Mitigating saturation effects in rice nitrogen estimation using Dualex measurements and machine learning

Peihua Shi[1], Yuan Wang[2], Congfei Yin[1], Kaiqing Fan[1], Yinfei Qian[3*†] and Gui Chen[4*†]

[1]Department of Agronomy and Horticulture, Jiangsu Vocational College of Agriculture and Forestry, Jurong, China, [2]State Key Laboratory of Soil and Sustainable Agriculture, Changshu National Agro-Ecosystem Observation and Research Station, Institute of Soil Science, Chinese Academy of Sciences, Nanjing, China, [3]Soil and Fertilizer & Resources and Environmental Institute, Jiangxi Academy of Agricultural Sciences, Nanchang, China, [4]Institute of Biotechnology, Jiaxing Academy of Agricultural Science, Jiaxing, China

Nitrogen is essential for rice growth and yield formation, but traditional methods for assessing nitrogen status are often labor-intensive and unreliable at high nitrogen levels due to saturation effects. This study evaluates the effectiveness of flavonoid content (Flav) and the Nitrogen Balance Index (NBI), measured using a Dualex sensor and combined with machine learning models, for precise nitrogen status estimation in rice. Field experiments involving 15 rice varieties under varying nitrogen application levels collected Dualex measurements of chlorophyll (Chl), Flav, and NBI from the top five leaves at key growth stages. Incremental analysis was performed to quantify saturation effects, revealing that chlorophyll measurements saturated at high nitrogen levels, limiting their reliability. In contrast, Flav and NBI remained sensitive across all nitrogen levels, accurately reflecting nitrogen status. Machine learning models, particularly random forest and extreme gradient boosting, achieved high prediction accuracy for leaf and plant nitrogen concentrations ($R^2 > 0.82$), with SHAP analysis identifying NBI and Flav from the top two leaves as the most influential predictors. By combining Flav and NBI measurements with machine learning, this approach effectively overcomes chlorophyll-based saturation limitations, enabling precise nitrogen estimation across diverse conditions and offering practical solutions for improved nitrogen management in rice cultivation.

KEYWORDS

rice nitrogen estimation, Dualex measurements, saturation effect, incremental analysis, machine learning, nitrogen balance index, SHAP analysis

1 Introduction

Nitrogen is a fundamental nutrient for rice growth and yield formation, playing a crucial role in processes such as photosynthesis, protein synthesis, and enzyme activity (Liu et al., 2021; Luo et al., 2021). Effective nitrogen management not only boosts rice yield and quality but also reduces environmental pollution from excessive fertilizer use, thereby supporting sustainable agriculture (Guo et al., 2020; Cai et al., 2023). Therefore, accurately monitoring the nitrogen nutritional status of rice is vital for developing effective fertilization strategies.

Traditional nitrogen assessment methods largely depend on laboratory analyses, which are often cumbersome, time-consuming, and expensive, making them impractical for large-scale, real-time monitoring (Berger et al., 2020). To address the need for fast, non-destructive measurements, devices such as the SPAD chlorophyll meter have been widely used (Tan et al., 2021; Singh et al., 2022). These tools estimate chlorophyll content by analyzing the optical properties of leaves, thus providing an indirect measure of nitrogen levels. However, chlorophyll measurements often show saturation effects at high nitrogen levels, limiting their ability to accurately capture differences in nitrogen content (Jiang et al., 2021; Gao et al., 2024). This drawback affects their precision under high-nitrogen conditions, hindering accurate nitrogen management.

To overcome these limitations, multispectral and hyperspectral remote sensing technologies at the canopy level have been introduced for agricultural nitrogen monitoring (Fu et al., 2021; Qian et al., 2022). These methods utilize the reflectance characteristics of vegetation across various spectral bands to develop indices such as the Normalized Difference Vegetation Index (NDVI) and the Optimized Soil-Adjusted Vegetation Index (OSAVI). These indices help estimate crop nitrogen content and biomass (Qian et al., 2022; Yang et al., 2023; Li et al., 2024). Although remote sensing is beneficial for large-scale, non-contact monitoring, it is also subject to limitations. Data reliability can be affected by weather conditions, observation angles, and variations in light, leading to inconsistencies (Ma et al., 2020; Wu et al., 2023). Furthermore, at high biomass or nitrogen levels, these indices may also exhibit saturation effects, resulting in reduced accuracy in nitrogen estimation (Zhang et al., 2019b; Chang et al., 2022).

Saturation effects present a major challenge in nitrogen estimation. When measurements saturate at high nitrogen levels, distinguishing between different nitrogen statuses becomes difficult, regardless of whether chlorophyll content or canopy-level indices are used. This leads to reduced accuracy in high-nitrogen conditions, complicating the implementation of precise nitrogen management. Therefore, new approaches that can mitigate or eliminate saturation effects are urgently needed to enhance the accuracy and reliability of nitrogen estimation.

Abbreviations: Chl, Chlorophyll; Flav, Flavonoid; NBI, Nitrogen Balance Index; LNC, Leaf Nitrogen Concentration; PNC, Plant Nitrogen Concentration; NNI, Nitrogen Nutrition Index; PLS, Partial Least Squares Regression; SVR, Support Vector Regression; RF, Random Forest; XGB, Extreme Gradient Boosting; NN, Neural Network; SHAP, SHapley Additive exPlanations; RMSE, Root Mean Square Error; RRMSE, Relative Root Mean Square Error.

Flavonoids (Flav) are a diverse group of phenolic compounds that are abundantly present in plants and play key roles in their responses to biotic and abiotic stresses (Gitelson et al., 2017; Hamdane et al., 2023). Under stress conditions, such as nitrogen deficiency, Flav levels increase, making them reliable indicators of nitrogen stress (Li et al., 2021; Waqas et al., 2023). The portable Dualex sensor facilitates the simultaneous measurement of chlorophyll (Chl) and Flav in leaves, allowing the calculation of the Nitrogen Balance Index (NBI), which is derived as the ratio of Chl to Flav (Cerovic et al., 2012, 2015). Measurements of Chl and Flav have been effective in differentiating genotypic responses to nitrogen and water stress in wheat (Raya-Sereno et al., 2024). Studies have shown that the Dualex sensor maintains a stronger linear correlation with actual chlorophyll concentration than devices like SPAD-502 and CCM-200, especially at higher chlorophyll levels. It also demonstrates higher reliability in tests on crops such as maize, soybean, and wheat, being less susceptible to the effects of uneven chlorophyll distribution (Dong et al., 2019, 2022). Dong et al. (2021) successfully estimated the Nitrogen Nutrition Index (NNI) in maize across various growth stages by applying multiple linear regression models that integrated Dualex readings with environmental variables. Overall, Dualex measurements prove to be valuable in estimating plant chlorophyll and nitrogen content, as well as in distinguishing different levels of nitrogen and water stress, particularly maintaining high sensitivity even under high nitrogen conditions (Rubo and Zinkernagel, 2022). However, there is still a lack of comprehensive research on the role of Flav and NBI in rice nitrogen estimation and their performance under varying nitrogen conditions.

Nitrogen distribution within rice plants shows variability across different leaf positions (Chang et al., 2019; Ouyang et al., 2021). Karaca et al. (2023) highlighted that optical sensor data collected from various leaf positions can improve the estimation of crop nitrogen status. In crops like melon and sweet pepper, data from lower leaves demonstrated superior predictive performance at certain growth stages. Another study reported that Dualex sensor measurements across different leaf positions performed similarly when estimating the NNI. However, data from the third fully expanded Leaf From the Top (LFT) often showed higher predictive accuracy, especially during the later growth stages of maize (Dong et al., 2021). Some studies also suggest that variations in leaf layers or specific measurement positions do not significantly impact the assessment of nitrogen content in wheat when using spectral reflectance (Röll et al., 2019). In rice, SPAD readings taken from the upper-middle section of the fourth LFT have been found to provide more accurate and stable reflections of nitrogen status (Yuan et al., 2016). These findings suggest that relying on a single leaf position may not comprehensively capture the overall nitrogen status of the plant. Therefore, using multi-leaf, multi-indicator data could improve nitrogen diagnosis accuracy. However, integrating data from multiple leaf positions and indicators to build accurate nitrogen estimation models remains a challenge that requires further investigation.

Machine learning models are particularly effective at handling complex nonlinear relationships and multivariate data, making them well-suited for agricultural data analysis and prediction tasks (Shi et al., 2021; Huang et al., 2023; Mandal et al., 2024).

Combining machine learning with Dualex measurements offers the potential to enhance nitrogen estimation accuracy and mitigate saturation effects. Based on this framework, our study systematically investigates the relationships between Dualex measurements and rice nitrogen indicators through multi-location, multi-variety, and multi-leaf-position testing. We assess the effectiveness of Flav and NBI in reducing saturation effects at high nitrogen levels. We also develop rice nitrogen nutritional indicator estimation models by integrating machine learning, focusing on optimizing feature selection and model structure to enhance interpretability and practical application. By addressing the limitations of traditional methods under high nitrogen conditions and extending applicability across different growth stages, we aim to ultimately improve the efficiency and sustainability of agricultural production.

2 Materials and methods

2.1 Study area and experimental design

Field experiments were conducted across Jiangsu, Zhejiang, and Jiangxi provinces in China. These experiments followed a randomized block design, incorporating 3 to 6 different nitrogen application rates, each with 3 to 4 replicates. A total of 15 rice varieties were tested, including 10 japonica, 4 indica, and 1 indica-japonica hybrid, covering a range of regions and cultivation systems. Basic details of these varieties are shown in Table 1. Except for variations in nitrogen application, all other agronomic practices followed local conventional standards.

2.2 Data collection

Samples were obtained every 10-15 days from the Tillering (TI) stage to the Heading (HD) stage, corresponding to key growth stages: TI, Stem Elongation (SE), Panicle Initiation (PI), and HD. Due to differences in growth durations among the rice varieties, some early-maturing varieties had entered the Grain Filling (GF) stage when most varieties were at HD. Consequently, a small number of samples were collected during the GF stage for these varieties. At each sampling time, 3 to 6 hills of uniformly growing plants were destructively sampled from each plot. Samples were separated into leaves, sheaths, and panicles, then oven-dried at 105°C for 30 minutes, followed by 75°C until a constant weight was achieved. Nitrogen concentration was determined using the Kjeldahl method.

