Artificial Intelligence and Digital Technology

Article

Research on Hyperspectral Imaging Detection Method of Nitrogen in Facility-Grown Lettuce

Yixue Zhang [1,*], Jingbo Zhi [2], Jialiang Zheng [3], Zhaowei Li [3] and Tiezhu Li [3]

[1] Basic Engineering Training Center, Jiangsu University, Zhenjiang, Jiangsu, China
[2] The University of Sydney, Sydney, NSW, Australia
[3] School of Agricultural Engineering, Jiangsu University, Zhenjiang, Jiangsu, China
* Correspondence: Yixue Zhang, Basic Engineering Training Center, Jiangsu University, Zhenjiang, Jiangsu, China

Abstract: Rapid non-destructive detection of crop nutrition serves as a crucial basis for water-fertilizer management and environmental regulation. This paper proposes a rapid nitrogen detection method for lettuce based on hyperspectral imaging technology. Hyperspectral image data of lettuce samples were acquired, processed using the SG smoothing algorithm, and analyzed with the RF algorithm to extract nitrogen-specific wavelengths. Finally, a KELM model under the RBF-Kernel function was established to predict lettuce nitrogen content. Results demonstrate that the KELM prediction model based on RF feature extraction achieves excellent performance, with an R^2 value exceeding 0.95 and RMSE below 0.27. This method provides scientific support for water and fertilizer irrigation decisions based on crop nitrogen requirements.

Keywords: Protected lettuce; Crop nutrition; Hyperspectral imaging; Feature extraction; Rapid detection

Published: 31 December 2025

1. Introduction

Lettuce, as one of the most important protected leafy vegetable, possesses high nutritional value and strong market demand. Nitrogen is a core nutrient for the growth and development of lettuce and plays a decisive role in photosynthesis, chlorophyll synthesis, and amino acid metabolism. Appropriate nitrogen supply can increase lettuce yield by 20-40%, whereas nitrogen deficiency leads to leaf yellowing and a sharp decline in biomass; excessive nitrogen, on the other hand, causes nitrate accumulation (with the EU standard limit being 2500-4500 mg/kg), posing a threat to food safety [1]. To address the need for green, efficient, and intelligent management in protected agriculture, breakthroughs in rapid and non-destructive nitrogen detection technology have become a critical technical bottleneck for efficient water and fertilizer management, as well as ensuring yield and quality [2,3].

Traditional manual judgment methods require years of accumulated experience and are highly subjective, resulting in relatively high error rates [4]. Chemical analysis methods can detect nitrogen content more accurately, but they are time-consuming and require destructive sampling [5]. In the field of crop nutrient detection, hyperspectral imaging (HSI) technology can perceive differences in the internal composition and nutrient macromolecule spectra of crops. Due to its rapid, non-destructive, and multi-parameter synchronous analysis capabilities, it has become a forefront approach in crop nutrient monitoring [6,7]. By capturing spectral-spatial information in the 900-2500 nm range, it is possible to quantify the relationship between leaf biochemical components and optical properties. Compared to traditional methods, its advantages lie in achieving spectral integration [8], efficient acquisition of multidimensional features [9], and spatial visualization of nitrogen [10]. It overcomes the limitations of traditional spectrometer point-source sampling, the lack of nutrient response in RGB imaging, and the insufficiency of spectral information in 3D imaging [11,12].

Nitrogenous compounds in lettuce leaves, such as proteins, chlorophyll and lutein, exhibit characteristic absorption peaks at specific wavelengths of 760 nm, 850 nm, 1040 nm separately. By establishing machine learning models (such as PLSR and CNN) that correlate hyperspectral images of these characteristic spectral bands with nitrogen concentration, quantitative prediction of crop nitrogen nutrition can be achieved [13]. Based on this, this study proposes a rapid lettuce nitrogen detection method based on hyperspectral imaging. By integrating a nitrogen feature band selection algorithm with deep learning models, the method enables rapid and accurate analysis of lettuce nitrogen content and the generation of spatial distribution maps of nitrogen nutrition. This technology overcomes the limitations of traditional methods, which are often subjective and time-consuming, and provides a scientific basis for efficient, intelligent water and fertilizer management in controlled environment agriculture, promoting the green, efficient, and intelligent development of facility agriculture [14].

