Evaluation of vertical C/N distribution in maize canopy using ensemble learning with hyperspectral data

Xinlei Xu[1,2], Xingang Xu[1,2*], Yang Meng[1], Guijun Yang[1], Hanyue Xue[1], Yuekun Song[1]

1、Key Laboratory of Quantitative Remote Sensing in Ministry of Agriculture and Rural Affairs, Information Technology Research Center, Beijing Academy of Agricultural and Forestry Sciences, Beijing 100097, China;

2、Beijing University of Agriculture, Beijing 102206, China;

Abstract: Metabolic status of carbon (C) and nitrogen (N) as two essential elements in crop plants has essential influence on the ultimate formation of yield and quality in crop production. Ratio of carbon to nitrogen (C/N), defined as the ratio of LCC (leaf carbon concentration) to LNC (leaf nitrogen concentration), is useful for understanding and quantifying carbon and nitrogen metabolism in crops, and is one metric for effectively evaluating the balance of carbon and nitrogen, nutrient status and growth vigor and in crop plants. Existing studies concentrated mostly on the horizontal variability of C/N, with few researches dedicated to assessing the vertical distribution of C/N. This study aimed to evaluate vertical C/N distribution in maize canopies by coupling ensemble learning with hyperspectral data. Utilizing leaf C/N data from the upper, middle, and lower layer based on maize height, along with hyperspectral reflectance data from maize canopies at diverse growth phases and planting densities, this study used the EL (ensemble learning) with one dual-level framework to estimate the vertically layered C/N in maize canopies. Firstly, the technique of selecting features, Simulated Annealing (SA) was utilized to select spectral features sensitive to C/N across various vertical layers from maize canopy. Then, EL with a dual-level framework was developed for C/N estimation in different layers, the five machine learning methods of Partial Least Squares Regression (PLS), Random Forest Regression (RF), Elastic Net, Lasso, and Support Vector Regression (SVR) were employed as initial base learners, and Ridge Regression functioned as the secondary meta-learner. the analysis results indicate that the ensemble learning approach obtained superior accuracy relative to singular machine learning methods. The SA-EL technique yielded R^2 of 0.745 and 0.702 for C/N estimation in the upper and middle layers and 0.674 for the lower layers, respectively, alongside rRMSE of 7.48% and 9.82% for the upper and middle layers, and 13.1% for the lower layer. It indicates that coupling EL with hyperspectral data can efficiently assesses vertical C/N distribution in maize canopies and the proposed method may serve as a benchmark for the hyperspectral evaluation of other biochemical indices in vertical layers of crop canopies.

Keywords: Ratio of carbon to nitrogen (C/N); Hyperspectral data; Ensemble Learning; Simulated Annealing Algorithm; Vertical Distribution

INTRODUCTION

Ratio of carbon to nitrogen (C/N) in crop canopy leaves is one vital indicator for evaluating the physiological state of carbon and nitrogen metabolism within plants. It serves as a comprehensive metric for diagnosing the carbon-nitrogen balance, assessing growth activity, and providing early warnings of nutrient stress. Real-time, effective, and accurate monitoring of C/N ratios in crop canopies has significant importance for diagnosing crop growth status and guiding field management practices (Xu et al., 2018).

As crops progress through their developmental stages, the vertical spatial distribution characteristics of the canopy become increasingly pronounced. Both nitrogen and carbon exhibit vertical stratification within the canopy, often displaying an inverse relationship (Wang et al., 2024). Utilizing these vertical spatial distribution characteristics allows for early detection of C/N ratio information in the middle and lower canopy layers, enabling timely diagnosis of nutrient deficiencies and supporting precise field fertilization management decisions. Recent advancements in hyperspectral remote sensing fields have facilitated studies on retrieving vegetation carbon and nitrogen contents and their ratios based on spectral characteristics. Previous research has primarily focused on overall canopy-scale C/N estimation using hyperspectral data combined with regression methods to monitor the C/N ratio in rice and other crops like wheat and barley (Xu et al., 2018). These studies confirm the feasibility of spectral techniques in C/N monitoring but typically model average C/N values at the canopy level or for homogeneous plots without considering the heterogeneity along vertical stratification.

Recent studies have begun focusing on the vertical distribution of nitrogen, constructing the models for assessing nitrogen distribution across different canopy layers (upper, middle, and lower) highlighting that ignoring vertical stratification can lead to systematic biases in nitrogen remote sensing estimation. However, there has been no study extending remote sensing estimation of C/N to a vertically stratified scale (Yang et al., 2024). In terms of modeling, single machine learning models are widely used to evaluate vegetation biochemical parameters but are susceptible to sample distribution, noise interference, and overfitting/underfitting issues. Ensemble learning methods, including Bagging, Boosting and Stacking, have shown promise in enhancing model robustness and generalization capability. This study adopts the Stacking ensemble learning framework to leverage multiple machine learning algorithms, aiming to improve stability and predictive capability for vertically stratified C/N ratio estimation. Despite current research gaps regarding the vertical stratification of vegetation C/N using remote sensing, existing models often lack sufficient stability. Therefore, this study combines hyperspectral data with Stacking ensemble learning method to investigate the vertical distribution of C/N in maize canopies. This study aims to explore the feasibility of using hyperspectral data combined with ensemble learning methods for estimating the C/N ratio in vertically stratified maize canopy leaves, providing new insights into vertical nutrient patterns.

