Hyperspectral–Region Aggregation Network for Maize Leaf Nitrogen Content Estimation via Spectral–Regional Joint Modeling

Fuheng Qu
Changchun University of Science and Technology
Mingyang Zhu
Changchun University of Science and Technology
Xiaofeng Li
lixiaofeng@iga.ac.cn

Chinese Academy of Sciences
Li Zhu
Changchun University of Science and Technology
Chunran Wu
Changchun University of Science and Technology
Peng Gao
Changchun University of Science and Technology

Article

Keywords: Maize leaf nitrogen content, Hyperspectral inversion, Hyperspectral-Region Aggregation Network (HSRAN)

Posted Date: September 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10174556/v1

Additional Declarations: No competing interests reported.

Hyperspectral–Region Aggregation Network for Maize Leaf Nitrogen Content Estimation via Spectral–Regional Joint Modeling

Fuheng Qu 1, Mingyang Zhu 1, Xiaofeng Li *2, Li Zhu 2, Chunran Wu 2, Peng Gao 2

[1] School of Computer Science and Technology, Changchun University of Science and Technology, Changchun 130000, China

[2] Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun 130102, China

[*] lixiaofeng@iga.ac.cn

[ǂ] Fuheng Qu and Mingyang Zhu contributed equally to this work

Abstract:

Accurate estimation of maize leaf nitrogen content is important for improving nitrogen-use efficiency and supporting precision crop management. However, leaf-level hyperspectral modeling is challenged by high spectral redundancy and heterogeneous spectral responses among local leaf regions. This study proposes a Hyperspectral–Region Aggregation Network (HSRAN) for maize leaf nitrogen content estimation from region-level hyperspectral spectra. HSRAN consists of a Spectral Adaptive Recalibration Encoder (SARE) and a Context-Aware Gated Aggregation Module (CAGM). SARE performs band-wise residual recalibration and extracts regional spectral representations, whereas CAGM models contextual dependencies among regional features and performs gated attention-based aggregation for leaf-level prediction.

Field experiments were conducted in 2024 and 2025 at the jointing, silking, and maturity stages. HSRAN was evaluated against PLSR, RF, XGBoost, SVR, 1D-CNN, MLP, and Transformer1D models. Across the stage-specific and pooled datasets, HSRAN achieved the highest R^2 and the lowest RMSE while maintaining competitive MAE values. On the pooled full-growth-period dataset, HSRAN achieved an R^2 of 0.84, an RMSE of 3.63 g kg^{-1}, and an MAE of 2.59 g kg^{-1}. At the jointing, silking, and maturity stages, the corresponding R^2 values were 0.56, 0.76, and 0.72, respectively. Ablation experiments indicated that integrating SARE and CAGM improved R^2 from 0.80 to 0.84. To interpret regional contributions, the learned attention weights were mapped back to the original leaf coordinates recorded during regional sampling. Regions near leaf veins, tips, and margins often received relatively higher attention weights, suggesting that their local spectra provided informative cues for model prediction.

These findings indicate that spectral–regional joint modeling can improve leaf-level hyperspectral estimation of maize nitrogen content. HSRAN provides a practical framework for non-destructive nitrogen assessment in maize.

Keywords: Maize leaf nitrogen content; Hyperspectral inversion; Hyperspectral-Region Aggregation Network (HSRAN)

1 Introduction

Nitrogen is an essential nutrient for maize growth and yield formation because it directly affects leaf photosynthesis, dry matter accumulation, and grain filling. Insufficient nitrogen supply can restrict crop growth and reduce yield, whereas excessive nitrogen application increases production costs and may lead to resource waste, soil acidification, nitrate leaching, and greenhouse-gas emissions. Therefore, accurate and non-destructive assessment of maize nitrogen status is important for improving nitrogen-use efficiency and supporting sustainable crop production.

Hyperspectral remote sensing provides dense and continuous spectral information and is more sensitive than conventional multispectral sensing to subtle changes in leaf biochemical composition and structure, particularly in the red-edge and near-infrared regions. Previous studies have demonstrated the potential of hyperspectral data for maize nitrogen monitoring at different spatial scales. Zhang et al. [1] estimated maize leaf nitrogen content using unmanned aerial vehicle hyperspectral imagery and machine-learning methods. Wang et al. [2] developed a hyperspectral model for estimating SPAD values in summer maize and showed that spectral responses can serve as indirect indicators of nitrogen status. Wei et al. [3] investigated the vertical distribution of leaf nitrogen in spring maize under different mulching and irrigation conditions and highlighted the effects of canopy structure on nitrogen estimation. Yun et al. [4] combined spectral and texture information to estimate maize nitrogen nutrition, whereas Zhai et al. [5] developed a dynamic nitrogen-diagnosis method for maize under drip irrigation.

In addition to direct nitrogen estimation, hyperspectral imaging has been widely used to assess maize chlorophyll content, water status, and other physiological traits associated with nitrogen metabolism. Wang et al. [6] estimated chlorophyll and water contents in maize leaves under drought stress using hyperspectral imaging. Zhou et al. [7] retrieved chlorophyll content in film-mulched maize using image segmentation and spectral analysis. Li et al. [8] explored chlorophyll estimation using multi-source remote sensing data, and Li et al. [9] compared PROSAIL- and Sentinel-2-based methods for estimating chlorophyll content in summer maize. These studies indicate that hyperspectral information can capture biochemical and structural variation relevant to maize nitrogen status.

Related research on winter wheat and rice has also demonstrated the value of hyperspectral sensing for nutrient estimation. Sun et al. [10] developed a nitrogen-concentration estimation model for winter wheat using UAV hyperspectral images. Xu et al. [11] proposed a CARS-RUN-ELM method for the simultaneous estimation of nitrogen and phosphorus contents in rice leaves. Guo et al. [12] showed that the red-edge and near-infrared regions were sensitive to changes in leaf nitrogen content in winter wheat under different nitrogen application levels. In addition, Tan et al. [13] used partial least squares regression to predict maize yield under different water and nitrogen conditions. With the development of machine-learning methods, hyperspectral crop-trait estimation has gradually shifted from empirical spectral-index regression toward data-driven nonlinear modeling [14].

Despite this progress, several limitations remain in leaf-level hyperspectral nitrogen

estimation. First, hyperspectral data contain a large number of highly correlated adjacent bands, which may introduce redundant information and increase the difficulty of stable feature extraction under limited sample sizes. Second, the spectral responses of different local regions within a leaf are not necessarily identical. Differences in tissue structure, leaf veins, illumination, water status, local senescence, and nitrogen transport may cause regional spectral heterogeneity. Existing methods often use the average spectrum of the whole leaf or treat pixels and local regions as independent samples, which may weaken complementary information among regional spectral features. Recent studies have shown that maize leaf veins and mesophyll regions may exhibit distinct spectral responses under nitrogen stress, suggesting that local regional information can be valuable for nitrogen estimation [19]. However, a unified framework that jointly addresses spectral redundancy and contextual relationships among regional spectral features remains limited.

To address these challenges, this study proposes a Hyperspectral–Region Aggregation Network (HSRAN) for maize leaf nitrogen content estimation. HSRAN consists of a Spectral Adaptive Recalibration Encoder (SARE) and a Context-Aware Gated Aggregation Module (CAGM). SARE performs band-wise residual recalibration and extracts compact regional spectral representations, whereas CAGM models contextual dependencies among regional features and aggregates them through gated attention pooling to generate a leaf-level representation. In addition, the learned regional attention weights are mapped back to the original leaf coordinates recorded during regional sampling to visualize the spatial distribution of model-relevant spectral information. The proposed framework was evaluated using maize leaf samples collected in 2024 and 2025 across the jointing, silking, and maturity stages. Its performance was compared with conventional machine-learning methods and representative deep-learning models, and the contributions of spectral recalibration and regional aggregation were further examined through ablation and aggregation-strategy analyses.

2 Materials and Methods

2.1 Field Experiment Design

The study site was located at the Agricultural Experiment Station of the Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences (43.998°N, 125.402°E). The experimental material was the silage maize variety "Jinboshi 825." Field experiments were conducted in 2024 and 2025 with varying nitrogen application levels and agricultural management practices to obtain samples with diverse maize leaf nitrogen status.

In 2024, 96 leaf samples were collected at each of the jointing, silking, and maturity stages. In 2025, 81 samples were collected at each stage. When sampling, attention was paid to the treatment conditions and collect the leaves at the upper, middle, and lower canopy positions of the plants to ensure the representativeness and diversity of the samples.