The NNI was derived by calculating the ratio of Plant Nitrogen Concentration (PNC) to the Critical Nitrogen concentration (N_c). NNI values greater than 1 indicate nitrogen surplus, 1 indicates sufficiency, and less than 1 indicates deficiency.

$$NNI = \frac{PNC}{N_c}$$

where N_c is calculated using the nitrogen dilution curve established from our dataset:

$$N_c = 3.44 \times W^{-0.43}$$

where W represents the above-ground dry matter (t·ha^{-1}). The parameters (3.44 and -0.43) were derived from our experimental data using the method described by Justes et al. (1994). This method involves plotting plant nitrogen concentration against biomass and

TABLE 1 Summary of field experiments and nitrogen treatments.

No.	Varieties	Subspecies	Experimental Site	Growth Period	Nitrogen Rate
1	NG3908	japonica	Nanjing	May–Oct	0, 60, 120, 180, 240, 300
2	NG5055	japonica	Nanjing	May–Oct	0, 100, 150, 200, 250, 300
3	NG46	japonica	Zhenjiang	May–Oct	0, 60, 120, 180, 240, 300
4	YG13	japonica	Zhenjiang	May–Oct	0, 75, 150, 225,300
5	NG3	japonica	Zhenjiang	May–Oct	0, 75, 150, 225,300
6	JH218	japonica	Jiaxing	May–Oct	0, 80, 120, 140, 160
7	JYZK6	japonica	Jiaxing	May–Oct	0, 200, 300
8	XS14	japonica	Jiaxing	May–Oct	0, 100, 200
9	YY15	Hybrid[1]	Jiaxing	May–Oct	0, 189, 230, 270
10	J67	japonica	Jiaxing	May–Oct	0, 127, 169
11	HHZ	indica	Yichun	Jun–Oct	0, 135, 180, 225
12	MXXZ	indica	Yichun	Jun–Oct	0, 135, 180, 225
13	CY5	japonica	Yichun	Jun–Oct	0, 135, 180, 225
14	TYXZ	indica	Yichun	Jun–Oct	0, 135, 180, 225
15	TYHZ	indica	Yichun	Jun–Oct	0, 135, 180, 225

[1]This rice variety is an indica-japonica hybrid bred by the Ningbo Academy of Agricultural Sciences.

fitting a power function to the minimum nitrogen concentration observed at each biomass level, representing the N_c required for optimal growth.

On the same day as the destructive sampling, 10 to 20 rice plants were selected from each plot to measure the 1-5 LFT using a Dualex Scientific+ instrument (Force-A Inc., France). Each leaf was measured at three positions: middle, upper one-third, and lower one-third, and the average value was calculated for that leaf. The Dualex Scientific+ is a portable optical sensor designed for rapid, non-destructive detection of various physiological parameters in plant leaves. By measuring the transmission and reflection of specific wavelengths of light, it provides readings for Chl, Flav, Anthocyanin content (Anth), and the NBI. NBI is the ratio of Chl to Flav, serves as an indirect measure of the plant's nitrogen nutritional status. Although the Dualex Scientific+ instrument also measures Anth, preliminary analyses indicated that anthocyanin levels in rice leaves were consistently low and showed minimal variation across treatments and growth stages. Consequently, Anth was not included in the subsequent analyses.

During early growth stages, such as the TI stage, it was sometimes not possible to identify five fully expanded leaves in a shoot due to the limited number of leaves. In these cases, we measured at least the 1-4 LFT and calculated the average values. This average was then used as the value for the 5 LFT. This data preprocessing step ensured consistency across samples and prevented errors due to unequal data when applying machine learning models.

2.3 Modeling methods

The study employed several modeling approaches including Simple Linear Regression (LR), Quadratic Curve Regression (QCR), Partial Least Squares Regression (PLS), and four machine learning models—Support Vector Regression (SVR), Random Forest (RF), Extreme Gradient Boosting (XGB), and Neural Network (NN)—to estimate rice Leaf Nitrogen Concentration (LNC), PNC, and NNI. The hyperparameters for each machine learning model were selected based on preliminary analyses that involved testing various parameter combinations using grid search and cross-validation. This approach helped balance model complexity and performance, ensuring robust generalization. By comparing the performance of these models, their applicability in estimating rice nitrogen nutrition indices was evaluated.

2.3.1 Simple linear regression and quadratic curve regression

Initial analyses used LR and QCR to examine the relationships between individual Dualex measurements (Chl, Flav, NBI) and nitrogen indices (LNC, PNC, NNI) to gain a preliminary understanding of variable interactions and fitting effectiveness.

2.3.2 Partial least squares regression

PLS regression, a linear method suitable for multivariate, highly correlated data, was employed. It extracts latent variables that maximize the covariance between independent and dependent variables. In this study, the PLS model used two principal components to balance complexity and explanatory power.

2.3.3 Support vector regression

SVR, an application of support vector machines in regression, is capable of handling nonlinear and high-dimensional data. The SVR model utilized a radial basis function kernel, with the following key hyperparameters:

- Penalty parameter $C = 1.0$: Controls model complexity to prevent overfitting.
- Epsilon $\varepsilon = 0.2$: Defines the width of the support vector boundary, influencing the model's robustness.

2.3.4 Random forest

RF is an ensemble learning method that constructs multiple decision trees and averages their outcomes to reduce overfitting. Key hyperparameters of the RF model included:

- Number of trees $n_{estimators} = 30$: Ensures sufficient model diversity.
- Maximum depth *max_depth* = 5: Limits tree depth to prevent overfitting.

2.3.5 Extreme gradient boosting

XGB, a tree-based ensemble model, employs gradient boosting for efficient and accurate predictions. Its key hyperparameters were set as follows:

- Number of trees $n_{estimators} = 30$.
- Maximum depth *max_depth* = 4: Limits tree depth to prevent overfitting.
- Learning rate *learning_rate* = 0.1: Regulates each tree's contribution; a smaller rate enhances generalization.
- Regularization parameters *reg_alpha* = 0.5 and *reg_lambda* = 0.5: Apply L1 and L2 regularization to prevent overfitting.

2.3.6 Neural network

The NN model was constructed as a multilayer perceptron regressor with the following structure and parameters:

- Hidden layer size *hidden_layer_sizes* = (100),: One hidden layer with 100 neurons.
- Maximum iterations *max_iter* = 500: Restricts training time.
- Activation function: Rectified Linear Unit (ReLU) to enhance nonlinear representation.
- Learning rate: Adaptive learning rate to optimize training.

2.3.7 Model training and validation

For LR and QCR, the models used average values of Chl, Flav, or NBI from the 1-5 LFT as single input variables. For PLS and machine learning models, four different variable combinations served as inputs:

- Combination 1: Chl measurements from the 1-5 LFT and their average (six variables).
- Combination 2: Flav measurements from the 1-5 LFT and their average.
- Combination 3: NBI measurements from the 1-5 LFT and their average.
- Combination 4: All variables from the three combinations above.

A 10-fold cross-validation was employed to randomly divide the dataset into training and validation sets, ensuring reliable and stable model evaluation. This method reduces the model's dependence on data partitioning and enhances generalization.

2.4 Saturation effect analysis

To evaluate the saturation effects of different spectral measurements on rice nitrogen indices, we conducted an incremental change analysis that included the following steps:

2.4.1 Data normalization
Normalization was performed using Min-Max scaling on both independent (X) and dependent (Y) variables to make them comparable.

$$X_{scaled} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}, \quad Y_{scaled} = \frac{Y - Y_{\min}}{Y_{\max} - X_{\min}}$$

2.4.2 Sorting
Data were sorted in ascending order based on X_{scaled} to facilitate the calculation of incremental changes, ensuring accurate reflection of X's impact on Y.

2.4.3 Increment calculation
We quantified the effect of X_{scaled} changes on Y_{scaled} by computing increments between adjacent data points.

$$\Delta X_i = X_{scaled,i+1} - X_{scaled,i}, \quad \Delta Y_i = Y_{scaled,i+1} - Y_{scaled,i}$$

2.4.4 Threshold for ΔX
A minimum $|\Delta X_i|$ threshold of 1×10^{-3} was applied to exclude data points with very small ΔX_i values. This step prevents excessively large rates of change in the subsequent calculations, ensuring stability and accuracy in the rate computations.

$$\Delta X_i = \begin{cases} \Delta X_i, & if |\Delta X_i| \geq 1 \times 10^{-3} \\ NaN, & if |\Delta X_i| < 1 \times 10^{-3} \end{cases}$$

2.4.5 Rate of change computation
The rate of change was determined to assess Y_{scaled} variations relative to X_{scaled}, with rates scaled to the range [-1, 1] using the maximum absolute rate.

$$Rate_i = \frac{\Delta Y_i}{\Delta X_i}, \quad Rate\_Scaled_i = \frac{Rate_i}{max(|Rate|)}$$

2.4.6 Data smoothing
A moving average with a window size of 5 was used to smooth the rates, reducing random noise and emphasizing overall trends.

$$Rate\_Smoothed_i = \frac{1}{5} \sum_{j=i-2}^{i+2} Rate\_Scaled_j$$

2.4.7 Density curve analysis
To examine the behavior at high nitrogen levels, we focused on the top 20% of Y_{scaled} values and used Gaussian Kernel Density Estimation (KDE) on the corresponding X_{scaled} to obtain a smooth estimate of their probability density function. This non-parametric method does not assume any specific underlying distribution, making it suitable for real-world data. It helps in assessing whether the measurements continue to respond to increases in nitrogen levels or exhibit saturation.

2.5 Feature importance analysis and model evaluation

To gain deeper insights into the contributions of each input variable to the model's predictions, we employed the SHAP (SHapley Additive exPlanations) method (Lundberg and Lee, 2017), which is based on game theory, to analyze feature importance for the RF and XGB models. SHAP values quantify the impact of each feature on the prediction outcomes by calculating its marginal contribution across all possible feature combinations. Compared to traditional feature importance metrics, the SHAP method offers consistency and additivity, enhancing model interpretability and providing guidance for optimizing measurement strategies in practical applications.

To comprehensively evaluate the predictive performance of each model, we used the coefficient of determination (R^2), Root Mean Square Error (RMSE), and Relative Root Mean Square Error (RRMSE) as evaluation metrics. The formulas for these indicators are as follows:

- **Coefficient of determination**:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

- **Root mean square error**:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y})^2}$$

- **Relative root mean square error**:

$$RRMSE = \frac{RMSE}{\bar{Y}} \times 100\%$$

where Y_i is the actual value, $\hat{Y}_i$ is the predicted value, $\bar{Y}$ is the mean of the actual values, and n is the number of samples.