1 MATERIALS AND METHODS

1.1 Study Area and Experimental Design

The maize experiment was conducted at the Xiaotangshan Experimental Base located in Changping District, Beijing (geographic coordinates: 40°18′ N, 116°39′ E). Situated in the northern part of North China Plain, this site is characterized by its flat terrain and fertile soil. The region experiences a temperate monsoon climate, marked by distinct seasons and concurrent rainfall and heat. With an annual average temperature ranging from approximately 11 to 13°C, the area provides optimal conditions for maize cultivation. This environment is particularly conducive to maize growth, making it an ideal location for conducting agricultural experiments.

The experimental area consists of 60 plots arranged in six rows with a spacing of 0.6 meters between rows. Each plot measures 3.6 meters in length and 2.5 meters in width, with a 1.5-meter gap between adjacent plots. The total dimensions of the experimental field are 39 meters north-south by 29.6 meters east-west. Four different planting densities were established: plots were numbered from west to east as 1-1 to 4-15.

The experiment was divided into two phases. The first phase took place on August 18, 2020, during late summer when sunlight was abundant. At this time, the maize was typically at the critical grain-filling stage. The second phase occurred on September 14, 2020, as autumn began and temperatures started to decline. By then, most maize in Beijing was still in the grain-filling period but transitioning from the rapid filling phase to the late filling stage. Although the rate of dry matter accumulation in the kernels may have begun to slow, the filling process continued until maturity. An overview of the experimental area is provided in Figure 1.