2.2 Research Methods

2.2.1 Hyperspectral Data Collection and Processing

A hyperspectral imaging system operating in the 390 to 1100 nm spectral range was used to acquire hyperspectral images of maize leaves. The system comprised a charge-coupled device (CCD) camera, an imaging spectrograph (Pika XC2, Reson Inc., Bozeman, MT, USA), a zoom lens (XENOPLAN, F/1.4, f = 23 mm, Schneider-Kreuznach, Bad Kreuznach, Germany), a four-lamp illumination unit (OSRAM, Munich, Germany; 35 W per lamp, total input power 140 W, total radiant power approximately 5 to 7 W), and a computer running data acquisition and control

software (SpectrononPro version 2.100, Fermi Resonance Inc., Bozeman, MT, USA). The imaging spectrograph had a slit width of 50 μm, a spectral resolution of 1.3 nm, and a spatial resolution of 0.15 mm per pixel. Measurements were conducted in a dark environment to eliminate the influence of ambient light. Dark current correction and black-white calibration were performed before each measurement. A total of 462 spectral bands were acquired per sample. The hyperspectral imaging system is shown in Figure 1, and 462 spectral bands are measured for each sample.

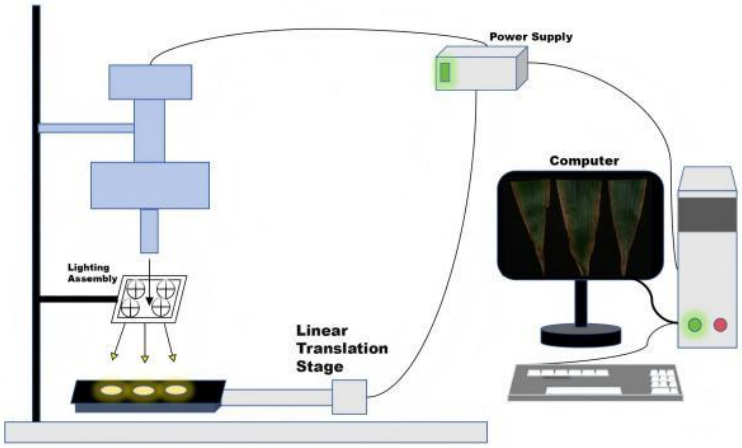


Figure 1. Hyperspectral Imaging System

Hyperspectral cubes and corresponding RGB images were exported using SpectrononPro and subsequently processed in MATLAB. In this study, the spectral data are applied to the independently developed precise segmentation system to carry out the segmentation operation of maize leaves, so as to ensure the stability and high signal-to-noise ratio of the spectral data input to the subsequent model, thereby improving the precision of segmentation.

Input hyperspectral imaging data for initialization, automatically read the width, height and number of bands of the image, and then set up the matrix. The system reads each input sample in sequence, and images failing the predefined criteria for signal integrity or exposure balance were excluded and reacquired when necessary.

After the sample quality meets the standard, the leaf area extraction work is carried out. This step relies on the near-infrared band and red band of the hyperspectral image to calculate NDVI:

$$NDVI = \frac{R_{nir} - R_{red}}{R_{nir} + R_{red}}$$

Based on the empirical threshold (NDVI > 0.4), a preliminary vegetation mask is generated, after which it enters the multi-feature fusion segmentation module.

This module constructs a composite feature map using hyperspectral, RGB images and texture features, integrates HSV color features, local standard deviation filtering and edge detection operators to show the image boundaries, and then removes pseudo-noise regions through morphological optimization (opening and closing operations) to obtain a complete leaf mask. After determining the mask, random sampling is performed on the target area and spectra are extracted. Randomly selected the reflection data of several central pixels in the leaf area, calculate the average reflectance of the 462 bands in the 20×20 pixel range each time, and perform 100 times of random average operations. Store the regional sample spectral data and the measured concentration of nitrogen content labels into a training sample table for modeling analysis. The details of the flow chart are shown in Figure 2.

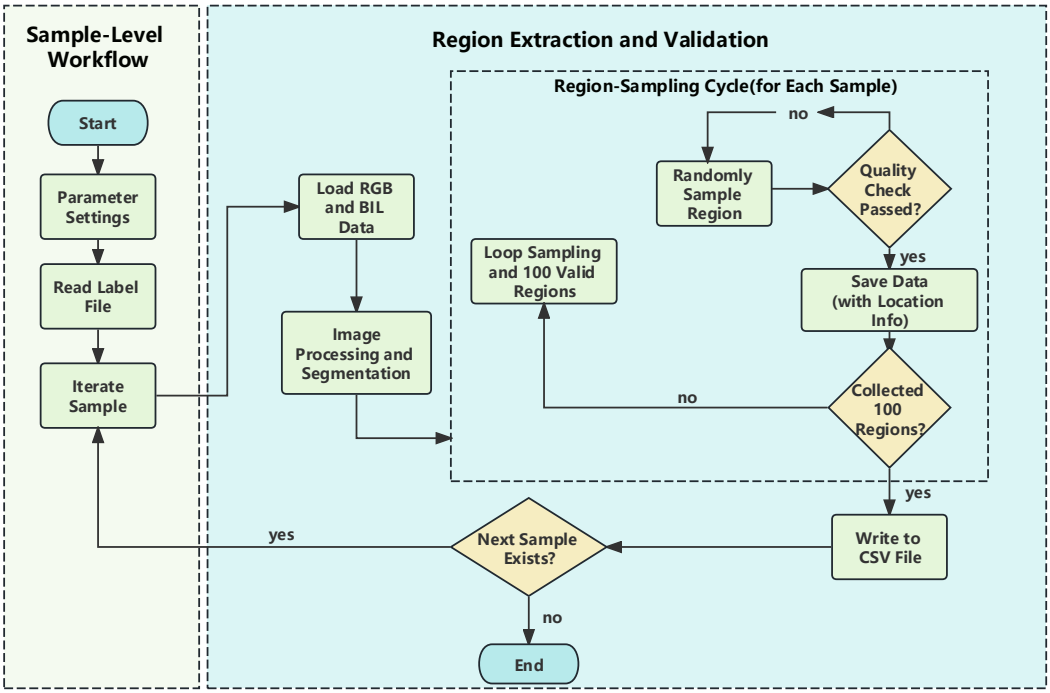


Figure 2. Flowchart of Hyperspectral Data Processing

2.2.2 Determination of Nitrogen Content: Leaf samples were enzyme-deactivated at 105 °C for 30 min and subsequently oven-dried at 75 °C to constant weight. The dried samples were ground into powder, sealed in sample bags, and stored under dry conditions. Total nitrogen concentration was determined using a continuous-flow analyzer according to LY/T 1228–2015. The content of nitrogen is within the range defined in Table 1. The total nitrogen conditions in different growth periods have been statistically included in Table 1. The nitrogen content distribution in each period is different. The overall range is 7.89 to 56.89 grams per kilogram. The nitrogen content is relatively high in the elongation period and flowering period, and the nitrogen content is significantly reduced in the maturity stage, showing a situation of nitrogen accumulation and redistribution. The data is comprehensive and can provide support for spectral modeling analysis. The data of each stage such as germination, flowering and maturity of plant periodic experiments were obtained from 2024 to 2025. The complete data set covers the data of each period in two years.

Table 1. Nitrogen Content Range

maize-Period	No.	Min ($g \cdot kg^{-1}$)	Max ($g \cdot kg^{-1}$)	Average ($g \cdot kg^{-1}$)
2024-Jointing Stage	96	20.06	56.89	41.56
2024-Silking Stage	96	21.37	45.46	32.1
2024-MatureStage	96	14.14	28.04	19.01
2025-JointingStage	81	20.92	40.24	35.04
2025-SilkingStage	81	8.83	42.14	31.8
2025-MatureStage	81	7.89	25.54	19.69

2.3 Spectral Data Preprocessing Process

(1) Standard Normal Variate Transformation: The first step is to employ the standard normal variate transformation to eliminate scattering effects and instrumental errors:

$$X_{SNV} = \frac{X_{D1} - \mu}{\sigma}$$

Where μ represents the mean of a single spectrum, and σ represents the standard deviation; the SNV transformation produces spectra with zero mean and unit variance.

(2) In this study, the Savitzky-Golay (SG) method is used to smooth the spectrum to eliminate random noise in the hyperspectral data while retaining the morphology of the original spectrum. SG filtering maintains the morphology of the absorption peak position by using the least squares fitting of local polynomials, reduces high-frequency noise, and then generates the original spectrum:

$$X_{raw} = \{x_1, x_2, ..., x_n\}$$

For every band i in the window, data is fit with a polynomial of order p in a window centered at i with a window width of w=2m+1:

$$x(i+k) \approx \sum_{j=0}^{p} a_j k^j, \quad k = -m, ..., m$$

The polynomial coefficients p, fitting coefficients a_j, and the number of bands m that the sliding window spans to both sides of the central band will be solved for, and the smoothed spectral values can be obtained:

$$X_{SG}(i) = \sum_{k=-m}^{m} c_k X_{raw}(i+k)$$

The determination of the convolution coefficient c_k is related to the polynomial coefficient p and the window width w, and it is completed through polynomial fitting. In this study, the parameter settings of SG filtering are as follows.

