

Article

Selection of Optimal Diagnostic Positions for Early Nutrient Deficiency in Cucumber Leaves Based on Spatial Distribution of Raman Spectra

Zhaolong Hou ¹ , Yaxuan Wang ², Feng Tan ³, Jiaxin Gao ¹ , Feng Jiao ⁴, Chunjie Su ⁵ and Xin Zheng ^{2,*}

¹ College of Engineering, Heilongjiang Bayi Agricultural University, Daqing 163319, China; houzhaolong@byau.edu.cn (Z.H.); a441380540@163.com (J.G.)

² College of Civil Engineering and Water Conservancy, Heilongjiang Bayi Agricultural University, Daqing 163319, China; wangyaxuan1980@byau.edu.cn

³ College of Information and Electrical Engineering, Heilongjiang Bayi Agricultural University, Daqing 163319, China; bayitf@byau.edu.cn

⁴ College of Agriculture, Heilongjiang Bayi Agricultural University, Daqing 163319, China; jiaofeng1980@163.com

⁵ College of Horticulture and Landscape Architecture, Heilongjiang Bayi Agricultural University, Daqing 163319, China; scj_1992@163.com

* Correspondence: zhxin@byau.edu.cn

Abstract: Accurate diagnosis of crop nutritional status is critical for optimizing yield and quality in modern agriculture. This study enhances the accuracy of Raman spectroscopy-based nutrient diagnosis, improving its application in precision agriculture. We propose a method to identify optimal diagnostic positions on cucumber leaves for early detection of nitrogen (N), phosphorus (P), and potassium (K) deficiencies, thereby providing a robust scientific basis for high-throughput phenotyping using Raman spectroscopy (RS). Using a dot-matrix approach, we collected RS data across different leaf positions and explored the selection of diagnostic positions through spectral cosine similarity analysis. These results provide critical insights for developing rapid, non-destructive methods for nutrient stress monitoring in crops. Results show that spectral similarity across positions exhibits higher instability during the early developmental stages of leaves or under short-term (24 h) nutrient stress, with significant differences in the stability of spectral data among treatment groups. However, visual analysis of the spatial distribution of positions with lower similarity values reveals consistent spectral similarity distribution patterns across different treatment groups, with the lower similarity values predominantly observed at the leaf margins, near the main veins, and at the leaf base. Excluding low-similarity data significantly improved model performance for early (24 h) nutrient deficiency diagnosis, resulting in higher precision, recall, and F1 scores. Based on these results, the efficacy of the proposed method for selecting diagnostic positions has been validated. It is recommended to avoid collecting RS data from areas near the leaf margins, main veins, and the leaf base when diagnosing early nutrient deficiencies in plants to enhance diagnostic accuracy.

Keywords: nutrient deficiency; cucumbers; Raman spectroscopy; precision agriculture; leaf position; nitrogen; phosphorus; potassium



Academic Editors: Maurizio Badiani and Georgia Ntatsi

Received: 6 March 2025

Revised: 23 March 2025

Accepted: 30 March 2025

Published: 12 April 2025

Citation: Hou, Z.; Wang, Y.; Tan, F.; Gao, J.; Jiao, F.; Su, C.; Zheng, X. Selection of Optimal Diagnostic Positions for Early Nutrient Deficiency in Cucumber Leaves Based on Spatial Distribution of Raman Spectra. *Plants* **2025**, *14*, 1199. <https://doi.org/10.3390/plants14081199>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As the tension between human activities and the environment intensifies, the transition from traditional extensive agriculture to precision agriculture has emerged as a pivotal direction in modern agricultural development. Precision agriculture requires rapid, reliable,

and non-destructive methods to capture crop information [1]. Nutrient diagnostics play a pivotal role in this transition. Nitrogen (N), phosphorus (P), and potassium (K) are essential nutrients for plant growth, each playing critical roles in metabolism and development. They are central to crop nutrient diagnostics and fertilizer management strategies [2]. N is a fundamental component of plant proteins and chlorophyll, and its deficiency leads to leaf yellowing and growth retardation [3,4]. P plays a critical role in energy transfer and photosynthesis, with early-stage P deficiency causing dark green leaf discoloration and impairing root development [5]. K functions as an activator of various enzymes involved in photosynthesis and significantly enhances crop stress resistance. A deficiency in K results in leaf yellowing and browning, while also reducing the plant's ability to resist diseases [6]. Although laboratory-based chemical analyses (e.g., Kjeldahl [7], spectrophotometry [8]) offer high accuracy, they are time-consuming, destructive, and risk environmental contamination. It is evident that high-throughput, cost-effective, and non-destructive detection technologies are poised to revolutionize crop nutrient diagnosis, significantly advancing the field of precision agriculture.

Raman spectroscopy (RS) is a non-invasive, non-destructive technique that resists moisture interference, does not require sample pretreatment, and allows for *in vivo* detection. This method is adept at identifying subtle changes in the biochemical composition of plant tissues, offering a robust tool for the early detection of plant nutritional stress [9]. Notably, RS can swiftly analyze plant samples, detecting pathogens or identifying sources of abiotic stress within seconds [10–12]. Importantly, experimental evidence indicates that RS lasers do not cause thermal or photodegradation of plant materials [13]. While Raman spectroscopy is effective in detecting visible nutrient deficiencies, its capability to diagnose latent starvation, where biochemical changes precede visible symptoms, remains an important research area.

Nutritional stress in crops induces changes in specific chemical components within leaf tissues, which ultimately manifest as visible deficiency symptoms [14]. Leaf-based diagnostics is a prevalent method for assessing plant nutritional status in agricultural practice [15]. While previous studies have explored the application of RS for diagnosing nutrient deficiencies in crops such as rice, *Arabidopsis thaliana*, Pak Choi, and Choy Sum [16,17], they have not specifically addressed the influence of leaf position selection on diagnostic accuracy during spectral data collection. Traditionally, repeated measurements or multi-point detection have been commonly employed to obtain averaged spectra, thereby mitigating the influence of anomalous spectral data [16,18]. Prior research has demonstrated significant spatial variability in nutrient distribution within the same leaf [19]. In comparison, systematically determining the optimal diagnostic positions on individual leaves is of greater practical significance [20,21]. Moreover, several reports have attempted to enhance the accuracy of RS diagnosis through combinations of various preprocessing methods [22], feature extraction techniques [23], or algorithmic improvements [24]. Nevertheless, these approaches often overlook a critical issue: the validity of the results heavily depends on the accuracy of the raw spectral data. Therefore, identifying specific positions on the leaf that yield the most stable and representative Raman spectra is critical for improving diagnostic accuracy in early nutrient deficiency detection.

Cosine Similarity Analysis (CSA) is a metric used to evaluate the cosine of the angle between two vectors within a vector space. Due to its computational simplicity and independence from absolute intensity, this method has been widely applied in fields such as spectral analysis and anomaly detection [25–27]. CSA not only reflects the similarity between vectors, but also captures the variations in their individual components, making it a useful tool for identifying relevant or anomalous samples in complex datasets [28].

In this study, we focused on the selection of optimal positions for the early diagnosis of N, P, and K deficiencies in cucumber leaves. Nutrient stress was induced in cucumber plants through soilless cultivation. Spectral data from cucumber leaves were collected using a Raman spectrometer in a dot-matrix pattern, referencing sensor array methodologies, across different stress durations and the control group [29]. Concurrently, chemical analysis was performed to quantify N, P, and K levels in the leaves, enabling the monitoring of nutrient deficiency progression. By analyzing the spatial variations in spectral similarity in different positions, we aimed to identify key positions on cucumber leaves with stable spectral responses that accurately reflect nutritional status. Finally, low-similarity spectral data were removed from each leaf, and the performance of the diagnostic model was compared before and after data cleansing to validate the effectiveness of the diagnostic positions selection method.