2.6 Statistical analysis tools

All statistical analyses and model construction in this study were performed using Python 3.12.4. Data processing and analysis utilized libraries such as Pandas, NumPy, and SciPy. Machine learning modeling was done using Scikit-learn, XGB, TensorFlow, and Keras, while visualization was carried out using Matplotlib. Statistical tests were conducted with Statsmodels and SciPy.stats, and feature importance analysis was performed using the SHAP library. To ensure reproducibility of results, all random processes were set with the same random seed during modeling and testing.

3 Results

3.1 Relationship between Dualex measurements and rice nitrogen indicators

Average Dualex values (Chl, Flav, NBI) from the first to fifth LFT were analyzed using LR and QCR models to explore their relationship with rice nitrogen nutritional indicators—LNC, PNC, and the NNI (Figure 1). The findings revealed notable differences in the effectiveness of these indicators for nitrogen estimation depending on the specific Dualex metrics used.

Chl displayed a nonlinear relationship with both LNC and PNC (Figures 1A, D). As Chl levels increased, both LNC and PNC rose, showing accelerated growth rates at Chl values around 30. However, a saturation effect was evident between Chl values of 35 to 40, where further increases in Chl did not lead to corresponding changes in LNC and PNC.
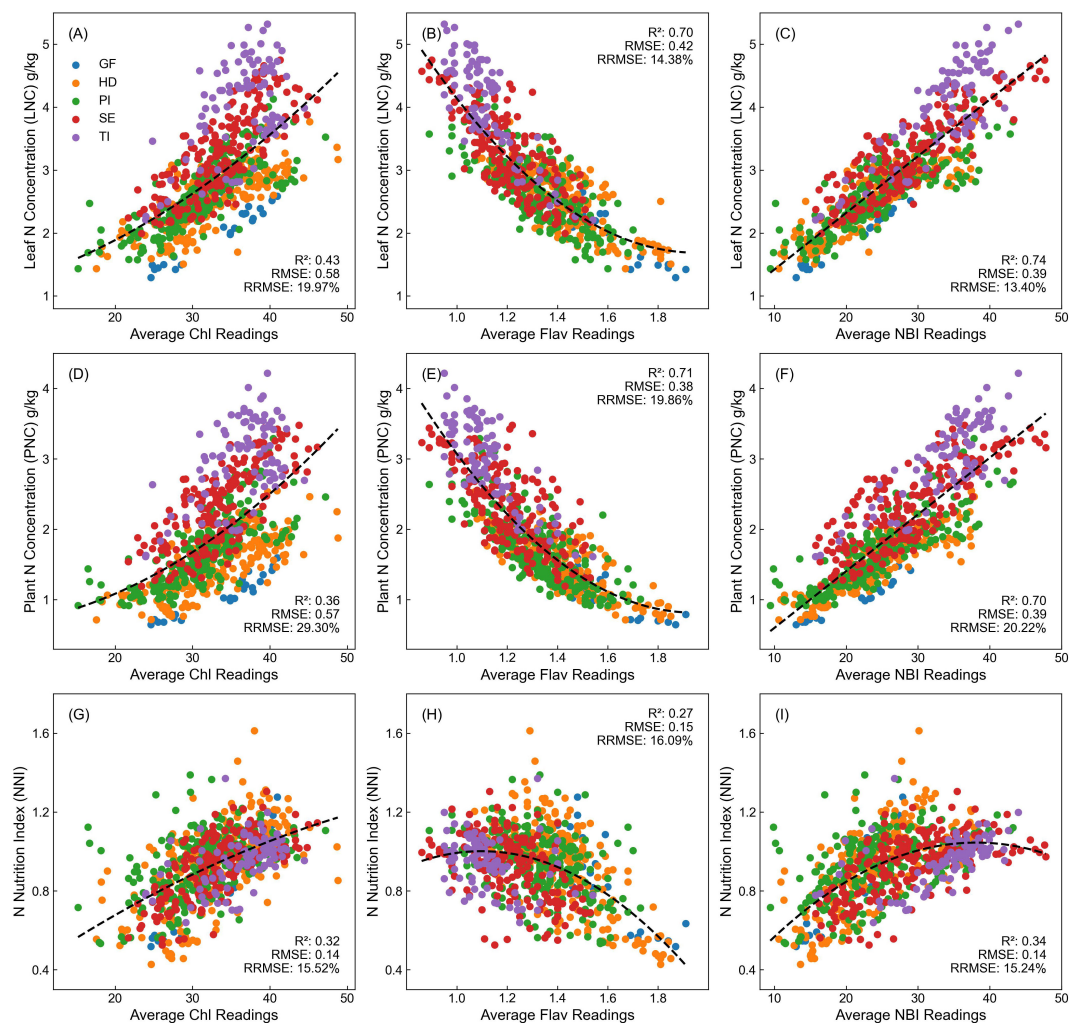

FIGURE 1
Relationship between Dualex values and rice nitrogen indicators across different varieties and growth stages. **(A–C)** show Chl, Flav, and NBI vs. LNC; **(D–F)** show Chl, Flav, and NBI vs. PNC; **(G–I)** show Chl, Flav, and NBI vs. NNI.

Conversely, Flav showed a stable and linear relationship with LNC and PNC, without notable saturation effects (Figures 1B, E). Flav levels increased steadily as nitrogen content decreased, demonstrating a strong negative correlation with LNC and PNC across the full range, with high regression fits (R^2 = 0.70 and 0.71). Even when accounting for differences across rice varieties and growth stages, Flav consistently provided an accurate assessment of nitrogen status.

NBI, defined as the ratio of Chl to Flav, exhibited a significant linear relationship with LNC and PNC (Figures 1C, F), achieving R^2 values of 0.74 and 0.70, respectively. This underscores NBI's robustness and stability in estimating nitrogen nutritional status. As a composite indicator, NBI effectively mitigates the saturation effect observed when using Chl alone, making it a highly reliable tool for nitrogen diagnosis.

For NNI estimation, which represents a relative value (the ratio of PNC to the N_c), measurements across different growth stages showed consistent NNI values, primarily ranging between 0.4 and 1.4. However, Dualex readings varied across stages, leading to lower estimation accuracy for NNI (Figures 1G–I). Although NNI tended to increase with changes in Chl, Flav, and NBI, this trend exhibited significant variability across growth stages, resulting in inconsistent patterns.

3.2 Saturation effect analysis of Dualex values in estimating nitrogen indicators

The saturation effect analysis of Dualex values in estimating rice nitrogen nutritional indicators is illustrated in Figure 2. To minimize the impact of model fitting differences when evaluating saturation effects across various Dualex indices, we applied an incremental analysis method using raw and standardized data. The standardized rate of change ($\Delta Y/\Delta X$) was calculated based on
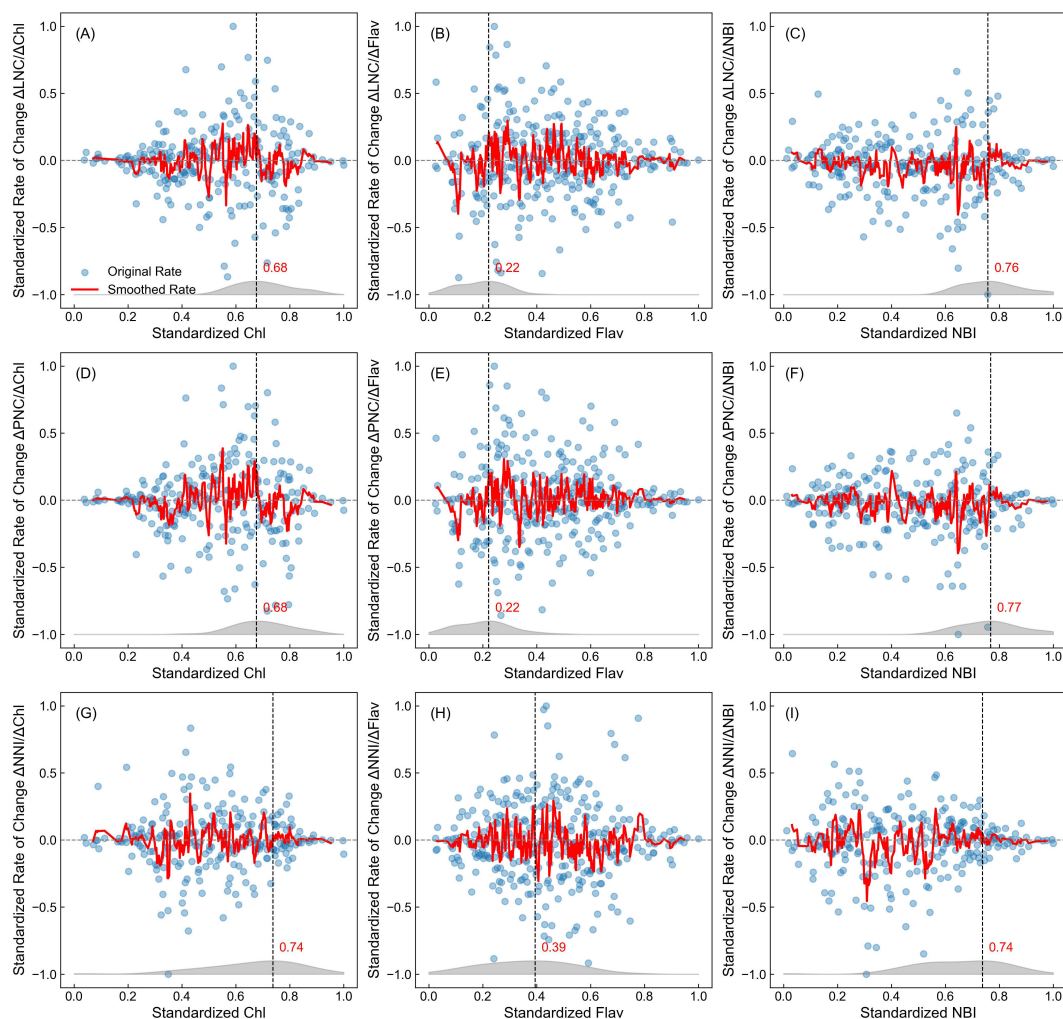