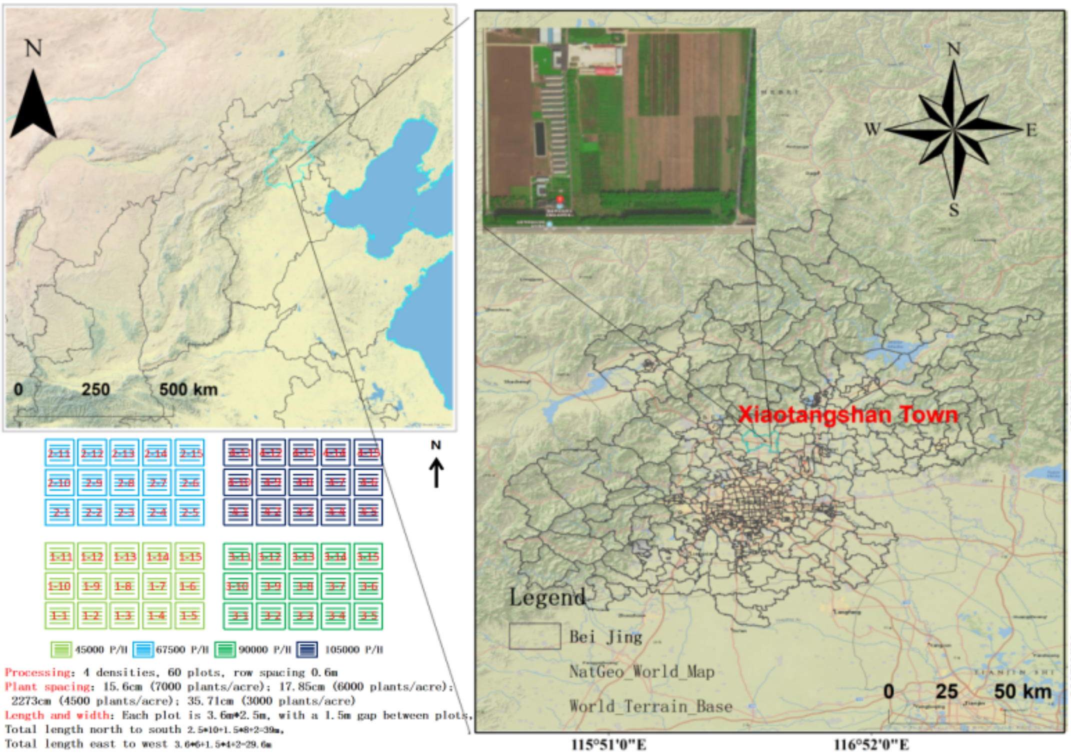

Figure 1. Experimental Area

1.2 Field Data Collection

1.2.1 Maize Hyperspectral Data Measurement

In this study, maize was selected as the subject of investigation, and a portable ASD (Analytical Spectral Devices) FieldSpec spectroradiometer was employed to measure the hyperspectral reflectance characteristics of its canopy. To ensure the accuracy and comparability of the measurements, the spectroradiometer probe was calibrated against a standard white reference panel before each measurement session. This calibration step was essential for system calibration and illumination optimization. Field measurements commenced only after the reference spectral curve displayed on the connected laptop stabilized, with the reflectance value approaching unity. During data collection, the probe was positioned vertically at a height of approximately 150 cm directly above the center of the maize canopy, maintaining a nadir viewing angle. Each experimental plot underwent ten repeated measurements, from which an average was calculated to serve as the representative spectral value for that plot. The acquired spectral data covered a wavelength range of 350–2500 nm, encompassing 2151 bands across the visible to short-wave infrared regions. To enhance data quality and minimize spectral noise, raw spectral data were preprocessed using a Uniform Filter smoothing algorithm. This preprocessing step also helped mitigate the influence of water absorption bands, thereby improving the accuracy and applicability of the hyperspectral data in analyzing carbon-nitrogen relationships within the maize canopy.

1.2.2 Acquisition of C/N Data from Different Vertical Layers of Maize

Following the measurement of maize canopy hyperspectral reflectance, a stratified sampling and analysis method was employed to systematically obtain carbon and nitrogen parameters from leaves at different heights within the maize canopy. This method ensured that vertical heterogeneity in canopy structure was accounted for by collecting samples from the same locations as the spectral measurements. Specifically, the maize canopy was divided into upper, middle, and lower layers, corresponding to the top one-third, middle one-third, and near-basal portions of the canopy, respectively. Leaf samples of equal area were collected from these three vertical positions using a punch tool.

Within each experimental plot, measurements were taken at three positions—left, center, and right—along the central row of maize plants. For sample collection, two adjacent maize plants in the third row of each plot were selected. From each plant, leaf samples were collected from the upper, middle, and lower canopy layers. Considering the large size of maize leaves and potential physiological variations along their length, three subsamples were taken from the base, middle, and tip of each leaf using a punch with a consistent area to ensure uniformity. After collection, all samples were labeled and sent to the laboratory for analysis.

Carbon and nitrogen contents were determined using a microelement analyzer. The values measured at the three leaf positions within the same vertical layer were averaged to represent the C/N ratio for that layer of the maize plant. This approach provided robust data support for analyzing vertical C/N distribution characteristics and their relationship with spectral information. Ultimately, a total of 62 upper-layer leaf samples, 62 middle-layer samples, and 59 lower-layer samples were obtained.

1.3 Technical Workflow for Hyperspectral Estimation of C/N in Different Vertical Layers of Maize Canopy

The technical workflow of this study is as follows: In the experimental data phase, ASD hyperspectral data and ground-measured carbon-to-nitrogen (C/N) content data were collected. These raw data underwent preprocessing, which included Uniform Filter smoothing to mitigate spectral noise and min-max normalization to standardize the data range. Subsequently, the processed data were augmented to generate an expanded dataset. From this augmented dataset, various spectral vegetation indices were extracted. To identify the most sensitive vegetation indices, feature selection methods—Simulated Annealing (SA) was employed. Each method retained vegetation indices that exhibited high correlation with C/N ratios, ensuring that only the most relevant features were used for subsequent modeling.

For estimating vertically stratified C/N within the maize canopy, an ensemble learning approach based on the stacking principle was adopted. This framework comprised two tiers: In the first tier, five machine learning models—Partial Least Squares Regression (PLS), Random Forest Regression (RF), Elastic Net, Lasso, and Support Vector Regression (SVR)—were utilized as base learners. These models were trained on the selected vegetation indices. In the second tier, Ridge Regression was employed as the meta-learner to integrate predictions from the base learners, ultimately estimating C/N ratios in different vertical layers of the maize canopy.

1.4 Calculation of Typical Vegetation Indices

Vegetation indices are indispensable tools in the field of remote sensing, utilized to detect and assess vegetation conditions. These indices encapsulate the combined influence of chlorophyll content and plant physiological status. High vegetation index values typically signify elevated chlorophyll levels, which can indirectly indicate a plant's nitrogen uptake efficiency. By analyzing specific spectral bands, vegetation indices provide critical insights into vegetation health and characteristics. In this study, we employed a range of vegetation indices related to chlorophyll content, water content, nitrogen content, red-edge characteristics, and pigments (refer to Table 1).

Different types of vegetation indices emphasize distinct aspects, thereby enabling a more comprehensive evaluation of crop growth status and facilitating diverse application requirements. Moreover, the use of vegetation indices not only reduces model training time but also conserves storage space, thus promoting efficient prediction capabilities. Feature selection, by eliminating irrelevant features and mitigating the risk of overfitting, enhances model adaptability to new data, thereby improving generalization capability

Through these steps, the study ensures that the augmented data are statistically robust and well-justified, providing a solid foundation for subsequent analyses and enhancing the overall reliability of the results. By employing advanced vegetation indices and rigorous feature selection methods, this research aims to offer deeper insights into crop health monitoring and predictive modeling. This approach not only streamlines the analysis process but also improves the accuracy and applicability of the findings.