2. Materials and methods

2.1. Experimental samples cultivation

The experimental sample is Italian year-round bolting lettuce potted and cultivated for 40 days in the Venlo type multi span greenhouse in the Key Laboratory of modern agricultural equipment and technology of Jiangsu University, Ministry of Education (Figure 1). The nutrient solution was selected from the improved Yamasaki formula [15] and configured according to different nutrient gradients (20%, 60%, 100% and 150%). 100 lettuce nitrogen samples were divided into four groups, 25 plants in each group. The 100% nitrogen group was used as the standard nutrient solution control group. In the experiment, from the lettuce samples with obvious differences between the growth of lettuce with different nutritional gradients, two pieces of lettuce leaves with good shape and intact leaves were selected from each sample. Each group of lettuce samples selected 50 pieces of lettuce leaves to be sealed and preserved, and the hyperspectral information was collected quickly. Four groups of samples were sampled in turn and then detected quickly to maintain the living characteristics.

Figure 1. Lettuce cultivation site.

2.2. Hyperspectral Imaging System

The experiment was conducted using the HSI-NIR push-broom hyperspectral imaging system developed by Shanghai Wuling Optoelectronics Technology Co., Ltd. (Figure 2), and the collected hyperspectral data were analyzed using the accompanying analysis software (Figure 3). The system has a wavelength range of 871.607-1766.322 nm, with a resolution of 3.5 nm.

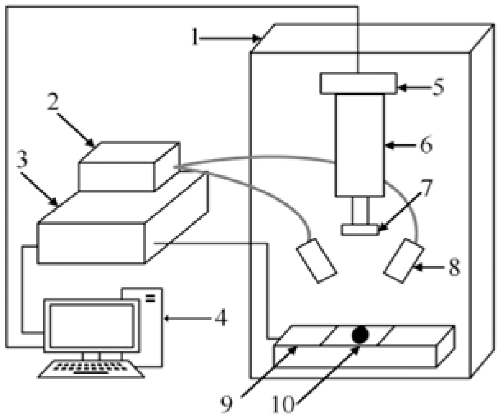

Figure 2. Experimental Platform. 1- Collection Box; 2- Fill Light; 3- Control Box; 4- Computer; 5- Near-Infrared Camera;.6- Spectrometer; 7- Lens; 8- Light Guide; 9- Sample Stage; 10- Sample

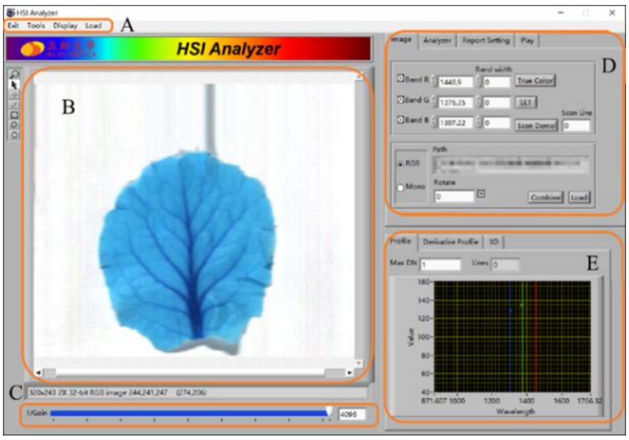

Figure 3. Hyperspectral Data Processing.

2.3. Determination of Leaf Nitrogen Content

In this experiment, the Kjeldahl method was employed to determine the reference nitrogen content of lettuce leaves. The measured nitrogen values were used as standard data to compare with hyperspectral measurement results, thereby enabling an objective evaluation of the accuracy and reliability of the hyperspectral quantitative analysis model. This comparative approach provides a solid basis for assessing the effectiveness of hyperspectral techniques in nitrogen content estimation.

After hyperspectral data acquisition, lettuce leaf samples that remained fresh were collected for chemical analysis. First, the samples were subjected to freeze-drying treatment to remove moisture while preserving the original chemical composition of the leaves. This step helps to minimize the influence of water content on subsequent nitrogen determination and ensures the stability of the sample material.