2. Materials and Methods

2.1. Plant Growth Conditions and Experimental Design

We hypothesized that spatial-spectral variations in leaves could be systematically analyzed to determine optimal diagnostic positions for early nutrient deficiency detection. The cucumber variety (Jin You 401), developed by the Tianjin Kerun Cucumber Research Institute, was used in this study. Perlite served as the substrate for indoor soilless cultivation, ensuring controlled and consistent cultivation conditions across all samples. The nutrient solution was based on the Hoagland formula, with normal nutrient solution applied during the early growth stages [30]. Throughout cultivation, plants were exposed to fluorescent light for 16 h per day, maintaining a relative humidity of $(50 \pm 5)\%$, and a temperature of $29\text{ }^{\circ}\text{C}/26\text{ }^{\circ}\text{C}$ (day/night). The hydroponic system was regularly monitored to maintain stable pH and electrical conductivity levels. Due to the high uniformity of cucumber seedlings during early growth and their heightened sensitivity to nutrient stress, 48 uniformly grown seedlings were selected for experimentation at the two-leaf, one-core stage. The cultivation substrate was first rinsed with deionized water before the seedlings were randomly assigned to four groups: control-check (CK), nitrogen-deficient (ND), phosphorus-deficient (PD), and potassium-deficient (KD), with twelve plants per group. Each group was cultivated with its respective nutrient solution. The CK nutrient solution followed the complete Hoagland formula, containing $\text{Ca}(\text{NO}_3)_2 \cdot 4\text{H}_2\text{O}$ at 945 mg/L, KNO_3 at 607 mg/L, $(\text{NH}_4)_2\text{H}_2\text{PO}_4$ at 115 mg/L, and $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ at 493 mg/L, along with trace elements, including Fe-EDTA at 2.5 mg/L, H_3BO_3 at 2.86 mg/L, $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$ at 2.13 mg/L, $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ at 0.22 mg/L, $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ at 0.08 mg/L, and $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$ at 0.02 mg/L. In the ND nutrient solution, $\text{Ca}(\text{NO}_3)_2 \cdot 4\text{H}_2\text{O}$ and KNO_3 were omitted and replaced with CaCl_2 at 520 mg/L and KCl at 450 mg/L, respectively [16]. In the PD nutrient solution, $(\text{NH}_4)_2\text{H}_2\text{PO}_4$ was removed and replaced with NH_4Cl at 53 mg/L to maintain N supply, while in the KD nutrient solution, NaNO_3 at 510 mg/L was used as a substitute for KNO_3 . Sampling time points were selected based on preliminary experiments and previous studies on nutrient deficiency responses in crops [17,31]. Leaf samples were collected at 24 (24 h), 72 (72 h), 120 (120 h), and 168 (168 h) hours after stress induction. At each time point, the first-node leaves from three cucumber plants per treatment group were excised, sealed in bags, and immediately brought back to the laboratory for RS data collection and chemical analysis of N, P, and K content. A total of 12 leaf samples were collected at each time point, resulting in 48 leaf samples across all time points. Although Raman spectroscopy is inherently a non-destructive technique, in this study, leaf excision was employed to ensure consistent measurement conditions and improve data reliability. This approach provides a foundation for future in vivo non-destructive diagnostics.

2.2. Measurement of Leaf NPK Content

After collecting the spectral data, cucumber leaves were immediately placed in an oven at 105 °C for 20 min to deactivate the enzymes [32]. Subsequently, the leaves were dried at 80 °C to a constant weight [33]. The dried samples were then stored in a desiccator for subsequent chemical analysis.

The samples were weighed, ground, and digested with concentrated H₂SO₄. The N content was determined using the Kjeldahl method, with the total N content (w_1) calculated using Equation (1) as follows:

$$w_1 = \frac{(V_2 - V_0) \times c \times 0.014}{m \times (V_1/V)} \times 100 \quad (1)$$

where c is the concentration of the sulfuric acid standard titration solution (1/2 H₂SO₄) in 0.01 mol/L, V_2 is the volume of the standard acid solution consumed by the sample (mL), V_0 is the volume consumed by the blank (mL), V_1 is the volume of liquid A tested during distillation (mL), V is the total volume of liquid A tested (mL), m is the sample mass (g), and 0.014 represents the mass of N in 1 mL of 1 mol/L sulfuric acid standard titration solution (g).

The P content was measured by molybdenum-antimony anti-absorption spectrophotometry, with the total P content (w_2) calculated using Equation (2), as follows:

$$w_2 = \frac{\rho \times V}{m} \times \frac{V_2}{V_1} \times 10^{-4} \quad (2)$$

where ρ is the mass concentration of P in liquid A (mg/L), V is the total volume of liquid A tested (mL), V_1 is the dispensed volume of liquid A tested (mL), V_2 is the volume of the color solution (mL), and m is the sample mass (g).

The K content was assessed using flame photometry, with the total K content (w_3) calculated using Equation (3), as follows:

$$w_3 = \frac{(\rho - \rho_0) \times V}{m} \times \frac{V_2}{V_1} \times 10^{-4} \quad (3)$$

where ρ is the mass concentration of K in liquid A (mg/L), ρ_0 is the mass concentration of K in the reagent blank digestion solution (mg/L), V is the total volume of liquid A tested (mL), V_1 is the dispensed volume of liquid A tested (mL), V_2 is the volume of the color solution (mL), and m is the sample mass (g).

The NPK content results were expressed as g·kg^{−1} on a dry weight basis.

2.3. Raman Spectroscopy Data Collection

To minimize fluorescence interference, a miniature Raman spectrometer (ATP3000P, OPTOSKY, Xiamen, China) equipped with a 785 nm fiber laser source was utilized to acquire spectral data. The spectrometer covered a wavelength range of 200–3400 cm^{−1} with a spectral resolution of ±4 cm^{−1}, conforming to standard specifications. The RS parameters were as follows: laser intensity set at 400 mW, integration time of 3.5 s, with three consecutive acquisitions per point, and the mean of these three measurements was used as the representative spectrum.

During the preparation of the spectral sampling dot-matrix template, we tested templates with different hole spacings of 0.5 cm, 0.75 cm, and 1.0 cm. Due to the susceptibility of the template material to deformation during high-temperature laser cutting, the maximum achievable hole diameters for the 0.5 cm and 0.75 cm hole spacing templates were limited to 0.3 cm and 0.35 cm, respectively. In contrast, the 1.0 cm hole spacing template allowed for a hole diameter of 0.5 cm. Larger hole diameters facilitate data acquisition, but

may reduce spatial resolution. Given the relatively large leaf area of cucumber plants, the template with a 1.0 cm hole spacing and a 0.5 cm hole diameter was therefore selected, as shown in Figure 1. Spectral data were collected in a dot-matrix pattern according to the template positions, focusing on the main vein side of the leaf. Due to variability in leaf size under different nutrient stresses and time points, the number of spectra collected varied slightly among groups. A total of 1367 spectra were acquired across four groups: 420 from CK, 303 from ND, 307 from PD, and 337 from KD.

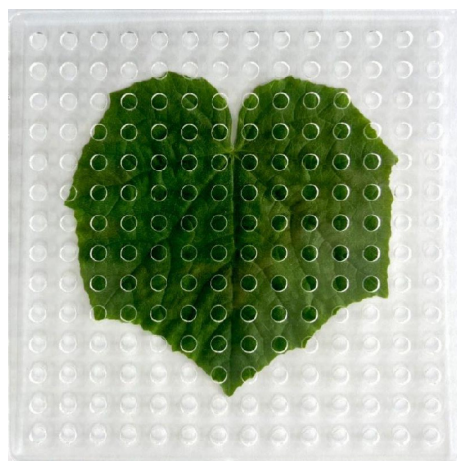


Figure 1. Template for spectral acquisition using a dot-matrix method on leaf surfaces.

2.4. Spectral Similarity Calculation Method

Cosine similarity is commonly used to evaluate the similarity between two multidimensional vectors. When the two vectors have the same orientation, the angle θ between them is 0° , and $\cos \theta$ equals 1. Conversely, when the vectors are orthogonal, θ is 90° , and $\cos \theta$ is 0. Therefore, a cosine value approaching 1 indicates that θ is near 0° , signifying a higher degree of similarity between the vectors. In this study, discrete spectral intensity distributions in the wavenumber domain are treated as multidimensional vectors. The cosine similarity is the inner product normalized by the norms of the vectors and can be expressed using Equation (4), as follows:

$$\cos \theta = \frac{\vec{A} \cdot \vec{B}}{\|\vec{A}\| \|\vec{B}\|} = \frac{\sum_i A_i \cdot B_i}{\sqrt{\sum_i A_i^2} \sqrt{\sum_i B_i^2}} \quad (4)$$

where the vectors $A = [A_1, A_2, A_3, \dots]$ and $B = [B_1, B_2, B_3, \dots]$ represent spectra with intensities A_i and B_i at the wavelength $1/\lambda_i$.