Table 1. Selected Vegetation Indices

Category	Vegetation Index Name	Calculation Formula
Basic Vegetation Indices (Related to Chlorophyll Content)	Normalized Difference Vegetation Index (NDVI)	(NIR−R)/(NIR+R)
	Ratio Vegetation Index (RVI)	NIR/R
	Difference Vegetation Index (DVI)	NIR−R
	Blue Normalized Difference Vegetation Index (BNDVI)	(NIR−B)/(BIR+B)
	Photochemical Reflectance Index (PRI)	$(R_{531}-R_{570})/(R_{531}+R_{570})$
	Perpendicular Vegetation Index (PVI)	(NIR−0.1×R)/0.12+1
	Enhanced Vegetation Index (EVI)	2.5×[(NIR−R)/(NIR+6×R−7.5×B+1)]
	Green Normalized Difference Vegetation Index (GNDVI)	(NIR−G)/(NIR+G)
	Red-Edge Normalized Difference Vegetation Index 2 (RENDVI2)	(NIR−RE)/(NIR+RE)
Water-Related Indices	Normalized Difference Infrared Index (NDII)	(NIR−SWIR1)/(NIR+SWIR1)
	Normalized Difference Water Index (NDWI)	(G−SWIR1)/(G+SWIR1)
	Moisture Stress Index (MSI)	SWIR1/NIR
	Normalized Short-Wave Infrared Index (NSWII)	(SWIR1−SWIR2)/(SWIR1+SWIR2)
Red-Edge Parameter Indices	Red-Edge Normalized Difference Vegetation Index (NDRE)	(NIR−RE)/(NIR+RE)
	Red-Edge Ratio Vegetation Index (RENDVI)	(RE−R)/(RE+R)
	Red-Edge Reflectance Ratio (RER)	RE/R
	Red-Edge Position (REP)	700nm-750nm (The wavelength corresponding to the maximum value of the band reflection rate derivative)
	Red-Edge Slope (RES)	(RE−R/(λRE−λR)
Nitrogen-Related	Nitrogen Index (NI)	NIR/(R×SWIR1)

Indices	Vegetation Nitrogen Index (VNI)	(740nm−720nm)/(740nm+720nm)
	Nitrogen Absorption Index (NAI)	(1/SWIR1)+(1/SWIR2)
	Crop Nitrogen Status Index (CNSI)	SWIR1/RE
	Normalized Carbon-Nitrogen Index (NCNI)	(SWIR1−RE)/(SWIR1+RE)
	Nitrogen Reflectance Ratio Index (NRRI)	NIR/RED
	Normalized Difference Nitrogen Index (NDNI)	(NIR−SWIR2)/(NIR+SWIR2)
	Red-Edge Nitrogen Sensitivity Index (RNSI)	(RED−G)/(RE+G)
	Modified Chlorophyll Absorption in Reflectance Index (MCARI)	1.2×[(NIR−R)−0.2×(NIR−G)× NIR/R)]
	Chlorophyll Index (CI)	(NIR/RE)−1
Pigment-Related Indices	Anthocyanin Reflectance Index (ARI)	(1/550nm)−(1/700nm)
	Photochemical Reflectance Index 2 (PRI2)	(531nm−570nm)/500nm
	Blue-Edge Slope (BES)	(B−R500)/(λ_B−λ_{500})
	Yellow-Edge Slope (YES)	(R_{570}−G)/(λ_{570}−λ_G)
	Soil-Adjusted Vegetation Index (SAVI)	(NIR−B)/(NIR+R)
	Spectral_Derivative (SD)	Average of the first derivative of reflectivity
Structural and Biochemical Parameter Indices	Spectral_2nd_Derivative (S2D)	Average second derivative of reflectivity
	Cumulative Red-Edge Reflectance Area Index (CRSAI)	700nm-750nm Band-limited reflectivity integral value (trapezoidal integral)

Note: NIR,RE,R,G,B,SWIR1 and SWIR2 are reflectance values of wheat in near-infrared,red-edge,red,green,blue,short-wave infrared 1 and short wave infrared 2 bands,respectively, from the multispectral cameras.

1.5 Vegetation Spectral Feature Variable Selection Methods

The Simulated Annealing (SA) was employed to extract appropriate spectral bands for target attribute features, facilitating the construction of vegetation indices. This approach aids in the development of accurate estimation models (Liang et al., 2025, Chen et al., 2023, Yang et al.,2025). The primary objective of this study was to identify sensitive spectral bands from 38 vegetation index features that exhibit strong responsiveness to the carbon-nitrogen ratio (C/N) in maize canopies. The feature selection method selected 13 sensitive vegetation index features for each canopy layer (upper, middle, and lower).