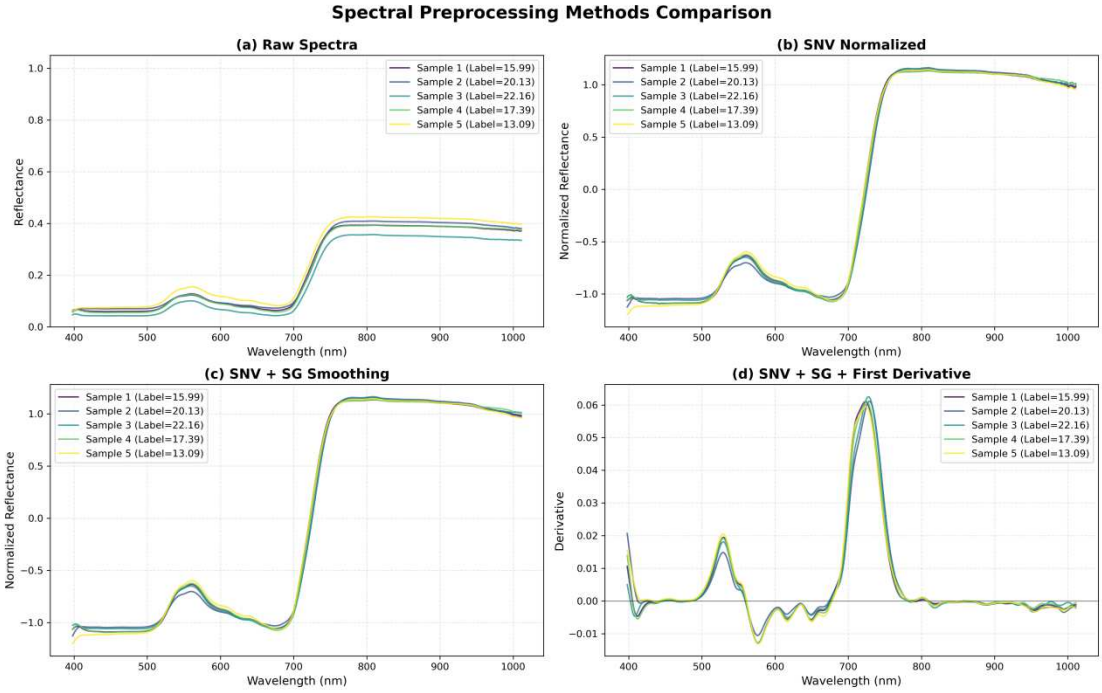
$$X_{SG} = SG(X_{raw}, w=11, p=2, d=0)$$

(3) First derivative transformation: After completing the SG smoothing process, in order to further weaken the effects of baseline drift and enhance the spectral detail features, a first derivative transformation is applied to the smoothed spectrum. Derivative transformation can highlight the changing trend of spectral curves and inflection point information, and can enhance the sensitivity to the absorption characteristics of spectral curves and the change of red edge position. The first-order derivative spectrum is as follows:

$$X_{D1} = SG(X_{SG}, w=11, p=2, d=1)$$

Set the width of the window to 11, the number of operations is 2, and take d as 1 to calculate the first-order derivative. Fit the local spectral curve and its first-order derivative with a cubic polynomial, highlight the strong slope change characteristics, and maintain the continuity of the spectrum.

The variation of nitrogen content is often accompanied by the change of the red edge position and absorption intensity. The first-order derivative processing of these spectral response characteristics



related to nitrogen can be optimized, and then the efficiency of the model to distinguish the difference in nitrogen content can be improved.

Figure 3: Spectral Data Preprocessing

Figure 3 shows the differences in hyperspectral data at each stage of maize leaf pretreatment. The original spectral band has the characteristics of vegetation spectrum, only the amplitudes are different. SNV transformation makes the morphology of sample spectra closer and can also reduce the influence of scattering. SG smoothing filters out random noise, and the first-order derivative transformation optimizes the response characteristics of key spectral bands such as the red edge. Overall, this preprocessing workflow effectively enhances the discriminatory ability of features related to nitrogen content while ensuring the preservation of spectral structural information, thereby providing a reliable data foundation for subsequent modeling and analysis.

2.4 Evaluation Metrics

This study quantitatively evaluates the performance of the nitrogen content regression model with three commonly used indicators. These three indicators are the coefficient of determination (R^2), the root mean square error (RMSE), and the mean absolute error (MAE). Suppose the actual value of nitrogen content of the i-th sample is $\hat{y}_i$, and its predicted value is $\bar{y}$.

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

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

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

2.5 HSRAN Model

2.5.1 Model Design Principles

There are two major challenges in modeling: the high-dimensional hyperspectral data, strong correlation and much redundancy between adjacent bands, and the contribution of absorption effects of different spectral bands to nitrogen absorption is different; the micro-structure of leaves, the spatial inhomogeneity of nitrogen distribution and light, bring difficulties to empirical data modeling, the spectrum of a single region or pixel cannot fully explain the nitrogen status of leaves, and if there is no model of inter-regional correlation, the model is unstable.

To address the above issues, this paper proposes a Hyperspectral–Region Aggregation Network that integrates spectral enhancement and regional correlation modeling concepts—HSRAN (Hyperspectral–Region Aggregation Network). The overall structure of the model is shown in Figure 4.

HSRAN takes the spectrally sampled regional data of the divided leaf area as input and employs a two-stage modeling strategy of "spectral feature enhancement + regional feature association aggregation" to achieve direct prediction from regional spectral data to leaf-level nitrogen content.

The overall architecture comprises three core modules:

1. SARE (Spectral Adaptive Recalibration Encoder Module)
2. CAGM (Context-Aware Gated Aggregation Module)
3. Regression Prediction Module

Among these: the SARE module primarily addresses the hyperspectral dimensional redundancy problem by embedding mappings and residual structures to control the degree of residual enhancement across spectral bands while enhancing the expression of key bands related to nitrogen content; the CAGM module introduces a cross-regional attention mechanism at the regional level and combines a Multi-Instance Learning (MIL) strategy for associative modeling and adaptive weighting of multiple regional features, thus achieving leaf-level feature representation; the regression module performs nonlinear mapping on the aggregated leaf-level features, outputting the final nitrogen content prediction results.

By the above design, HSRAN realizes "de-redundancy and key enhancement" in the spectral dimension, and "regional correlation and adaptive aggregation" in the spatial dimension, thus balancing the spectral information and regional structural information, which enhances its stability and adaptability performance under the current data setting to a certain extent.

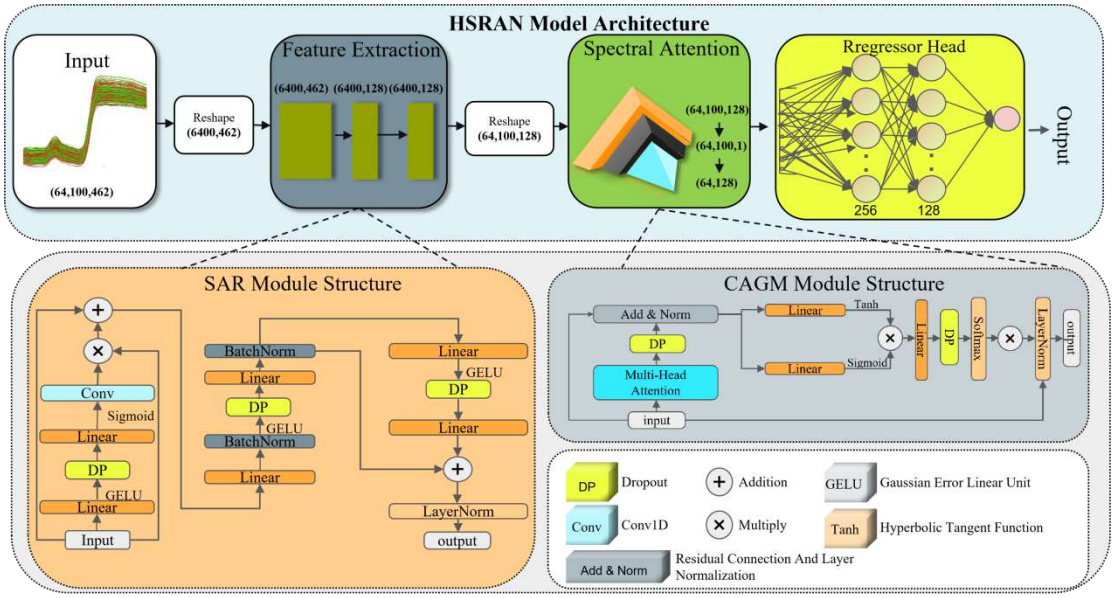


Figure 4 HSRAN maize Leaf Nitrogen Content Estimation Model Overall Architecture. The model input is regional-level hyperspectral data, with dimensions represented as (B, R, L), indicating batch size, the number of leaf regions, and the number of spectral bands, respectively. The overall process includes: (1) Spectral feature extraction module, incorporating the SARE mechanism, recalibrating bands and reducing dimensionality; (2) Context-Aware Gated Aggregation Module(CAGM), aggregating regional-level features into leaf-level representations using multi-head self-attention and gated attention pooling; (3) Regression head, mapping aggregated features to nitrogen content predictions. This architecture facilitates direct prediction from hyperspectral reflectance to nitrogen content (g·kg⁻¹).