FIGURE 2
Standardized rate of change relative to Dualex values. This figure illustrates the standardized rate of change ($\Delta Y/\Delta X$) of LNC **(A–C)**, PNC **(D–F)**, and NNI **(G–I)** in relation to standardized Dualex values: Chl, Flav, and NBI. In each subplot, blue scatter points represent the original calculated standardized rate of change data, while the red curve shows the smoothed trend line obtained using a 5-point moving average, reflecting the overall trend. The gray curve at the bottom indicates the density distribution of measurement values corresponding to the highest 20% of nitrogen indices (Y values). The vertical black dashed line marks the point of highest density, annotated with the corresponding X-axis value.

these original data points. Each subplot displays the rate of change of nitrogen nutritional indicators relative to standardized Dualex values. In the analysis of Chl versus nitrogen indicators (Figures 2A, D, G), an initial low rate of change was observed in the lower Chl range, followed by a gradual increase, peaking at moderate Chl values. However, as Chl values continued to rise, the rate of change began to decline, particularly in the high Chl region. In the low Chl region, the low nitrogen indicator values were primarily responsible for the reduced rate, while in the high Chl region, a declining rate suggested that further increases in Chl no longer led to corresponding rises in nitrogen indicators, indicating a clear saturation effect. The density curves at the bottom further validated this: standardized Chl values peaked around 0.68 (for LNC and PNC) and 0.74 (for NNI), showing a marked decline in estimation accuracy at these levels.
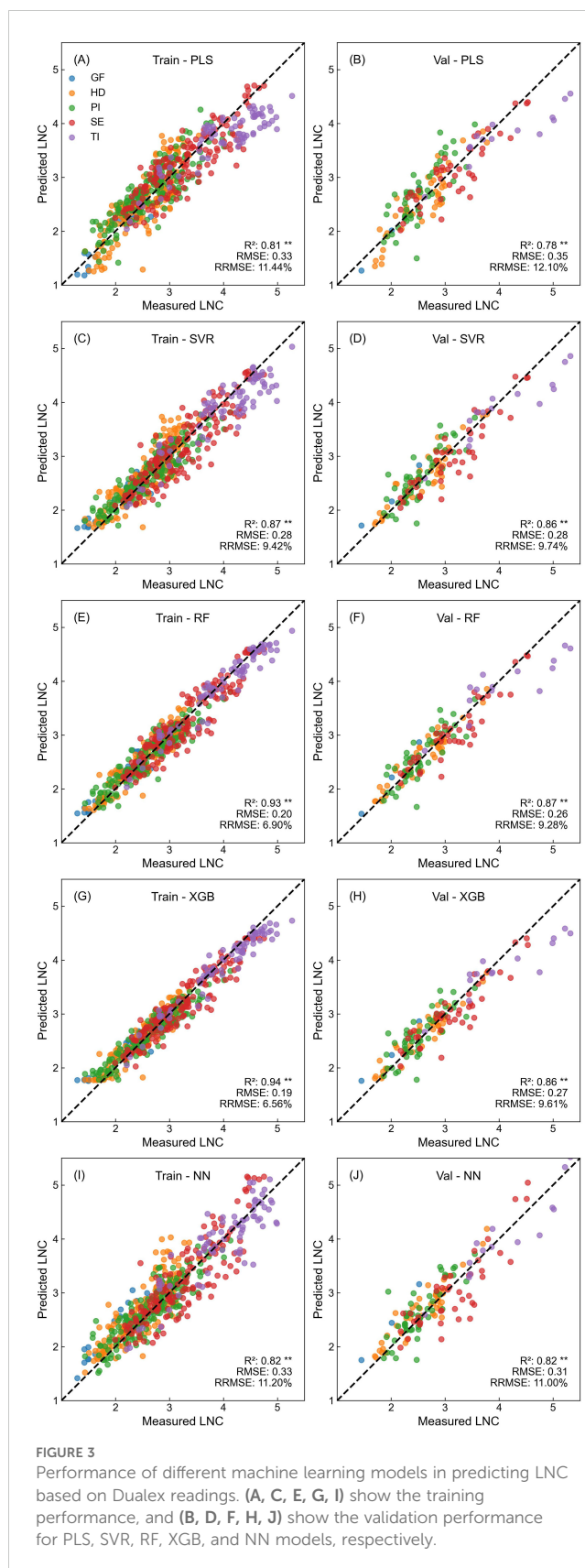
In contrast, Flav exhibited a negative correlation with nitrogen indicators (Figures 1B, E, H). The rate of change remained relatively stable in the low Flav region (Figures 2B, E, H), indicating effectiveness in estimating nitrogen indicators. The density peaks for Flav values were observed at 0.22 (for LNC and PNC) and 0.39 (for NNI), with a gradual distribution, suggesting that even at high nitrogen indicator levels, Flav measurements were evenly spread without a significant saturation effect. However, for NNI, as Flav values decreased, a rapid decline in the rate of change was noted, with the density peak at 0.39 indicating a more pronounced saturation. Comparing this to Figures 1H, NNI exhibited almost horizontal distribution in the low Flav range, indicating diminished sensitivity to changes in Flav.

As a composite indicator, NBI (Chl to Flav ratio) displayed a more consistent trend in estimating LNC, PNC, and NNI (Figures 2C, F, I). The rate of change initially increased, then decreased with rising NBI, resembling the pattern seen in Chl but with density peaks closer to boundary values (0.76, 0.77, and 0.74, respectively). This suggests that NBI effectively estimates nitrogen nutritional status even at high nitrogen levels, exhibiting significantly less saturation than Chl alone.
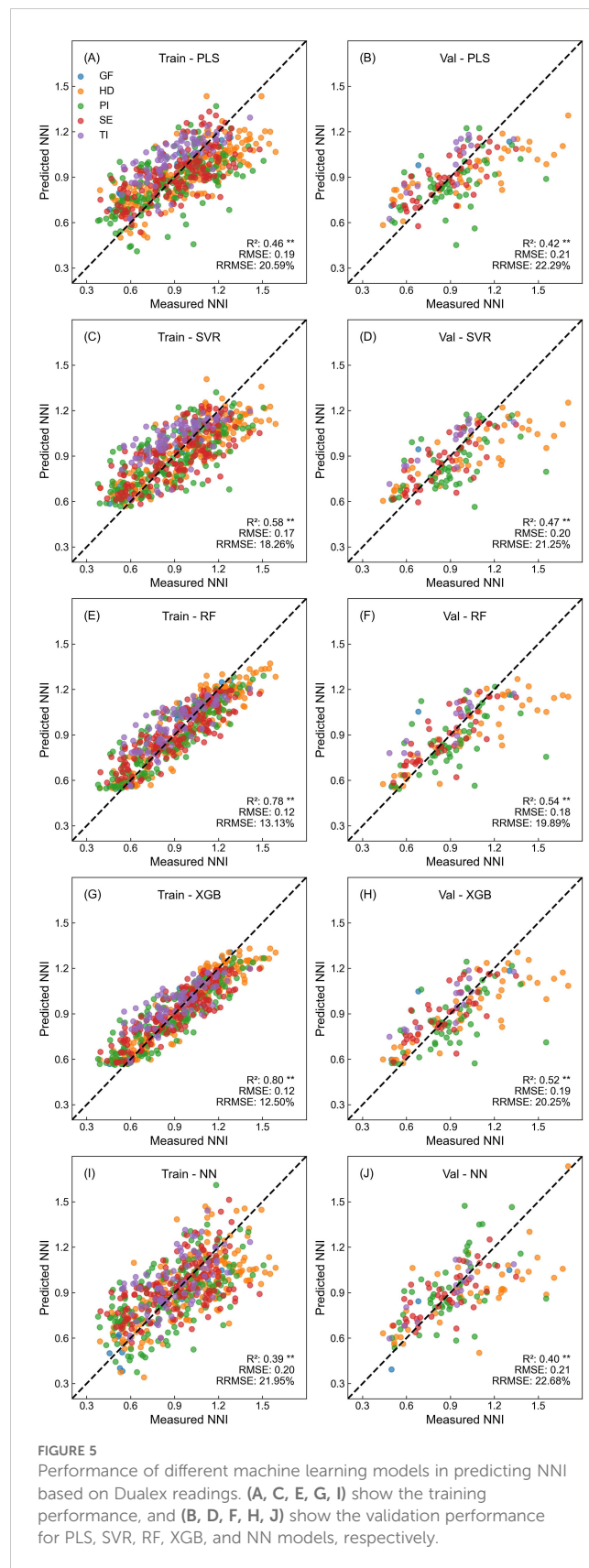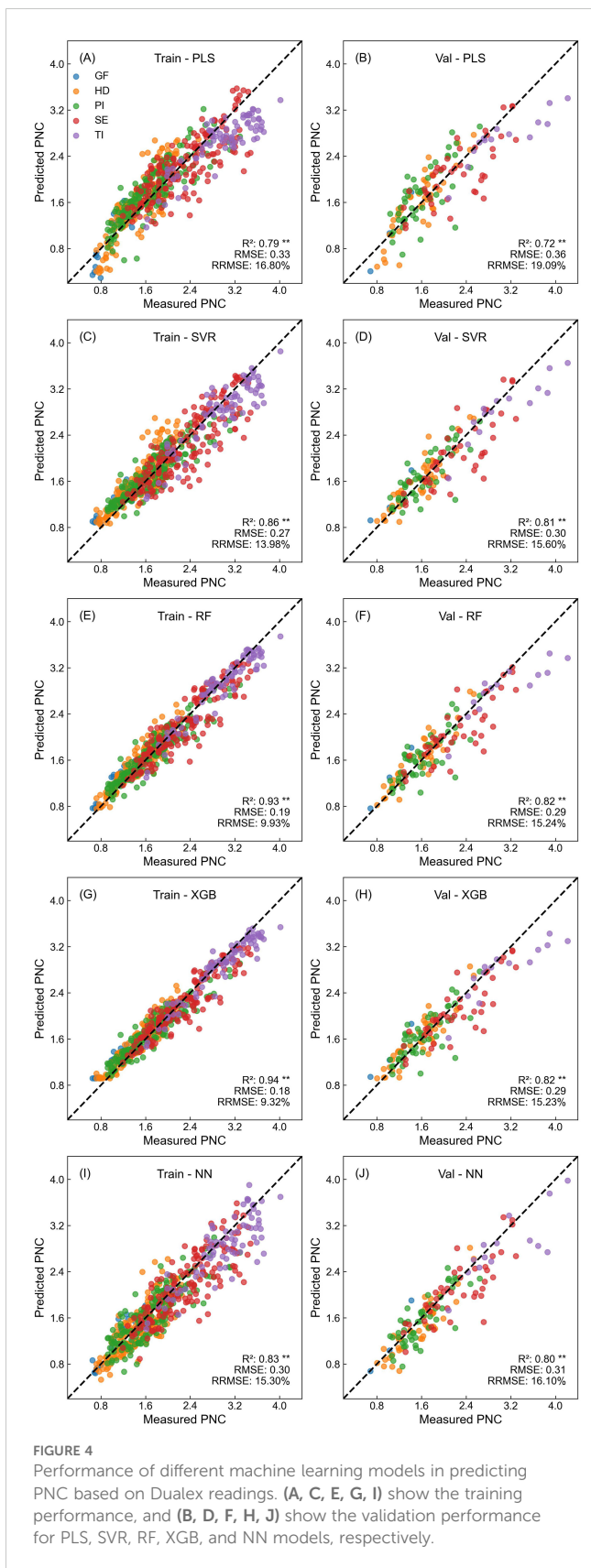
In summary, Chl demonstrated a clear saturation effect, particularly at high nitrogen levels, where its sensitivity diminished. Conversely, Flav and NBI retained considerable variability in their rates of change at high nitrogen levels, showing enhanced sensitivity to changes in nitrogen status and providing a more accurate reflection of variations in rice nitrogen nutrition.

