

Article

Estimation of Cadmium Content in *Lactuca sativa* L. Leaves Using Visible–Near-Infrared Spectroscopy Technology

Lina Zhou, Leijinyu Zhou, Hongbo Wu, Tingting Jing, Tianhao Li, Jinsheng Li, Lijuan Kong and Limei Chen *

College of Engineering and Technology, Jilin Agricultural University, Changchun 130118, China; zhoulina976430@163.com (L.Z.); zhouleijinyu8514@163.com (L.Z.); whb05082023@163.com (H.W.); 13274422035@163.com (T.J.); 18131571123@163.com (T.L.); jinshengl@jiau.edu.cn (J.L.); konglijuan@jiau.edu.cn (L.K.)

* Correspondence: chenlimei@jiau.edu.cn; Tel.: +86-130-3901-0328

Abstract: In order to monitor cadmium contamination in lettuce quickly, non-invasively, and accurately, and to understand the growth status of lettuce under cadmium pollution, lettuce was used as the test material to detect and analyze the visible–near-infrared reflectance spectra and leaf cadmium content under different concentrations of cadmium stress. A model for estimating lettuce leaf cadmium content was established. For model establishment, firstly, the original spectra were preprocessed using smoothing (Savitzky–Golay, SG), SG combined with multiplicative scatter correction (MSC), SG combined with standard normal variable transformation (SNV), SG combined with mean normalization (MN), SG combined with the first derivative (FD), SG combined with the second derivative (SD), SG combined with the baseline offset (B), and SG combined with de-trending (D). Then, the principal component analysis (PCA) was applied to perform dimensionality reduction on the data. Finally, the reduced dataset was divided into training and testing sets in a 2:1 ratio, and separate models for estimating the lettuce leaf cadmium content were built using partial least squares regression (PLSR), the backpropagation neural network (BP-NN), and support vector regression (SVR) in combination. The results showed that the accumulated cadmium content in lettuce leaves increased with an increase in the soil cadmium concentration. In the visible light range, the spectral reflectance of lettuce leaves increased with an increase in the cadmium concentration. In the near-infrared range, the spectral reflectance of the lettuce leaves under 10 mg/kg and 20 mg/kg cadmium stress was lower than that of the control group. The PLSR models established using the SG + MSC and SG + SNV preprocessing methods exhibited the strongest estimation capability for lettuce leaf cadmium content, with R_p^2 and RMSEP values of 0.92 and 1.53 mg/kg, respectively, for the testing dataset. This study demonstrated that visible–near-infrared spectroscopy has great potential in monitoring cadmium contamination in lettuce.

Keywords: *Lactuca sativa* L.; estimation model; Cd; spectral preprocessing; machine learning; principal component analysis



Citation: Zhou, L.; Zhou, L.; Wu, H.; Jing, T.; Li, T.; Li, J.; Kong, L.; Chen, L. Estimation of Cadmium Content in *Lactuca sativa* L. Leaves Using Visible–Near-Infrared Spectroscopy Technology. *Agronomy* **2024**, *14*, 644. <https://doi.org/10.3390/agronomy14040644>

Academic Editor: Mercedes Del Río Celestino

Received: 1 February 2024

Revised: 12 March 2024

Accepted: 21 March 2024

Published: 22 March 2024



Copyright: © 2024 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

Cadmium (Cd) contamination is a global environmental issue, with many countries and regions facing varying degrees of cadmium pollution. In the 1930s, Japan experienced “Itai-itai disease”, which was caused by Cd contamination. Surveys conducted in the United States and the European Community have shown that 82–94% of cadmium released into the environment ends up in the soil through various pathways, with agricultural soils making up a significant portion of this [1]. In India, the average levels of Cd in different types of soil have been found to exceed the prescribed limits [2]. Roughly 20% of China’s arable land is affected by heavy metal pollution [3], with cadmium contamination exceeding standards in 7% of cases, making it the primary pollutant in contaminated soil areas [4]. Cadmium has strong biotoxicity and is difficult to degrade. Once absorbed and enriched

by plants, it not only affects the growth and development of plants but also poses a threat to human health through the food chain [5–7]. Thus, swiftly and accurately monitoring or identifying the extent of heavy metal stress in plants holds significant importance for ensuring food safety. Traditional chemical detection methods, while accurate, are complex and destructive, making it difficult to achieve rapid and large-scale monitoring. This poses challenges in terms of manpower and resources [8,9]. However, spectroscopy offers the advantages of high resolution and information content [10], providing the possibility for the fast and non-destructive detection of heavy metal pollution in plants.