After drying, the samples were ground into fine powder using a ball mill. The homogenization of sample particles improves the consistency of chemical reactions during digestion and enhances the reproducibility of nitrogen measurements. Subsequently, the powdered samples were transferred to a digestion furnace, where they were prepared into test solutions through a digestion process. This procedure ensures that nitrogen-containing compounds in the samples are fully decomposed and converted into detectable forms.

Finally, the nitrogen content of each sample was determined using the Kjeldahl method in combination with a flow analyzer. Through this process, the nitrogen content percentage of each lettuce leaf sample was accurately obtained. As shown in Table 1, the measured nitrogen content values provide reliable reference data for model calibration and validation. These results serve as an essential benchmark for evaluating the predictive performance of hyperspectral models developed in this study.

Table 1. Actual Nitrogen Content.

Sample gradient	Minimum(%)	Maximum(%)	Average(%)	Median(%)	Standard deviation(%)
20 %	1.2046	2.2759	1.7357	1.7382	0.3556
60 %	2.0475	3.2587	2.6423	2.6021	0.3971
100 %	2.8908	4.2683	3.6264	3.5894	0.4039
150 %	3.8489	5.1172	4.5558	4.5113	0.4047

3. Results and Analysis

3.1. Data Preprocessing

The raw data of samples collected by the hyperspectral imaging system are shown in Figure 4.

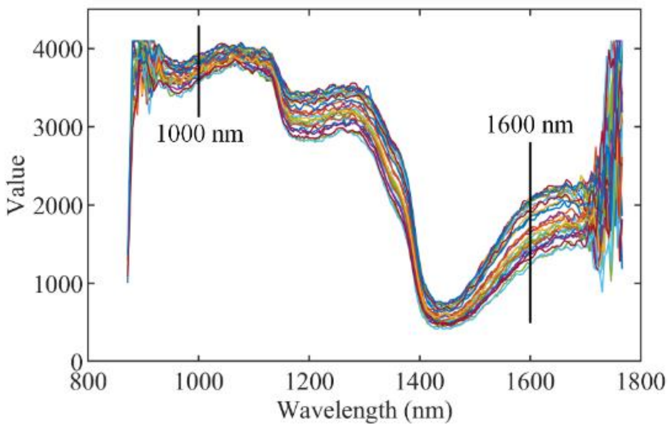

Figure 4. Raw Data and Data Range Selection Schematic Diagram.

In order to eliminate the influence of edge noise, the effective spectral range was truncated to 1000 - 1600 nm, and 180 characteristic wavelength points were evenly extracted in this range.

After determining the spectral range, the experiment needs to implement savitzky Golay (SG) smoothing pretreatment on the original data. Through parameter optimization verification, it is determined that the optimal window width is 7 points (polynomial order is fixed to 2), and the performance of the model is optimal: the determination coefficient R^2 is 0.9889, and the root mean square error RMSE is reduced to 4.4987 × 10-9 (significantly lower than other parameter combinations). The same batch of sample data were further corrected by MSC.

Figure 5 shows the comparison of the data before and after preprocessing. Based on the preprocessing results, SPXY algorithm is used to divide the data set, and the goodness of fit index is R^2=0.9107, RMSE=0.1351.

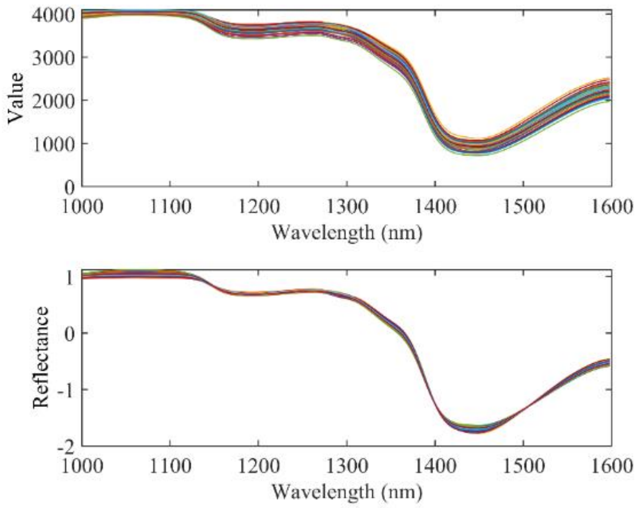

Figure 5. Comparison of data before and after pretreatment.