3.3 Model performance comparison for LNC, PNC, and NNI estimation

Figures 3–5 illustrate the performance of five different models in estimating rice nitrogen nutritional indicators using Dualex readings. Each model utilized measurements from the 1-5 LFT, including Chl, Flav, and NBI values for individual leaves and their averages, totaling 18 input variables. The models showed distinct differences in their performance across training and validation sets, particularly when estimating different nitrogen indicators. Overall, estimation accuracy was highest for LNC, followed by PNC, with NNI being the most challenging to estimate accurately.

FIGURE 3
Performance of different machine learning models in predicting LNC based on Dualex readings. **(A, C, E, G, I)** show the training performance, and **(B, D, F, H, J)** show the validation performance for PLS, SVR, RF, XGB, and NN models, respectively.

For LNC estimation (Figure 3), all models demonstrated strong fitting accuracy, achieving R² values above 0.81 in training and 0.78–0.87 in validation datasets. SVR, RF, and XGB models

FIGURE 4
Performance of different machine learning models in predicting PNC based on Dualex readings. **(A, C, E, G, I)** show the training performance, and **(B, D, F, H, J)** show the validation performance for PLS, SVR, RF, XGB, and NN models, respectively.

FIGURE 5
Performance of different machine learning models in predicting NNI based on Dualex readings. **(A, C, E, G, I)** show the training performance, and **(B, D, F, H, J)** show the validation performance for PLS, SVR, RF, XGB, and NN models, respectively.

demonstrated the highest validation accuracy, with R^2 values between 0.86 and 0.87, and RRMSE below 10%. Data from all sampling periods, except for minor overestimations at the tillering stage, were evenly distributed around the 1:1 line.

In estimating PNC (Figure 4), model performance was similar to that observed for LNC, despite with slightly reduced accuracy. RF and XGB remained the top performers, with validation R^2 values of 0.82 and RMSE at 0.29. The SVR model followed closely with an R^2

of 0.81. Conversely, PLS regression model demonstrated slightly lower performance.

For NNI estimation (Figure 5), all models showed a noticeable decline in fitting accuracy, which was consistent with regression analysis results from quadratic models. Despite the challenges, RF and XGB continued to lead, with validation R^2 values ranging from 0.52 to 0.54, marking a 71% improvement over QCR models. However, estimating NNI proved to be more difficult than LNC and PNC, with apparent biases in both low and high-value regions.

Radar charts were used to visualize the estimation accuracy of different models and input variable combinations on the validation dataset (Figure 6). The R^2 values decreased progressively from LNC to PNC and then to NNI (Figures 6A-C). While LNC and PNC estimations showed near-equal accuracy, R^2 values for NNI dropped substantially. Across all input variable combinations, R^2 for NNI estimation declined by an average of 57% and 55% compared to LNC and PNC, respectively.
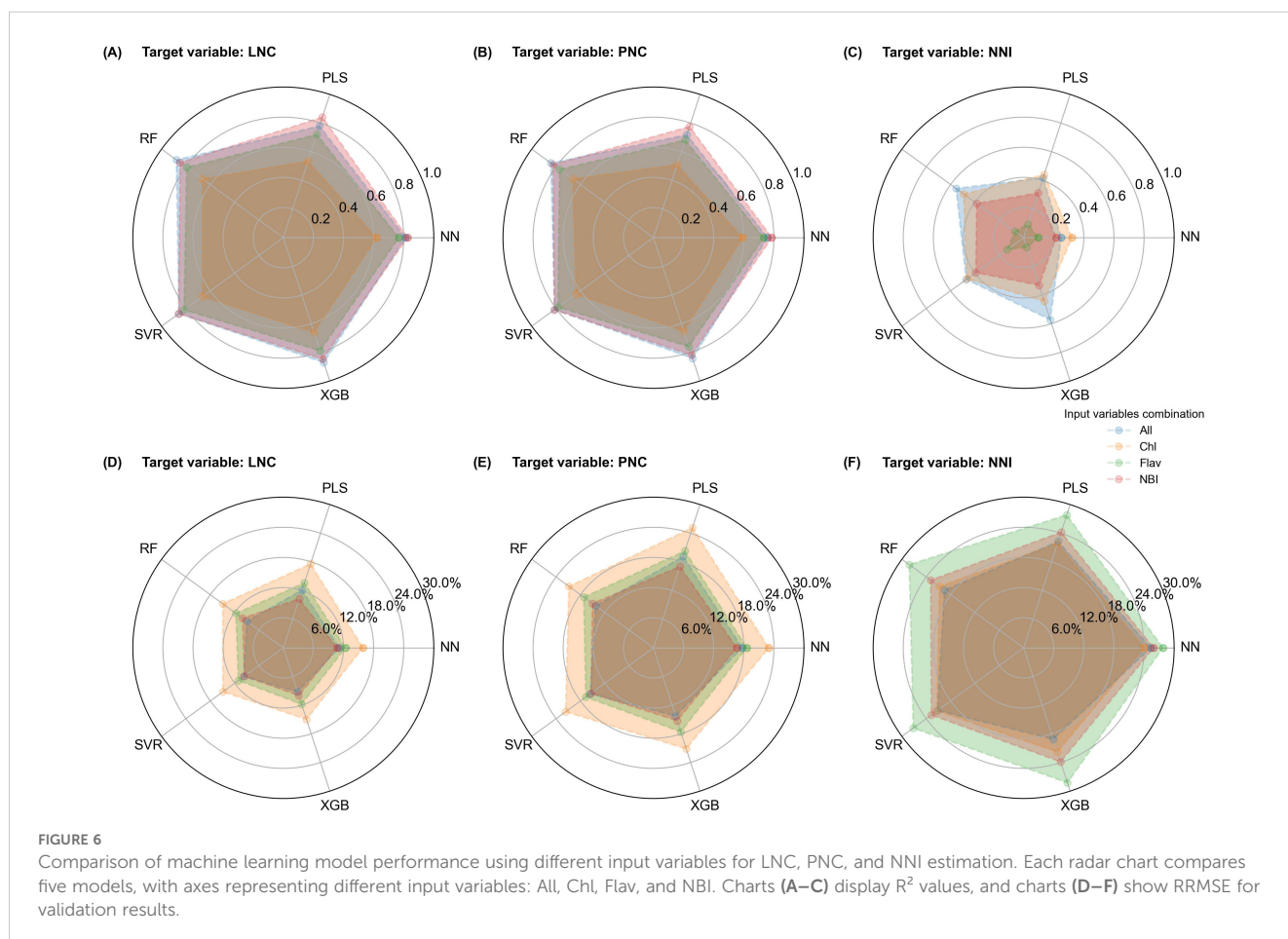
In comparing input variable combinations, the PLS model using multi-leaf NBI value was most effective in estimating LNC and PNC. For the other models, the all-variable combination (Combination 4) and NBI combination consistently provided high accuracy. Changes in RRMSE mirrored R^2 trends inversely (Figures 6D-F); models with higher R^2 exhibited lower RRMSE. RRMSE increased sequentially when estimating LNC, PNC, and

NNI. Among the input variable combinations, models using only the Chl combination had the highest RRMSE for LNC and PNC estimations, indicating lower accuracy in these cases. In contrast, for NNI estimation, models using only the Flav combination had the highest RRMSE (Figure 6F). The NBI and all-variable combinations consistently achieved similar, superior accuracy across all nitrogen indicators.

3.4 SHAP analysis of feature importance in nitrogen estimation

The feature importance analysis results based on SHAP values for the RF and XGB models in nitrogen estimation are presented in Figure 7. Both models used Combination 4 as input to assess the impact of each variable on their predictions.