Simulated Annealing (SA) is a probabilistic global optimization technique suitable for finding near-optimal solutions in complex search spaces. In this study, SA was integrated with a linear regression model, and Grid Search Cross-Validation (CV) was introduced for hyperparameter tuning to automatically determine whether to include an intercept term (fit_intercept). This process optimized model parameters, enhancing both model stability and generalization capability. Under the constraint of selecting 13 features, the SA algorithm identified the optimal combination of vegetation index features.

1.6 Ensemble Learning Modeling Method

Stacking ensemble learning, also known as Stacked Generalization, is an advanced machine learning integration technique that leverages the collective strengths of multiple base models to enhance overall predictive performance (Li et al., 2025, Jiang et al., 2024). The core principle involves aggregating the predictions from several base models and employing a meta-model to generate the final prediction (Roy et al., 2022, Sakhrawi et al., 2022). This hierarchical approach not only improves accuracy but also enhances generalization capability. To optimize the base models for superior performance, a combination of 5-fold cross-validation and grid search is utilized to identify the optimal configurations. Each base model undergoes this rigorous tuning process to ensure robustness and precision. The meta-model, constructed using a 5-fold cross-validation mechanism, integrates the predictions from the base models as new input features. Ridge Regression is then applied to refine and integrate these predictions, thereby producing the final output.

The architecture of this stacking model (illustrated in Figure 2) clearly delineates the hierarchical relationship between the base models and the meta-model, as well as the data flow mechanism. This framework effectively combines the advantages of both linear and nonlinear modeling strategies. By fostering multi-model collaboration, it significantly enhances the fitting capability for complex spectral-physicochemical relationships. Consequently, this approach provides a robust and efficient solution for hyperspectral inversion of vertically stratified carbon-nitrogen (C/N) ratios in maize canopies.

Through this sophisticated ensemble method, the model not only achieves higher predictive accuracy but also ensures better generalization across diverse datasets. The structured integration of base models and meta-models facilitates a more nuanced understanding of the underlying data patterns, ultimately leading to enhanced performance and reliability in estimating C/N ratios within maize canopies. This comprehensive approach underscores the potential of stacking ensemble learning in advancing remote sensing applications and crop health monitoring.

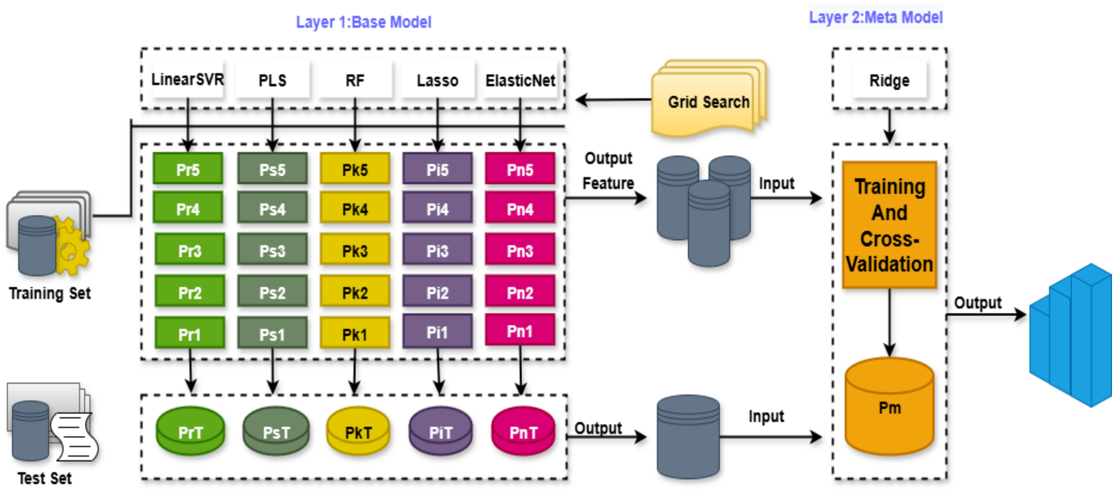

Figure 2. Design of the Two-Tiered Stacking-Based Ensemble Learning Workflow First-layer base models

First-level base models

Partial Least Squares Regression (PLS) is a powerful multivariate statistical technique designed for high-dimensional datasets, adept at mitigating multicollinearity. Random Forest Regression (RF), a non-parametric ensemble learning method, builds multiple decision trees and averages their predictions. Elastic Net regression integrates Lasso's L1 regularization and Ridge's L2 regularization, balancing the two penalty terms to perform feature selection and handle multicollinearity effectively, even when the number of features surpasses the number of samples. Lasso regression employs an L1-norm penalty term to shrink less significant regression coefficients to zero, facilitating feature selection and model simplification, ideal for scenarios with many variables but only a few key influencers. Linear Support Vector Regression (SVR) leverages Support

Vector Machines for regression tasks by constructing an optimal linear hyperplane to minimize prediction errors, demonstrating robustness to outliers and suitability for large-scale datasets with clear linear relationships. Together, these methods provide a comprehensive toolkit for advanced predictive modeling and analysis, ensuring robustness, accuracy, and generalizability across diverse datasets.