2.5.2 SARE Module

Hyperspectral data shows strong continuity, just like there are close correlations between adjacent bands. However, the reaction intensity of maize leaves to nitrogen content varies. Directly modeling the original spectral vector is easily disturbed by redundant bands, which is not conducive to the presentation and promotion of features.

To address this, this paper designs the Spectral Adaptive Recalibration Encoder Module (SARE), which is utilized for feature enhancement and band weight redistribution of the input spectra, achieving adaptive emphasis on key spectral regions.

(1) Feature Mapping

Let the input region-level spectral representation be:

$$X \in \mathscr{R}^{B \times L}$$

Number L, batch size B. The feature embedding contains two linear mapping operations:

$$F = W_2 \sigma(W_1 X)$$

The two parameter vectors W1 and W2 need to be learned, and σ(·) belongs to the type of nonlinear activation function. This method maps the original spectrum to a high−dimensional feature space, which can help detect the hidden correlation between bands.

(2) Spectral Recalibration

After obtaining the intermediate feature representation F, convolution operations are introduced to model local spectral correlations:

$$G = \mathrm{Conv1D}(F)$$

The convolution kernel traverses in the spectral dimension, which is a way to capture the local structural features of adjacent frequency bands, and then a frequency band weight vector is constructed:

$$A=\sigma(G)$$

A belongs to the category of RB × L, representing the adaptive weights of each frequency band. So far, the spectral calibration work is announced to be over:

$$X_{recal}=X\odot A$$

This process achieves the enhancement of crucial spectral bands while applying different degrees of residual enhancement across spectral bands.

(3) Residual Fusion

To prevent excessive modulation leading to the loss of original information, SARE employs a residual fusion structure:

$$X_{out}=X+X_{recal}$$

Add LayerNorm at the output end to improve stability. This module helps strengthen key information, maintains the original spectral structure, and promotes the model training to be stable.

2.5.3 CAGM Module

There are obvious differences in the distribution of nitrogen elements in maize leaves and the spectral responses within different spatial ranges. In some areas, spectral features are shifted or noise is generated due to factors such as light and structure. If only the features of a single area or pixel are adopted, it is easy to cause fluctuations in prediction and it is difficult to stably reflect the state of total nitrogen in leaves.

This paper proposes Multiple Instance Learning (MIL) at the regional level, and proposes a Context-Aware Gated Aggregation Module module to realize inter-regional relation modeling and adaptive feature aggregation.

(1) MIL formal modeling

A single leaf sample consists of N regional instances:

$$B=\{r_1,r_2,...,r_N\}$$

where: $r_i \in \mathcal{R}^d$ represents the spectral feature of the i-th region; the entire sample set B shares the same leaf-level label y.

Unlike traditional supervised learning, which directly regresses on a single region, Multiple Instance Learning (MIL) aims to learn a holistic representation from the instance-level ensemble:

$$Z=f(B)$$

where Z represents the global feature representation at the leaf level.

(2) Region Relation Modeling (Multi-Head Attention)

To capture the contextual dependencies between regions, the CAGM module employs a multi-head attention mechanism for associating region features.

First, a linear mapping is performed on the region features:

$$Q=RW_Q,\ K=RW_K,\ V=RW_V$$

where $R \in \mathcal{R}^{N\times d}$ is the region feature matrix, and W_Q,W_K,W_V are learnable parameters. The attention weights are calculated as:

$$A=\text{Softmax}(\frac{QK^T}{\sqrt{d}})$$

The region features are updated as:

$$R' = AV$$

The process permits the model to set up global dependencies among the different models of the regions ensemble, allowing for the inclusion in the representation of each region of contextual information from other regions, which improves the robustness of the model.

(3) Adaptive Modulation Aggregation

The context-enhanced regional representation R' is obtained and an adaptive weight modulation mechanism is added to compute the importance of the region:

$$\alpha_i = \frac{\exp(w^T r_i')}{\sum_{j=1}^{N} \exp(w^T r_j')}$$

where α_i is the contribution weight of the i-th region to prediction at the leaf level. The final representation of the leaf is:

$$z = \sum_{i=1}^{N} \alpha_i r_i'$$

Compared to simple averaging or fixed-weight aggregation, this mechanism can dynamically enhance the contribution of key regions while suppressing the influence of anomalous regions disturbed by noise.

(4) Residuals and Normalization Stabilization

To improve training stability, an Add & Norm structure is adopted internally in the module:

$$R_{out} = \text{LayerNorm}(R + R')$$

Residual connections prevent loss of original region information due to excessive modulation, while LayerNorm enhances convergence stability.

2.6 Training Strategy and Optimization

The software and hardware experimental platform for deep learning is shown in Table 2.

Table 2 Deep Learning Experimental Environment

Items	Detail
Operating System	Windows
CPU	Intel I5-12600KF
GPU	NVIDIA GEFORCE RTX 5060TI
Acceleration Environment	CUDA 12.8
Language	Python 3.9.23
Framework	Pytorch 2.7.1

The deep learning models involved in this paper (HSRAN, 1DCNN, MLP, Transformer1D) are implemented in PyTorch, and the traditional machine learning models (PLSR, RF, XGBoost, SVR) are implemented in scikit-learn. Except for the deep learning architecture part, the training process is consistent, so as to ensure that different models can be compared fairly.

The infrastructure of MLP is a three-layer fully-connected spectrum encoding device with dimensions 4622, 5612, 8128 in sequence, and then the nitrogen content is estimated by the MIL method of gated attention mechanism. The DCNN baseline adopted in 1 is a network with three-layer one-dimensional convolution structure, the channels are changed from 1 to 32, then to 64, and finally to 128, the size of the convolution kernel is 7, and then it is aggregated and connected to the regression head through gated attention MIL. The basic mode of Transformer1D

is to convert the 462-dimensional spectrum into a 128-dimensional embedded form, and then build the model through three-layer Transformer encoders, including 4 heads, 4 times the expansion ratio of the feed-forward part and CLS token, and finally carry out prediction operations by using the regression head. HSRAN operates with the spectrum enhancement encoder SARE, the region association aggregation module CAGM and the regression head. The SARE regression head has functions such as spectrum recalibration gating, and the CAGM regression head contains multi-head self-attention and gated attention aggregation.

All the deep learning models use the same division and optimisation of the data in terms of training strategy. First establish quality standards for all areas, and then correct the misplaced labels. Then a stratified partitioning strategy is used at the leaf level, 20% of the samples are set aside as an independent test set, and it is ensured that the partition is done at the leaf level; the remaining samples are used for 5-fold cross-validation. The AdamW optimizer is used in the training process; the initial learning rate is 1×10^{-4}, the initial weight decay is 3×10^{-2}, and a warmup and cosine annealing learning rate scheduling strategy with a warmup epoch of 10 are employed. Batch size is 16, and the maximum number of training epochs is 300. Early stopping is carried out based on the RMSE of the validation set with a patience of 30 and a minimum improvement of 1×10^{-4}. A maximum value of 1.0 is used for gradient clipping in training to improve training stability.

To solve the problem of overfitting in a small sample size, Gaussian noise perturbation and uniform scaling are applied uniformly to the training set, setting the noise standard deviation to 0.01, and the range of scaling is between 0.97 and 1.03; no noise perturbation or scaling is performed on the validation and test sets. At the test stage, Test-Time Augmentation is employed with 8 perturbation predictions per sample, and then the average of these is used to enhance the robustness of the test results.

The traditional machine learning baselines use a number of latent variables between 5 and 15 for PLSR, 200 trees with the maximum depth automatically set by the algorithm for RF, a maximum depth of 6 and a learning rate of 0.1 for XGBoost, and a penalty parameter C of 10 for SVR with a radial basis function kernel. All of the experiments are carried out at work stations with NVIDIA GPUs.

2.7 Experimental Setup and Validation Strategy

In order to systematically evaluate the performance of the model, this paper designs two verification methods. The first is the cross-year mixed sample verification strategy. The specific operation is to mix the samples of jointing stage, silking stage and maturity stage in the two growing seasons of 2024 and 2025, and divide them into three sets of data sets according to the growth period (jointing, silking and maturity). At the same time, all the samples in the two years are merged into a set of FULL data sets. This configuration is used for model comparison experiments, ablation experiments, spatial attention tests, and regional mechanism verification. The main purpose is to examine the predictive performance and structural effectiveness of the model under current sample conditions.