Cadmium stress can inhibit plant root growth, disrupt plant water balance and nutrient uptake, reduce photosynthetic efficiency, and lead to oxidative damage of leaves and chlorophyll reduction [11–13]. These symptoms can affect the absorption or reflection of light by plants, resulting in changes in the reflectance spectrum of plant leaves [14]. Therefore, reflectance spectroscopy can be used to identify cadmium pollution in plants.

Currently, extensive research has been conducted by scholars utilizing the variation characteristics of plant reflectance spectra to monitor heavy metal pollution in plants. Gu et al. [15] utilized correlation analysis and stepwise regression statistical methods to conduct a statistical analysis on the relationship between the leaf original spectrum, first-order derivative spectrum, spectral parameters, and cadmium content. The results showed that cadmium pollution has a significant impact on the spectral reflectance of small bok choy leaves in the visible, near-infrared, and far-infrared regions. Shi et al. [16] investigated the spectral response of citrus leaves under soil cadmium stress and developed a regression prediction model for spectral indices. They concluded that visible–near-infrared spectroscopy has great potential in monitoring heavy metal pollution in citrus trees. Abdel-Rahman et al. [17] utilized hyperspectral data and partial least squares (PLSs) regression to establish a quantitative monitoring model for cadmium content in Swiss chard leaves. Paresh H et al. [18] investigated the reflectance spectra of barley leaves under cadmium stress and found that there were significant differences in the spectral reflectance under different levels of cadmium stress. Spectroscopic techniques can accurately monitor cadmium stress in crops. Li et al. [19] studied the hyperspectral reflectance of lettuce leaves under cadmium stress and found that the partial least squares (PLSs) regression model based on the first-order differential spectrum (FDR) could accurately predict the cadmium mass ratio in lettuce leaves.

Therefore, in this study, lettuce was used as the test material, and visible–near-infrared spectral technology was employed as the research approach. By utilizing a pot cultivation method with externally added cadmium, the impact of cadmium stress on the visible–near-infrared reflectance spectra of lettuce leaves was investigated. Visible–near-infrared spectroscopy technology not only enables rapid, accurate, non-destructive testing but also possesses advantages, such as easy operation, environmental friendliness, and versatility. The aim was to establish a monitoring model for cadmium stress in lettuce, providing a theoretical basis and reference for the rapid, non-destructive, and accurate monitoring of cadmium stress in lettuce.

2. Materials and Methods

2.1. Materials and Experimental Design

The experiment was conducted in July 2023 on the campus of Jilin Agricultural University in Changchun, Jilin Province (125°42′ E and 43°82′ N), by using a pot cultivation method with externally added cadmium. The test material that was used was *Lactuca sativa* L., and it was purchased from Kuishou Agriculture and Technology Company (Weifang, China). The test soil is an all-purpose seedling substrate, consisting of pure natural black peat, and it was purchased from Changchun Saishi Agricultural Development Co., Ltd. (Changchun, China) It had an organic matter content of 12.59 g/kg, a total nitrogen content of 0.727 g/kg, an available phosphorus of 0.007 g/kg, and an available potassium content of 0.15 g/kg. Impurities in the soil were screened out and it was ground into fine particles and then left to rest in a dry and ventilated place for 3 days. Distilled water was