In the experiment, the presence of experimental errors in spectral data is unavoidable; averaging the spectra helps to eliminate biases that may arise from individual samples and provides a more accurate representation of the overall nutritional status of the leaf. Kelly et al. have explored the rationale of species mean values from a mathematical perspective [34]. Therefore, we used the mean spectrum of individual leaves as a reference spectrum. The optimal sampling areas for leaf nutrient diagnosis were analyzed by comparing the similarity between preprocessed spectra from different positions on the leaf and the reference spectrum.

2.5. Diagnostic Model Development and Evaluation Methods

The cucumber variety (Jin You 401), developed by the Tianjin Kerun Cucumber Research Institute, was used in this study. Perlite served as the substrate for indoor soilless cultivation, ensuring controlled and consistent cultivation conditions across all samples.

Data preprocessing is a critical step in spectral analysis, which helps to reduce the potential interference of instrumental errors and environmental factors on spectral data while maximizing spectral differences [35]. In this study, the iterative improved moving average method was used to correct the baseline drift due to background fluorescence, with the window size set to 31 and the iteration count to 5. After baseline correction, the spectral data were normalized in order to adjust the eigenvalues to a uniform scale, which solved potential issues arising from discrepancies in eigenvalue scales, thereby optimizing model performance.

Partial Least Squares Discriminant Analysis (PLS-DA) is a multivariate statistical technique extensively utilized in chemometrics [36,37]. Compared with traditional discriminant analysis methods, PLS-DA is adept at managing data characterized by strong correlations and multicollinearity, making it particularly effective in scenarios where the number of samples is fewer than the number of variables. To validate the diagnostic position selection method, we developed a PLS-DA model using the aforementioned preprocessing techniques for early detection of N, P, and K deficiencies in cucumber leaves.

To evaluate the model's performance, we utilized several metrics, including precision (macro-P), recall (macro-R), F1 score (macro-F1), the number of latent variables (LVs), and the number of misdiagnoses (MDs). F1 score was selected as a key metric due to its strengths in handling imbalanced datasets and its ability to balance precision and recall through the harmonic mean. The optimal latent variable setting for the model was determined using K-fold cross-validation. A higher number of LVs may improve model fitting, but also increases the risk of overfitting, reducing generalizability. To achieve a balance, we determined the optimal LVs as the smallest number at which the model reached stable performance, preferably below 10, as suggested in previous studies [38,39]. In nutrient deficiency diagnosis, misclassifications can lead to incorrect fertilization strategies, potentially harming crop growth and reducing yield. Therefore, among these metrics, lower LVs and MDs values, along with an F1 score closer to 1, indicate superior diagnostic performance.

The spectral data from different stress durations and treatment groups were evenly divided into the training and test sets at an interval of 1 based on the collection sequence to validate the diagnostic position selection method. All data processing was conducted using Python 3.10.4, with plotting performed in Origin 2021.

3. Results

3.1. Identification of Nutrient Deficiency in Cucumber

Visual observations of leaf color and growth status across different treatment groups under varying stress durations revealed that after 72 h of stress, ND and PD plants exhibited slight stress responses in leaf color. However, by 168 h, although the leaf color in KD plants showed no significant changes, their growth rate was noticeably slower compared to CK. These preliminary observations suggest that early deficiencies of different nutrients may need to be detected through comparison at different times.

To further validate these visual observations, chemical analyses were conducted to measure the N, P, and K content in the samples from each group, as shown in Figure 2. The N content in the CK group leaves exhibited an increase from 24 h to 72 h, followed by a decline from 72 h to 168 h. In contrast, P content remained relatively stable, while K content showed a decreasing trend after 72 h. This phenomenon was associated with the functional transition of leaves from a “sink” to a “source” during development and maturation [40,41]. At all observed time points (24 h, 72 h, 120 h, and 168 h) after introduction to stress, N, P, and K content in the leaves of the stress group were lower than those of CK. As the stress duration increased, N and K content in the leaves showed a marked decrease, while the decline in P content was relatively moderate. Notably, an anomalous increase in P

content was observed at 120 h. This phenomenon was likely attributed to the inherently narrow fluctuation range of P content, which, combined with individual variations among sampled plants, resulted in certain samples exhibiting higher P content at 120 h compared to their 72 h counterparts. However, despite this variation, the P content in stressed samples remained distinguishable from that of the CK group, confirming their nutrient-deficient status. Therefore, this deviation was considered to be within an acceptable range.

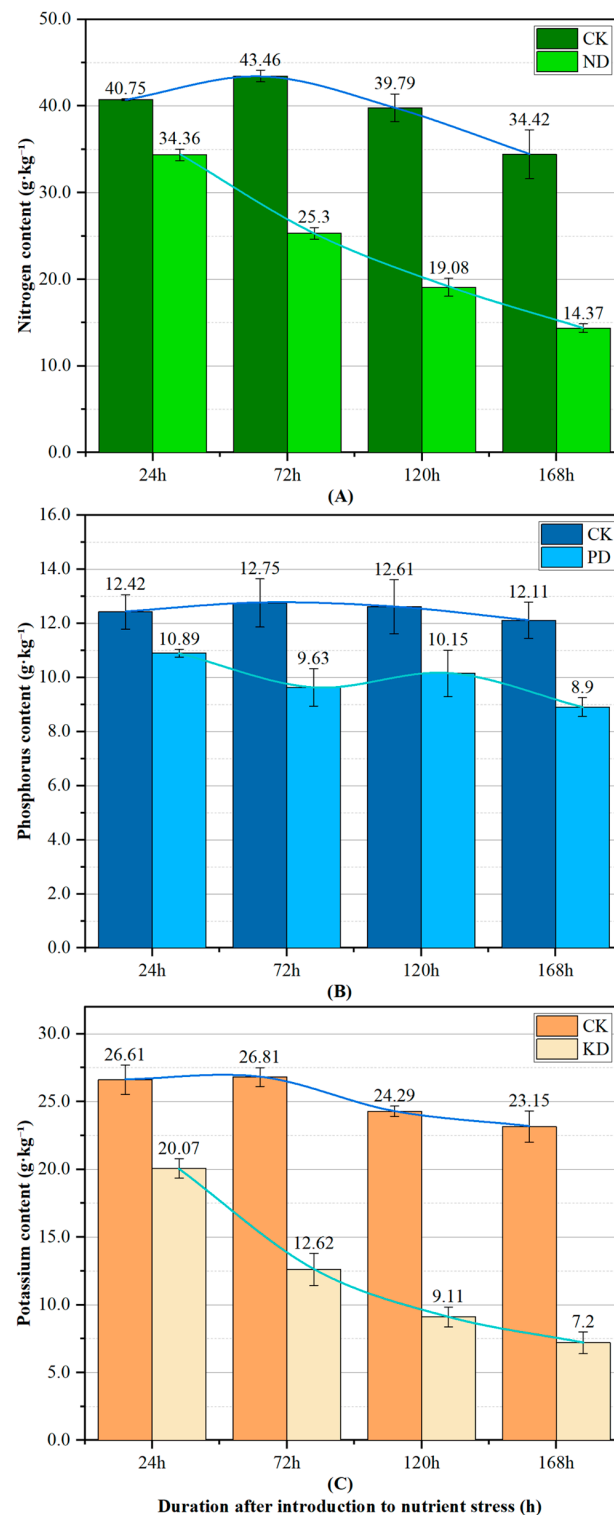


Figure 2. Trends in N, P, and K content in cucumber leaves over time after stress induction. Each bar represents the mean \pm S.E.M. ($n = 3$). (A) N content in ND vs. CK. (B) P content in PD vs. CK. (C) K content in KD vs. CK.

A comparative analysis of the chemical measurements and visual observations suggests that, within 24 h of nutrient stress in cucumber, an internal physiological response is initiated, despite the absence of significant morphological changes or color abnormalities in the leaves. This underscores the critical importance of timely intervention during the early stages of stress.

3.2. Evaluation and Analysis of Spectral Similarity at the Same Position

The cosine similarity of spectra is influenced by multiple factors, including the stability of the spectrometer, variations in the detection environment, and the ability of specific positions to represent the plant's nutrient deficiency status. Analyzing the similarity of spectra collected repeatedly from the same position is valuable for assessing spectral consistency. In this study, a total of 1367 spectra were acquired from 48 cucumber leaves, so the average number of acquisitions per leaf was about 30. Given the relatively controlled conditions of the laboratory environment, we aimed to evaluate the potential impact of spectrometer stability on spectral similarity. To this end, three random positions on the left were selected, and 30 spectra were collected from each position. These spectra were compared to the mean spectrum of their respective positions to evaluate similarity. Table 1 summarizes the minimum similarity values for 10, 20, and 30 spectral collections at each position compared to the mean spectrum.