In addition, a set of independent year verification analysis was also set up. The practice was to do five-fold cross-validation within the training set, and then to pull out the 2025 samples separately as the external test set. It should be noted that this group of experiments focuses on the impact of distribution differences between years on the migration ability of the model, and is not the main evaluation scheme for the model comparison in this paper. For performance changes, the

paper discusses from the perspectives of spectral distribution and label distribution to sort out the reasons behind the changes. The spectral distribution was calculated by principal component analysis (PCA), maximum mean difference and Fréchet distance, and the label distribution was measured by Kolmogorov-Smirnov test.

3. Results and Analysis

3.1 Comparison Experiments

The Coefficient of Determination (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) were used to evaluate the predictive performance of the various models during the jointing stage, silking stage, maturity stage and the entire growth period (FULL) on the datasets (Table 3). It should be noted that the data of each period are constructed by combining the samples of the same period in 2024 and 2025, so the results of each period reflect the predictive ability and stability of the model under the condition of mixed data in the same year.

The performance of each model in different growth stages was different. From the jointing stage, the R^2 of PLSR, RF, XGBoost and SVR models fell roughly between 0.35 and 0.42, and the overall accuracy was not ideal. It may be due to the low degree of nitrogen differentiation in the early stage of vegetative growth and the weak spectral response. In the jointing stage, 1D-CNN performed the worst, and R^2 was only 0.24, indicating that in the case of limited sample size and high spectral redundancy, it is difficult to ensure stable expression by using convolution structure for feature extraction. The R^2 of most traditional methods is not as good as MLP and Transformer1D (both are 0.45 and 0.43, respectively), but they are not comparable to HSRAN. HSRAN performed best at jointing stage, R^2 reached 0.56, RMSE and MAE were 4.16 g/kg^{-1} and 2.96 g/kg^{-1}, respectively, indicating that it did have a better ability to describe the weak spectral differences in the early stage.

At the silking stage, the performance of each model is generally improved. The R^2 of RF, XGBoost, 1D-CNN, and Transformer1D reached 0.58, 0.60, 0.52, and 0.58, respectively, and the MLP was increased to 0.73. That is to say, the more complex the canopy structure and the more obvious the nitrogen differentiation, the more significant the prediction contribution of nonlinear feature modeling. However, the RMSE of most baseline models is at least 2.67 g/kg^{-1}, and the robustness is not satisfactory in the case of enhanced spatial heterogeneity. The highest results are obtained in this step and the R^2 is 0.76, RMSE is 2.57 g/kg^{-1}, and the MAE is 2.12 g/kg^{-1}, where HSRAN is still the best. The R^2 and RMSE values of HSRAN are higher than the remaining models, and are closer to the R^2 and RMSE of the MLP model showing that HSRAN can model comprehensively very well in the region of information filtering in the spectrum and in feature aggregation.

The results of the maturity stage show that the overall prediction error continues to go down, but the R^2 of most traditional baseline models still does not exceed 0.50. Compared with most traditional machine learning methods, the performance of MLP is relatively good, R^2 is 0.60, MAE is 1.40 g/kg^{-1} ; the R^2 of Transformer1D is 0.52, RMSE is 1.96 (MAE is 1.57 g/kg^{-1}), which is better than 1D-CNN. This shows that in the maturity stage, spectral information can still be used in part by means of attention-based sequence modeling. However, even in the case of more obvious physiological differentiation in the late growth stage, HSRAN still remained the best, R^2 reached 0.72, RMSE and MAE decreased to 1.50 g/kg^{-1} and 1.20 g/kg^{-1}, respectively, and the three indicators were better than other comparison methods. That is to say, under the more obvious

differentiation conditions in the later stage, the framework combining spectral enhancement and regional correlation modeling has outstanding advantages.

Looking at the FULL dataset, HSRAN got the best overall result, with R^2 of 0.84, RMSE and MAE of 3.63 g/kg^{-1} and 2.59 g/kg^{-1}, respectively, which exceeded all traditional machine learning and baseline deep learning models. However, under the condition of multi-period fusion data, the R^2 of MLP is 0.82, and the Transformer1 D is 0.79, which shows that the deep model can learn more features than traditional machine learning. For the FULL setting, the RMSE of Transformer1 D is 4.12g/kg^{-1}, and the MAE is 2.95 g/kg^{-1}, which is better than 1D-CNN and most traditional methods, indicating that attention-based spectral sequence modeling does have certain benefits in multi-period fusion scenarios. Nevertheless, HSRAN continues to maintain optimal performance across the three metrics, indicating that relying solely on spectral sequence modeling is still insufficient to fully characterize the issues of Spectral Redundancy and spatial heterogeneity in leaf nitrogen content estimation, while integrating spectral enhancement with regional association modeling frameworks can provide more stable feature representations.

Overall, HSRAN demonstrates good predictive stability across different growth periods and the entire growth period data under conditions of cross-year mixed samples. The results indicate that combining spectral enhancement with regional association modeling contributes to improving the accuracy of hyperspectral nitrogen content inversion and enhances the model's stable performance under the current data settings.

Table 3. Predictive Effects of Different Models on maize Leaves at Various Stages

Model	Jointing Stage			Silking Stage			Maturity stage			FULL Stage		
	R²	**RMSE**	**MAE**	**R²**	**RMSE**	**MAE**	**R²**	**RMSE**	**MAE**	**R²**	**RMSE**	**MAE**
RF	0.42	4.21	3.29	0.58	3.94	3.21	0.50	2.76	2.33	0.60	6.05	3.81
1DCNN	0.24	5.50	3.80	0.52	3.53	2.75	0.32	2.34	1.88	0.64	5.42	3.90
XGBoost	0.41	4.25	3.21	0.60	3.89	2.93	0.50	2.76	2.30	0.69	6.10	3.88
PLSR	0.35	4.47	3.37	0.43	4.62	3.60	0.35	3.14	2.58	0.74	4.81	3.59
SVR	0.35	4.48	3.34	0.30	5.12	3.67	0.44	2.93	2.37	0.77	4.57	2.99
Transformer1D	0.43	4.77	3.29	0.58	3.29	2.43	0.52	1.96	1.57	0.79	4.12	2.95
MLP	0.45	4.68	3.27	0.73	2.67	1.99	0.60	1.79	1.40	0.82	3.78	2.70
Ours(HSRAN)	**0.56**	**4.16**	**2.96**	**0.76**	**2.57**	**2.12**	**0.72**	**1.50**	**1.20**	**0.84**	**3.63**	**2.59**

3.2 Ablation Experiments

To systematically evaluate the contribution of each module to the model's performance, a series of ablation experiments were conducted by gradually introducing the SARE and CAGM

modules based on the baseline network, with the results shown in Table 4.

The baseline model achieved $R^2 = 0.80$, RMSE was 3.98 g/kg^{-1} and MAE was 2.72 g/kg^{-1} for the test set in the baseline model that only contained the basic structure of the feature extraction. This is a sign that the network can reflect the mapping relationship between spectral data and nitrogen content to a certain extent in the absence of explicit introduction of spectral modeling or regional modeling mechanism.

With the addition of the SARE module, there was a slight improvement in model performance (R^2 increased to 0.81, RMSE decreased to 3.90 g/kg^{-1}, and MAE decreased to 2.68 g/kg^{-1}). This result suggests that Spectral Adaptive Recalibration Encoder can suppress some redundant bands' interference with regression while moderately enhancing the spectral responses more correlated with nitrogen content, all while maintaining the overall structure.

When only the CAGM module is introduced, model performance improves further to R^2=0.82, RMSE=3.90 g/kg^{-1}, and MAE=2.60 g/kg^{-1}. Compared to the baseline, both R^2 and MAE have shown improvement, indicating that explicitly modeling the contextual relationships between regions in cases where leaves are composed of multiple areas, and aggregating them through adaptive weights, helps to mitigate the influence of local noise and spatial heterogeneity on overall predictions.

When both SARE and CAGM were integrated, the overall performance is optimal : R^2 increases to 0.84, RMSE and MAE decrease to 3.63 g/kg^{-1} and 2.59 g/kg^{-1}, respectively. Compared with the baseline model, R^2 is increased by about 4 percentage points. This shows that spectral recalibration and Context-Aware Gated Aggregation Module are functionally complementary-the former mainly enhances the distinguishability of spectral features, while the latter improves the stability and global consistency of region-level feature expression.

In summary, the results of ablation experiments show that spectral recalibration (SARE) and Context-Aware Gated Aggregation Module (CAGM) play a substantial role in improving the accuracy and robustness of maize leaf nitrogen content inversion.