utilized as the solvent, and cadmium nitrate served as the exogenous cadmium source. A 200 mL solution containing varying concentrations of cadmium was sprayed layer by layer onto the respective experimental soil. Based on the variation range of 0.2 to 20.0 mg/kg for the environmental standard values of cadmium in agricultural soils, obtained from surveys across 16 countries and 2 international organizations, the soil cadmium content was established at 0 (CK), 1, 5, 10, and 20 mg/kg, adhering to the principle of incremental increase [20]. After thorough mixing, the soil was left to age for 10 days before being loaded into pots. The dimensions of each pot were 480 mm × 230 mm × 160 mm, and 1.5 kg of soil was loaded into each pot [21]. Each treatment level was set with 3 replicates. When the lettuce seedlings grew to the stage of two leaves and one heart, the seedlings with consistent growth and good development were selected and transplanted into pots, with 3 seedlings per pot. To ensure the healthy growth of the lettuce, an adequate water supply was maintained consistently throughout the entire experiment. Watering was conducted daily to ensure that the soil moisture remained between 60% and 70%. The moisture level of the soil was measured using the Y 315 soil temperature, moisture, and pH triple-function meter (YANI, Jinan, China). Furthermore, the pots were repositioned every two days during the experiment to ensure even exposure to light. The experiment was conducted under natural light conditions, with the daytime temperatures at 25 ± 2 °C and the nighttime temperatures at 18 ± 2 °C. The relative humidity was maintained between 60% and 70%. The environmental temperature and humidity were measured using a DL 9010 hygrometer (Deli, Ningbo, China).

2.2. Spectral Data Acquisition

The visible–near-infrared reflectance spectra data of the lettuce leaves were collected under five different treatments following 45 days of cadmium stress. The measurements were obtained using the AvaSpec-ULS2048 multi-purpose fiber optic spectrometer manufactured by Avantes (Apeldoorn, Netherlands) in the Netherlands, featuring a wavelength range of 200–1100 nm and a spectral resolution of 0.05–20 nm. The light source utilized was the AvaLight-DHc full-spectrum compact light source, also produced by Avantes, comprising a deuterium lamp wavelength range of 200–400 nm and a tungsten halogen lamp wavelength range of 400–2500 nm. The spectrometer measured a spectrum with 1610 data points.

The measurements were taken on a sunny day with no wind and few clouds, between 10:00 and 14:00. During the data collection, the reflective probe's fiber optic cables were connected to the spectrometer and light source. Then, the reflective probe was fixed using a probe holder, ensuring it formed a 45° angle with the leaf surface. Finally, a deuterium–halogen lamp was selected as the light source. After preheating for 8 min, a white calibration was performed, followed by the measurements. Throughout the entire experiment, a white calibration was conducted every 30 min. The top 2 or 3 leaves of the lettuce samples were selected for measurement, avoiding the main leaf veins. Three spectral data readings were taken from each lettuce leaf, resulting in a total of 135 spectral curves. The spectral information was collected as shown in Figure 1.

Spectral data can be affected by high-frequency random noise, baseline drift, and light scattering. To mitigate these interferences, preprocessing of the raw spectra is essential [22]. In this research, the original spectra underwent preprocessing using various methods such as smoothing (Savitzky–Golay, SG), which was combined with multiplicative scatter correction (MSC), standard normal variable transformation (SNV), mean normalization (MN), the first derivative (FD), the second derivative (SD), the baseline offset (B), and de-trending (D).

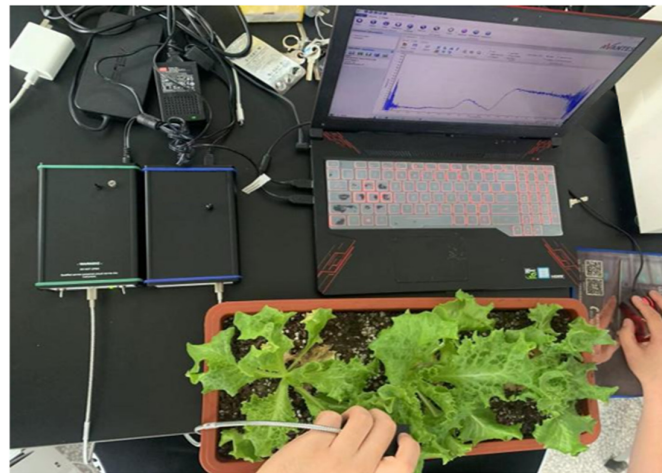


Figure 1. Spectrum data collection.

