training is terminated when the validation set loss

fails to decrease for 20 consecutive rounds; CNN and CNN-BiLSTM-Attention: the learning rate decay factor is 0.1 with a decay cycle of 1000 rounds; the dataset is shuffled before each training round to enhance generalization). Additionally, the training-test R^2 gap was calculated as a quantitative indicator of overfitting severity, where a gap < 0.05 indicates no significant overfitting, 0.05–0.15 indicates mild overfitting, and > 0.15 indicates severe overfitting.

To further evaluate the spatial generalization ability of the optimal model and address potential concerns regarding spatial autocorrelation, an additional Leave-One-Location-Out Cross-Validation (LOLOCV) was performed for the GA-BPNN model. In this validation scheme, all five samples from each of the 60 sampling locations were sequentially held out as a test set, with the model trained on the remaining 295 samples from 59 locations. This approach ensures that the test set consists entirely of geographic locations unseen during training, providing a robust assessment of spatial extrapolation capability.

2.5 Model accuracy verification

To evaluate and compare the accuracy of the STN prediction models established by the eight modeling methods, four metrics were used: R^2 , Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Bias Error (MBE) (20). A higher R^2 value indicates that the RMSE, MAE, and MBE values are closer to zero, suggesting a more accurate prediction model. Specifically:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (3)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (4)$$

$$MBE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i) \quad (5)$$

where y_i , $\hat{y}_i$, and $\bar{y}$ are the measured, model-predicted, and measured average values of the soil element content, respectively; and n is the number of samples.

TABLE 3 Comparison of precision metrics across three modeling experiments.

Methons\metrics	Train				Test			
	R ²	MAE	MBE	RMSE	R ²	MAE	MBE	RMSE
Lab-Hyperspectral Only	0.4422	0.1502	0.0170	0.2124	0.2458	0.0974	0.0380	0.1323
Sentinel-Simulated Only	0.6729	0.1239	0.0028	0.1634	0.3731	0.0918	0.0199	0.1229
Combined method	0.5975	0.1349	0.0015	0.1814	0.6284	0.0693	0.0213	0.0879

3 Results and analysis

3.1 Results of SSB method

The correlation between the two-dimensional spectral indices (DI, RI, and NDI) and three-dimensional spectral indices (TBI1, TBI2, TBI3, TBI4, and TBI5) of *Hemerocallis citrina* Baroni throughout its full growth period, as well as STN content was calculated using the SSB selection method, as shown in Figures 3 and 4. The results indicated that for the two-dimensional spectral indices, the best band combination for DI was B (2, 3), with a correlation coefficient of $R = 0.5074$. For RI, bands 8A and 8 had the best band combination, with a correlation coefficient of $R = 0.5337$. The highest correlation coefficient between NDI and STN was obtained from the B (8A, 8) combination, with a correlation coefficient of $R = 0.5335$. There were differences in the correlation between the three-band spectral indices and STN. The best band combination for TBI1 was bands 2, 4, and 4, with a correlation coefficient of $R = 0.4972$. The band combination with the highest correlation between TBI2 and STN was B (4, 7, 2), with a correlation coefficient of $R = -0.7349$. The best band combination for TBI3 was bands 7, 2, and 4, with a correlation coefficient of $R = 0.7354$. The band combination with the highest correlation between TBI4 and STN was B (3, 2, 8A), with a correlation coefficient of $R = -0.6136$. The best band combination for TBI5 was bands 5, 3, and 4, with a correlation coefficient of $R = 0.7012$. In summary, among the eight spectral indices throughout the full growth period, the highest

correlation with STN was TBI3, with a correlation coefficient of $R = 0.7354$. The highest correlation between the two-dimensional spectral indices and STN was observed for RI, with a correlation coefficient of $R = 0.5337$.

Further analysis of the correlation coefficients between spectral feature bands and STN during different growth stages exhibited the highest correlation of the TBI4 index constructed from B (2, 3, 8A) with STN content during the spring seedling stage, with an R of 0.7497. The TBI2 index constructed from B (4, 6, 1) showed a strong negative correlation ($R = -0.7238$). During the leaf expansion stage, all indices had an absolute correlation coefficient value greater than 0.60, with the highest correlation coefficient for the TBI4 index constructed from B (2, 3, 8) ($R = 0.8245$). The TBI2 index constructed from B (4, 6, 2) showed a strong negative correlation ($R = -0.7878$). During the bolting stage, the TBI3 index constructed from B (9, 7, 8A) showed the highest negative correlation ($R = -0.8398$), while the B (8A, 9, 7) combination of the TBI2 index showed the highest positive correlation ($R = 0.8397$). During the bud emergence stage, the TBI5 index constructed from B (6, 7, 8) showed the highest correlation coefficient (0.6743), and the combination of DI, RI, and NDI with B (8A, 8) showed the highest correlation coefficients. During the flowering stage, the TBI3 index constructed from B (6, 2, 4) showed the highest correlation coefficient ($R = 0.8137$), while the TBI2 index constructed from B (4, 6, 3) showed the highest negative correlation coefficient ($R = -0.8088$). Comparing the indicators of different growth stages, it was evident that the three-dimensional spectral

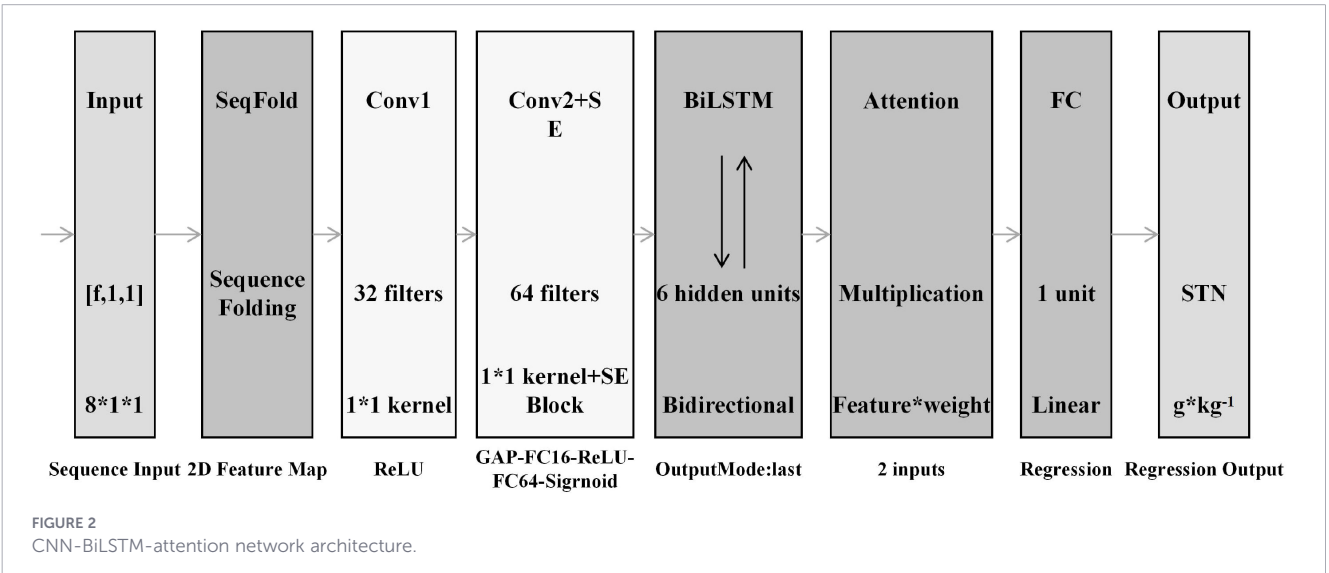


TABLE 4 Hyperparameter configurations for machine learning models.

Model	Hyperparameter	Value
SVM	best_c	4
	best_g	0.8
	-p	0.05
RF	ntree	500
	mtry	3
	extra_options.nodesize	5
PLSR	k	band num
XGBoost	num_trees	500
	params.eta	0.01
	params.max_depth	3
BPNN	net = newff(p_train, t_train, 8)	8
	net.trainParam.epochs	1000
	net.trainParam.goal	1.00E-06
	net.trainParam.lr	0.01
GA-BPNN	S1 = 8	8
	net.trainParam.epochs	1000
	net.trainParam.goal	1.00E-06
	net.trainParam.lr	0.01
	gen	50
	pop_num	5
CNN	p_train reshape	[8, 1, 1, sample number]
	p_test reshape	[8, 1, 1, sample number]
	outdim	1
	MaxEpochs	100
	LearnRateDropFactor	0.1
	LearnRateDropPeriod'	1000
CNN-BiLSTM-Attention	f_	size(res, 2)-outdim
	p_train reshape	[f_, 1, 1, sample number]
	p_test reshape	[f_, 1, 1, sample number]
	MaxEpochs	100
	LearnRateDropFactor	0.1
	LearnRateDropPeriod'	1000

index had a better correlation with STN than the two-dimensional spectral index, with an average correlation coefficient that was 18.44% higher than that of the two-dimensional spectral index. As shown in Table 5.