Second-layer meta-model

Ridge Regression is a linear modeling technique that incorporates an L2 regularization term into the loss function. By penalizing the squared magnitude of the regression coefficients, Ridge Regression shrinks these coefficients towards smaller values, thereby mitigating multicollinearity and enhancing the stability of model estimates (Ahmed et al., 2022). Each of the aforementioned modeling methods—Partial Least Squares Regression (PLS), Random Forest Regression (RF), Elastic Net regression, Lasso regression, and Linear Support Vector Regression (SVR)—exhibits unique characteristics suited to different types of data and modeling challenges. When combined with vegetation indices selected through rigorous feature selection processes, these methods are employed to construct hyperspectral inversion models for estimating the carbon-nitrogen ratio (C/N) in maize canopies. The predictive performance of these models is rigorously evaluated using cross-validation techniques. This evaluation process aims to identify the optimal modeling strategy by assessing the accuracy, robustness, and generalizability of each method under various conditions.

Through this comprehensive approach, the study not only leverages the strengths of diverse modeling techniques but also ensures that the resulting models are well-calibrated and reliable. The integration of advanced feature selection methods with sophisticated regression techniques provides a robust framework for hyperspectral inversion, facilitating precise and dependable estimations of C/N ratios in maize canopies. This systematic evaluation underscores the importance of selecting appropriate modeling strategies tailored to specific datasets and analytical goals, ultimately enhancing the overall effectiveness and applicability of the models.

1.7 Model Evaluation Methods

To validate the reliability of the established maize canopy carbon-nitrogen ratio (C/N) estimation model, this study employs a suite of statistical metrics to systematically assess the model's predictive accuracy. These metrics include the coefficient of determination (R^2), relative Root Mean Square Error (rRMSE), Mean Relative Error (MRE), and Root Mean Square Error (RMSE). rRMSE is a normalized variant of the root mean square error (RMSE), typically calculated by dividing RMSE by the mean or range of observed values. This relative error metric eliminates the influence of units, thereby facilitating comparisons across different models or datasets. The formulas for R^2, RMSE, rRMSE, and MRE are as follows:

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

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

$$rRMSE = \frac{RMSE}{\bar{y}} \times 100\% \quad (3)$$

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

2 RESULTS AND ANALYSIS

2.1 Analysis of Feature Vegetation Index Selection

This study integrated the feature selection method—Sequential Forward Selection (SA)—to screen 13 vegetation indices for each of the upper, middle, and lower layers within a Stacking ensemble model. The results indicate that although the specific combinations of the selected vegetation indices vary across layers, indices such as NDVI, REP, NDNI, and SAVI consistently exhibit high correlations across different layers. This demonstrates strong stability and cross-method consistency. These high-contribution indices are primarily concentrated in the red-edge to near-infrared spectral region (680–900 nm), which is highly sensitive to plant physiological and biochemical parameters.

Further analysis reveals that these key indices have been widely used in existing remote sensing studies to monitor chlorophyll content, canopy nitrogen status, or photosynthetic activity. For instance, NDVI and BNDVI are classical greenness indices that show significant responses to changes in chlorophyll concentration.NDNI and SAVI are extensively applied in remote estimation of vegetation nitrogen content and possess strong robustness against soil background interference REP (Red-Edge Position) effectively reflects photosynthetic potential and chlorophyll gradient dynamics,while indices such as CNSI and NCNI are specifically designed for pigments related to carbon-nitrogen metabolism and are directly linked to the dynamic changes in leaf carbon and nitrogen components Notably, NCNI and BNDVI, which perform prominently in the SA algorithm; all reflect characteristics of synergistic responses involving red-edge slope, nitrogen uptake, and canopy structure.

2.2 Hyperspectral Estimation of C/N in Different Vertical Layers of Maize Canopy Based on Ensemble Learning

The predicted C/N ratio outputs from the base models were used as input features to construct a Stacking ensemble learning model with Ridge Regression as the meta-model (Zhao et al., 2024). Compared to the base models, the Stacking model achieved the best estimation accuracy under all three feature selection methods. SA performed the best. Similarly, in the middle and lower layers, the SA-based models also achieved the highest performance.

After feature selection using the SA algorithm, the Stacking model achieved the following metrics (Figure 3): in the upper layer, $R^2 = 0.745$, RMSE = 1.249, rRMSE = 7.48%; in the middle layer, $R^2 = 0.702$, RMSE = 1.672, rRMSE = 9.82%; and in the lower layer, $R^2 = 0.674$, RMSE = 2.431, rRMSE = 13.1%. the Stacking ensemble model improved the upper-layer R^2 by 0.046, middle-layer R^2 by 0.076, and lower-layer R^2 by 0.106; reduced upper-layer RMSE by 0.107, middle-layer RMSE by 0.201, and lower-layer RMSE by 0.365; and lowered upper-layer rRMSE by 0.64%, middle-layer rRMSE by 1.2%, and lower-layer rRMSE by 1.96%. The improvement in prediction accuracy was most significant in the lower canopy layer.