Table 1. Statistical summary of minimum similarity values for multiple spectral collections at the same position.

Positions	10 Times	20 Times	30 Times
P1	0.99211	0.99139	0.99135
P2	0.99191	0.99153	0.99134
P3	0.99273	0.99215	0.99203
Mean value	0.99225	0.99169	0.99157

The results showed that the mean value of the minimum similarity for 10 spectra collected from the three positions was 0.99225, for 20 was 0.99169, and for 30 was 0.99157. Although the spectrometer exhibited high reproducibility, a slight decline in similarity was observed as the number of spectral collections from the same position increased. This trend should be attributed to the cumulative effects of spectrometer stability. Notably, the minimum similarity values of spectra repeatedly collected at each position, when compared with their respective mean spectra, remained above 0.991. Therefore, when identifying outliers in spectral similarity across positions, we prioritized spectral data with similarity below 0.991, enabling a more targeted focus on spectra that may be anomalous in characterizing the plant's nutrient status. It is important to emphasize that 0.991 is not a fixed global threshold, but rather a reference benchmark for outlier analysis, aimed at minimizing biases stemming from instrument stability and enhancing the reliability of the analysis.

3.3. Analysis of Spectral Similarity Outliers at Different Positions

Box plots are highly effective for identifying outliers, providing the distinct advantage of remaining unaffected by them, which allows for an accurate and stable depiction of data dispersion. Additionally, box plots facilitate data cleansing processes. In this study, we systematically analyzed the cosine similarity between individual positions on cucumber leaves and the mean spectrum using a box plot, as shown in Figure 3. The results indicated that the minimum value of spectral similarity was 0.93328, the maximum value was 0.99724, the mean value was 0.99244, and the standard deviation was 0.00434. Points below the

lowest value in the box plot were labeled as similarity outliers, resulting in the detection of 63 anomalous values. Table 2 presents the distribution of these outliers, while Table 3 summarizes the data distribution for spectral similarities below the overall lower quartile.

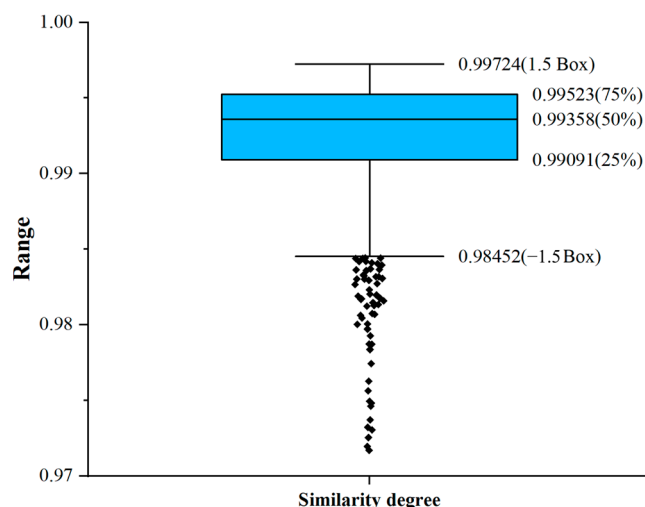


Figure 3. Box plot of spectral similarities across cucumber leaf positions. The boxes represent the interquartile range, and the lines inside the boxes represent the medians. The whiskers denote the lowest and highest values within 1.5 times the interquartile range, while points below the lower whisker represent outliers in similarity values.

Table 2. Statistical distribution of anomalous spectral similarity data in cucumber leaves.

Duration of Stress	Treatment Groups				Total
	CK	ND	PD	KD	
24 h	37	9	4	1	51
72 h	2	3	0	0	5
120 h	1	2	1	1	5
168 h	1	0	0	1	2
Total	41	14	5	3	63

Table 3. Statistical distribution of spectral similarity values below the overall lower quartile in cucumber leaves.

Duration of Stress	Treatment Groups				Total
	CK	ND	PD	KD	
24 h	107	35	23	6	171
72 h	22	28	19	7	76
120 h	21	23	11	6	61
168 h	15	11	4	4	34
Total	165	97	57	23	342

The lower quartile is 0.99091.

A comprehensive analysis of Tables 2 and 3 reveals a discernible pattern in the distribution of similarity values. Notably, the number of spectral similarity anomalies and those below the overall lower quartile were highest in the 24 h group. The number of anomalies decreased gradually as the stress duration increased. This trend suggests that spectral similarity at various positions on cucumber leaves exhibits greater instability during the early developmental stages or under shorter periods of nutrient stress. Further comparison among treatment groups indicates that the spectral similarity in CK leaves was the most unstable, followed by ND and PD, with KD showing relatively more stability.

3.4. Analysis of Spatial Distribution Characteristics of Spectral Similarity on Cucumber Leaves

To further investigate the distribution characteristics of spectral similarity on cucumber leaves, we focused on the similarity values at positions below the lower quartile and replaced the similarity values at other positions with 1.0. By visualizing these processed data as a heat map, we clearly depicted the distribution of lower similarity positions on the leaves, as illustrated in Figure 4. In the heat map, the closer the similarity value is to 1.0, the darker the corresponding color appears. This color gradient effectively highlights the distribution pattern of spectral similarity across different treatment groups and stress durations. Notably, the spectral data with lower similarity were primarily concentrated along the leaf margins, followed by areas near the main veins and the leaf base, in both control plants and those subjected to stresses. The observed spectral variations likely correspond to biochemical adjustments in leaf tissues during early nutrient stress adaptation. These findings validate our earlier hypothesis that spectral data from individual positions on cucumber leaves may not consistently characterize early nutrient deficiencies, providing a critical reference for optimizing subsequent diagnostic models.

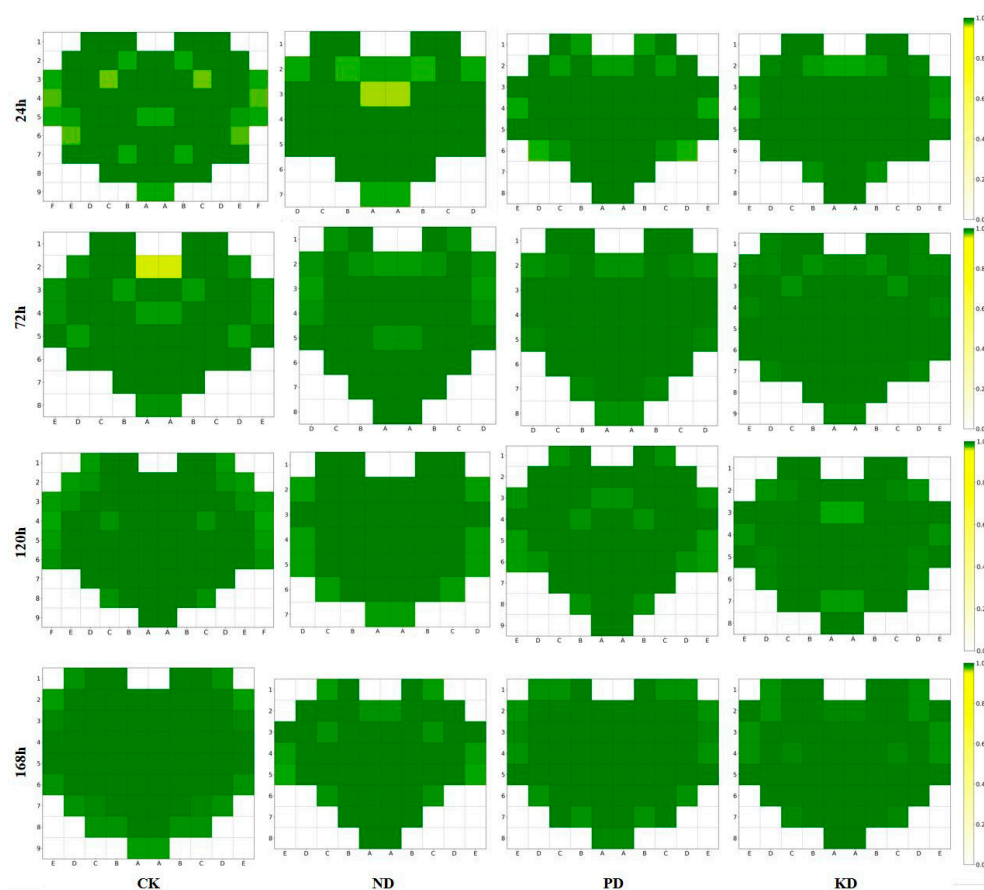


Figure 4. Spatial distribution of low-similarity spectral positions across cucumber leaves. CK (control-check), ND (nitrogen-deficient), PD (phosphorus-deficient), and KD (potassium-deficient).