3.2 Comparison of different machine learning methods

Through a comparative analysis of the performance of the two feature band selection methods—the Successive Projections Algorithm (SPA) and the SSB method—in predicting the spectral response mechanism of STN, it was found that the SSB method

achieved an average test-set R^2 of 0.5093, surpassing the SPA method (average test-set R^2 of 0.3429). This represents an improvement of 48.52% in the average test-set R^2 and a reduction of 16.56% in RMSE. The optimal models for the two methods yielded test-set R^2 values of 0.5531 (SPA method) and 0.6284 (SSB method), respectively (Figure 5). The SSB method demonstrated particular effectiveness in mitigating overfitting and enhancing the performance of complex models. Therefore, this study adopted the SSB method for feature band selection and employed it to analyze the spectral response mechanism of STN content. The SSB-SPXY method was used to predict STN content using various modeling techniques, and Figure 5 presents the prediction results of the different models throughout the full growth period. After a comprehensive comparison of the models' performance in predicting STN content, it was established that the GA-BPNN model performed the best, with the highest R^2 (0.6284) on the test set, the lowest RMSE (0.0879 g·kg⁻¹), and the lowest MAE (0.0693 g·kg⁻¹). The R^2 improvement rate from the training set to the test set was 5.17%, whereas the RMSE and MAE decreased by 51.55% and 48.63%, respectively. Statistical testing confirmed that the GA-BPNN model's R^2 on the test set was significantly higher than that of the next best performing models, CNN-BiLSTM-Attention and PLSR ($p < 0.05$). The GA-BPNN model was tested by 5-fold cross-validation. The results showed that the model had good generalization ability on the independent validation, with an average R^2 of 0.3549 ± 0.1381 and a RMSE of 0.1530 ± 0.0108 g·kg⁻¹. The average R^2 of the test set is 0.5408 ± 0.1171 , and the composite test set R^2 is 0.5395. The average R^2 attenuation rate between the test set and the training set was 16.08%. The model demonstrated high precision (MAE < 0.07 g·kg⁻¹ on the test set) and strong generalization (RMSE < 0.09 g·kg⁻¹ on the test set), enhancing its capability to effectively detect spectral mechanisms of STN content in Hemerocallis citrina Baroni planting areas. The CNN-BiLSTM-Attention, PLSR, and BPNN models followed, with both the training set and test set R^2 values exceeding 0.56. The CNN-BiLSTM-Attention model showed the best generalization ability on the test set, with the highest R^2 (0.6012), a 6.00% increase from the training set, the lowest RMSE (0.0910 g·kg⁻¹), a 51.60% decrease from the training set, and the test set MAE (0.0720 g·kg⁻¹), a 47.99% decrease from the training set. The MBE was positive and less than 0.04, indicating a controllable system positive bias, and suggesting the architectural design had potential for complex feature extraction. The PLSR model demonstrated the best stability between the training and test sets with R^2 values of 0.5824 and 0.5779, respectively. The R^2 decay rate was only 0.78% and the MBE for the training and test sets were 0.0000 g·kg⁻¹ and 0.0137 g·kg⁻¹, respectively, indicating a slight positive systematic error. This suggests that the PLSR model has a strong ability to capture high-value samples. The BPNN model's test set R^2 (0.5843) and RMSE (0.0930 g·kg⁻¹) were between those of the PLSR and the CNN-BiLSTM-Attention models. However, the training set R^2 (0.6029) was the highest among the three models. Notably, the training set MBE was -0.0093 g·kg⁻¹, indicating a slight underestimation during the training phase, which was effectively corrected in the test phase (test set MBE: 0.0273 g·kg⁻¹), reflecting the model's strong adaptability to changes in data distribution.

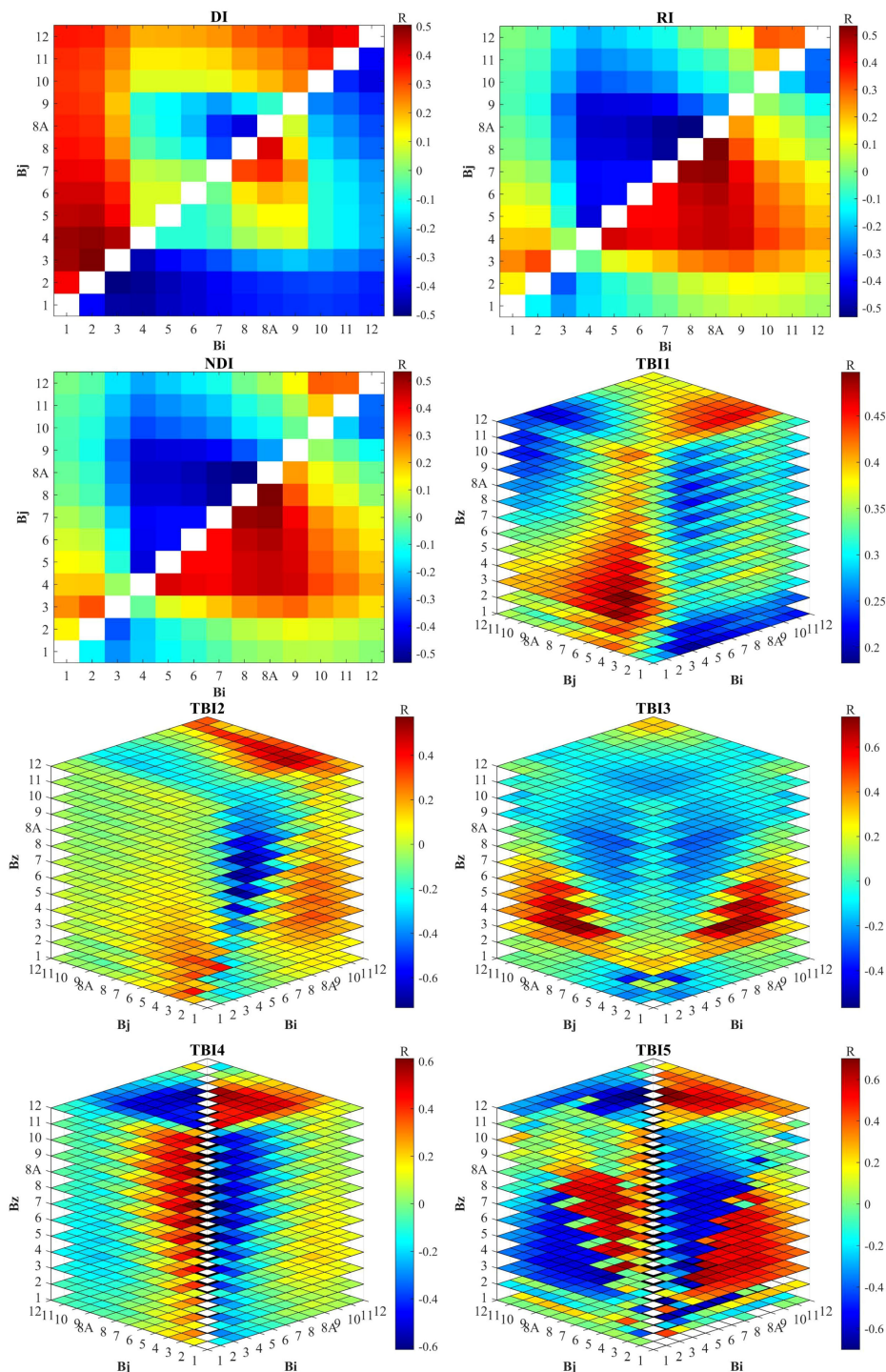


FIGURE 3
Correlation between STN and different spectral indices throughout the full growth period.