3.2. Characteristic Wavelength Selection

If the preprocessed spectral data is directly used for full-spectrum modeling, the high dimensionality and strong correlation among adjacent bands may easily lead to overfitting, which in turn reduces the generalization performance of the model. In hyperspectral analysis, redundant spectral information not only increases computational burden but also interferes with the effective extraction of key information related to target variables. Therefore, it is necessary to reduce the data dimension through characteristic wavelength selection in order to improve model stability and prediction reliability.

In this study, the Random Frog (RF) algorithm was applied to screen key wavelength variables from hyperspectral data. This method evaluates the contribution of each spectral variable through repeated random sampling and probability statistics, enabling the identification of wavelengths that are most closely related to nitrogen content. By selecting representative wavelength variables, the accuracy and robustness of the quantitative model can be effectively improved.

As shown in Figure 6, the RF algorithm was implemented to analyze the importance distribution of spectral variables. Based on the RF algorithm, 180 spectral variables within the wavelength range of 1000–1600 nm were initially extracted for analysis. This spectral interval contains abundant information related to nitrogen content and is suitable for subsequent feature screening.

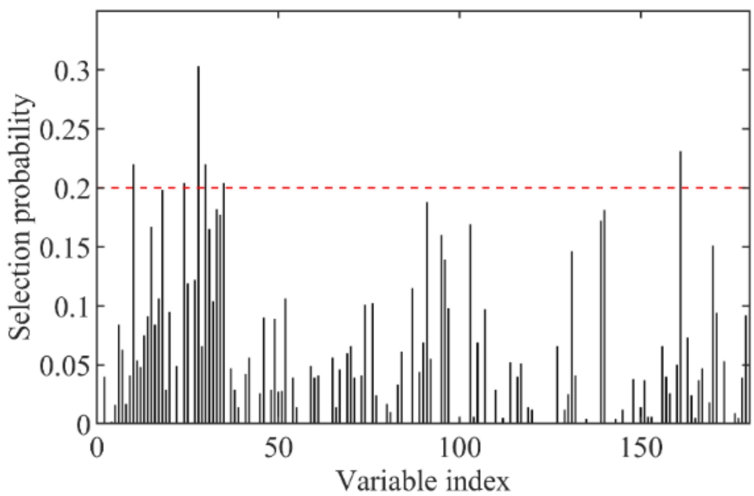

Figure 6. The results of RF operation.

To reduce the influence of random interference and ensure the stability of variable selection results, the RF algorithm was executed for 1000 iterations. During this process, the importance probability of each spectral variable was calculated. A higher selection probability indicates a greater contribution of the corresponding wavelength to nitrogen content estimation. By statistically analyzing the selection frequency of each variable, the relative importance of spectral bands was quantitatively evaluated.

A significance threshold of 0.2 was set to further refine the selection results. Variables with importance probabilities exceeding this threshold were retained as characteristic wavelengths. Finally, six key wavelength variables were selected, as summarized in Table 2. This result demonstrates that the RF algorithm can effectively reduce spectral dimensionality while preserving critical information. The distribution of the selected characteristic wavelengths along the spectral curve is illustrated in Figure 6, which shows that the retained wavelengths are mainly concentrated in regions with strong spectral response characteristics.

Table 2. Feature extraction results.

Algorithm	Number of variables	Number	Characteristic wavelength (nm)
RF	6	10,24,28,30,35,161	1035.136,1086.922,1101.293,1108.413,1126.031,1533.281

Overall, the characteristic wavelength selection process based on the RF algorithm provides a reliable foundation for subsequent quantitative modeling. By reducing redundant spectral information and highlighting key wavelength features, the constructed model can achieve improved prediction accuracy and enhanced generalization performance.

4. Discussion

In this study, the Random Frog (RF) algorithm was employed to optimize the feature variables of hyperspectral images, and a spectral quantitative analysis model for nitrogen content in lettuce was subsequently established. The purpose of applying feature optimization was to reduce spectral redundancy and enhance the relevance between input variables and nitrogen content. On this basis, the Kernel Extreme Learning Machine (KELM) was selected as the modeling algorithm. Due to its ability to efficiently handle nonlinear relationships and high-dimensional data, this model demonstrates clear advantages in hyperspectral quantitative analysis.