2.3. Leaf Cadmium Content Determination

The determination of the cadmium content in the leaves was carried out through a digestion method [23]. Firstly, the collected leaves were washed, dried, and ashed. Then, 0.10–0.70 g of the sample was placed in a dry polyvinyl chloride microwave tube, and 10 mL of superior pure concentrated nitric acid was added. The mixture was oscillated for 1 min. Subsequently, the polyvinyl chloride digestion tube was placed in a 120-degree Celsius reaction furnace for pre-digestion for 20 min. The process was considered complete when a small amount of yellow smoke emerged from the digestion vessel and cooled to room temperature. Next, 5 mL of superior pure concentrated nitric acid was added to the pre-digested microwave tube, which was then placed in a Touchwin 2.0 microwave digestion instrument for microwave digestion. Finally, the digestion solution was transferred to a centrifuge tube with a total volume of 50 mL through three additions of ultrapure water. The mixture was oscillated for 1 min and filtered using a 0.22 μ m filter membrane to remove impurities. Ultimately, the cadmium content in the leaves was quantified via inductively coupled plasma–mass spectrometry (ICP-MS) (300D, Perkinelmer, Waltham, MA, USA).

2.4. Data Processing and Analysis

The spectral data were collected using AvaSoft 8 (Avantes, Apeldoorn, The Netherlands) spectral acquisition software, with the basic data input and analysis conducted in Excel 2010 (Microsoft, Redmond, DC, USA). The preprocessing of the collected spectral data was performed using The Unscrambler X 10.4 (CAMO, Oslo, Norway), while the model training and validation were carried out in Matlab 2023b (MathWorks, Natick, MA, USA). The graphs were produced using Origin 2021 (Electronic Arts Inc. (EA), Redwood City, DC, USA).

A significance test is utilized to ascertain if there are notable variances in the average cadmium content within lettuce leaves across various groups. The variance analysis was conducted, followed by Duncan's multiple range test, with the LSD (Least Significant Difference) used for post hoc testing. Typically, a significance level of 5% or 1% is employed as the threshold for establishing significance. When the significance level falls below 5% or 1%, it indicates a notable or highly notable difference.

Partial least squares regression (PLSR) integrates the characteristics of multiple linear regression analysis, principal component analysis, and canonical correlation analysis. It can eliminate the issue of multicollinearity among independent variables, thereby enhancing the stability and predictive power of the constructed model [24]. This makes it particularly suitable for spectral analysis, which involves a large number of independent variables, offering an effective approach for addressing multivariate to multivariate problems [25].

The BP neural network (BPNN) refers to a multilayer neural network that utilizes error backpropagation, comprising an input layer, hidden layers, and an output layer. These layers are interconnected through corresponding activation functions, weights, and thresholds [26]. The BPNN possesses robust capabilities for nonlinear representation, along with significant adaptive capacity and generalization performance, making it commonly used in research on pattern recognition and classification problems [27,28].

Support vector regression (SVR) is an effective predictive algorithm for addressing regression issues. It operates by identifying a hyperplane within the original feature space to map input variables to output variables. To tackle nonlinear problems, support vector machines utilize kernel functions to perform nonlinear mapping in high-dimensional spaces [29,30]. SVR balances the training error and generalization capability, demonstrating numerous unique advantages in solving pattern recognition problems characterized by small samples, nonlinearity, high dimensionality, and local minima [31].

Therefore, this study employed three algorithms, PLSR, BPNN, and SVR, to construct the predictive models for lettuce Cd contamination. The accuracy of the models was evaluated using the coefficient of determination (R^2) and root mean square error (RMSE). A higher R^2 , closer to 1, and an RMSE closer to 0 indicate greater precision of the established models.

3. Results and Discussion

3.1. Effects of Cadmium Stress on Cadmium Content in Lettuce Leaves

According to Figure 2, it can be observed that the cadmium content in lettuce leaves increases with the increasing cadmium concentration in the soil. This indicates that lettuce exhibits sensitivity to cadmium stress and possesses a pronounced capability for cadmium accumulation. Furthermore, the accumulation effect intensifies as the concentration of cadmium increases. This phenomenon can be attributed to the substantial presence of organic acids, cellulose, pectin, and similar substances within the plant cell wall, which have the ability to chelate cadmium. With escalating cadmium concentrations, there is a corresponding increase in cadmium content within the cellular organelles, leading to its accumulation [32,33].