The SVM, CNN, and XGBoost models yielded poor results. The SVM model exhibited training and test sets R^2 s of 0.6440 and 0.5116, respectively, with an attenuation rate of 20.56%, indicating a mild overfitting. The test set MBE was $0.0305 \text{ g}\cdot\text{kg}^{-1}$, showing a positive bias, possibly due to the kernel space mapping causing high-value sample predictions to be skewed. The CNN model showed a training set-test set R^2 attenuation rate of 30.42%, with a significantly underestimated training set MBE ($-0.0250 \text{ g}\cdot\text{kg}^{-1}$) and a test set MBE of $0.0169 \text{ g}\cdot\text{kg}^{-1}$, suggesting partial correction of

system bias through deep feature learning, with overestimation remaining. Compared with the SVM and CNN models, the XGBoost model exhibited severe overfitting, with training and test sets R^2 s of 0.6752 and 0.4286 (attenuation rate of 36.53%), respectively, and an overfit index (training set RMSE/test set RMSE) of 1.50, indicating a weak generalization ability. The test set MBE was $0.0466 \text{ g}\cdot\text{kg}^{-1}$, revealing the model's strong memory effect on high nitrogen samples. The RF model had the poorest modeling accuracy, with a 72.16% difference between the training (0.8975)

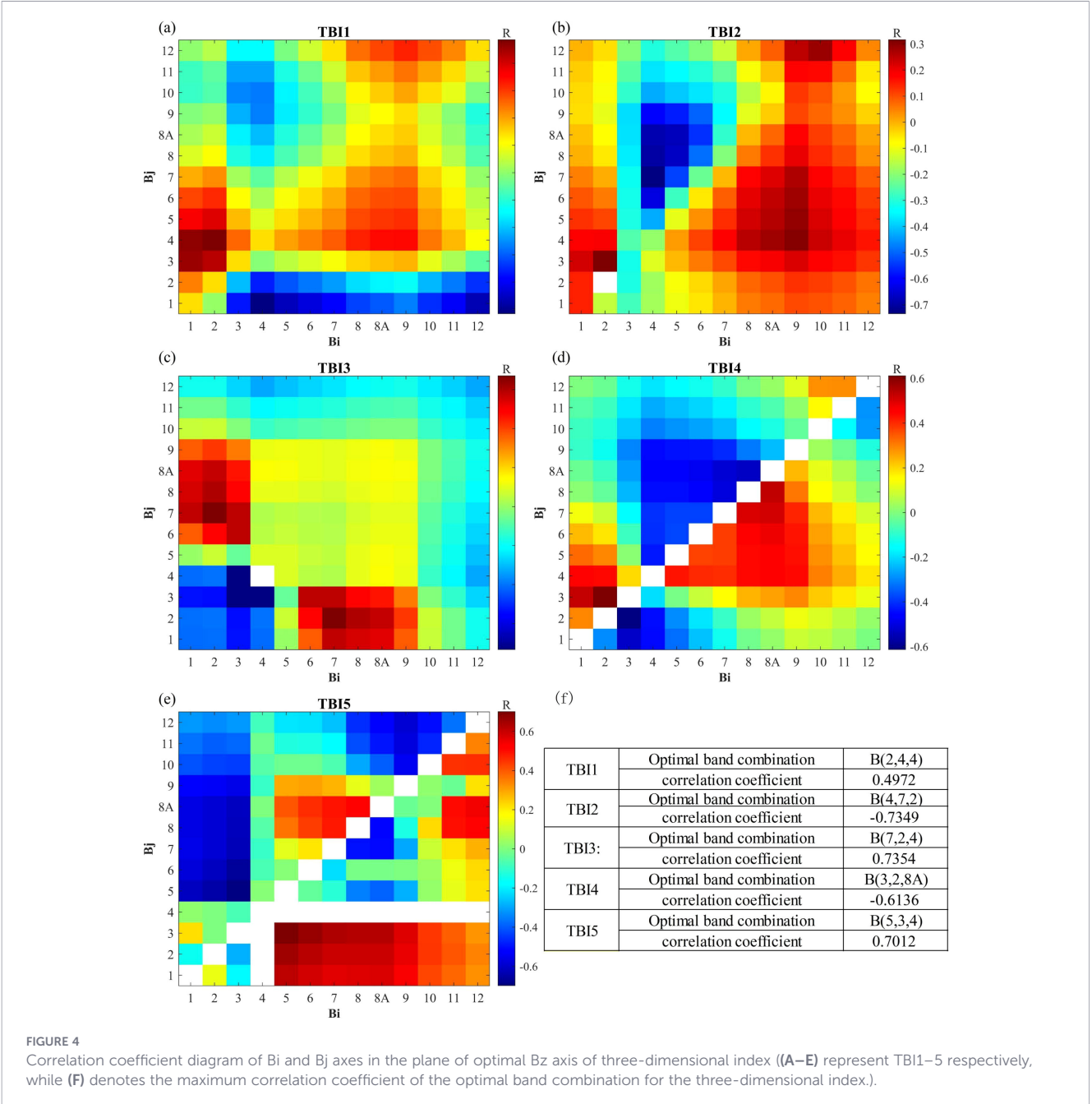


TABLE 5 Optimal band combination and correlation coefficient of different growth periods.

Period	Index	DI	RI	NDI	TB11	TB12	TB13	TB14	TB15
Spring seedling	Optimal band combination	B(2, 3)	B(2, 3)	B(2, 3)	B(1, 4, 4)	B(4, 6, 1)	B(6, 1, 4)	B(2, 3, 8A)	B(8, 7, 11)
	correlation coefficient	0.6114	0.5071	0.5067	0.6056	-0.7238	0.7222	0.7497	0.6760
Leaf expansion	Optimal band combination	B(2, 4)	B(3.1)	B(1, 3)	B(1, 4, 4)	B(4, 6, 2)	B(6, 1, 4)	B(2, 3, 8)	B(12, 8A, 8)
	correlation coefficient	0.7604	-0.6377	0.6359	0.7978	-0.7878	0.7857	0.8245	0.7168
Bolting	Optimal band combination	B(8A, 8)	B(5, 4)	B(5.4)	B(1, 4, 4)	B(8A, 9, 7)	B(9, 7, 8A)	B(3, 2, 9)	B(4, 5, 3)
	correlation coefficient	0.5807	0.5945	0.5928	0.3551	0.8397	-0.8398	-0.7311	0.7984
Bud emergence	Optimal band combination	B(8A, 8)	B(8A, 8)	B(8A, 8)	B(9, 4, 4)	B(4, 7, 2)	B(7, 2, 4)	B(3, 2, 9)	B(6, 7, 8)
	correlation coefficient	0.4795	0.5393	0.5391	0.4255	-0.6281	0.6260	-0.6079	0.6743
Flowering	Optimal band combination	B(10, 12)	B(2, 3)	B(2, 3)	B(1, 4, 4)	B(4, 6, 3)	B(6, 2, 4)	B(3, 2, 7)	B(12, 9, 7)
	correlation coefficient	0.6130	0.6859	0.6815	0.5693	-0.8088	0.8137	-0.7664	0.7838

At a 0.01 level (two - tailed), the correlation was significant (P<0.01).