The Extreme Learning Machine (ELM) generates the input layer weights and hidden layer biases randomly during the initialization phase and keeps them fixed throughout the entire training process. This mechanism effectively avoids the time-consuming iterative parameter adjustment required by traditional gradient-based learning algorithms, thereby significantly improving training efficiency. As a result, ELM is particularly suitable for applications involving large-scale spectral data and repeated modeling processes.

In the stage of solving the output weights, ELM does not rely on iterative optimization strategies. Instead, it directly computes the minimum-norm least-squares solution based on the Moore–Penrose generalized inverse theory. This analytical solution simplifies the training procedure and reduces computational complexity. More importantly, this approach ensures that the model reaches a unique optimal solution under the given conditions, effectively suppressing the risk of overfitting and enhancing the generalization capability of the model when applied to unknown samples.

On the basis of ELM, the Kernel Extreme Learning Machine (KELM) introduces a kernel-based nonlinear mapping mechanism. By predefining a kernel function and its corresponding parameters, the original input space is implicitly mapped into a high-dimensional feature space. In this space, complex nonlinear relationships between spectral variables and nitrogen content can be more effectively captured. By predefining a kernel function, such as the radial basis function (RBF) or Sigmoid function, the hidden layer output matrix is uniquely determined by the kernel function, eliminating the need to explicitly construct hidden layer nodes.

This improvement greatly reduces model complexity. The training process of KELM only requires a single-step solution of the output layer weights, while retaining the global optimization characteristics inherited from ELM. Consequently, KELM achieves a balance between computational efficiency and enhanced generalization performance, making it well suited for hyperspectral quantitative modeling tasks.

In this study, the KELM model adopts the RBF kernel function. The regularization coefficient C was set to 100, and the kernel parameter S was set to 10. These parameter settings aim to achieve an appropriate balance between fitting accuracy and model stability. The detailed modeling performance results are presented in Table 3. As shown in Table 3, when the KELM model is constructed using spectral variables selected by the Random Frog (RF) algorithm, the coefficient of determination exceeds 0.95. This result indicates a strong consistency between the predicted nitrogen content values and the reference measurements. The modeling results demonstrate that the proposed RF–KELM approach is reasonable and effective for nitrogen content estimation based on hyperspectral data.

Table 3. KELM modeling.

Characteristic Selection Method	Number of variables	RC2	RMSEC	RP2	RMSEP
RF	6	0.9534	0.2427	0.9541	0.2631

5. Conclusion

This paper proposes a method for rapid detection of crop nutrition based on hyperspectral imaging. Data preprocessing and analysis were conducted using various algorithms such as SG smoothing, SNV, and SPXY, followed by the extraction and optimization of nitrogen characteristic wavelengths using the RF algorithm. On this basis, KELM models with RBF-Kernel functions were established to achieve rapid evaluation of nitrogen nutrition in lettuce. The results indicate that the nitrogen detection model established using the Random Frog (RF) algorithm performed excellently, with a coefficient of determination $R^2 > 0.95$ and RMSE less than 0.27, achieving high-precision and rapid analysis of nitrogen nutrition. Based on hyperspectral characteristic images of lettuce nitrogen, the spatial distribution of nitrogen content in lettuce was visualized by setting the threshold of nitrogen content feature spectra, providing an intuitive reference for evaluating lettuce nutritional levels and water-fertilizer supply requirements.

Therefore, the results of this paper have strong guiding significance and reference value for the rapid detection of crop nutrition and the decision-making of water-fertilizer irrigation based on crop demand.

Funding: This paper was funded by National Key Research and Development Program of China (No.2022YFD2002302); Jiangsu Province Industrial Forward-looking Technology Research and Development Project (No.BE2023017); National Key Research and Development Program for Young Scientists (No.2022YFD2000200).