3.5. Establishment of an Early Diagnostic Model for Nutrient Deficiency in Cucumber

Figure 5 presents the complete process of raw spectral data preprocessing. Initially, the characteristic peaks in the raw spectra are obscured by the fluorescence background. To address this issue, we applied an iterative improved moving average method for baseline correction, as depicted in Figure 5B. The effect of this correction is illustrated in Figure 5C, where the corrected spectra more accurately reflect the original characteristic peaks, ensuring that all spectral curves are positioned above the zero-coordinate axis. To further examine the relationship between the corrected spectra and the original peaks, both the

raw and baseline-corrected spectra were normalized, as shown in Figure 5D,E. Although the spectra between 200 and 700 cm^{-1} were not fully baseline-corrected due to strong fluorescence interference, the spectral characteristics in this range were rendered more clearly, which does not compromise the reliability of subsequent model training. In the range of 700 to 3400 cm^{-1} , the spectra displayed a series of prominent peaks, with distinct characteristic peaks observed at 747, 917, 1005, 1048, 1080, 1117, 1155, 1185, 1218, 1265, 1288, 1301, 1327, 1387, 1440, 1488, 1528, 1611, 1674, and 3191 cm^{-1} in the Raman spectra of cucumber leaves.

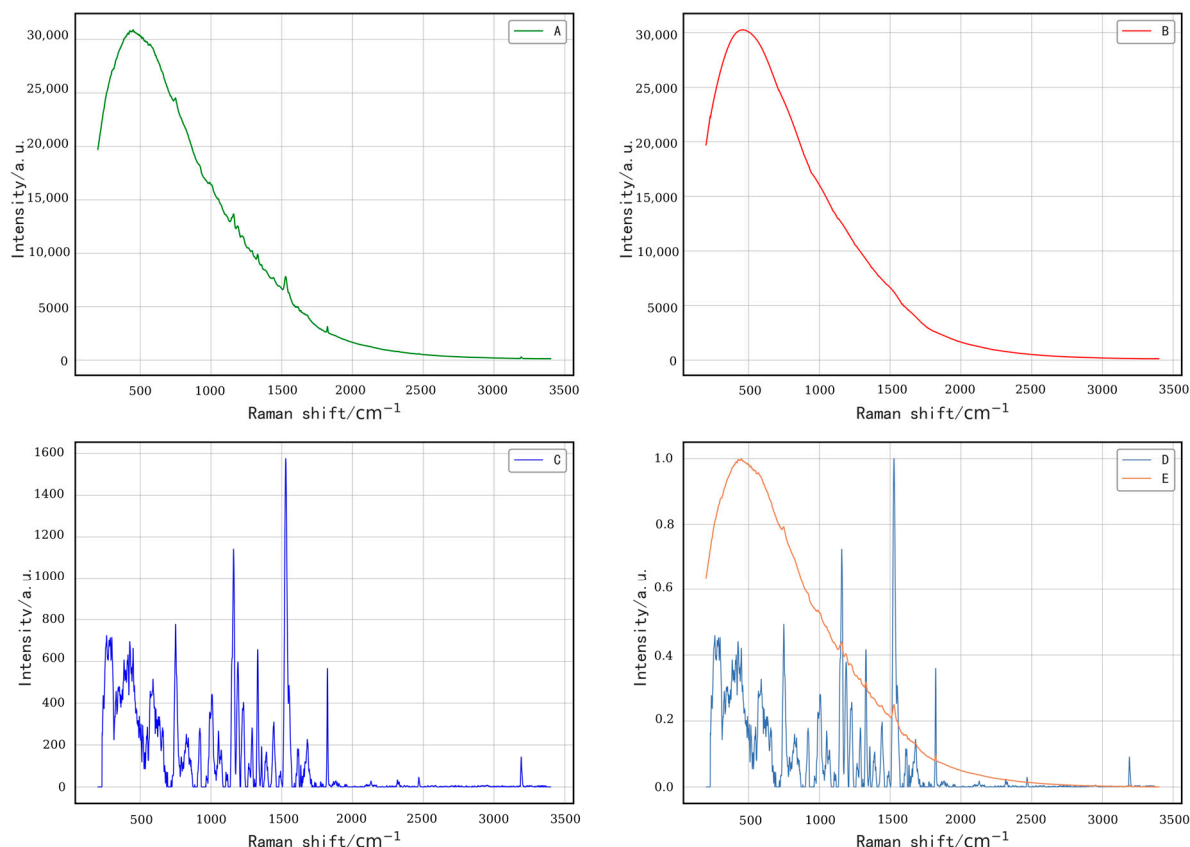


Figure 5. Workflow of spectral data preprocessing. (A) Raw spectra; (B) Fitted baseline; (C) Baseline-corrected spectra; (D) Normalized spectra after baseline correction; (E) Normalized raw spectra.

We integrated the aforementioned preprocessing methods with PLS-DA to construct an early diagnostic model for N, P, and K deficiencies in cucumber leaves. To assess the diagnostic performance of the model, a confusion matrix was generated, with the horizontal axis representing the predicted labels of the samples and the vertical axis representing the true labels. The diagonal values indicate the number of correctly predicted samples, providing a clear visualization of the diagnostic results across different treatment groups. As depicted in Figure 6, the model effectively classified the spectral data of cucumber leaves from the test set into CK, ND, PD, and KD categories across different stress durations, providing empirical support for further optimization of the model.

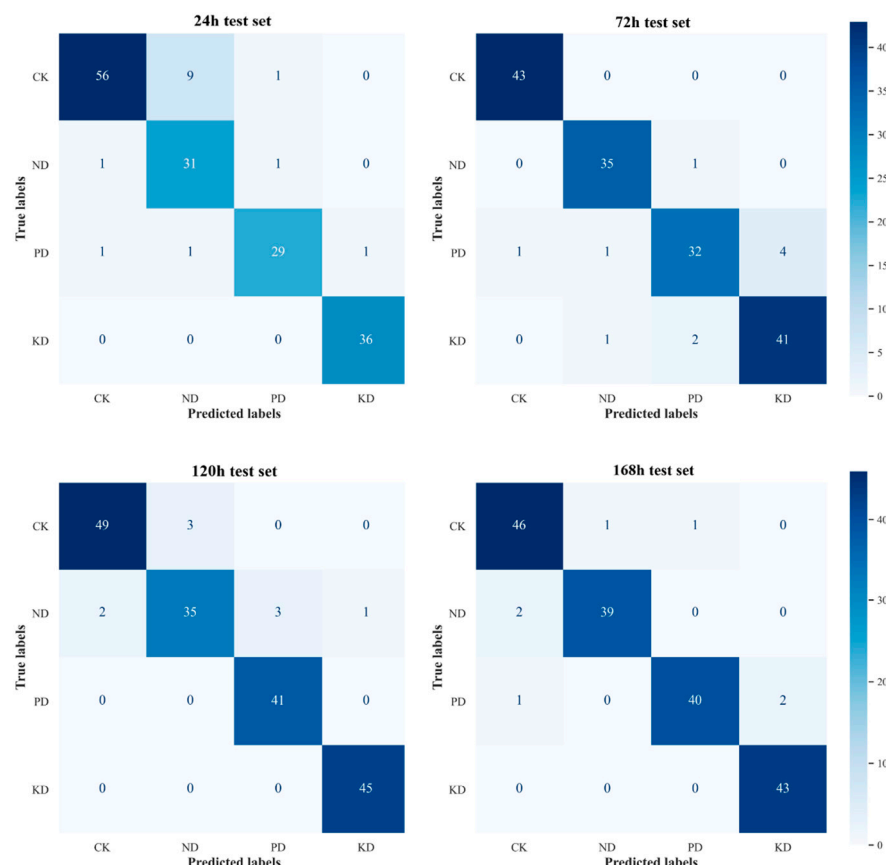


Figure 6. Confusion matrices for test sets under different stress durations.