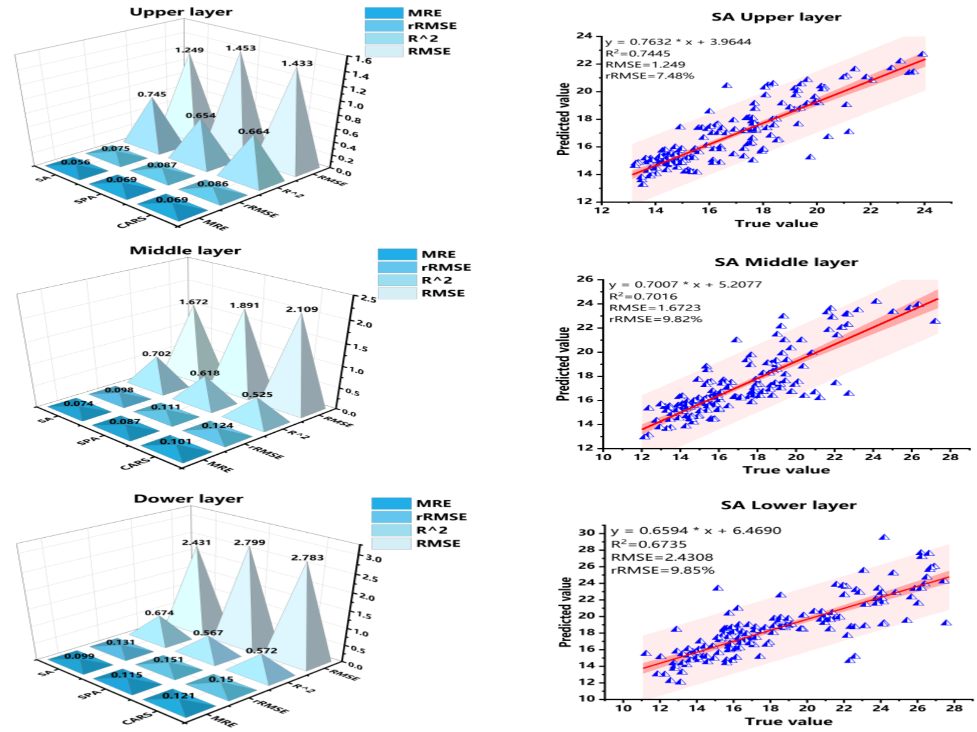

Figure 3. Evaluation metrics of ensemble models for C/N vertical distribution in maize canopies

3 DISCUSSIONS

With continuous advancements in computer technology, an increasing number of studies in recent years have attempted to estimate the carbon-nitrogen ratio (C/N) in crops by combining various modeling approaches with spectral data. Spectral technology, due to its advantages of speed, non-destructiveness, and capacity for batch processing, has been widely applied in agriculture, ecology, and environmental monitoring. However, to date, no universally applicable and highly stable C/N estimation model has been established. It indicates that traditional single-model approaches may have inherent limitations when dealing with complex spectral response relationships, such as insufficient ability to capture non-linear patterns, weak generalization, or susceptibility to noise. Therefore, improving overall prediction accuracy and model robustness remains a critical challenge in current research.

To address these issues, this study introduces the Stacking ensemble learning algorithm to construct a multi-model fusion framework for C/N estimation, aiming to fully leverage the strengths of individual models across different feature spaces to optimize final predictions. Experimental results show that, compared to individual base models, the Stacking ensemble model achieves significant improvements in both overall prediction accuracy and stability, maintaining strong performance across different feature selection methods. This finding aligns with current research trends and further validates the potential of ensemble learning in spectral modeling.

However, the study also found that the Stacking model does not exhibit an absolute advantage over all individual base models in every scenario. This suggests that, although ensemble learning enhances generalization overall, there remains considerable room for improvement in areas such as model architecture optimization, feature contribution analysis, and fusion strategies. Therefore, the full potential of ensemble learning for estimating C/N using spectral data requires further exploration and experimental validation.

4 CONCLUSIONS

This study addresses the impact of vertical structural heterogeneity in maize canopy on the accuracy of remote sensing inversion of carbon-nitrogen ratio (C/N). It proposes a canopy inversion modeling approach that integrates multi-spectral remote sensing data with vertically stratified field-measured C/N information. To enhance model performance and interpretability, the feature selection strategies—Simulated Annealing (SA) was used to select the sensitive VIs, and a Stacking-based ensemble learning model was constructed to assess the vertical distribution of C/N in maize canopies. The SA algorithm demonstrated superior global search capability in feature selection, effectively avoiding local optima and identifying more representative combinations of vegetation indices. The Stacking ensemble model significantly outperformed individual base learners, achieving high-precision inversion of C/N in the upper (UP), middle (MID), and lower (LOW) canopy layers, with validation set coefficients of determination (R^2) reaching 0.745, 0.702, and 0.674.