For LNC prediction, Figures 7A, D show the SHAP analysis for the RF and XGB models, respectively. In the RF model (Figure 7A), the NBI value from the 2 LFT was the most influential variable, followed by the Flav value from the 1 LFT and the NBI value from the 1 LFT. The XGB model (Figure 7D) also highlighted these three variables as having the highest impact, with their SHAP values significantly surpassing those of other variables. This suggests that these variables were critical for accurate LNC prediction.
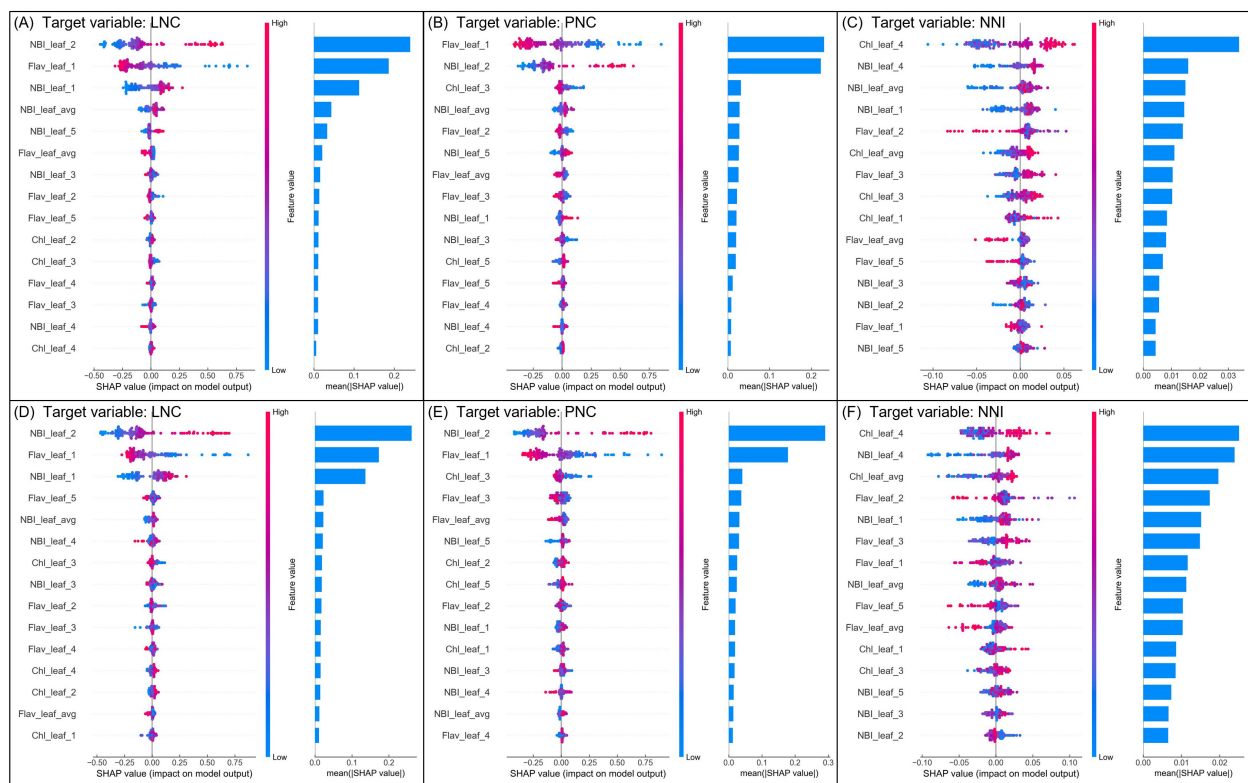

FIGURE 6
Comparison of machine learning model performance using different input variables for LNC, PNC, and NNI estimation. Each radar chart compares five models, with axes representing different input variables: All, Chl, Flav, and NBI. Charts (A–C) display R^2 values, and charts (D–F) show RRMSE for validation results.

FIGURE 7
SHAP value analysis of feature importance for RF **(A–C)** and XGB **(D–F)** models in predicting rice nitrogen indicators. Each plot presents the SHAP values for the 15 most important features. In the SHAP scatter plots on the left, red indicates higher feature values and blue indicates lower feature values. Positive SHAP values indicate that the feature increases the predicted nitrogen indicator, while negative SHAP values indicate that the feature decreases it. The right side ranks feature importance based on the mean absolute SHAP value.

In the case of PNC estimation, Figures 7B, E illustrate the SHAP analysis for the RF and XGB models. The RF model (Figure 7B) indicated that the Flav value from the 1 LFT and the NBI value from the 2 LFT were the most significant variables. Similarly, in the XGB model (Figure 7E), the NBI of the 2 LFT and Flav of the 1 LFT were also identified as the most influential factors. This consistency across models underscores the importance of NBI and Flav as key indicators for both LNC and PNC estimation.

For NNI prediction, Figures 7C, F display the SHAP analysis results for the RF and XGB models. The RF model (Figure 7C) identified the Chl value from the 4 LFT as the most important variable, followed by the NBI value from the 4 LFT. The XGB model (Figure 7F) similarly showed these two variables as dominant; however, there were differences in their SHAP value distributions. In the RF model, the Chl value from the 4 LFT had a SHAP value notably higher than the other variables, whereas, in the XGB model, SHAP values of variables decreased in a stepwise manner, indicating that no single variable was overwhelmingly dominant.

Overall, the SHAP analysis results for LNC and PNC predictions revealed certain dominant variables that enabled the models to achieve high-precision estimates. However, for NNI prediction, the lack of clearly dominant variables and weaker associations between input variables and the target presented challenges in achieving satisfactory model accuracy.

4 Discussion

4.1 Saturation effect and its implications

Incremental analysis highlighted distinct differences in the saturation effects of different Dualex indices when estimating rice nitrogen indicators. The results showed that Chl experiences significant saturation at high measurement values, consistent with findings from studies using SPAD-502 (Yuan et al., 2016; Zhang et al., 2019a; Jiang et al., 2021) and chlorophyll-based multispectral/hyperspectral indices (Fu et al., 2021; Yao et al., 2023; Gao et al., 2024). Previous research has noted that SPAD values become less responsive to nitrogen at higher concentrations, complicating the differentiation of nitrogen levels (Yue et al., 2020). This saturation effect is documented across various crops, including rice (Wang et al., 2023), wheat (Jiang et al., 2021), and maize (Dong et al., 2021). Our findings reaffirm that Chl and SPAD measurements exhibit saturation under high nitrogen conditions.

When leaf chlorophyll content is low, an increase can enhance the leaf's net photosynthetic rate (Tantray et al., 2020). However, as nitrogen availability rises and chlorophyll reaches a threshold level, photosynthetic rates plateau due to limits in light absorption area (Hou et al., 2019). At this point, excess nitrogen is allocated to non-photosynthetic reserves, altering the ratio of chlorophyll to leaf nitrogen (Hou et al., 2019; Mu and Chen, 2021).

Consequently, while Chl is a sensitive indicator at low nitrogen levels, its effectiveness decreases at higher levels, introducing errors into predictive models. This trend was observed in our analysis, particularly in LNC predictions, where models relying solely on Chl values achieved lower R^2 scores, failing to capture nitrogen variations accurately. Therefore, Chl's utility as a single diagnostic variable is limited, especially under high nitrogen conditions.

The saturation effect in this study was mainly observed when rice plants exhibited high nitrogen content. Since nitrogen concentration in rice leaves or plants gradually decreases during growth—a pattern known as the nitrogen dilution effect (Yao et al., 2021; Li et al., 2022a)—with the impact most noticeable around the tillering and jointing stages. This differs from canopy-level saturation observed after canopy closure, where high leaf area can cause remote sensing indices (e.g., NDVI) to saturate due to optical signal limitations, reducing their sensitivity to nitrogen changes (Tenreiro et al., 2021). Conversely, Chl saturation is primarily driven by physiological limits on chlorophyll content within leaves, necessitating a distinction between these two types of saturation in research.

In contrast, Flav and NBI showed no significant saturation effects, maintaining their effectiveness in estimating nitrogen indicators, even at high nitrogen levels. Flav is closely associated with plant stress responses, increasing under conditions of nitrogen stress (Li et al., 2021). When nitrogen is scarce, plants produce more flavonoids as a protective response (Gitelson et al., 2017). NBI combines the divergent responses of these indicators. In nitrogen-rich conditions, Chl levels are high while Flav remains low, leading to a high NBI. Under nitrogen deficiency, Chl decreases and Flav increases, resulting in a low NBI. By integrating complementary data from both indicators, NBI provides a robust reflection of nitrogen status, offering distinct advantages over Chl alone. This aligns with studies showing that NBI can overcome saturation and retain sensitivity under high nitrogen levels (Dong et al., 2022; Song et al., 2022). Our dataset, encompassing 15 rice varieties across different growth stages, confirmed that NBI and Flav maintained high predictive accuracy across various environmental conditions, suggesting these indicators can reliably reflect nitrogen status across diverse environmental conditions.

4.2 Comparison of machine learning models

We compared the performance of five machine learning models —PLS, SVR, RF, XGB, and NN—in predicting nitrogen indicators. Overall, XGB and RF models showed the highest accuracy for LNC, PNC, and NNI predictions, whereas PLS regression and SVR models had noticeably lower performance, lagging behind the tree-based approaches. This trend is consistent with findings from other studies comparing similar models (Fan et al., 2022; Kganyago et al., 2024). While PLS regression is effective at explaining linear relationships, it struggles with complex nonlinear patterns (Yang et al., 2023). SVR, though capable of managing nonlinear data, is highly sensitive to parameter selection and less efficient with high-

dimensional inputs (Ben Ishak, 2016). Conversely, tree-based ensemble models like RF and XGB excel at handling nonlinear, high-dimensional datasets (Nzekwe et al., 2024), with XGB particularly excelling due to its gradient boosting optimization, which iteratively improves model performance.

The NN model, however, was less stable than the XGB and RF models in this study. This may be due to the relatively smaller dataset size, as neural networks generally require larger datasets to train effectively and achieve good generalization. While the NN performed adequately in estimating LNC and PNC, its predictions for NNI were less accurate than those of the other models. The RF and XGB models demonstrated strong performance, although the training R^2 values were slightly higher than the validation R^2 values, especially for NNI estimation, the differences were minimal. This highlights the importance of model selection and appropriate complexity in achieving reliable predictions (Moradi et al., 2020; Yates et al., 2023).

The lower estimation accuracy for NNI compared to LNC and PNC can be attributed to several factors. NNI is a composite index involving both PNC and N_c, where N_c is dependent on plant biomass and derived from the nitrogen dilution curve. This adds layers of complexity and potential sources of error. The dynamic nature of biomass accumulation and nitrogen dilution across growth stages introduces variability that is difficult to capture with current models. Additionally, the relatively narrow range of NNI values reduces the models' ability to distinguish subtle differences. Measurement errors in biomass and nitrogen content further compound the challenge. Future research could explore incorporating growth stage information, more precise biomass measurements, and additional physiological parameters to enhance NNI prediction accuracy.