3.6. Selection and Validation Analysis of Early Diagnostic Positions for Nutrient Deficiencies in Cucumber

To evaluate the correlation between the spectral similarity of cucumber leaves and the selection of early diagnostic positions for nutrient deficiencies, we excluded spectral data with similarity values below the lower quartile of each leaf, as these are shown in Figure 7, where the gray color indicates that the data from these positions were culled. Table 4 compares the diagnostic model's evaluation results across different stress durations before and after data cleansing.

Before data cleansing, model evaluation metrics for cross-validation and the test set indicated that the optimal number of LVs across different stress durations ranged from 6 to 11, while the number of MDs varied between 7 and 15. Notably, in the 24 h group, the optimal number of LVs reached 11, the highest number of MDs (15) was recorded, and the F1 scores for cross-validation and the test set were the lowest, at 89.56% and 91.20%, respectively. These findings suggest that the model faces substantial challenges in the early (24 h) stress phase, potentially exhibiting a degree of overfitting. As stress duration increased, the number of MDs gradually declined, while macro-P, macro-R, and macro-F1 scores improved. In the 168 h group, F1 scores for cross-validation and the test set reached 91.45% and 96.03%, respectively, highlighting the significant impact of stress duration on the stability of spectral signals.

After data cleansing, the optimal LVs and MDs at each time point were reduced, particularly in the 24 h group, where the number of LVs decreased by 4 (from 11 to 7) and the number of MDs dropped by 7 (from 15 to 8), indicating a notable enhancement in model reliability during the early stage (24 h). Furthermore, Test set results demonstrated that the F1 scores for the 24 h, 72 h, 120 h, and 168 h groups increased by 2.10%, 2.16%, 1.53%, and 1.81%, respectively, surpassing the pre-cleaning outcomes. Cross-validation results further

corroborated the trends observed in the test set, confirming that data cleansing not only optimized overall model performance, but also enhanced its generalization capability.

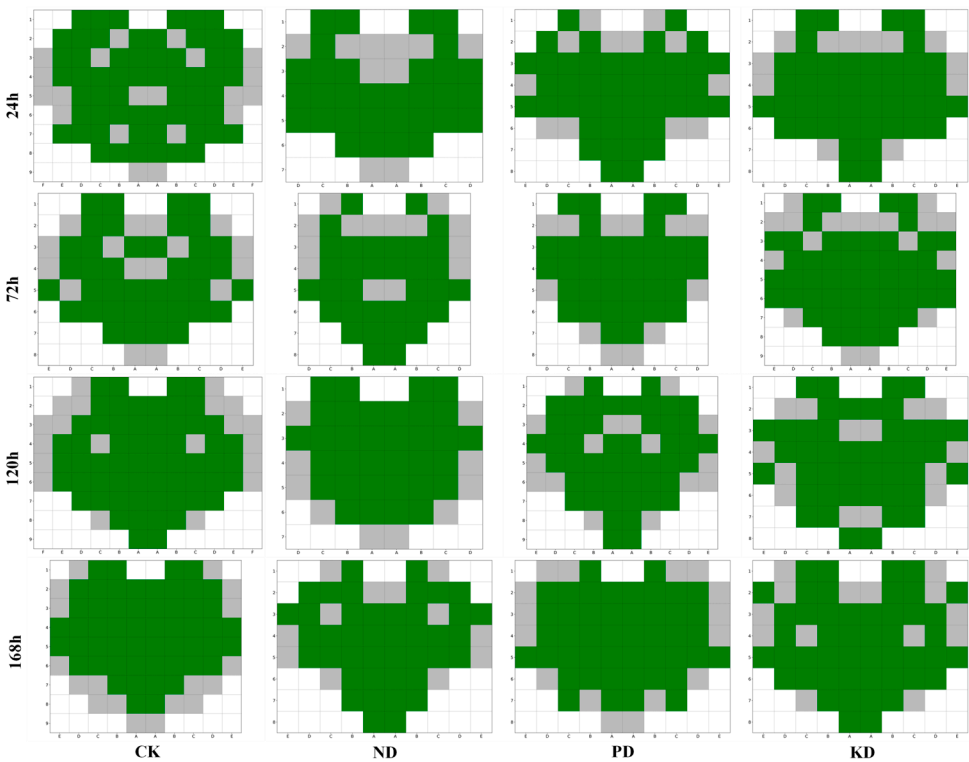


Figure 7. Distribution of excluded low-similarity spectral positions. Green indicates retained spectral data points, while gray indicates excluded positions.

Table 4. Comparison of diagnostic model evaluation metrics across different stress durations before and after data cleansing.

Group Names	Sample Size	LVs	MDs	Cross-Validation			Test Set Evaluation		
				Macro-P (%)	Macro-R (%)	Macro-F1 (%)	Macro-P (%)	Macro-R (%)	Macro-F1 (%)
24 h	335	11	15	90.99	89.31	89.56	90.75	92.35	91.20
24 h-cleaned	254	7	8	92.32	91.52	90.57	92.70	94.76	93.30
72 h	323	7	10	91.71	90.55	90.70	93.72	93.65	93.64
72 h-cleaned	243	7	5	92.68	91.92	91.67	95.78	95.87	95.80
120 h	358	8	9	91.21	90.92	90.91	94.80	94.90	94.78
120 h-cleaned	270	6	5	91.80	90.97	91.00	96.34	96.30	96.31
168 h	351	6	7	92.21	92.55	91.45	96.12	95.99	96.03
168 h-cleaned	265	5	3	93.64	94.24	92.95	97.84	97.93	97.84

LVs, the number of latent variables; MDs, the number of misdiagnoses.

In summary, data cleansing significantly improved diagnostic accuracy, with the reduction in MDs (by up to 7) and the increase in F1 scores (by up to 2.16%) demonstrating its effectiveness. These findings suggest that removing low-similarity spectral data from areas near the margins, main veins, and base of cucumber leaves can significantly enhance the accuracy and reliability of the early diagnostic model for nutrient deficiencies. Consequently, it can be inferred that the level of spectral similarity on cucumber leaves is closely related to the reliability of the diagnostic positions.

4. Discussion

While RS has demonstrated significant potential for monitoring plant nutritional stress [42], research into identifying optimal diagnostic positions under varying nutritional conditions remains limited. Therefore, to effectively utilize RS for the early diagnosis of nutrient deficiencies, it is essential to identify key diagnostic positions that accurately represent the nutritional status of crop leaves across different conditions. Removing low-similarity data improved model accuracy by eliminating spectral noise and inconsistent signals from unstable leaf regions. The diagnostic position selection method proposed in this study effectively identifies anomalous spectral data on cucumber leaves, circumventing the limitations associated with the previous reliance on averaged spectral data. Our study found that avoiding the collection of RS data from areas near the cucumber leaf margins, main veins, and the leaf base can effectively improve the reliability of the spectral data. Furthermore, optimizing the sampling data through the analysis of the spatial distribution of spectral similarity obviously enhanced the PLS-DA model's ability to diagnose early-stage N, P, and K deficiencies in cucumber, resulting in more stable and reliable diagnostic performance.

In contrast to Hu et al. [20], who primarily investigated variations in SPAD values across different positions on cucumber leaves, our study demonstrates significant variations in spectral similarity, particularly near the leaf margins, main veins, and the leaf base, where greater fluctuations and instability were observed. This finding aligns with Hu's results in terms of positional variability; however, their study focused solely on identifying optimal sampling positions for N diagnosis, without addressing the influence of other key nutrients. Meanwhile, prior research has used hyperspectral imaging to diagnose N deficiency in cucumber plants by mapping chlorophyll distribution [43]. Nonetheless, this approach primarily targets later stages of chlorophyll degradation, which can be influenced by various biotic and abiotic stresses, thereby limiting the specificity and accuracy of hyperspectral imaging [16]. In comparison, RS, which directly samples stable spectral areas on the leaf, offers a distinct advantage for early diagnosis of nutrient deficiencies.

The preprocessed spectral results demonstrated that the spectral features within the 700–1800 cm^{-1} range in cucumber leaves closely resemble those observed in rice leaves, as reported by Sanchez et al. [17]. This similarity highlights the potential of RS for diagnosing early nutrient deficiencies across different plant species. The identified spectral peaks correspond to pectin [44], cellulose [45], carotenoids [46], phenylpropanoids [47,48], protein [49], and aliphatic vibrations [50]. This indicates that variations in spectral similarity at different leaf positions are closely related to the changes in their contents, which is crucial for the understanding of the diagnostic mechanism of plant nutrient deficiencies.