This study demonstrates that the proposed maize canopy vertical stratification C/N estimation model, based on canopy hyperspectral data and ensemble learning, can effectively achieve accurate estimation and evaluation of leaf C/N across different canopy layers. The proposed methodological framework not only provides a high-precision, interpretable technical pathway for vertical nutrient monitoring in maize but also offers a transferable modeling paradigm for stratified spectral inversion of other physicochemical parameters in crop canopies.

ACKNOWLEDGEMENTS

This research was funded by the National Natural Science Foundation (No. 42371323) and the National Key Research and Development Program (Nos. 2023YFD1701001-03 and 2022YFD2001103).

REFERENCES

Ahmed, A. A. M., Sharma, E., Jui, S. J. J., Deo, R. C., Nguyen-Huy, T., & Ali, M. (2022). Kernel Ridge Regression Hybrid Method for Wheat Yield Prediction with Satellite-Derived Predictors. Remote Sensing, 14(5), 1136. https://doi.org/10.3390/rs14051136

Chen, X., Li, F., & Chang, Q. (2023). Combination of Continuous Wavelet Transform and Successive Projection Algorithm for the Estimation of Winter Wheat Plant Nitrogen Concentration. Remote Sensing, 15(4), 997. https://doi.org/10.3390/rs15040997

Jiang, X., Gao, L., Xu, X., Wu, W., Yang, G., Meng, Y., Feng, H., Li, Y., Xue, H., & Chen, T. (2024). Combining UAV Remote Sensing with Ensemble Learning to Monitor Leaf Nitrogen Content in Custard Apple (Annona squamosa L.). Agronomy, 15(1), 38. https://doi.org/10.3390/agronomy15010038

Li, Y., Xu, X., Wu, W., Zhu, Y., Yang, G., Gao, L., Meng, Y., Jiang, X., & Xue, H. (2024). Hyperspectral Estimation of Leaf Nitrogen Content in White Radish Based on Feature Selection and Integrated Learning. Remote Sensing, 16(23), 4479. https://doi.org/10.3390/rs16234479

Liang, J., Yin, X., & Lin, M. (2025). An enhanced microstate clustering algorithm based on canopy, K-means, and genetic simulated annealing. Biomedical Physics & Engineering Express, 11(4), 045002. https://doi.org/10.1088/2057-1976/adda50

Roy, J., & Saha, S. (2022). Ensemble hybrid machine learning methods for gully erosion susceptibility mapping: K-fold cross validation approach. Artificial Intelligence in Geosciences, 3, 28–45. https://doi.org/10.1016/j.aiig.2022.07.001

Sakhrawi, Z., Sellami, A., & Bouassida, N. (2022). Software enhancement effort estimation using correlation-based feature selection and stacking ensemble method. Cluster Computing, 25(4), 2779–2792. https://doi.org/10.1007/s10586-021-03447-5

Wang, B., Gu, S., Wang, J., Chen, B., Wen, W., Guo, X., & Zhao, C. (2024). Maximizing the Radiation Use Efficiency by Matching the Leaf Area and Leaf Nitrogen Vertical Distributions in a Maize Canopy: A Simulation Study. Plant Phenomics, 6, 0217. https://doi.org/10.34133/plantphenomics.0217

Xu, X., Yang, G., Yang, X., Li, Z., Feng, H., Xu, B., & Zhao, X. (2018). Monitoring ratio of carbon to nitrogen (C/N) in wheat and barley leaves by using spectral slope features with branch-and-bound algorithm. Scientific Reports, 8(1). https://doi.org/10.1038/s41598-018-28351-8

Yang, C., Feng, M., Bai, J., Sun, H., Bi, R., Song, L., Wang, C., Zhao, Y., Yang, W., Xiao, L., Zhang, M., & Song, X. (2025). Hyperspectral estimation of chlorophyll density in winter wheat using fractional-order derivative combined with machine learning. Frontiers in Plant Science, 15. https://doi.org/10.3389/fpls.2024.1492059

Yang, Z., Tian, J., Zhang, L., Zan, O., Yan, X., & Feng, K. (2024). Spectral detection of leaf carbon and nitrogen as a proxy for remote assessment of photosynthetic capacity for wheat and maize under nitrogen stress. Computers and Electronics in Agriculture, 224, 109174. https://doi.org/10.1016/j.compag.2024.109174

Zhao, H., Wang, J., Guo, J., Hui, X., Wang, Y., Cai, D., & Yan, H. (2024). Detecting Water Stress in Winter Wheat Based on Multifeature Fusion from UAV Remote Sensing and Stacking Ensemble Learning Method. Remote Sensing, 16(21), 4100. https://doi.org/10.3390/rs16214100