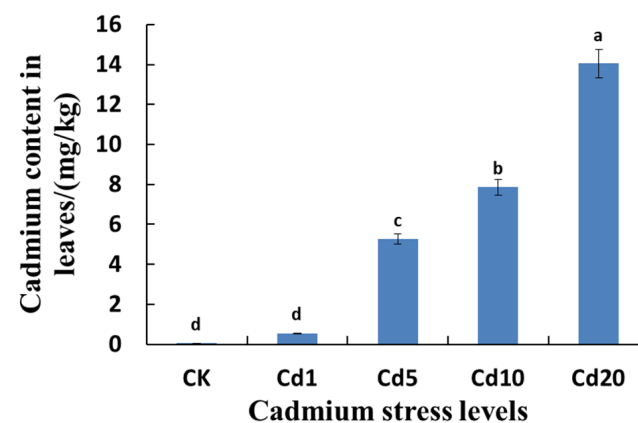


Figure 2. Effects of cadmium stress on cadmium content in lettuce leaves. Note: Different lowercase letters indicate significant differences at the $p < 0.05$ level, with an F value of 226.77.

3.2. Analysis of Visible–Near-Infrared Reflectance Spectroscopy

The wavelength range of the collected spectra in the experiment is 177.480 nm to 1100.316 nm. Plants are sensitive to heavy metal stress in the visible–near-infrared wavelength range, and within the near-infrared range, the spectral response of plants to heavy metal stress is mainly observed in the red edge (670 nm to 780 nm) and red valley bands (600 nm to 720 nm) [19,34]. In order to improve model accuracy, reduce noise interference, and consider significant spectral jitter beyond 850 nm, this study selects the spectral bands

ranging from 400 nm to 850 nm for analysis. The average spectral curves for each treatment are obtained, resulting in the visible–near-infrared reflectance spectrum shown in Figure 3.

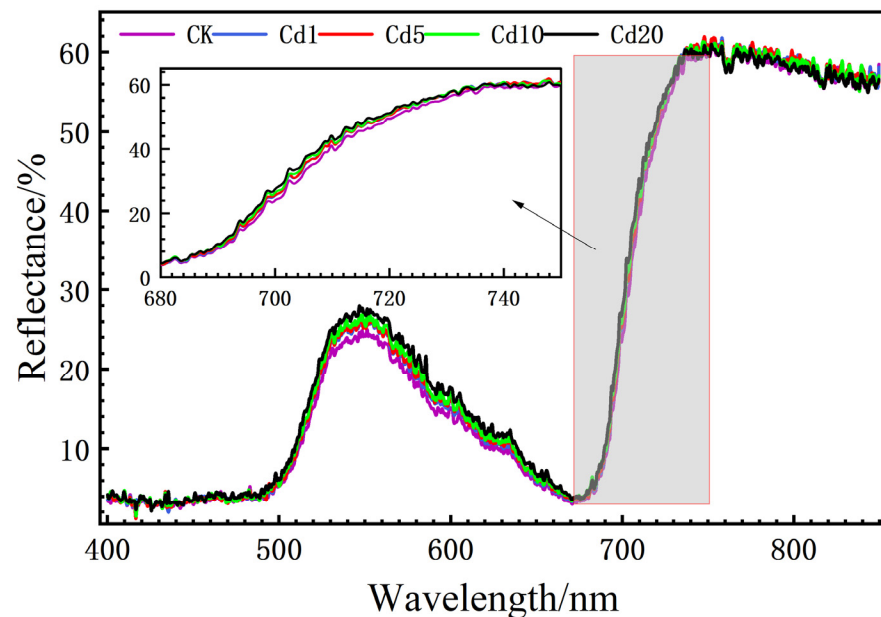


Figure 3. Visible–near-infrared reflectance spectral characteristics in lettuce leaves under cadmium stress.