While we anticipate that the method proposed in this study will enhance the accuracy of RS for detecting nutrient status in various crops, further research is required to validate its efficacy. Notably, the dataset utilized here reflects the impact of a single variable on spectral similarity across different positions on leaves under hydroponic conditions. In practical agricultural settings, environmental factors are far more complex and dynamic, potentially introducing additional uncertainties into the selection of diagnostic positions. Moreover, variability in leaf size and environmental conditions may limit the generalizability of these spatial patterns. Future work should consider multi-crop trials and real-field conditions to validate the robustness of these diagnostic positions. Additionally, validating the model across diverse environmental conditions will be essential to enhance its practical use in real-world agricultural settings, ensuring broader applicability in precision agriculture.

5. Conclusions

This study introduces a novel analytical method for selecting early diagnostic positions of nutrient deficiencies in plant leaves by analyzing the spatial distribution characteristics of spectral similarity. We examined the impact of four nutrient statuses (ND, PD, KD, and CK) and four stress durations (24 h, 72 h, 120 h, and 168 h) on the selection of diagnostic positions. We systematically explored reliable areas for spectral data collection from cucumber leaves at the early stage of nutrient deficiency, considering aspects such as outlier identification, low-similarity patterns, and model validation. Our findings demonstrate that, under the tested nutrient stresses and stress durations, excluding RS data from areas near leaf margins, main veins, and the leaf base yields more representative and reliable spectral data. Integrating this diagnostic approach into automated sensor networks could facilitate real-time, large-scale monitoring of plant nutritional status. Additionally, this method effectively identifies anomalous spectral data on cucumber leaves, leading to a marked improvement in diagnostic model performance across all stress durations. Notably, during the early stress stage at 24 h, the model achieves higher diagnostic accuracy. This method not only enhances early nutrient deficiency detection in cucumbers but also lays the groundwork for non-destructive diagnostics in other crops, contributing to the advancement of precision agriculture.

Author Contributions: Z.H.: data curation, formal analysis, methodology, writing—original draft. Y.W.: funding acquisition, investigation. F.T.: conceptualization, funding acquisition, writing—review, and editing. J.G.: data curation, investigation. F.J.: data curation. C.S.: methodology. X.Z.: funding acquisition, project administration, resources, supervision, writing—review, and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the National Key Research and Development Program (2023YFD2301605), Heilongjiang Provincial Natural Science Foundation of China (PL2024E024), Heilongjiang Provincial Natural Science Foundation Joint Guidance Program (LH2019E072), the Key Research and Development Program of Heilongjiang Province (GZ20220020), the Heilongjiang Natural Science Foundation (LH2023F043), the Natural Science Talent Support Program of Heilongjiang Bayi Agricultural University (ZRCPY202015), the Higher Education Teaching Reform Research Project of Heilongjiang Province (SJGY20210622), the San-Zong Scientific Research Support Program of Heilongjiang Bayi Agricultural University (ZRCPY202120) and the 2022 Doctoral Startup Fund of Heilongjiang Bayi Agricultural University (XDB202211).

Data Availability Statement: The datasets presented in this article are not readily available because the data are part of an ongoing study. Requests to access the datasets should be directed to Z.H.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Gebbers, R.; Adamchuk, V.I.J. Precision agriculture and food security. *Science* **2010**, *327*, 828–831. [[PubMed](#)]
2. Lemaire, G.; Tang, L.; Bélanger, G.; Zhu, Y.; Jeuffroy, M.-H. Forward new paradigms for crop mineral nutrition and fertilization towards sustainable agriculture. *Eur. J. Agron.* **2021**, *125*, 126248.
3. Kant, S.; Bi, Y.-M.; Rothstein, S.J. Understanding plant response to nitrogen limitation for the improvement of crop nitrogen use efficiency. *J. Exp. Bot.* **2011**, *62*, 1499–1509. [[PubMed](#)]
4. Ohyama, T. Nitrogen as a major essential element of plants. *Nitrogen Assim. Plants* **2010**, *37*, 1–17.
5. Ellsworth, D.S.; Crous, K.Y.; Lambers, H.; Cooke, J. Phosphorus recycling in photorespiration maintains high photosynthetic capacity in woody species. *Plant Cell Environ.* **2014**, *38*, 1142–1156. [[CrossRef](#)] [[PubMed](#)]
6. Sardans, J.; Peñuelas, J. Potassium Control of Plant Functions: Ecological and Agricultural Implications. *Plants* **2021**, *10*, 419. [[CrossRef](#)]
7. Sáez-Plaza, P.; Navas, M.J.; Wybraniec, S.; Michałowski, T.; Asuero, A.G.J. An overview of the Kjeldahl method of nitrogen determination. Part II. Sample preparation, working scale, instrumental finish, and quality control. *Crit. Rev. Anal. Chem.* **2013**, *43*, 224–272.