References

1. T. Wu, H. Zhang, Y. Xiao, Z. Tang, F. Yang, H. Qi, and C. Zhang, "Quantification of multiple elements in Anji white tea using hyperspectral imaging combined with machine learning regression," Journal of Food Composition and Analysis, vol. 142, p. 107520, 2025. doi: 10.1016/j.jfca.2025.107520

2. J. Hao, Y. Zhang, Y. Zhang, and L. Wu, "Prediction of antioxidant enzyme activity in tomato leaves based on microhyperspectral imaging technique," Optics & Laser Technology, vol. 179, p. 111292, 2024. doi: 10.1016/j.optlastec.2024.111292
3. M. Mahmoodi-Eshkaftaki, M. Mahbod, and H. R. Ghenaatian, "Non-destructive estimation of biomass characteristics: Combining hyperspectral imaging data with neural networks," Renewable Energy, vol. 224, p. 120137, 2024. doi: 10.1016/j.renene.2024.120137
4. G. A. O. Jingxia, P. E. I. Hongxia, Z. H. A. N. G. Qingyun, X. I. Weina, and X. I. E. Hua, "The effect of water and fertilizer coupling on the growth development and yield of day lily," Journal of Irrigation & Drainage, 2024.
5. H. Y. Gao, H. P. Mao, and X. D. Zhang, "Measurement of nitrogen content in lettuce canopy using spectroscopy combined with BiPLS-GA-SPA and elm," Guang pu xue yu Guang pu fen xi= Guang pu, vol. 36, no. 2, pp. 491-495, 2016.
6. I. Orrillo, J. P. Cruz-Tirado, A. Cardenas, M. Oruna, A. Carnero, D. F. Barbin, and R. Siche, "Hyperspectral imaging as a powerful tool for identification of papaya seeds in black pepper," Food Control, vol. 101, pp. 45-52, 2019.
7. H. Luo, X. Chai, and C. Chen, "Based on multi-scale hyperspectral near ground remote to sensing the quality of Southern Xinjiang jujube," In Fifth Symposium on Novel Optoelectronic Detection Technology and Application, March, 2019, pp. 612-618.
8. H. Feng, Y. Chen, J. Song, B. Lu, C. Shu, J. Qiao, and W. Yang, "Maturity classification of rapeseed using hyperspectral image combined with machine learning," Plant Phenomics, vol. 6, p. 0139, 2024. doi: 10.34133/plantphenomics.0139
9. S. Vignati, A. Tugnolo, V. Giovenzana, A. Pampuri, A. Casson, R. Guidetti, and R. Beghi, "Hyperspectral imaging for fresh-cut fruit and vegetable quality assessment: Basic concepts and applications," Applied Sciences, vol. 13, no. 17, p. 9740, 2023. doi: 10.3390/app13179740
10. A. Femenias, F. Gatius, A. J. Ramos, V. Sanchis, and S. Marín, "Standardisation of near infrared hyperspectral imaging for quantification and classification of DON contaminated wheat samples," Food Control, vol. 111, p. 107074, 2020. doi: 10.1016/j.foodcont.2019.107074
11. J. C. Yang, M. Everett, C. Buehler, and L. McMillan, "A real-time distributed light field camera," Rendering Techniques, vol. 2002, no. 77-86, p. 2, 2002.
12. G. Z. Gu ZhengMin, L. Z. Li ZhenFeng, S. F. Song FeiHu, Z. J. Zhang JunSheng, and Z. W. Zhuang Wei, "Investigation of measurement method of soybean canopy leaf area based on light field camera," 2018.
13. N. S. Gruda, J. Dong, and X. Li, "From salinity to nutrient-rich vegetables: Strategies for quality enhancement in protected cultivation," Critical Reviews in Plant Sciences, vol. 43, no. 5, pp. 327-347, 2024. doi: 10.1080/07352689.2024.2351678
14. M. A. Cho, A. Ramoelo, P. Debba, O. Mutanga, R. Mathieu, H. Van Deventer, and N. Ndlovu, "Assessing the effects of subtropical forest fragmentation on leaf nitrogen distribution using remote sensing data," Landscape ecology, vol. 28, no. 8, pp. 1479-1491, 2013.
15. L. I. Trejo-Téllez, and F. C. Gómez-Merino, "Nutrient solutions for hydroponic systems," Hydroponics-a standard methodology for plant biological researches, vol. 1, pp. 1-22, 2012.