The visible–near-infrared reflectance spectra of lettuce leaves show distinct absorption valleys, known as “blue valleys” and “red valleys”, around 430 nm and 670 nm, respectively. A reflection peak, referred to as the “green peak”, appears around 550 nm. Between 680 nm and 750 nm, there is a sharp increase in leaf reflectance, indicating the presence of the typical “red edge effect”. Additionally, a small absorption valley is observed around 760 nm, possibly caused by a narrow water absorption band near 760 nm due to water vapor absorption [35,36]. In the visible light spectrum, the leaf spectral reflectance increases with the rise in cadmium concentration. However, under cadmium stress at 1 mg/kg and 5 mg/kg, the changes in the leaf spectral reflectance are less pronounced compared to the CK group. This could be attributed to the fact that cadmium stress at 1 mg/kg and 5 mg/kg promotes plant growth and development, leading to an increase in chlorophyll content, thereby resulting in less noticeable variations in the leaf spectral reflectance compared to the CK group [37]. Under cadmium stress at concentrations of 10 mg/kg and 20 mg/kg, the leaf spectral reflectance is significantly higher than that of the CK group. This increase in reflectance could be attributed to the inhibitory effect of cadmium stress at these concentrations on the chlorophyll content in lettuce leaves. The chlorophyll content decreases, leading to leaf yellowing and causing changes in the leaf spectral reflectance [38].

In the near-infrared wavelength range, under cadmium stress of 10 mg/kg and 20 mg/kg, the leaf spectral reflectance is lower than that of the CK group. This could be attributed to the fact that cadmium stress at 10 mg/kg and 20 mg/kg can lead to the destruction of the internal cellular structures in the leaves, resulting in a reduced reflectance ability. Consequently, the leaf spectral reflectance is lower than that of the CK group [39].

3.3. Spectral Preprocessing

According to Figure 4, it can be observed that the SG treatment technique effectively reduces noise and interference signals in the data. The SNV treatment standardizes the spectral curves through normalization. The MSC treatment alleviates the impact of scattering on the spectral curves. The MN treatment helps achieve consistency in the spectral intensity. The FD and SD treatment techniques address overlapping peaks in the spectral curves, enhancing the discrimination between spectra. The B treatment increases the signal resolution, making features and peaks more prominent. The D treatment removes the trend

signals, thereby improving the data analysis accuracy and reinforcing signal characteristics. These methods not only enhance the signal-to-noise ratio but also maintain the effectiveness of spectral information [40].