8. Peverill, K.I. *Soil Analysis: An Interpretation Manual*; CSIRO Publishing: Clayton, Australia, 1999.
9. Altangerel, N.; Ariunbold, G.O.; Gorman, C.; Alkahtani, M.H.; Borrego, E.J.; Bohlmeier, D.; Hemmer, P.; Kolomiets, M.V.; Yuan, J.S.; Scully, M.O. In vivo diagnostics of early abiotic plant stress response via Raman spectroscopy. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 3393–3396.
10. Heraud, P.; Wood, B.R.; Beardall, J.; McNaughton, D. Effects of pre-processing of Raman spectra on in vivo classification of nutrient status of microalgal cells. *J. Chemom. A J. Chemom. Soc.* **2006**, *20*, 193–197.
11. Sanchez, L.; Pant, S.; Xing, Z.; Mandadi, K.; Kurouski, D. Rapid and noninvasive diagnostics of Huanglongbing and nutrient deficits on citrus trees with a handheld Raman spectrometer. *Anal. Bioanal. Chem.* **2019**, *411*, 3125–3133. [[CrossRef](#)]
12. Higgins, S.; Biswas, S.; Goff, N.K.; Septiningsih, E.M.; Kurouski, D.J. Raman spectroscopy enables non-invasive and confirmatory diagnostics of aluminum and iron toxicities in rice. *Front. Plant Sci.* **2022**, *13*, 754735.
13. Butler, H.J.; McAinsh, M.R.; Adams, S.; Martin, F.L. Application of vibrational spectroscopy techniques to non-destructively monitor plant health and development. *Anal. Methods* **2015**, *7*, 4059–4070.
14. de Bang, T.C.; Husted, S.; Laursen, K.H.; Persson, D.P.; Schjoerring, J.K.J. The molecular–physiological functions of mineral macronutrients and their consequences for deficiency symptoms in plants. *New Phytol.* **2021**, *229*, 2446–2469.
15. de Mello Prado, R.; Rozane, D.E. Leaf analysis as diagnostic tool for balanced fertilization in tropical fruits. In *Fruit Crops*; Srivastava, A.K., Hu, C., Eds.; Elsevier: Amsterdam, The Netherlands, 2020; pp. 131–143.
16. Huang, C.H.; Singh, G.P.; Park, S.H.; Chua, N.-H.; Ram, R.J.; Park, B.S. Early Diagnosis and Management of Nitrogen Deficiency in Plants Utilizing Raman Spectroscopy. *Front. Plant Sci.* **2020**, *11*, 663. [[CrossRef](#)]
17. Sanchez, L.; Ermolenkov, A.; Biswas, S.; Septiningsih, E.M.; Kurouski, D. Raman Spectroscopy Enables Non-invasive and Confirmatory Diagnostics of Salinity Stresses, Nitrogen, Phosphorus, and Potassium Deficiencies in Rice. *Front. Plant Sci.* **2020**, *11*, 573321.
18. Zhao, X.; Cai, L. Early detection of zinc deficit with confocal Raman spectroscopy. *J. Raman Spectrosc.* **2018**, *49*, 1706–1712.
19. Zhang, X.; Liu, F.; He, Y.; Gong, X. Detecting macronutrients content and distribution in oilseed rape leaves based on hyperspectral imaging. *Biosyst. Eng.* **2013**, *115*, 56–65.
20. Hu, J.; Li, C.; Wen, Y.; Gao, X.; Shi, F.; Han, L. Spatial distribution of SPAD value and determination of the suitable leaf for N diagnosis in cucumber. *IOP Conf. Ser. Earth Environ. Sci.* **2018**, *108*, 022001. [[CrossRef](#)]
21. Yuan, Z.; Cao, Q.; Zhang, K.; Ata-Ul-Karim, S.T.; Tian, Y.; Zhu, Y.; Cao, W.; Liu, X.J. Optimal leaf positions for SPAD meter measurement in rice. *Front. Plant Sci.* **2016**, *7*, 719.
22. Mostafapour, S.; Dörfer, T.; Heinke, R.; Rösch, P.; Popp, J.; Bocklitz, T.J. Investigating the effect of different pre-treatment methods on Raman spectra recorded with different excitation wavelengths. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2023**, *302*, 123100.
23. Zhao, X.; Xu, M.; Zhang, W.; Liu, G.; Tong, L.J. Identification of zinc pollution in rice plants based on two characteristic variables. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2021**, *261*, 120043. [[CrossRef](#)] [[PubMed](#)]
24. Zhou, W.; Tang, Y.; Qian, Z.; Wang, J.; Guo, H.J. Deeply-recursive convolutional neural network for Raman spectra identification. *RSC Adv.* **2022**, *12*, 5053–5061. [[CrossRef](#)] [[PubMed](#)]
25. Kim, B.; Min, C.; Kim, H.; Cho, S.; Oh, J.; Ha, S.-H.; Yi, J.-h. Structural Health Monitoring with Sensor Data and Cosine Similarity for Multi-Damages. *Sensors* **2019**, *19*, 3047. [[CrossRef](#)]
26. Kwak, B.I.; Han, M.L.; Kim, H.K. Cosine similarity based anomaly detection methodology for the CAN bus. *Expert Syst. Appl.* **2021**, *166*, 114066. [[CrossRef](#)]
27. Irimatsugawa, T.; Shimizu, Y.; Okubo, S.; Inaba, H. Cosine similarity for quantitatively evaluating the degree of change in an optical frequency comb spectra. *Opt. Express* **2021**, *29*, 35613–35622. [[CrossRef](#)]
28. Smith, I.; Ortmann, J.; Abbas-Aghababazadeh, F.; Smirnov, P.; Haibe-Kains, B. On the distribution of cosine similarity with application to biology. *arXiv* **2023**, arXiv:2310.13994.
29. Dong, W.; Sheng, K.; Huang, B.; Xiong, K.; Liu, K.; Cheng, X. Stretchable Self-Powered TENG Sensor Array for Human-Robot Interaction Based on Conductive Ionic Gels and LSTM Neural Network. *IEEE Sens. J.* **2024**, *24*, 37962–37969. [[CrossRef](#)]
30. Hoagland, D.R.; Arnon, D.I. The water-culture method for growing plants without soil. *Circ. Calif. Agric. Exp. Stn.* **1938**, *347*, 39.
31. Antoszewski, G.; Guenther, J.F.; Roberts, J.K.; Adler, M.; Dalle Molle, M.; Kaczmar, N.S.; Miller, W.B.; Mattson, N.S.; Grab, H. Non-Invasive Detection of Nitrogen Deficiency in Cannabis sativa Using Hand-Held Raman Spectroscopy. *Agronomy* **2024**, *14*, 2390. [[CrossRef](#)]
32. Yu, M.; Bai, X.; Bao, J.; Wang, Z.; Tang, Z.; Zheng, Q.; Zhi, J. The Prediction Model of Total Nitrogen Content in Leaves of Korla Fragrant Pear Was Established Based on Near Infrared Spectroscopy. *Agronomy* **2024**, *14*, 1284. [[CrossRef](#)]
33. Peng, X.; Chen, D.; Zhou, Z.; Zhang, Z.; Xu, C.; Zha, Q.; Wang, F.; Hu, X. Prediction of the Nitrogen, Phosphorus and Potassium Contents in Grape Leaves at Different Growth Stages Based on UAV Multispectral Remote Sensing. *Remote Sens.* **2022**, *14*, 2659. [[CrossRef](#)]
34. Kelly, C.; Price, T.D. Comparative methods based on species mean values. *Math. Biosci.* **2004**, *187*, 135–154. [[CrossRef](#)]

35. Lieber, C.A.; Mahadevan-Jansen, A. Automated Method for Subtraction of Fluorescence from Biological Raman Spectra. *Appl. Spectrosc.* **2016**, *57*, 1363–1367. [[CrossRef](#)]
36. Wong, K.H.; Razmovski-Naumovski, V.; Li, K.M.; Li, G.Q.; Chan, K.J. Differentiation of *Pueraria lobata* and *Pueraria thomsonii* using partial least square discriminant analysis (PLS-DA). *J. Pharm. Biomed. Anal.* **2013**, *84*, 5–13. [[PubMed](#)]
37. Huang, X.; Li, H.; Ruan, Y.; Li, Z.; Yang, H.; Xie, G.; Yang, Y.; Du, Q.; Ji, K.; Yang, M.J. An integrated approach utilizing raman spectroscopy and chemometrics for authentication and detection of adulteration of agarwood essential oils. *Front. Chem.* **2022**, *10*, 1036082.
38. Guo, P.; Li, T.; Gao, H.; Chen, X.; Cui, Y.; Huang, Y. Evaluating calibration and spectral variable selection methods for predicting three soil nutrients using Vis-NIR spectroscopy. *Remote Sens.* **2021**, *13*, 4000. [[CrossRef](#)]
39. Xiao, Y.; Zhang, X.; Liu, J.; Li, H.; Jiang, J.; Li, Y.; Diao, S. Prediction of cyanidin 3-rutinoside content in *Michelia crassipes* based on near-infrared spectroscopic techniques. *Front. Plant Sci.* **2024**, *15*, 1346192.
40. Smith, A.P.; Fontenot, E.B.; Zahraeifard, S.; DiTusa, S.F. Molecular components that drive phosphorus-remobilisation during leaf senescence. *Annu. Plant Rev. Phosphorus Metab. Plants* **2015**, *48*, 159–186.
41. Ingestad, T.; Ågren, G.I. Plant nutrition and growth: Basic principles. *Plant Soil* **1995**, *168*, 15–20.
42. Payne, W.Z.; Kurouski, D.J. Raman spectroscopy enables phenotyping and assessment of nutrition values of plants: A review. *Plant Methods* **2021**, *17*, 78.
43. Shi, J.-Y.; Zou, X.-B.; Zhao, J.-W.; Wang, K.-L.; Chen, Z.-W.; Huang, X.-W.; Zhang, D.-T.; Holmes, M. Nondestructive diagnostics of nitrogen deficiency by cucumber leaf chlorophyll distribution map based on near infrared hyperspectral imaging. *Sci. Hortic.* **2012**, *138*, 190–197.
44. Synytsya, A.; Čopíková, J.; Matějka, P.; Machovič, V.J. Fourier transform Raman and infrared spectroscopy of pectins. *Carbohydr. Polym.* **2003**, *54*, 97–106.
45. Edwards, H.; Farwell, D.; Webster, D.J. FT Raman microscopy of untreated natural plant fibres. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **1997**, *53*, 2383–2392.
46. Schulz, H.; Baranska, M.; Baranski, R.J. Potential of NIR-FT-Raman spectroscopy in natural carotenoid analysis. *Biopolym. Orig. Res. Biomol.* **2005**, *77*, 212–221.
47. Agarwal, U.P.J. Raman imaging to investigate ultrastructure and composition of plant cell walls: Distribution of lignin and cellulose in black spruce wood (*Picea mariana*). *Planta* **2006**, *224*, 1141–1153.
48. Lei, K.; Kai, W.; Xiaodong, L.; Bo, Z. High Pressure Structural Investigation of Benzoic Acid: Raman Spectroscopy and X-ray Diffraction. *J. Phys. Chem. C* **2016**, *120*, 14758–14766.
49. Devitt, G.; Howard, K.; Mudher, A.; Mahajan, S.J. Raman spectroscopy: An emerging tool in neurodegenerative disease research and diagnosis. *ACS Chem. Neurosci.* **2018**, *9*, 404–420.
50. Yu, M.M.; Schulze, H.G.; Jetter, R.; Blades, M.W.; Turner, R.F.J. Raman microspectroscopic analysis of triterpenoids found in plant cuticles. *Appl. Spectrosc.* **2007**, *61*, 32–37.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.