Table 4. Ablation Experiment Table

Baseline	SARE	CAGM	Test-set performance		
			R^2	RMSE	MAE
✓	—	—	0.8	3.98	2.72
✓	✓	—	0.81	3.9	2.68
✓	—	✓	0.82	3.9	2.6
✓	✓	✓	**0.84**	**3.63**	**2.59**

3.3 Attention-Based Regional Importance Visualization

To examine the regional spectral features emphasized by HSRAN, the attention weights generated by CAGM were mapped back to the corresponding locations on the original leaf images using the coordinates recorded during regional sampling. For each sampled region, the attention weight represented its relative contribution to the leaf-level nitrogen prediction within the corresponding leaf sample.

The visualization results showed that regions near leaf veins, tips, and margins often received relatively higher attention weights, whereas some interveinal regions received lower weights. This pattern was observed across samples from different growth stages, suggesting that local spectra from these regions provided comparatively informative cues for nitrogen estimation.

The attention maps should be interpreted as model-dependent regional importance maps rather than direct measurements of local nitrogen concentration. Higher attention weights do not necessarily indicate higher absolute nitrogen concentrations in the corresponding regions. Instead, they indicate that the spectral features of these regions were more useful for the fitted model when estimating leaf-level nitrogen content. The observed spatial pattern is qualitatively consistent with potential differences in tissue structure, nitrogen transport, local senescence, and optical properties across leaf regions.

Figure 5. Attention-based regional importance maps of maize leaves at different growth stages. The attention weights generated by CAGM were mapped to the corresponding sampling locations on the original leaf images. Warmer colors indicate regions with relatively higher attention weights within each leaf sample.

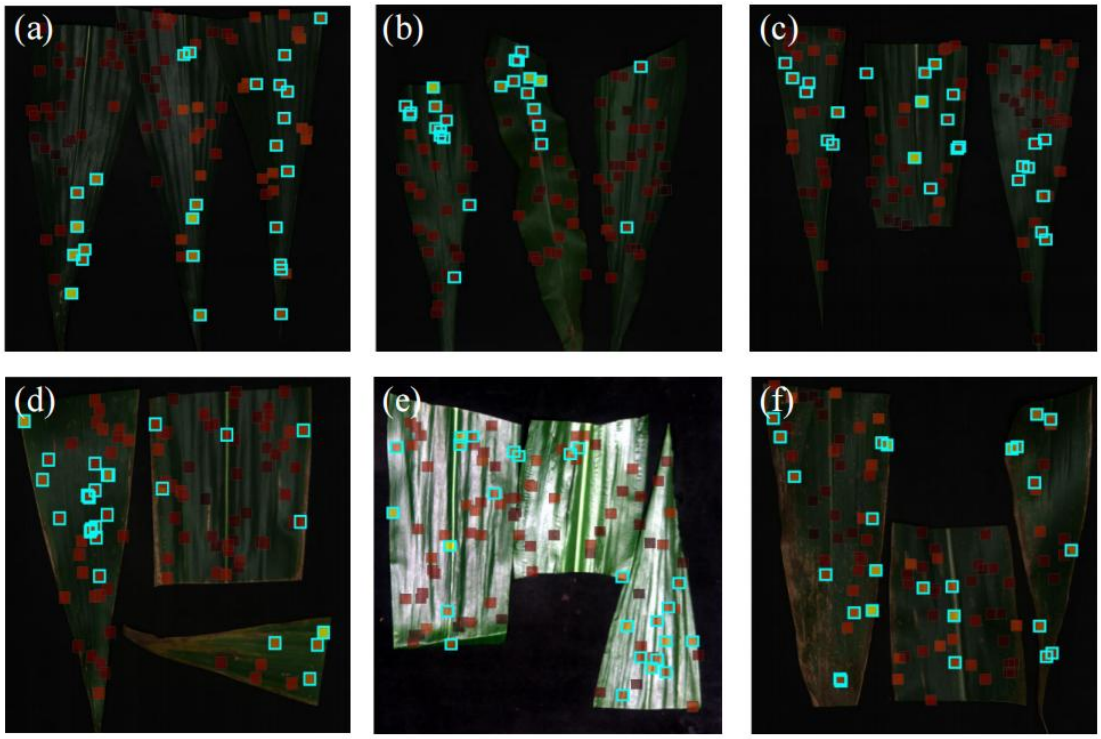


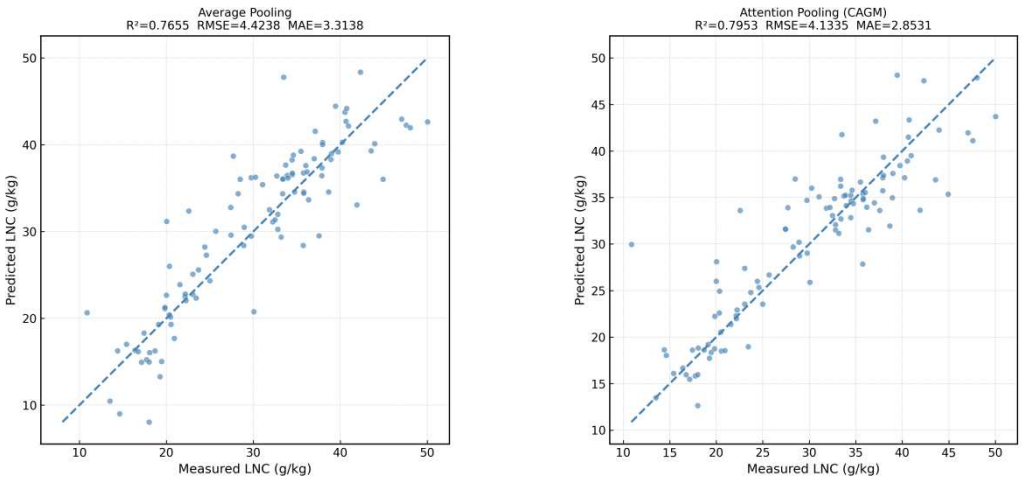
Figure 5 Regional Attention Distribution of the Model at Different Periods

3.4 Comparison of Regional Aggregation Strategies

To evaluate the effectiveness of adaptive regional aggregation, the proposed Context-Aware Gated Aggregation Module (CAGM) was compared with mean pooling. Mean pooling treats all regional spectral features as equally important and calculates the leaf-level representation by direct averaging. In contrast, CAGM first models contextual dependencies among regional features through multi-head self-attention and then performs gated attention pooling to assign adaptive weights to different regions.

As shown in Figure 6, CAGM achieved an R^2 of 0.7953, an RMSE of 4.1335 g/kg^{-1}, and an MAE of 2.8531 g/kg^{-1}. In contrast, mean pooling yielded an R^2 of 0.7655, an RMSE of 4.4238 g/kg^{-1}, and an MAE of 3.3138 g/kg^{-1}. The improved performance of CAGM indicates that treating all regional spectra equally is less effective than context-aware adaptive aggregation. This result suggests that different local leaf regions provide unequal spectral information for leaf-level nitrogen estimation, and that adaptive regional aggregation can better exploit these differences.

Figure 6. Comparison of mean pooling and context-aware gated aggregation for leaf-level nitrogen estimation. Mean pooling assigns equal importance to all regional spectral features, whereas CAGM performs contextual feature interaction and gated attention-based aggregation. Higher R^2 and lower RMSE and MAE indicate improved predictive performance.



(a) (b)

Figure 6: Comparison of mean pooling and context-aware gated aggregation for leaf-level nitrogen estimation

4. Discussion

4.1 Effects of Spectral–Regional Joint Modeling on Nitrogen Estimation

This study developed HSRAN to improve maize leaf nitrogen estimation by jointly modeling band-specific spectral responses and contextual relationships among regional spectral features. Across the jointing, silking, maturity, and pooled full-growth-period datasets, HSRAN outperformed the conventional machine-learning models, including PLSR, RF, XGBoost, and SVR, and achieved stronger overall performance than the deep-learning baselines, including 1D-CNN, MLP, and Transformer1D.

From the spectral perspective, hyperspectral data contain a large number of highly correlated adjacent bands, and different wavelengths may contribute unequally to nitrogen estimation. SARE addresses this issue through band-wise residual recalibration. Specifically, the model generates a band-wise gating profile, applies one-dimensional convolution to locally smooth the learned gate along the spectral dimension, and then performs residual modulation of the original spectrum. This design preserves the original spectral signal while allowing different wavelengths to receive different degrees of enhancement. The ablation results showed that adding SARE improved R^2 from 0.80 to 0.81 and reduced RMSE from 3.98 to 3.90 g/kg^{-1},, indicating that adaptive spectral recalibration improved regional spectral representation.