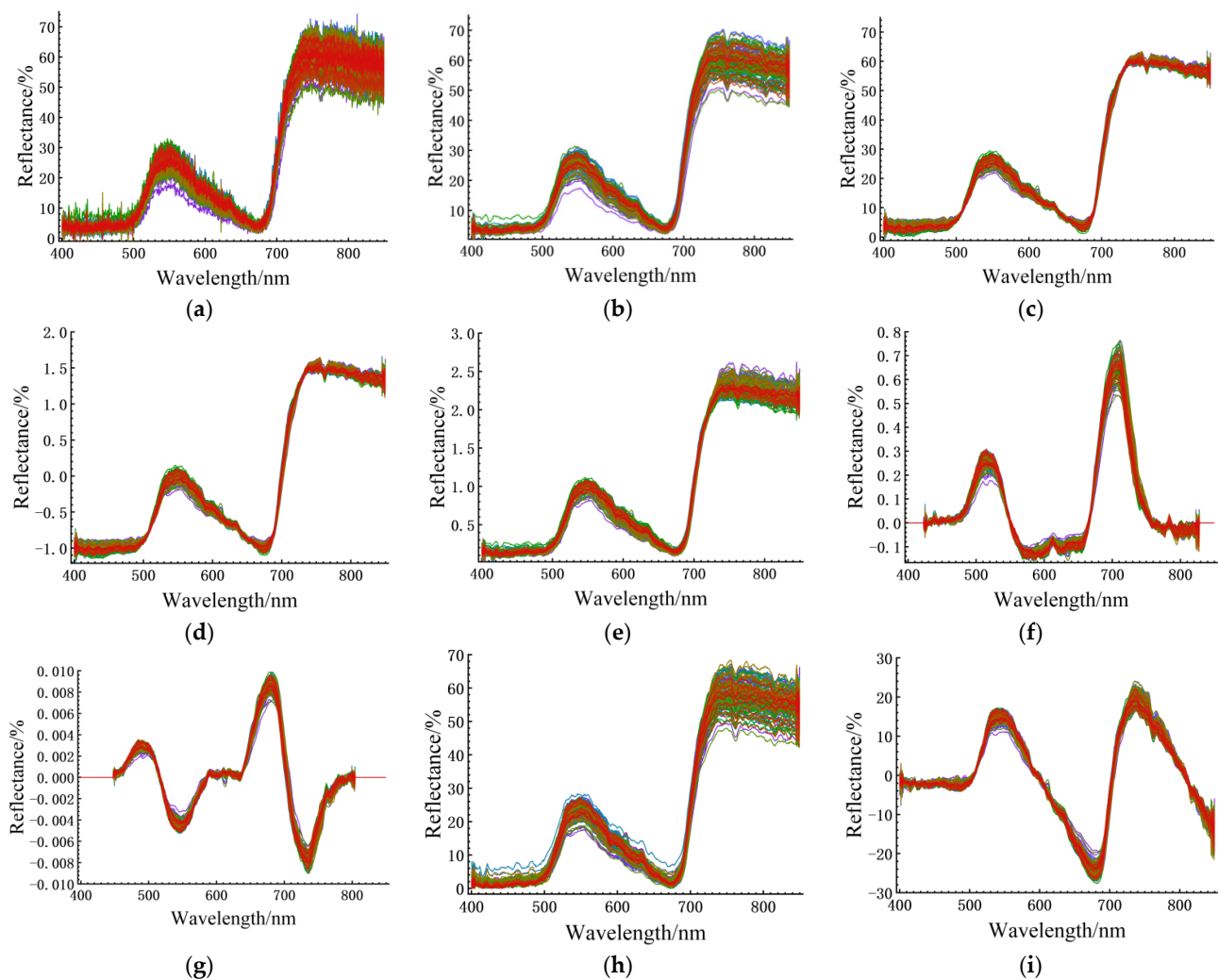


Figure 4. The spectral curves after preprocessing: (a) the original spectral curves; (b) the spectra preprocessed with SG; (c) the spectra preprocessed with SG + MSC; (d) the spectra preprocessed with SG + SNV; (e) the spectra preprocessed with SG + MN; (f) the spectra preprocessed with SG + FD; (g) the spectra preprocessed with SG + SD; (h) the spectra preprocessed with SG + B; and (i) the spectra preprocessed with SG + D.

3.4. Dimensionality Reduction Processing

Due to the original spectral information having 781 data nodes, forming a high-dimensional matrix of 781×135 , and this information having certain correlations that can lead to information overlap, dimensionality reduction processing is necessary. In this study, principal component analysis (PCA) was used for dimensionality reduction processing to map the spectral data from high dimension to low dimension. This not only preserves the most important features of the spectral data but also removes redundant information and reduces the impact of collinearity, thereby speeding up subsequent data analysis and modeling. The number of principal components for the original spectrum and various preprocessed spectra are determined based on the principle of cumulative contribution rate $\geq 85\%$ and the principle of variance percentage of each principal component $\geq 1\%$ [29] (Table 1).

Table 1. Determination of the number of principal components.

Preprocessing Method	Number of Principal Components	Cumulative Contribution Rate (%)
original spectral	10	87.37
SG	11	97.42
SG + MSC	9	90.66
SG + SNV	9	90.56
SG + MN	11	95.99
SG + FD	11	86.46
SG + SD	15	92.73
SG + B	12	97.76
SG + D	13	91.01

3.5. Model Establishment

Estimation models for cadmium content in lettuce leaves are built using partial least squares regression (PLSR), the backpropagation neural network (BP-NN), and support vector regression (SVR). In the PLSR model, the preprocessed spectral data are used as input variables and the cadmium content in the leaves is the output variable. In the BP-NN and SVR models, the reduced-dimensional principal component spectral data are used as the input variables, and the cadmium content in the leaves is the output variable. The training set comprises 90 samples, and the test set contains 45 samples. One sample is randomly selected for validation out of every three, based on the range of cadmium content.

3.5.1. Model Establishment Based on Partial Least Squares Regression (PLSR)

PLSR modeling was performed on the original spectra and various preprocessed spectral data. The training set accounts for two-thirds of the dataset, while the test set accounts for one-third of the dataset.