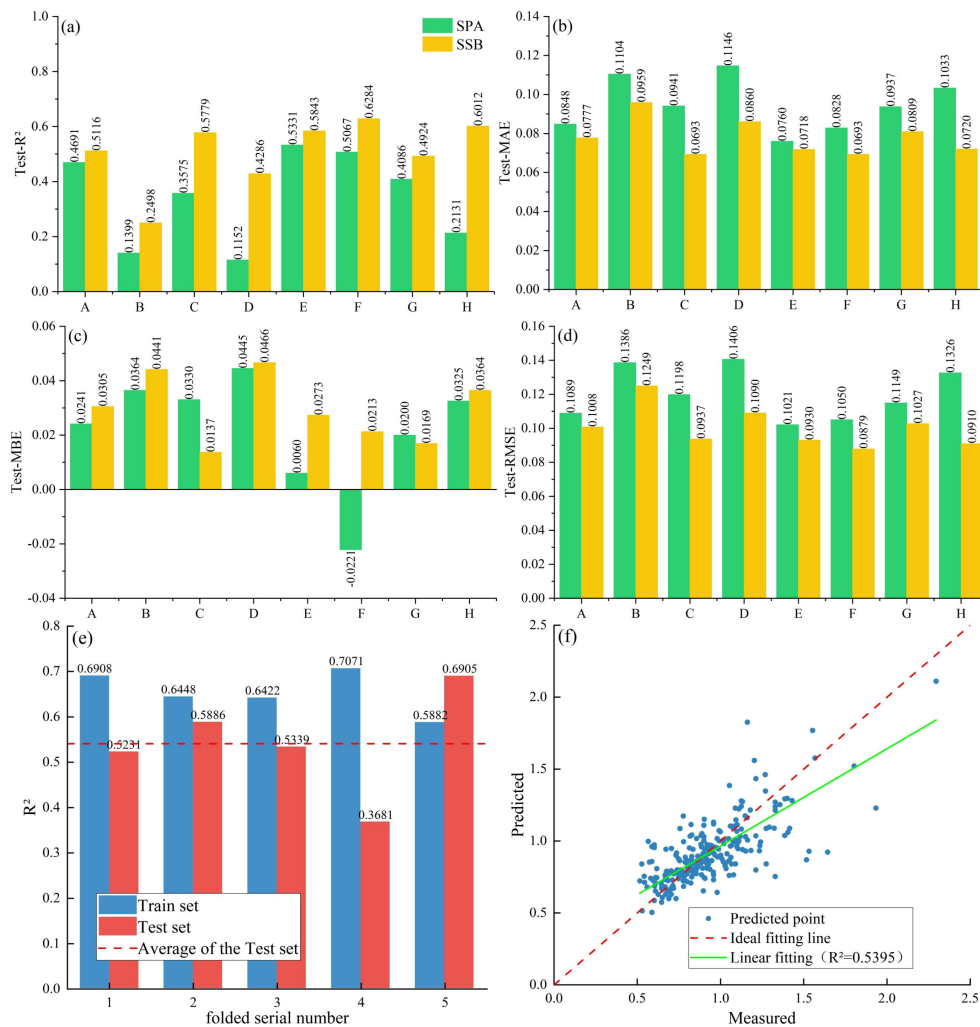


FIGURE 5

Comparison of accuracy of machine learning models across the full growth cycle of *Hemerocallis citrina* Baroni and cross-validation results of the optimal model throughout the entire growth period. ((A–D) respectively correspond to the four model performance metrics of the SPA method and SSB method in eight models (A, B, C, D, E, F, G, H respectively corresponding to SVM, RF, PLSR, XGBoost, BPNN, GA-BPNN, CNN, CNN-BiLSTM-Attention); (E, F) respectively demonstrate the multiple cross-validation R^2 value comparison of the GA-BPNN model under the SSB method, as well as the comprehensive cross-validation performance of this model on all 300 test samples).

and test (0.2498) sets R^2 s, which was the highest among all the models. This indicates that the model was heavily influenced by noise in the training data. The RMSE of the test set ($0.1265 \text{ g}\cdot\text{kg}^{-1}$) was 36.40% higher than that of the training set ($0.0864 \text{ g}\cdot\text{kg}^{-1}$), highlighting a significant lack of generalization ability. The MBE of the training set was nearly zero ($0.0003 \text{ g}\cdot\text{kg}^{-1}$), while the bias on the test set sharply increased to $\text{MBE} = 0.0441 \text{ g}\cdot\text{kg}^{-1}$, indicating overfitting to high-value samples. This severe overfitting is likely due to the model's tendency to overlearn from noisy or redundant features in the high-dimensional spectral data, which is a known limitation of tree-based models when noised features are present. In summary, in STN inversion modeling using hyperspectral data from the full growth period of the *Hemerocallis citrina* Baroni planting areas, the GA-BPNN model demonstrated the best overall performance, achieving the highest R^2 on the test set. It exhibited a 5.17% improvement in R^2 between the training and test sets, and the MAE on the test set ($0.0693 \text{ g}\cdot\text{kg}^{-1}$) was 3.45% lower than that of the BPNN model. Its unique global optimization mechanism

effectively addresses the issue of local convergence in high-dimensional spectral data, maintaining prediction accuracy ($\text{RMSE} < 0.10 \text{ g}\cdot\text{kg}^{-1}$) while demonstrating strong adaptability to the spatiotemporal heterogeneity of farmland STN.

To rigorously evaluate the spatial generalization ability of the GA-BPNN model, we conducted an additional Leave-One-Location-Out Cross-Validation (LOLOCV) (Figure 6). The results showed that the model achieved a combined test R^2 of 0.5055 and RMSE of $0.1770 \text{ g}\cdot\text{kg}^{-1}$ across all 300 samples. Compared to the original SPXY-based validation results ($R^2 = 0.6284$, $\text{RMSE} = 0.0879 \text{ g}\cdot\text{kg}^{-1}$), this represents a 19.56% reduction in R^2 and a 101.34% increase in RMSE. The performance decline under LOLOCV confirms that spatial autocorrelation influences model evaluation. However, the model maintained a moderate predictive capability ($R^2 = 0.5055$) under this stringent spatial validation, indicating that the GA-BPNN has learned meaningful spectral-STN relationships rather than merely memorizing site-specific patterns. The relatively large standard deviations across the 60 locations ($R^2 \text{ std} = 1.0350$,

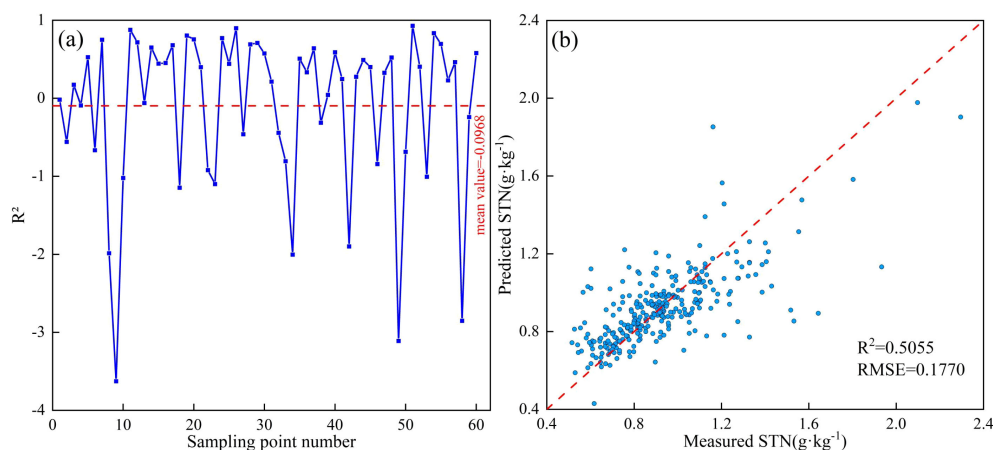


FIGURE 6

Cross-validation results of the leave-one-out method: (A) R^2 values for each sampling point; (B) Scatter plot of predicted values versus measured values ($n=300$).

RMSE std = 0.0859 $\text{g}\cdot\text{kg}^{-1}$) highlight the substantial spatial heterogeneity of soil properties in the study area.

3.3 Comparison and comprehensive analysis of model accuracy in different growth periods

Based on the SSB-SPXY method and various modeling techniques, STN predictions were conducted for different growth stages of *Hemerocallis citrina* Baroni. Figure 7 illustrates the prediction results of the different models at the various growth stages. Compared with other machine learning models, the GA-BPNN model achieved the highest accuracy in predicting STN at different growth stages of *Hemerocallis citrina* Baroni. During the spring seedling stage, the GA-BPNN model had a test set R^2 of 0.6343, with the smallest difference from the training set R^2 (0.6255), which was only 0.0088. There was no significant overfitting, and the test set RMSE and MAE were the lowest at 0.1040 and 0.0841 $\text{g}\cdot\text{kg}^{-1}$, respectively. By contrast, the RF and XGBoost models exhibited severe overfitting; the training set R^2 exceeded 0.77, whereas the test set R^2 was less than 0.23. The CNN and CNN-BiLSTM-Attention models had a test set R^2 of < 0.38. During the leaf expansion stage, the GA-BPNN model had the highest test set R^2 (0.7859) and the lowest RMSE and MAE at 0.0687 $\text{g}\cdot\text{kg}^{-1}$ and 0.0588 $\text{g}\cdot\text{kg}^{-1}$, respectively. The CNN, BPNN, and SVM models followed, with test set R^2 values of 0.7781, 0.7526, and 0.7425, respectively. The XGBoost model showed significant overfitting, with a test set R^2 of only 0.5882, which was significantly lower than that of training set (0.8302). During the bolting stage, the R^2 values for the training and test sets of the GA-BPNN model were 0.6293 and 0.4732, respectively, performance optimal. The test set had the lowest MAE of 0.0665 $\text{g}\cdot\text{kg}^{-1}$ with an MBE of 0.0041 $\text{g}\cdot\text{kg}^{-1}$, which was significantly lower than the |MBE| of 0.0320 $\text{g}\cdot\text{kg}^{-1}$ for the BPNN model. In the RF and XGBoost test sets, R^2 exhibited extremely negative values of -0.5307 and -0.9087, respectively, losing the predictive ability of the models. During bud emergence, the R^2 values of the training and test sets of the GA-BPNN model were 0.5384 and 0.6223, respectively. This test set exhibited the lowest RMSE and MAE values (0.0708 and 0.0570 $\text{g}\cdot\text{kg}^{-1}$, respectively). The

BPNN model demonstrated a secondary performance, with its training set R^2 (0.6113) exceeding the test set R^2 by 0.0538. The R^2 differences among the RF, PLSR, and CNN models for both the training and test sets were greater than 0.4, indicating severe overfitting. During the flowering stage, the GA-BPNN model showed the smallest difference of 0.1523 between the test set R^2 (0.3416) and the training set R^2 (0.4939), indicating no overfitting. The BPNN model had an R^2 of 0.3649 on the test set, which was 0.3919 lower than the R^2 value of the training set, indicating overfitting. The CNN model ($R^2=0.8654$ and 0.1620 for the training and test sets, respectively) exhibited significant overfitting. Additionally, the CNN-BiLSTM-Attention model ($R^2=0.0214$ on the test set) and the PLSR model ($R^2=-0.2272$ on the test set) failed completely. In summary GA-BPNN's genetic algorithm optimization effectively prevented overfitting across all stages, maintaining |training-test R^2 difference| < 0.15, whereas other models showed differences up to 0.70. To effectively demonstrate and analyze the generalization capabilities of different models, the study compared the R^2 performance of training and test sets across various growth stages and models. Figure 7C presents the overfitting diagnostic metrics of the top-performing GA-BPNN model across five growth stages. The training-test R^2 gap ranges from -0.0088 to 0.1561. Notably, during the spring seedling and bud stage, the test set R^2 actually exceeds the training set R^2 , yielding negative gaps of -0.0088 and -0.0839 respectively, indicating exceptional generalization capability. At the leaf expansion stage, the gap is 0.0582 (mild overfitting); while during the heading and flowering stages, the gaps reach 0.1561 and 0.1523 respectively, approaching the "severe overfitting" threshold of 0.15.