4.3 Feature importance across different input variables and leaf positions

SHAP analysis was instrumental in interpreting the models by quantifying the contribution of each input variable to the predictions (Lundberg et al., 2017). This is essential for enhancing model transparency, which is particularly valuable in agriculture, where interpretable models facilitate practical application (Ryo, 2022). The SHAP results highlighted that spectral data from different leaf positions do not contribute equally to nitrogen prediction. For LNC and PNC, the most significant contributions came from measurements taken at the 1 LFT and 2 LFT, corroborating findings from related studies where spectral reflectance from top leaves was found to be most sensitive for nitrogen assessment (Dong et al., 2021; Lu et al., 2022; Wang et al., 2023). For NNI prediction, SHAP analysis pointed to the importance of Chl measurements from the 4 LFT, suggesting that chlorophyll content in middle to lower leaves is particularly relevant. This may stem from the generally lower estimation accuracy of NNI, which tends to make the significance of variables more uniform. Additionally, nitrogen translocation in middle and lower leaves closely aligns with the plant's overall nitrogen nutritional status (Sun et al., 2018; Li et al., 2022b).

To improve NNI prediction accuracy, future strategies could include incorporating additional variables closely related to NNI, such as plant leaf area index (Tuccio et al., 2022), growth stages, and other physiological and environmental factors (Dong et al., 2022, 2024). Moreover, combining these models with crop growth simulations or radiative transfer models, as well as using multi-source data, could enhance the model's ability to account for complex interactions (Raya-Sereno et al., 2024; Wang et al., 2024). Expanding the dataset's size and diversity will also be crucial for improving generalization and predictive accuracy.

The differences in contributions from various leaf positions suggest the potential for optimizing measurement strategies in practical applications. By focusing on key leaf positions that provide the most significant input to the models, measurement efficiency can be improved. Reducing the number of measurement variables not only lowers costs but also enhances the computational efficiency of the models, making them more viable for real-world use.

5 Conclusion

This study systematically assessed the effectiveness of Dualex measurements for estimating rice nitrogen nutritional indicators. The results indicated that Chl experiences significant saturation at high nitrogen levels, limiting its utility as a nitrogen diagnostic indicator. In contrast, Flav and NBI did not exhibit saturation effects and remained highly sensitive under high nitrogen conditions, accurately reflecting rice nitrogen status. Among the machine learning models tested, RF and XGB were the most effective in predicting LNC and PNC. SHAP analysis identified the most influential feature variables and leaf positions, further enhancing model interpretability and accuracy.

Future research should focus on expanding the dataset to include more diverse rice varieties and growth stages and incorporating additional crop and environmental variables. Additionally, exploring the applicability of Flav and NBI across different crops and environmental contexts could provide broader technical support for nitrogen management in precision agriculture.

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 authors.

Author contributions

PS: Conceptualization, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing. YW: Investigation, Methodology, Project administration, Resources, Supervision, Writing – review & editing. CY: Investigation, Methodology, Resources, Writing – review & editing. KF: Investigation, Methodology, Resources, Writing – review & editing. YQ: Conceptualization, Methodology, Project administration, Resources, Supervision, Writing – review & editing. GC: Conceptualization, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Conflict of interest

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

Generative AI statement

The authors declare that Generative AI was used in the creation of this manuscript. GENERATIVE AI was used for the creation of this Manuscript. This article was edited and refined using generative AI.

Publisher's note

References

Ben Ishak, A. (2016). Variable selection using support vector regression and random forests: A comparative study. *Intell. Data Anal.* 20, 83–104. doi: 10.3233/IDA-150795

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

Cai, S., Zhao, X., Pittelkow, C. M., Fan, M., Zhang, X., and Yan, X. (2023). Optimal nitrogen rate strategy for sustainable rice production in China. *Nature* 615, 73–79. doi: 10.1038/s41586-022-05678-x

Cerovic, Z. G., Ghozlen, N. B., Milhade, C., Obert, M., Debuisson, S., and Le Moigne, M. (2015). Nondestructive diagnostic test for nitrogen nutrition of grapevine (Vitis vinifera L.) based on dualex leaf-clip measurements in the field. *J. Agric. Food Chem.* 63, 3669–3680. doi: 10.1021/acs.jafc.5b00304

Cerovic, Z. G., Masdoumier, G., Ghozlen, N. B., and Latouche, G. (2012). A new optical leaf-clip meter for simultaneous non-destructive assessment of leaf chlorophyll and epidermal flavonoids. *Physiol. Plant* 146, 251–260. doi: 10.1111/j.1399-3054.2012.01639.x

Chang, G. J., Oh, Y., Goldshleger, N., and Shoshany, M. (2022). Biomass estimation of crops and natural shrubs by combining red-edge ratio with normalized difference vegetation index. *J. Appl. Remote Sens.* 16, 14501. doi: 10.1117/1.JRS.16.014501

Chang, T., Zhao, H., Wang, N., Song, Q., Xiao, Y., Qu, M., et al. (2019). A three-dimensional canopy photosynthesis model in rice with a complete description of the canopy architecture, leaf physiology, and mechanical properties. *J. Exp. Bot.* 70, 2479–2490. doi: 10.1093/jxb/ery430

Dong, R., Miao, Y., Wang, X., Chen, Z., and Yuan, F. (2021). Improving maize nitrogen nutrition index prediction using leaf fluorescence sensor combined with environmental and management variables. *Field Crops Res.* 269, 108180. doi: 10.1016/j.fcr.2021.108180

Dong, L., Miao, Y., Wang, X., Kusnierek, K., Zha, H., Pan, M., et al. (2024). In-season dynamic diagnosis of maize nitrogen status across the growing season by integrating proximal sensing and crop growth modeling. *Comput. Electron. Agric.* 224, 109240. doi: 10.1016/j.compag.2024.109240

Dong, R., Miao, Y., Wang, X., Yuan, F., and Kusnierek, K. (2022). Combining leaf fluorescence and active canopy reflectance sensing technologies to diagnose maize nitrogen status across growth stages. *Precis. Agric.* 23, 939–960. doi: 10.1007/s11119-021-09869-w

Dong, T., Shang, J., Chen, J. M., Liu, J., Qian, B., Ma, B., et al. (2019). Assessment of portable chlorophyll meters for measuring crop leaf chlorophyll concentration. *Remote Sens.* 11, 2706. doi: 10.3390/rs11222706

Fan, K., Li, F., Chen, X., Li, Z., and Mulla, D. J. (2022). Nitrogen balance index prediction of winter wheat by canopy hyperspectral transformation and machine learning. *Remote Sens.* 14, 3504. doi: 10.3390/rs14143504

Fu, Y., Yang, G., Pu, R., Li, Z., Li, H., Xu, X., et al. (2021). An overview of crop nitrogen status assessment using hyperspectral remote sensing: Current status and perspectives. *Eur. J. Agron.* 124, 126241. doi: 10.1016/j.eja.2021.126241

Gao, S., Yan, K., Liu, J., Pu, J., Zou, D., Qi, J., et al. (2024). Assessment of remote-sensed vegetation indices for estimating forest chlorophyll concentration. *Ecol. Indic.* 162, 112001. doi: 10.1016/j.ecolind.2024.112001

Gitelson, A., Chivkunova, O., Zhigalova, T., and Solovchenko, A. (2017). *In situ* optical properties of foliar flavonoids: Implication for non-destructive estimation of flavonoid content. *J. Plant Physiol.* 218, 258–264. doi: 10.1016/j.jplph.2017.08.009

Guo, Y., Chen, Y., Searchinger, T. D., Zhou, M., Pan, D., Yang, J., et al. (2020). Air quality, nitrogen use efficiency and food security in China are improved by cost-effective agricultural nitrogen management. *Nat. Food* 1, 648–658. doi: 10.1038/s43016-020-00162-z

Hamdane, Y., Segarra, J., Buchaillot, M. L., Rezzouk, F. Z., Gracia-Romero, A., Vatter, T., et al. (2023). Using ground and UAV vegetation indexes for the selection of fungal-resistant bread wheat varieties. *Drones* 7, 454. doi: 10.3390/drones7070454

Hou, W., Tränkner, M., Lu, J., Yan, J., Huang, S., Ren, T., et al. (2019). Interactive effects of nitrogen and potassium on photosynthesis and photosynthetic nitrogen allocation of rice leaves. *BMC Plant Biol.* 19, 302. doi: 10.1186/s12870-019-1894-8

Huang, C., Zhang, Z., Lu, Z., Zhang, X., Zhu, L., and Yang, W. (2023). Leaf grading for cotton verticillium wilt based on VFNet-Improved and Deep Sort. *J. Intell. Agric. Mech.* 4, 12–21. doi: 10.12398/j.issn.2096-7217.2023.02.002

Jiang, J., Wang, C., Wang, H., Fu, Z., Cao, Q., Tian, Y., et al. (2021). Evaluation of three portable optical sensors for non-destructive diagnosis of nitrogen status in winter wheat. *Sensors* 21, 5579. doi: 10.3390/s21165579

Justes, E., Mary, B., Meynard, J. M., Machet, J. M., and Thelier-Huche, L. (1994). Determination of a critical nitrogen dilution curve for winter wheat crops. *Ann. Bot.* 74, 397–407. doi: 10.1006/anbo.1994.1133

Karaca, C., Thompson, R. B., Peña-Fleitas, M. T., Gallardo, M., and Padilla, F. M. (2023). Evaluation of absolute measurements and normalized indices of proximal optical sensors as estimators of yield in muskmelon and sweet pepper. *Remote Sens.* 15, 2174. doi: 10.3390/rs15082174

Kganyago, M., Adjorlolo, C., Mhangara, P., and Tsoeleng, L. (2024). Optical remote sensing of crop biophysical and biochemical parameters: An overview of advances in sensor technologies and machine learning algorithms for precision agriculture. *Comput. Electron. Agric.* 218, 108730. doi: 10.1016/j.compag.2024.108730

Li, X., Ata-UI-Karim, S. T., Li, Y., Yuan, F., Miao, Y., Kato, Y., et al. (2022a). Advances in the estimations and applications of critical nitrogen dilution curve and nitrogen nutrition index of major cereal crops. A review. *Comput. Electron. Agric.* 197, 106998. doi: 10.1016/j.compag.2022.106998

Li, Z., Jiang, H., Qin, Y., Yan, H., Jiang, X., and Qin, Y. (2021). Nitrogen deficiency maintains the yield and improves the antioxidant activity of Coreopsis tinctoria Nutt. *Biosci. Biotechnol. Biochem.* 85, 1492–1505. doi: 10.1093/bbb/zbab048

Li, Y. Y., Ming, B., Fan, P. P., Liu, Y., Wang, K. R., Hou, P., et al. (2022b). Effects of nitrogen application rates on the spatio-temporal variation of leaf SPAD readings on the maize canopy. *J. Agric. Sci.* 160, 32–44. doi: 10.1017/S0021859621001052

Li, S., Xu, C., Wang, N., Cao, H., and Yu, F. (2024). Research on rice nitrogen unmanned aerial vehicle hyperspectral inversion based on BWO-ELM. *J. Intell. Agric. Mech.* 5, 14–21. doi: 10.12398/j.issn.2096-7217.2024.03.002

Liu, Y., Wang, H., Jiang, Z., Wang, W., Xu, R., Wang, Q., et al. (2021). Genomic basis of geographical adaptation to soil nitrogen in rice. *Nature* 590, 600–605. doi: 10.1038/s41586-020-03091-w

Lu, J., Lu, J., Nie, W., Song, J., Zhan, Q., Wang, M., et al. (2022). Estimation of nitrogen nutrition index in chrysanthemum using chlorophyll meter readings. *Ind. Crops Prod.* 187, 115459. doi: 10.1016/j.indcrop.2022.115459

Lundberg, S. M., and Lee, S. I. (2017). "A unified approach to interpreting model predictions," in *Advances in neural information processing systems*. Eds. I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, et al (Curran Associates, Inc, Long Beach, CA, USA). Available at: https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf (accessed September 20, 2024).