From the regional perspective, a maize leaf is composed of local regions with potentially different spectral responses because of differences in tissue structure, illumination, leaf veins, local senescence, and nitrogen transport. Treating each regional spectrum independently may ignore complementary information among regions, whereas direct averaging assumes that all regions contribute equally to leaf-level prediction. CAGM addresses this limitation by first

modeling contextual dependencies among regional features through multi-head self-attention and then aggregating the features using gated attention pooling. In the comparison of aggregation strategies, CAGM achieved an R^2 of 0.7953, an RMSE of 4.1335 g/kg^{-1},, and an MAE of 2.8531 g/kg^{-1},, whereas mean pooling achieved an R^2 of 0.7655, an RMSE of 4.4238 g/kg^{-1},, and an MAE of 3.3138 g/kg^{-1},. These results indicate that context-aware adaptive aggregation provides a more informative leaf-level representation than direct averaging of regional spectra.

The ablation experiments further showed that SARE and CAGM provided complementary benefits. SARE improved the representation of individual regional spectra, whereas CAGM improved the integration of multiple regional features. When both modules were included, HSRAN achieved the best overall performance on the pooled dataset, with an R^2 of 0.84, an RMSE of 3.63 g/kg^{-1},, and an MAE of 2.59 g/kg^{-1},.

The performance advantage of HSRAN varied across growth stages. At the jointing and maturity stages, HSRAN achieved the largest improvement in R^2 over MLP, with increases of approximately 0.11 and 0.12, respectively. At the silking stage, HSRAN achieved the highest R^2 and the lowest RMSE, whereas MLP achieved a slightly lower MAE. These results suggest that HSRAN provided the most balanced predictive performance across growth stages. The relatively larger gains at the jointing and maturity stages may be associated with stronger variation in local leaf spectral responses during rapid vegetative development and nitrogen remobilization, although this interpretation should be validated using additional physiological measurements.

4.2 Interpretation of Attention-Based Regional Importance

The attention-based regional importance maps provide an intuitive interpretation of how HSRAN aggregates local spectral information for leaf-level nitrogen estimation. In this study, the attention weights generated by CAGM were mapped back to the corresponding locations on the original leaf images using the coordinates recorded during regional sampling. Therefore, each highlighted region represents a local spectral instance that received a relatively higher contribution weight during the prediction of nitrogen content for the corresponding leaf sample.

The visualization results showed that regions near leaf veins, leaf tips, and margins often received relatively higher attention weights, whereas some interveinal regions received lower weights. This pattern was observed in representative samples from different growth stages, indicating that the local spectra from these regions may provide more informative cues for leaf-level nitrogen estimation. From a physiological perspective, this observation is plausible. Leaf veins are important pathways for water and nutrient transport, and the surrounding tissues differ from interveinal mesophyll in anatomical structure, water status, and optical properties. In particular, vascular bundles and thick-walled tissues may produce distinctive near-infrared spectral responses [18]. In addition, leaf tips and margins may exhibit relatively strong spectral variation because of local senescence, transpiration gradients, tissue-structure differences, and variable light exposure. Previous hyperspectral studies have also reported that maize leaf veins and mesophyll regions exhibit different spectral responses under nitrogen stress [19].

Nevertheless, the attention maps should be interpreted cautiously. Higher attention weights do not necessarily indicate higher absolute nitrogen concentrations in the corresponding regions. Instead, they indicate that the local spectral features of these regions were more useful for the fitted model when estimating leaf-level nitrogen content. Because regional coordinates were used only for post-hoc visualization and were not directly incorporated into the model input, the attention maps should not be interpreted as evidence that HSRAN explicitly learned leaf geometry

or spatial topology.

The comparison between CAGM and mean pooling further supports the value of adaptive regional aggregation. Mean pooling assumes that all regional spectra contribute equally to leaf-level prediction, whereas CAGM models contextual dependencies among regional features through multi-head self-attention and performs gated attention-based aggregation. CAGM achieved an R^2 of 0.7953, an RMSE of 4.1335 g/kg^{-1},, and an MAE of 2.8531 g/kg^{-1},, whereas mean pooling achieved an R^2 of 0.7655, an RMSE of 4.4238 g/kg^{-1}, and an MAE of 3.3138 g/kg^{-1},. These results indicate that local regions within a leaf provide unequal spectral information for the fitted prediction model and that context-aware adaptive aggregation can construct a more informative leaf-level representation than direct averaging of regional spectra.

4.3 Research Limitations and Future Work

Although HSRAN achieved promising performance for maize leaf nitrogen estimation, several limitations should be acknowledged.

First, cross-year generalization was constrained by distribution shifts between the two experimental years. In the year-independent evaluation, the model achieved an R^2 of 0.81 during within-year cross-validation on the 2024 dataset, whereas its performance declined to an R^2 of 0.47 when directly transferred to the 2025 dataset. As shown in Figure 7, samples from the two years exhibited visible separation in principal component space, indicating differences in their spectral distributions. The maximum mean discrepancy values across growth stages ranged from 0.129 to 0.179, further suggesting spectral covariate shifts. In addition, the nitrogen-content distributions differed significantly between years, with a Kolmogorov–Smirnov statistic of 0.211 and a p-value below 0.001. The difference was particularly evident at the jointing stage, where the mean nitrogen concentration differed between 2024 and 2025. These spectral and label shifts likely limited direct cross-year transferability. The current SARE and CAGM modules focus on spectral recalibration and contextual aggregation of regional spectral features within the source domain and do not explicitly address domain alignment. Future work should investigate domain-adaptation methods, few-shot calibration strategies, and transfer-learning approaches to improve model robustness across years and field conditions.

Second, the biological interpretation of regional attention requires further validation. The attention maps suggested that regions near leaf veins, tips, and margins often provided informative spectral cues for model prediction. However, attention weights are model-derived importance estimates and cannot be directly equated with local nitrogen concentration or physiological activity. Future studies should combine micro-area chemical measurements, tissue-level nitrogen analysis, chlorophyll measurements, or high-resolution physiological imaging to determine whether high-attention regions also exhibit stronger nitrogen-related biochemical variation.

Third, the representativeness of the experimental data remains limited. The current dataset was collected at a single experimental site from one silage maize cultivar, Jinboshi 825, over two growing seasons. Although the samples covered different nitrogen treatments and three growth stages, differences in climate, soil properties, cultivar characteristics, and crop-management practices may alter leaf spectral responses. Future experiments should include multiple cultivars, sites, seasons, and management conditions to provide a more rigorous assessment of model generalizability.

Fourth, the present study was conducted at the leaf scale under controlled imaging conditions. Translating the proposed framework to canopy-scale or field-scale monitoring remains

challenging because field observations are affected by variable illumination, canopy occlusion, leaf-angle variation, background interference, and mixed-pixel effects. Future work could combine the proposed regional modeling strategy with canopy radiative-transfer models, UAV hyperspectral imagery, or multi-source remote sensing data to explore pathways for scaling leaf-level nitrogen estimation to field applications.

Finally, the current framework still relies on full-spectrum hyperspectral measurements and the construction of multiple local regions for each leaf, which may limit computational efficiency and practical deployment. Future work should investigate lightweight network architectures and assess whether wavelengths receiving consistently high recalibration weights across samples and repeated runs can serve as candidates for simplified spectral-input models. Such efforts may help reduce acquisition and inference costs while preserving reliable nitrogen-estimation performance.

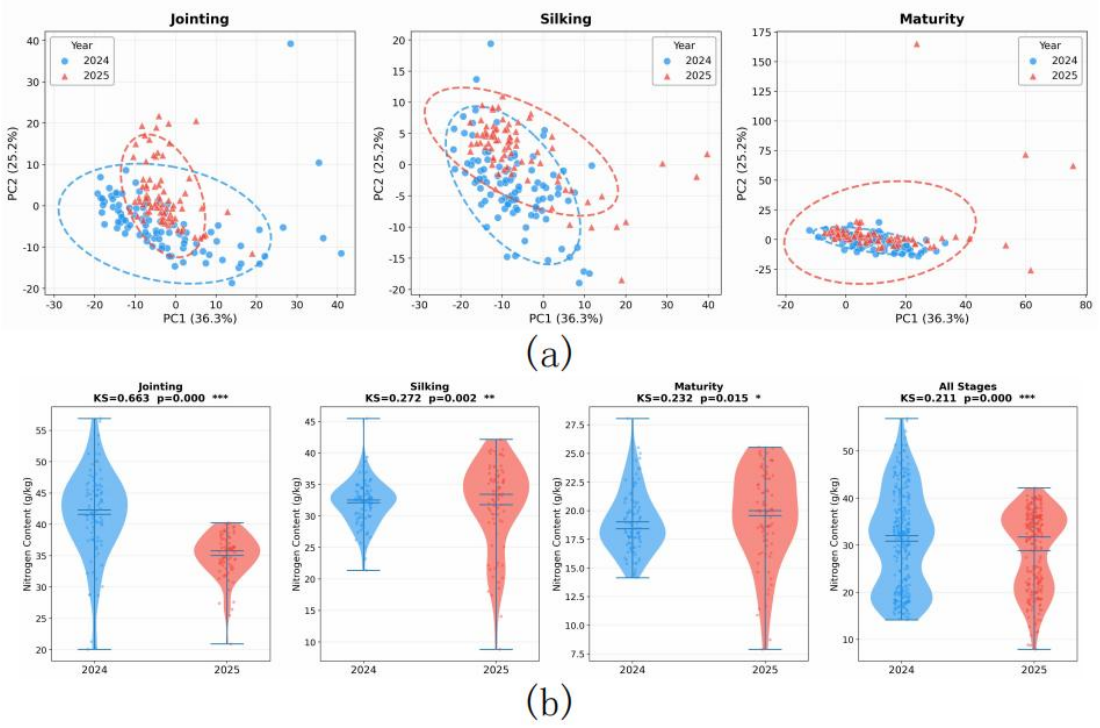


Figure 7 Distribution difference analysis between the samples from 2024 and 2025

5. Conclusion

This paper proposed the HSRAN model to address the concurrent challenges of spectral redundancy and spatial heterogeneity in hyperspectral nitrogen content inversion for maize leaves.