We constructed a STN prediction model for five growth stages of *Chrysanthemum morifolium* by integrating SSB-SPXY-GA-BPNN. Specifically, we divided the full growth cycle dataset (300 samples) into five independent subsets (60 samples each) corresponding to five key growth stages (spring seedling stage, leaf expansion stage, heading stage, bud stage, and flowering stage). For each subset, we built an optimal GA-BPNN model, forming a prediction system of "staged modeling (dividing training and test sets) -integrated modeling (directly combining training and test sets for each growth stage)." Comparative analysis with the full growth cycle model revealed that

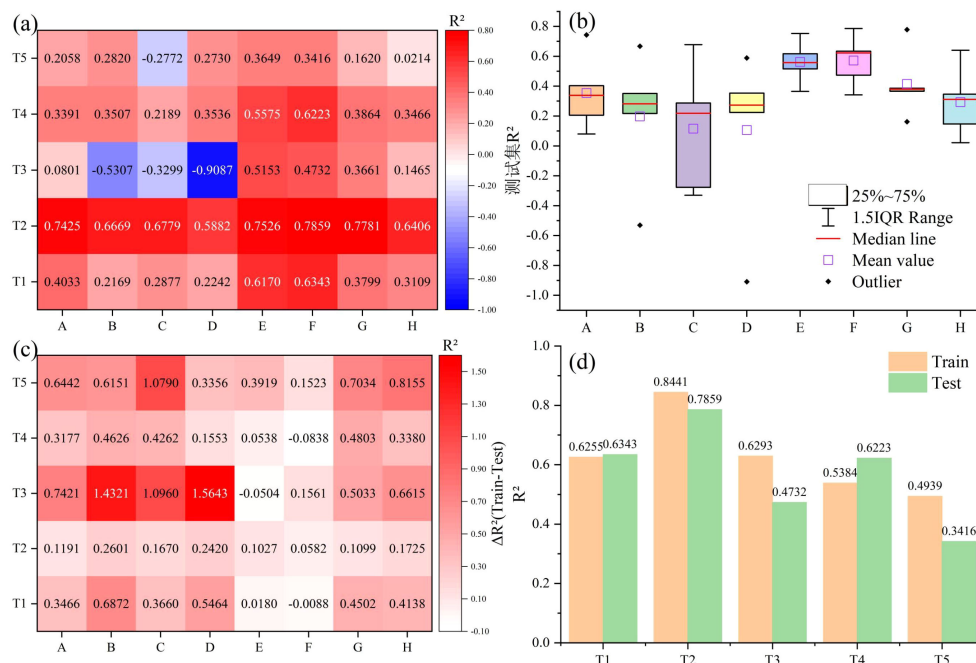


FIGURE 7

Precision performance of machine learning methods across different growth stages (measured by test set R^2) (T1–T5 correspond to spring seedling stage, leaf expansion stage, bolting stage, bud emergence stage, and flowering stage, respectively; A, B, C, D, E, F, G, H correspond to SVM, RF, PLSR, XGBoost, BPNN, GA-BPNN, CNN, and CNN-BiLSTM-Attention, respectively). (A) Heatmaps of test set R^2 performance for different machine learning methods across stages; (B) Stability distribution maps of models across growth stages; (C) Heatmaps of overfitting severity for different models at various stages; (D) Performance comparison of the optimal model (GA-BPNN) across different growth stages.

this growth stage-specific integration model (not a traditional ensemble method like Bagging or Boosting, but a data partitioning and model deployment strategy based on crop physiological ecology principles) achieved significantly better RMSE and MAE accuracy for both training and test sets compared to the full growth cycle model. Specifically, the RMSE, MAE, and MBE of the test set were 2.02%, 2.79%, and 42.65% lower than those of the full growth cycle model, indicating reduced systematic bias. Additionally, the R^2 of the growth stage-specific integration model's training set (0.6134) was higher than that of the full growth cycle training set (0.5975), with a significant reduction in performance difference between training and test sets (R^2 difference of only 0.0064, a decrease of 79.28%), as shown in Figure 8. This demonstrates that the growth-stage-specific integrated modeling approach exhibits higher prediction accuracy and stronger resistance to overfitting. All growth-stage models employ globally optimized hyperparameters determined through grid search (50 generations of genetic algorithm, population size 5, and 8 hidden layer nodes) to ensure model comparability and stability.

4 Discussion

4.1 Physicochemical mechanism of soil total nitrogen spectral response

This study employs the SSB method, utilizing post-processing spectral data of laboratory-measured soil samples (air-dried,

ground, and sieved). The approach eliminates confounding factors such as soil moisture, surface roughness, living vegetation, and atmospheric interference, enabling a pure analysis of the intrinsic correlation mechanism between soil total nitrogen (STN) and spectral reflectance. However, it is crucial to distinguish between laboratory conditions and actual satellite remote sensing: mixed pixel effects, dynamic moisture variations, and atmospheric conditions significantly influence field spectral signals (20, 31). Therefore, the findings of this study should be regarded as a theoretical framework and band selection guide for STN monitoring, providing prior knowledge for subsequent field applications. However, their practical utility requires validation through real Sentinel-2 imagery under field conditions. Future research will focus on establishing radiative transfer models to bridge the gap between laboratory spectra and satellite observations, while integrating moisture correction algorithms and end-member decomposition techniques.

The sensitive band combination identified by the SSB method can be systematically explained from an exponential dimension perspective (15). The correlation between the TBI series and STN is generally superior to that of two-dimensional spectral indices, consistent with the findings of Gao et al. (13). The TBI3 (B7, B2, B4) series, which spans the entire growth period, demonstrates the optimal spectral index. This three-band combination (red edge 3-773–793 nm, blue light 458–523 nm, red light 650–680 nm) exhibits the highest correlation with STN ($R = 0.7354$), attributed to the multi-spectral response mechanism of soil organic matter (SOM) (32–35). First, the characteristic absorption mechanism: the blue light band (B2) selectively absorbs aromatic structures in soil

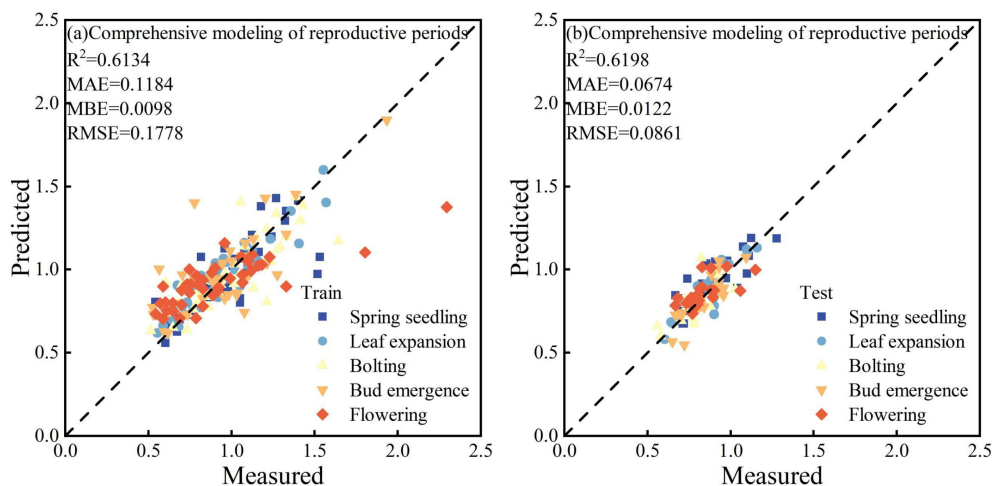


FIGURE 8

Precision index of the GA-BPNN model with the growth stage-specific integrated modeling method. [(A) and (B) are the prediction results of the training set and test set respectively].

humus, with its electronic transition energy closely related to the polymerization degree of humus. Second, the mineral co-variation mechanism: the red light band (B4) primarily responds to charge transfer transitions of soil iron oxides (e.g., pyrite, hematite), whose content often forms a co-variation relationship with STN through organic-matter-iron complexation. Third, the interfacial response mechanism: the red edge band (B7) lies at the transition zone between the strong absorption edge of chlorophyll and the weak reflection of soil organic matter, exhibiting extreme sensitivity to the interfacial effects of soil organic-mineral complexes (36, 37). The nonlinear combination of these three bands ($TBI3 = (B7+B2)/B4$) effectively separates organic matter signals from the mineral matrix background, thereby enhancing the specificity of STN estimation. In contrast, while two-dimensional indices (e.g., RI and NDI) can suppress some environmental noise ($R > 0.53$) through the ratio of red edge 4 (B8A) to NIR (B8), their sensitivity is constrained by the analytical capacity of linear combinations in complex systems (14).