Luo, X., Keenan, T. F., Chen, J. M., Croft, H., Colin Prentice, I., Smith, N. G., et al. (2021). Global variation in the fraction of leaf nitrogen allocated to photosynthesis. *Nat. Commun.* 12, 4866. doi: 10.1038/s41467-021-25163-9

Ma, X., Huete, A., Tran, N. N., Bi, J., Gao, S., and Zeng, Y. (2020). Sun-angle effects on remote-sensing phenology observed and modelled using Himawari-8. *Remote Sens.* 12, 1339. doi: 10.3390/rs12081339

Mandal, D., Siqueira, R., de, Longchamps, L., and Khosla, R. (2024). Machine learning and fluorosensing for estimation of maize nitrogen status at early growth-stages. *Comput. Electron. Agric.* 225, 109341. doi: 10.1016/j.compag.2024.109341

Moradi, R., Berangi, R., and Minaei, B. (2020). A survey of regularization strategies for deep models. *Artif. Intell. Rev.* 53, 3947–3986. doi: 10.1007/s10462-019-09784-7

Mu, X., and Chen, Y. (2021). The physiological response of photosynthesis to nitrogen deficiency. *Plant Physiol. Biochem.* 158, 76–82. doi: 10.1016/j.plaphy.2020.11.019

Nzekwe, C. J., Kim, S., and Mostafa, S. A. (2024). Interaction selection and prediction performance in high-dimensional data: A comparative study of statistical and tree-based methods. *J. Data Sci.* 22, 259–279. doi: 10.6339/24-JDS1127

Ouyang, W., Yin, X., Yang, J., and Struik, P. C. (2021). Roles of canopy architecture and nitrogen distribution in the better performance of an aerobic than a lowland rice cultivar under water deficit. *Field Crops Res.* 271, 108257. doi: 10.1016/j.fcr.2021.108257

Qian, B., Ye, H., Huang, W., Xie, Q., Pan, Y., Xing, N., et al. (2022). A sentinel-2-based triangular vegetation index for chlorophyll content estimation. *Agric. For. Meteorol.* 322, 109000. doi: 10.1016/j.agrformet.2022.109000

Raya-Sereno, M. D., Camino, C., Pancorbo, J. L., Alonso-Ayuso, M., Gabriel, J. L., Beck, P. S. A., et al. (2024). Assessing wheat genotype response under combined nitrogen and water stress scenarios coupling high-resolution optical and thermal sensors with radiative transfer models. *Eur. J. Agron.* 154, 127102. doi: 10.1016/j.eja.2024.127102

Röll, G., Hartung, J., and Graeff-Hönninger, S. (2019). Determination of plant nitrogen content in wheat plants via spectral reflectance measurements: Impact of leaf number and leaf position. *Remote Sens.* 11, 2794. doi: 10.3390/rs11232794

Rubo, S., and Zinkernagel, J. (2022). Exploring hyperspectral reflectance indices for the estimation of water and nitrogen status of spinach. *Biosyst. Eng.* 214, 58–71. doi: 10.1016/j.biosystemseng.2021.12.008

Ryo, M. (2022). Explainable artificial intelligence and interpretable machine learning for agricultural data analysis. *Artif. Intell. Agric.* 6, 257–265. doi: 10.1016/j.aiia.2022.11.003

Shi, P., Wang, Y., Xu, J., Zhao, Y., Yang, B., Yuan, Z., et al. (2021). Rice nitrogen nutrition estimation with RGB images and machine learning methods. *Comput. Electron. Agric.* 180, 105860. doi: 10.1016/j.compag.2020.105860

Singh, V., Kunal,, Singh, M., and Singh, B. (2022). Spectral indices measured with proximal sensing using canopy reflectance sensor, chlorophyll meter and leaf color chart for in-season grain yield prediction of basmati rice. *Pedosphere* 32, 812–822. doi: 10.1016/j.pedsph.2022.06.015

Song, X., Yang, G., Xu, X., Zhang, D., Yang, C., and Feng, H. (2022). Winter wheat nitrogen estimation based on ground-level and UAV-mounted sensors. *Sensors* 22, 549. doi: 10.3390/s22020549

Sun, Y., Zhu, S., Yang, X., Weston, M. V., Wang, K., Shen, Z., et al. (2018). Nitrogen diagnosis based on dynamic characteristics of rice leaf image. *PloS One* 13, e0196298. doi: 10.1371/journal.pone.0196298

Tan, L., Zhou, L., Zhao, N., He, Y., and Qiu, Z. (2021). Development of a low-cost portable device for pixel-wise leaf SPAD estimation and blade-level SPAD distribution visualization using color sensing. *Comput. Electron. Agric.* 190, 106487. doi: 10.1016/j.compag.2021.106487

Tantray, A. Y., Bashir, S. S., and Ahmad, A. (2020). Low nitrogen stress regulates chlorophyll fluorescence in coordination with photosynthesis and Rubisco efficiency of rice. *Physiol. Mol. Biol. Plants* 26, 83–94. doi: 10.1007/s12298-019-00721-0

Tenreiro, T. R., García-Vila, M., Gómez, J. A., Jiménez-Berni, J. A., and Fereres, E. (2021). Using NDVI for the assessment of canopy cover in agricultural crops within modelling research. *Comput. Electron. Agric.* 182, 106038. doi: 10.1016/j.compag.2021.106038

Tuccio, L., Massa, D., Cacini, S., Iovieno, P., and Agati, G. (2022). Monitoring nitrogen variability in two Mediterranean ornamental shrubs through proximal fluorescence-based sensors at leaf and canopy level. *Sci. Hortic.* 294, 110773. doi: 10.1016/j.scienta.2021.110773

Wang, Z., Li, H., Xia, Y., Fan, M., Zhao, H., Lu, J., et al. (2023). Relationship between blade SPAD value and protein content of rice in cold region. *Agric. Res. Arid Areas* 41, 150–161. doi: 10.7606/j.issn.1000-7601.2023.05.16

Wang, D., Struik, P. C., Liang, L., and Yin, X. (2024). Developing remote sensing- and crop model-based methods to optimize nitrogen management in rice fields. *Comput. Electron. Agric.* 220, 108899. doi: 10.1016/j.compag.2024.108899

Waqas, M., Wang, L., Jones, J. J., Josef Turetschek, R., Engelmeier, D., Geilfus, C.-M., et al. (2023). Short-term phosphorus deficiency induces flavonoid accumulation in the lamina of pak choi: A finishing treatment that influences inner quality. *Sci. Hortic.* 314, 111953. doi: 10.1016/j.scienta.2023.111953

Wu, B., Zhang, M., Zeng, H., Tian, F., Potgieter, A. B., Qin, X., et al. (2023). Challenges and opportunities in remote sensing-based crop monitoring: a review. *Natl. Sci. Rev.* 10, nwac290. doi: 10.1093/nsr/nwac290

Yang, H., Yin, H., Li, F., Hu, Y., and Yu, K. (2023). Machine learning models fed with optimized spectral indices to advance crop nitrogen monitoring. *Field Crops Res.* 293, 108844. doi: 10.1016/j.fcr.2023.108844

Yao, B., Ata-Ul-Karim, S. T., Li, Y., Ye, T., Zhu, Y., Cao, W., et al. (2023). Plant nitrogen status at phenological stages can well estimate wheat yield and its components. *Field Crops Res.* 297, 108950. doi: 10.1016/j.fcr.2023.108950

Yao, B., Wang, X., Lemaire, G., Makowski, D., Cao, Q., Liu, X., et al. (2021). Uncertainty analysis of critical nitrogen dilution curves for wheat. *Eur. J. Agron.* 128, 126315. doi: 10.1016/j.eja.2021.126315

Yates, L. A., Aandahl, Z., Richards, S. A., and Brook, B. W. (2023). Cross validation for model selection: A review with examples from ecology. *Ecol. Monogr.* 93, e1557. doi: 10.1002/ecm.1557

Yuan, Z., Cao, Q., Zhang, K., Ata-Ul-Karim, S. T., Tian, Y., Zhu, Y., et al. (2016). Optimal leaf positions for SPAD meter measurement in rice. *Front. Plant Sci.* 7, 719. doi: 10.3389/fpls.2016.00719

Yue, X., Hu, Y., Zhang, H., and Schmidhalter, U. (2020). Evaluation of both SPAD reading and SPAD index on estimating the plant nitrogen status of winter wheat. *Int. J. Plant Prod.* 14, 67–75. doi: 10.1007/s42106-019-00068-2

Zhang, K., Liu, X., Ma, Y., Zhang, R., Cao, Q., Zhu, Y., et al. (2019a). A comparative assessment of measures of leaf nitrogen in rice using two leaf-clip meters. *Sensors* 20, 175. doi: 10.3390/s20010175

Zhang, L., Shao, Z., Liu, J., and Cheng, Q. (2019b). Deep learning based retrieval of forest aboveground biomass from combined LiDAR and Landsat 8 data. *Remote Sens.* 11, 1459. doi: 10.3390/rs11121459