The model uses the Spectral Adaptive Recalibration Encoder (SARE) module to enhance the response of the nitrogen-sensitive band, and uses the Context-Aware Gated Aggregation Module module (CAGM) to model the interdependence between different regions of the leaf, so as to directly predict the nitrogen content at the leaf scale from the regional spectrum.

On field experiment data spanning 2024 to 2025, HSRAN achieved $R^2 = 0.84$ and RMSE = 3.63 $g \cdot kg^{-1}$ in the comprehensive data set of the whole growth period, which exceeded the traditional methods such as PLSR, RF, XGBoost and SVR, and was also superior to deep learning baseline models such as 1D-CNN, MLP and Transformer1D. Ablation experiments confirm that SARE and CAGM have complementary effects, which improve the prediction accuracy from the two dimensions of spectral enhancement and regional aggregation, respectively. Regional

attention visualization showed that the model gave higher weights to areas such as veins, leaf tips, and leaf margins, which was consistent with the physiological laws of nitrogen transport and redistribution, and also enhanced the interpretability of the model.

This study shows that spectral-regional joint modeling is an effective way to improve the inversion accuracy of leaf-scale nitrogen content. Subsequent work will verify the generalization ability of the model under multi-variety and multi-environment conditions, and introduce domain adaptation strategies to promote its application in the actual field.The HSRAN framework may also be adapted for estimating other foliar biochemical traits using hyperspectral data.

Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests (mandatory)

The authors declare no competing interests.

Author contributions

M.Z. conducted the experiments and drafted the manuscript. F.Q. and X.L. performed the experiments and critically reviewed the manuscript. L.Z., C.W., P.G. collected and processed the data, and participated in drafting the manuscript. All authors approved the final manuscript as submitted and agreed to be responsible for all aspects of the work.

Funding Declaration

This work was supported by the Xing'an League Science and Technology Cooperation Project (Grant No. MBHZ2025023).

References:

[1] Xingyu ZHANG, Yue ZHANG, Chenzhen XIA, Xiaoyan ZHANG, Yuxi LI, Xiaoyu LI. Estimation of Leaf Nitrogen Content in Maize Based on UAV Hyperspectral Image[J]. Remote Sensing Technology and Application, 2024, 39(4): 927-939 https://doi.org/10.11873/j.issn.1004-0323.2024.4.0927

[2] Wang Yuxi, Han Nana, Zhou Qingyun, Zhang Baozhong, Peng Zhigong. Study on the Estimation Model of Summer Maize Leaf SPAD Value Based on Hyperspectrum[J]. Journal of Tianjin Agricultural University, 2022, 29(3): 1-8.

[3] WEI Xiayong, HUANG Xi, BO Liyuan, et al. Vertical Distribution of Nitrogen Content in Spring Maize Leaves and Its Hyperspectral Inversion Under Different Film Mulching and Irrigation Levels in Northwest Arid Region[J]. Journal of China Agricultural University, 2022, 27(11). DOI: 10.11841/j.issn.1007-4333.2022.11.02.

[4] YUN BinYuan, XIE TieNa, LI Hong, YUE Xiang, LÜ MingYue, WANG JiaQi, JIA Biao. Nitrogen Nutrition Estimation of Maize Based on UAV Spectrum and Texture Information[J]. Scientia Agricultura Sinica, 2024, 57(16): 3154-3170.

[5] ZHAI Yong-quan, WEI Xue, YUN Bin-yuan, MA Jian-zhen, JIA Biao. Dynamic Diagnosis of Nitrogen Nutrition in Maize under Drip Irrigation Condition Based on UAV Image Parameters[J]. Chinese Journal of Agrometeorology, 2022, 43(04): 308-320.

[6] WANG Jingyong, ZHANG Mingzhen, LING Huarong, WANG Ziting, GAI Jingyao. A Hyperspectral Image-Based Method for Estimating Water and Chlorophyll Contents in Maize Leaves under Drought Stress[J]. Smart Agriculture, 2023, 5(3): 142-153.

[7] ZHOU ZhiHui, GU XiaoBo, CHENG ZhiKai, CHANG Tian, ZHAO TongTong, WANG YuMing, DU YaDan. Inversion of Chlorophyll Content of Film-Mulched Maize Based on Image Segmentation[J]. Scientia Agricultura Sinica, 2024, 57(6): 1066-1079.

[8] Li Xuan, Song Kaishan, Liu Jiping, et al. Study on chlorophyll content inversion method in maize leaves based on remote sensing fusion data[J]. Science Technology and Engineering, 2025, 25(11): 4428-4437.

[9] Zhuolin LI, Jinguo YUAN, Ziyan YANG, Wenchao WANG, Yancui LI, Bohan LIU. Comparative Study on Chlorophyll Content Inversion of Summer Maize Based on PROSAIL Model and Sentinel-2A Image[J]. Remote Sensing Technology and Application, 2025, 40(3): 621-635 https://doi.org/10.11873/j.issn.1004-0323.2025.3.0621

[10] SUN Fafu, LAI Ning, GENG Qinglong, LI Yongfu, LV Caixia, XIN Huinan, LI Na, CHEN Shuhuang. Estimation of nitrogen contentration in winter wheat leaves based on hyperspectral images of UAV[J]. Arid Zone Research, 2024, 41(6): 1069-1078.

[11] Xu Tongyu, Jin Zhongyu, Guo Zhonghui, Yang Liu, Bai Juchi, Feng Shuai, Yu Fenghua. Simultaneous inversion method of nitrogen and phosphorus contents in rice leaves using CARS-RUN-ELM algorithm[J]. Transactions of the Chinese Society of Agricultural Engineering, 2022, 38(10): 148-155. DOI: 10.11975/j.issn.1002-6819.2022.10.018

[12] Xinhui Guo, Xingxing QIAO, Yu ZHAO, Chao WANG, Meichen FENG, Lujie XIAO, Xiaoyan SONG, Meijun ZHANG, Wude YANG, Guangxin LI. Hyperspectral Remote Sensing Monitoring of Leaf Nitrogen Content in Winter Wheat under Different Nitrogen Application Levels[J]. Journal of Shanxi Agricultural Science, 2025, 53(05): 92-100 DOI:10.26942/j.cnki.issn.1002-2481.2025.05.11

[13] TAN XianMing, ZHANG JiaWei, WANG ZhongLin, CHEN JunXu, YANG Feng, YANG WenYu. Prediction of Maize Yield in Relay Strip Intercropping Under Different Water and Nitrogen Conditions Based on PLS[J]. Scientia Agricultura Sinica, 2022, 55(6): 1127-1138.

[14] ZHAO Jin-long, ZHANG Xue-yi, LI Yang. Hyperspectral Remote Sensing of Crop Information Based on Machine Learning Algorithm: State of the Art and Beyond[J]. Chinese Journal of Agrometeorology, 2023, 44(11): 1057-1071.

[15] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

[16] Ilse M, Tomczak J, Welling M. Attention-based deep multiple instance learning[C]//International conference on machine learning. PMLR, 2018: 2127-2136.

[17] Mulla D J. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps[J]. Biosystems engineering, 2013, 114(4): 358-371.

[18] Slaton M R, Raymond Hunt Jr E, Smith W K. Estimating near-infrared leaf reflectance from leaf structural characteristics[J]. American journal of botany, 2001, 88(2): 278-284.

[19] Song Z, Wei X, Zhang J, et al. Spatial-spectral feature mining in hyperspectral corn leaf venation structure and its application in nitrogen content estimation[J]. Computers and Electronics in Agriculture, 2024, 227: 109495.

[20] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.