4.2 Applicability analysis of machine learning models

Different modeling methods exhibited varying levels of accuracy in STN inversion for *Hemerocallis citrina* Baroni cultivation areas. The GA-BPNN model demonstrated the highest accuracy, with an R^2 of 0.6284 for the test set, representing a 5.17% improvement over the training set (Equation 2). Using genetic algorithms to globally optimize the initial weights of a network, this model overcomes the limitations of traditional BP neural networks, which are prone to local optima. This method enhances both the precision of nonlinear data predictions and the generalization ability of the model. In modeling the main growth stages of *Hemerocallis citrina* Baroni, the GA-BPNN model achieved high fitting accuracy and strong stability, consistent with Liu et al. (38) findings on the modeling and prediction of arable land quality. The traditional BPNN model achieved basic performance based on the nonlinear mapping ability (test set $R^2=0.5843$), however, the initial

weight randomness led to fluctuations in the results (MBE: -0.0093 – 0.0273 $g \cdot kg^{-1}$), and the stability was weaker than that of the optimized model (39) (Equations 3–5). The CNN-BiLSTM-Attention fusion model integrates the strengths of the CNN for spatial feature extraction, BiLSTM for temporal modeling, and attention mechanisms for focusing, thereby demonstrating its potential in multi-dimensional data processing (40). However, in hyperspectral time-series analysis, the attention mechanism may oversuppress effective features due to weight imbalance or insufficient samples, leading to the risk of losing STN information. By introducing latent variables, the traditional linear PLSR model effectively eliminates multicollinearity in hyperspectral data, achieving a test set R^2 of 0.5779, confirming a significant linear response of the spectral-STN (41). However, because of its linear framework, the PLSR model struggles to represent threshold effects, such as the STN saturation point and water content, as well as dynamic interactions, such as nitrogen-water synergy and humidity-temperature effects, which aligns with the findings of Li et al. (41). In other models, CNN suffers from continuity disruption during band selecting and the insufficient capture of nitrogen response features by local convolutional kernels (9), resulting in a test set R^2 of only 0.4924; SVM relies on fine-tuning parameters (penalty coefficients, kernel functions) and exhibits systematic bias (MBE = 0.0305 $g \cdot kg^{-1}$) (21); XGBoost controls complexity and parameter tuning through lambda, but the resolution decreases after spectral resampling (42), which weakens the ability to identify the absorption features of all-nitrogen specificity; RF suffers from serious overfitting due to noise band interference (B1, B10) and redundancy feature dependence (test set $R^2=0.2498$) (43) (Equations 2–5). The study showed that GA-BPNN has the best comprehensive performance in STN inversion, and the deep learning model needs to optimize the architecture by combining hyperspectral features. Traditional machine learning is limited by its ability to capture data, and this study provides an important basis for the selection of *Hemerocallis citrina* Baroni STN-monitoring model. Research indicates that deep learning models require architecture optimization through the integration of hyperspectral

features. Traditional machine learning is constrained by data acquisition capabilities, while GA-BPNN demonstrates optimal comprehensive performance in STN inversion, providing a critical theoretical framework for selecting the *Hemerocallis citrina* Baroni STN monitoring model. Notably, the performance disparity between SPXY segmentation validation ($R^2 = 0.6284$) and LOLOCV ($R^2 = 0.5055$) highlights the importance of considering spatial autocorrelation in soil attribute modeling. The 19.56% R^2 reduction under LOLOCV represents the extent of performance degradation when applying the model to new regions. It is worth emphasizing that these two validation strategies serve distinct purposes: SPXY evaluates the model's predictive capability across different phenological periods within the sampling area, whereas LOLOCV provides a more conservative estimate for spatial prediction tasks. For practical agricultural management applications, the actual model performance may lie between these two estimates. The LOLOCV result ($R^2 = 0.5055$) confirms that the model learns authentic spectral-soil relationships rather than spatial artifacts, supporting its potential application value beyond the sampling region.

4.3 Soil mechanism of growth stage-specific modeling accuracy improvement

Compared to the full growth cycle model, the growth stage-specific integrated modeling for *Hemerocallis citrina* Baroni demonstrates improved prediction accuracy. This enhancement is not merely a statistical fit improvement but stems from the stage-specific characteristics of STN (sodium thiosulfate) dynamic turnover in the soil-root interaction system (44). Our study reveals spectral monitoring heterogeneity of STN across different growth stages of *Hemerocallis citrina*, similar to findings by Hassan et al. (45) in wheat. However, the underlying mechanism should be attributed to differences in rhizosphere dynamics and organic matter decomposition stages rather than simple canopy changes. The dominant spectral response mechanisms coupled with rhizosphere processes during different growth stages are as follows: Spring seedling stage (TBI4 optimal: B2, B3, B8A): The soil is in the post-freeze-thaw recovery phase with low microbial activity and predominantly stable macromolecular humus. The blue-green light combination (B2, B3) is sensitive to stable humus-clay complexes under low temperatures, while red edge 4 (B8A) responds to electron transitions in the soil surface's minor oxidation layer (32, 46). Leaf expansion stage: With rising soil temperature and increased root exudates, rhizosphere microbial activity significantly increases and organic matter decomposition accelerates. TBI4 optimal (band combinations: B2, B3, B8) shows extreme sensitivity to soil aggregate structure reorganization caused by organic matter decomposition (33, 47, 48). At this stage, the correlation coefficient reaches its peak throughout the entire growth cycle ($R = 0.8245$), indicating the strongest spectral correlation between rhizosphere-active organic matter and STN. During the bolting stage (TBI3 optimal: B9, B7, B8A), the root nitrogen uptake rate peaks, and the soil undergoes rapid mineralization. The water vapor band (B9) is sensitive to the combined frequency vibrations of soil residual

moisture and hydroxyl compounds, indirectly reflecting microbial metabolites related to nitrogen mineralization (34). The combination of Red Edge 3 (B7) and Red Edge 4 (B8A) captures the electronic transition changes during the rapid transformation of organic matter. At the bud emergence stage (TBI5 optimal: B6, B7, B8), reproductive growth begins, with changes in root secretion composition (increased carbohydrate proportion), alterations in microbial community structure, and nitrogen transformation shifting toward fixation and slow release. The Red Edge 2–3 combination (B6, B7) primarily responds to organic matter transformation products induced by rhizosphere sediments (24, 35). The relatively low correlation coefficient at this stage ($R = 0.6743$) may reflect the weakening of the STN-spectral relationship in the rhizosphere due to soil nitrogen transfer to plants. During the flowering stage [TBI3 (B6, B2, B4)], soil nitrogen consumption peaks, the proportion of easily decomposable organic matter decreases, and the proportion of stabilized humus increases. Red Edge 2 (B6) is sensitive to stabilized organic matter with high aromatic condensation degree; the combination of blue light (B2) and red light (B4) re-emerges as dominant, reflecting the entry of soil organic matter into the stabilization phase (36, 49). Despite the high vegetation coverage (>80%) in the field, the red-edge-visible light combination still penetrated the canopy to capture soil background stabilization signals. In summary, the dynamic changes of spectral indices across different growth stages (visible light, red-edge, near-infrared, and water vapor bands) (37) validated the rationality of sampling design based on distinct growth phases. By establishing a soil-spectral coupling relationship model for growth stages, each phase could more accurately capture key spectral characteristics of the rhizosphere environment (49), thereby indirectly enhancing the structural adaptability and predictive accuracy of the GA-BPNN model. Under the constraint of 40 training set samples per growing season, the growth-stage-specific integrated modeling demonstrates adaptive advantages within each phase (test set $R^2=0.6198$, representing a 2.6% improvement over the full-growth-season model). However, its generalization capability requires validation with phases containing ≥ 60 samples per growing season. These results should be regarded as proof of methodological feasibility rather than final performance benchmarks.

This study aims to elucidate the spectral response characteristics of soil nitrogen (STN) in the soil. Although environmental variables such as temperature and humidity were not incorporated into the model (50), we acknowledge its critical role in nitrogen transformation. Existing research has been limited to explaining soil spectral characteristics without integrating soil microbial community structure, chemical composition of root exudates, and STN dynamics. Furthermore, the generalizability of this framework still requires further validation through cross-regional and cross-soil-type validation sets. Future research should develop a multi-source heterogeneous data fusion framework to integrate parameters such as soil microbiome, root metabolic profiles, and enzyme activities, and construct a growth cycle-adaptive hybrid architecture model (e.g., GA-BiLSTM) to enhance the interpretability of soil processes and ecological adaptability in STN monitoring systems.

5 Conclusions

This study focused on the cultivation areas of *Hemerocallis citrina* Baroni, obtaining high-resolution spectral reflectance data and STN content for different growth stages. Using the SSB method to select feature bands, a high-spectral inversion model for STN content was constructed using eight machine learning algorithms: RF, SVM, PLSR, XGBoost, BPNN, GA-BPNN, CNN, and CNN-BiLSTM-Attention. The main conclusions are as follows.

(1) By analyzing the correlation coefficients between different spectral indices and STN, we established that TBI3 had the highest correlation with STN throughout the growth period, with a correlation coefficient of $R = 0.7354$. The highest correlation coefficients between the spectral indices constructed based on different growth periods and STN were observed in the TBI series. Overall, three-dimensional spectral indices outperformed the two-spectral index dimensions.

(2) By comparing the precision of various indicators across the full growth period, it was evident that the GA-BPNN model had the best prediction accuracy. It achieved the highest R^2 value (0.6284) on the test set and showed a 5.17% improvement in R^2 between the training and test sets. In addition, GA-BPNN had the lowest MAE (0.0693 g·kg⁻¹) and RMSE (0.0879 g·kg⁻¹), demonstrating both high precision and strong generalization capabilities. This model effectively detected STN content in the *Hemerocallis citrina* Baroni planting area.

(3) In the comparison of the prediction accuracy for different growth periods, the GA-BPNN model had the highest prediction accuracy. Compared with the modeling of the full growth period, the comprehensive modeling method of the growth stage-specific periods had a higher prediction accuracy and anti-overfitting ability (the training and test sets were both greater than 0.61, with a difference of 0.0064).

The “SSB-SPXY-GA-BPNN-Growth Stage Adaptation” technical framework proposed in this study demonstrates innovation by integrating the physical mechanisms of remote sensing bands with the adaptive capabilities of machine learning. It provides an interpretable and transferable intelligent decision-making paradigm for precise nitrogen management throughout the entire crop growth cycle. However, the practical effectiveness and economic benefits of implementing this framework for variable-rate fertilization in real-world agricultural settings require further validation and quantification through subsequent field trials. The next step is to collaborate with agricultural extension departments to conduct field trials, verifying the framework’s practical value in real-world production. This will advance regional precision fertilization and soil health management while providing scientific support for precision agriculture implementation.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

PL: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Writing – original draft. XL: Conceptualization, Formal analysis, Methodology, Supervision, Writing – review & editing. PH: Conceptualization, Formal analysis, Methodology, Supervision, Writing – review & editing. JW: Data curation, Investigation, Methodology, Software, Writing – original draft. CY: Data curation, Investigation, Methodology, Software, Writing – original draft. ZY: Data curation, Methodology, Software, Writing – original draft. RB: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. LX: Methodology, Supervision, Writing – review & editing. HZ: Methodology, Supervision, Writing – review & editing. FY: Investigation, Methodology, Resources, Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This research was supported by Shanxi Basic Research Program (202403021212308), Innovation Project of Major State Basic Research Development Program (2021YFD1600301), Science and Technology Innovation Fund of Shanxi Agricultural University (2024BQ74), and 2024 Provincial Doctoral Innovation Project (2024KY301), National Natural Science Foundation of China (42301465).

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

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.

References

- Yang XM, Ma SH, Huang EH, Zhang DH, Chen GP, Zhu JL, et al. Nitrogen addition promotes soil carbon accumulation globally. *Sci China Life Sci.* (2025) 68:284–93. doi: 10.1007/s11427-024-2752-2
- Kopittke PM, Menzies NW, Wang P, McKenna BA, Lombi E. Soil and the intensification of agriculture for global food security. *Environ Int.* (2019) 132:105078. doi: 10.1016/j.envint.2019.105078
- Li ZW, Wang GY, Khan K, Yang L, Chi YX, Wang Y, et al. Irrigation combines with nitrogen application to optimize soil carbon and nitrogen, increase maize yield, and nitrogen use efficiency. *Plant Soil.* (2024) 499:605–20. doi: 10.1007/s11104-024-06480-6
- Litke L, Gaile Z, Ruza A. Effect of nitrogen fertilization on winter wheat yield and yield quality. *Agron Res.* (2018) 16:500–9. doi: 10.15159/AR.18.064
- Liu Y, Liao YC, Liu WZ. High nitrogen application rate and planting density reduce wheat grain yield by reducing filling rate of inferior grain in middle spikelets. *Crop J.* (2021) 9:412–26. doi: 10.1016/j.cj.2020.06.013
- Yao XQ, Yang W, Li MZ, Zhou P, Liu Z. Prediction of total nitrogen content in different soil types based on spectroscopy. *IFAC-PapersOnLine.* (2019) 52:270–6. doi: 10.1016/j.ifacol.2019.12.533
- Gruszczynski S, Gruszczynski W. Supporting soil and land assessment with machine learning models using the Vis-NIR spectral response. *Geoderma.* (2022) 405:115451. doi: 10.1016/j.geoderma.2021.115451
- Cundill SL, van der Werf HMA, Van der Meijde M. Adjusting spectral indices for spectral response function differences of very high spatial resolution sensors simulated from field spectra. *Sensors-Basel.* (2015) 15:6221–40. doi: 10.3390/s150306221
- Yang CB, Feng MC, Song LF, Jing BH, Xie YK, Wang C, et al. Study on hyperspectral monitoring model of soil total nitrogen content based on fractional-order derivative. *Comput Electron Agric.* (2022) 201:107307. doi: 10.1016/j.compag.2022.107307
- Liu SG, Hu ZQ, Han JZ, Li YY, Zhou T. Predicting grain yield and protein content of winter wheat at different growth stages by hyperspectral data integrated with growth monitor index. *Comput Electron Agric.* (2022) 200:107235. doi: 10.1016/j.compag.2022.107235
- Zhu HF, Hu W, Jing YD, Cao Y, Bi RT, Yang WD. Soil organic carbon prediction based on scale-specific relationships with environmental factors by discrete wavelet transform. *Geoderma.* (2018) 330:9–18. doi: 10.1016/j.geoderma.2018.05.029
- Bao YL, Ustin S, Meng XT, Zhang XL, Guan HX, Qi BS, et al. A regional-scale hyperspectral prediction model of soil organic carbon considering geomorphic features. *Geoderma.* (2021) 403:115263. doi: 10.1016/j.geoderma.2021.115263
- Gao CR, Yan XB, Qiao XX, Wei KX, Zhang XM, Yang S, et al. Multivariate prediction of soil aggregate-associated organic carbon by simulating satellite sensor bands. *Comput Electron Agric.* (2023) 209:107859. doi: 10.1016/j.compag.2023.107859
- Geng J, Lv JW, Pei J, Liao CH, Tan QY, Wang TX, et al. Prediction of soil organic carbon in black soil based on a synergistic scheme from hyperspectral data: Combining fractional-order derivatives and three-dimensional spectral indices. *Comput Electron Agric.* (2024) 220:108905. doi: 10.1016/j.compag.2024.108905
- Gao CR, Li H, Wang JC, Zhang X, Huang KM, Song XY, et al. Combined use of spectral resampling and machine learning algorithms to estimate soybean leaf chlorophyll. *Comput Electron Agric.* (2024) 218:108675. doi: 10.1016/j.compag.2024.108675
- Yang YC, Zhao YJ, Qin K, Zhao NB, Yang C, Zhang DH, et al. Prediction of black soil nutrient content based on airborne hyperspectral remote sensing. *Trans Chin Soc Agric Eng.* (2019) 35:94–101. doi: 10.11975/j.issn.1002-6819.2019.20.012
- Zhong L, Guo X, Ding M, Ye YC, Zhu Q, Guo J, et al. Spatial mapping of topsoil total nitrogen in mountainous and hilly areas of southern China using a continuous convolution neural network. *Catena.* (2023) 229:107228. doi: 10.1016/j.catena.2023.107228
- Xiao W, Chen WQ, He TT, Ruan LL, Guo JW. Multi-temporal mapping of soil total nitrogen using google earth engine across the shandong province of China. *Sustainability-Basel.* (2020) 12:10274. doi: 10.3390/su122410274
- Mashaba-Munghomezulu Z, Chirima GJ, Munghomezulu C. Modeling the spatial distribution of soil nitrogen content at smallholder maize farms using machine learning regression and sentinel-2 data. *Sustainability-Basel.* (2021) 13:11591. doi: 10.3390/su132111591
- Zhang QW, Liu MY, Zhang YB, Mao DH, Li F, Wu FP, et al. Comparison of machine learning methods for predicting soil total nitrogen content using landsat-8, sentinel-1, and sentinel-2 images. *Remote Sens.* (2023) 15:2907. doi: 10.3390/rs15112907
- Liu ZF, Lei HC, Lei L, Sheng HY. Spatial prediction of total nitrogen in soil surface layer based on machine learning. *Sustainability-Basel.* (2022) 14:11998. doi: 10.3390/su141911998
- Bennett ND, Croke BF, Guariso G, Guillaume JH, Hamilton SH, Jakeman AJ, et al. Characterising performance of environmental models. *Environ Model Softw.* (2013) 40:1–20. doi: 10.1016/j.envsoft.2012.09.011
- Liu WM, Cheng ZY, Li J, Li G, Pan NH. Assessment of ecological asset quality and its drivers in Agro-pastoral Ecotone of China. *Ecol Indic.* (2025) 170:113072. doi: 10.1016/j.ecolind.2025.113072
- Li WJ, Zhang K, Liu JX, Wu J, Zhang Y, Henke M. Optimizing daylily (*Heemerocallis citrina* Baroni) cultivation: integrating physiological modeling and planting patterns for enhanced yield and resource efficiency. *Front Plant Sci.* (2024) 15:1442485. doi: 10.3389/fpls.2024.1442485
- Ma XR, Wang LD, Zhu HF, Peng JJ, Bi RT. Impact of soil factors on the yield and agronomic traits of *Heemerocallis citrina* Baroni in the agro-pastoral ecotone of Northern China. *Agron J.* (2024) 14:967. doi: 10.3390/agronomy14050967
- Wang Z, Yao ZL, Zhu HF, Bi RT. Soil organic carbon stocks under daylily cultivation and their influencing factors in the agro-pastoral ecotone of Northern China. *Agron J.* (2025) 15:756. doi: 10.3390/agronomy15030756
- Peng JJ, Li S, Ma XR, Ding HX, Fang WJ, Bi RT. Spatial distribution and influencing factors of daylily cultivation in the farming-pastoral ecotone of Northern China. *Land.* (2024) 13:439. doi: 10.3390/land13040439
- Li J, Han ZP, Li YQ, Zhang HX. Biological characteristics of Daylily in Datong and its high-yield cultivation technology. *Horticult Seed.* (2019) 5:5–10. doi: 10.3390/horticult.2019.12.135
- Xu SX, Wang MY, Shi XZ, Yu QB, Zhang ZQ. Integrating hyperspectral imaging with machine learning techniques for the high-resolution mapping of soil nitrogen fractions in soil profiles. *Sci Tot Environ.* (2021) 754:142135. doi: 10.1016/j.scitotenv.2020.142135
- Lu GY, Liu YL, Wang J, Wu HP. CNN-BiLSTM-Attention: A multi-label neural classifier for short texts with a small set of labels. *Inform Proc Manag.* (2023) 60:103320. doi: 10.1016/j.ipm.2023.103320
- Zeng YJ, Verhoef A, Vereecken H, Ben-Dor E, Veldkamp T, Shaw L, et al. Monitoring and modeling the soil-plant system toward understanding soil health. *Rev Geophys.* (2025) 63:e2024RG000836. doi: 10.1029/2024RG000836
- Frampton WJ, Dash J, Watmough G, Milton EJ. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS J Photogramm.* (2013) 82:83–92. doi: 10.1016/j.isprsjprs.2013.04.007
- Hu MH, Wang JS, Yang P, Li P, He P, Bi RT. Estimation of daylily leaf area index by synergy multispectral and radar remote-sensing data based on machine-learning algorithm. *Agron J.* (2025) 15:456. doi: 10.3390/agronomy15020456
- Ustin SL, Middleton EM. Current and near-term advances in Earth observation for ecological applications. *Ecol Proc.* (2021) 10:1. doi: 10.1186/s13717-020-00255-4
- Feng W, Guo BB, Zhang HY, He L, Zhang YS, Wang YH, et al. Remote estimation of above ground nitrogen uptake during vegetative growth in winter wheat using hyperspectral red-edge ratio data. *Field Crops Res.* (2015) 180:197–206. doi: 10.1016/j.fcr.2015.05.020
- Fan YG, Feng HK, Jin XL, Yue JB, Liu Y, Li ZH, et al. Estimation of the nitrogen content of potato plants based on morphological parameters and visible light vegetation indices. *Front Plant Sci.* (2022) 13:1012070. doi: 10.3389/fpls.2022.1012070
- Mulla DJ. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosyst Eng.* (2013) 114:358–71. doi: 10.1016/j.biosystemseng.2012.08.009
- Liu SS, Peng YP, Xia ZQ, Hu YM, Wang GX, Zhu AX, et al. The GA-BPNN-based evaluation of cultivated land quality in the PSR framework using Gaofen-1 satellite data. *Sensors-Basel.* (2019) 19:5127. doi: 10.3390/s19235127
- Peng YP, Zhao L, Hu YM, Wang GX, Wang L, Liu ZH. Prediction of soil nutrient contents using visible and near-infrared reflectance spectroscopy. *ISPRS Int J Geo-Inf.* (2019) 8:437. doi: 10.3390/ijgi8100437
- Jia ZH, Wu KM, Wang HT, Zeng WH, Guo Y, Liang D. An improved crop yield prediction using CNN-BiLSTM model with attention mechanism. *J ASABE.* (2024) 67:1459–67. doi: 10.13031/ja.15629
- Li HY, Jia SY, Le ZC. Quantitative analysis of soil total nitrogen using hyperspectral imaging technology with extreme learning machine. *Sensors-Basel.* (2019) 19:4355. doi: 10.3390/s19204355
- Salunke R, Nobahar M, Alzaghoul OE, Khan S, La Cour I, Amini F. Near-surface soil moisture characterization in Mississippi's highway slopes using machine learning methods and UAV-captured infrared and optical images. *Remote Sens.* (2023) 15:1888. doi: 10.3390/rs15071888
- Zhou T, Geng YJ, Chen J, Pan JJ, Haase D, Lausch A. High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM derivatives, Sentinel-1 and Sentinel-2 data based on machine learning algorithms. *Sci Tot Environ.* (2020) 729:138244. doi: 10.1016/j.scitotenv.2020.138244
- Gastal F, Lemaire G. N uptake and distribution in crops: an agronomical and ecophysiological perspective. *J Exp Bot.* (2002) 53:789–99. doi: 10.1093/jexbot/53.370.789
- Hassan MA, Yang MJ, Rasheed A, Yang GJ, Reynolds M, Xia XC, et al. A rapid monitoring of NDVI across the wheat growth cycle for grain yield prediction using a multi-spectral UAV platform. *Plant Sci.* (2019) 282:95–103. doi: 10.1016/j.plantsci.2018.10.022
- Qi DL, Hu TT, Liu TT. Biomass accumulation and distribution, yield formation and water use efficiency responses of maize (*Zea mays* L.) to nitrogen supply methods

- under partial root-zone irrigation. *Agric Water Manage.* (2020) 230:105981. doi: 10.1016/j.agwat.2019.105981
47. Zhang Y, Li WJ, Han ZP, Zhang K, Liu JW, Henke M. Three-dimensional dynamic growth and yield simulation model of daylily plants. *Smart Agric.* (2024) 6:140–53. doi: 10.12133/j.smartag.SA202310011
48. Yang CB, Bai J, Sun H, Bi RT, Song LF, 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. doi: 10.1016/j.compag.2024.109851
49. Zhu MB, Liu SS, Xia ZQ, Wang GX, Hu YM, Liu ZH. Crop growth stage GPP-driven spectral model for evaluation of cultivated land quality using GA-BPNN. *Agriculture-Basel.* (2020) 10:318. doi: 10.3390/agriculture10080318
50. Mao JH, Wang JS, Liao JQ, Xu XL, Tian DS, Zhang RY, et al. Plant nitrogen uptake preference and drivers in natural ecosystems at the global scale. *New Phytol.* (2025) 246:972–83. doi: 10.1111/nph.70030