According to Table 2, it can be observed that the PLSR models established using different processing methods exhibit good predictive performance. The determination coefficients (R^2) for the training set are all above 0.95, and the root mean square errors (RMSE) range from 0.33 mg/kg to 0.81 mg/kg. For the test set, the prediction coefficients (R_p^2) are all above 0.80, and the root mean square errors of prediction (RMSEP) range from 1.53 mg/kg to 2.08 mg/kg. The PLSR models established using the SG + MSC and SG + SNV preprocessing methods demonstrate the best performance. The R^2 values for the training set are both 0.98, while the R_p^2 values for the test set are both 0.92. Additionally, the RMSEP for the test set is the lowest among all the models, at 1.53 mg/kg, representing an improvement of 0.46 mg/kg compared to the model established using the original spectra. These results indicate that the combination of the SG + MSC and SG + SNV preprocessing methods with PLSR modeling is effective in accurately estimating cadmium content in lettuce leaves.

Table 2. The effectiveness of the PLSR models.

Preprocessing Method	PCA	Training Set ($n = 90$)		Testing Set ($n = 45$)	
		R^2	RMSE	R_p^2	RMSEP
original spectral	10	1.00	0.33	0.86	1.99
SG	11	0.99	0.61	0.90	1.68
SG + MSC	9	0.98	0.68	0.92	1.53
SG + SNV	9	0.98	0.67	0.92	1.53
SG + MN	11	0.99	0.55	0.90	1.70
SG + FD	11	0.98	0.81	0.84	2.08
SG + SD	15	0.99	0.51	0.87	1.92
SG + B	12	0.99	0.50	0.91	1.56
SG + D	13	1.00	0.36	0.89	1.71

3.5.2. Model Establishment Based on BP Neural Network (BP-NN)

The BP-NN algorithm is a multilayer feedforward perceptron neural network that possesses strong self-learning and nonlinear mapping capabilities. During the establishment of the BP-NN model, the training set accounts for two-thirds of the dataset, and the testing set accounts for one-third of the dataset. The learning rate is set to 0.1, and the number of iterations is set to 1000.

The modeling results are shown in Table 3. The BP-NN method was used for modeling, and the R^2 values of the training set are all above 0.80, with the RMSE ranging from 1.50 mg/kg to 3.01 mg/kg. The R^2 values of the testing set are all above 0.65, and the RMSE_p values range from 1.74 mg/kg to 3.01 mg/kg. Among them, the BP-NN model established using SG + SD preprocessing has the best performance, with an R^2 of 0.86 and an RMSE of 1.93 mg/kg on the training set. The testing set achieves an R_p^2 of 0.88 and an RMSE_p of 1.87 mg/kg, which is a 0.16 mg/kg improvement compared to the model established using the original spectra. This indicates that the model constructed using SG + SD preprocessing combined with the BP-NN method can accurately estimate the cadmium content in lettuce leaves.

Table 3. The effectiveness of the BP-NN models.

Preprocessing Method	Training Set ($n = 90$)		Testing Set ($n = 45$)	
	R^2	RMSE	R_p^2	RMSE _p
original spectral	0.87	1.91	0.84	2.03
SG	0.91	1.50	0.80	2.46
SG + MSC	0.87	1.82	0.74	2.93
SG + SNV	0.82	2.20	0.69	3.01
SG + MN	0.89	3.01	0.68	1.74
SG + FD	0.87	1.84	0.85	2.11
SG + SD	0.86	1.93	0.88	1.87
SG + B	0.87	1.87	0.81	2.35
SG + D	0.88	1.87	0.82	2.19

3.5.3. Model Establishment Based on Support Vector Regression (SVR)

To ensure the SVR model performs optimally, it is necessary to determine the best penalty factor (c), kernel function parameter (g), and optimal kernel function within the SVR model. A grid search is able to comprehensively explore a specified parameter space. It achieves this by exhaustively searching through all possible combinations of parameters, ensuring that no potential combination is overlooked [41]. Therefore, in this study, a grid search was employed to determine the optimal penalty factor (c) and the best kernel function parameter (g) for SVR, as shown in Table 4. Through employing the brute force search technique, it is identified that, out of the Linear Kernel, Polynomial Kernel, Radial Basis Function Kernel, and Sigmoid Kernel within the SVR model, the Radial Basis Function (RBF) Kernel emerges as the most effective kernel function.

Table 4. SVR optimal parameters.

Preprocessing Method	c	g	MSE
original spectral	11.3137	5.6569	7.6856
SG	16.0000	2.8284	6.3072
SG + MSC	16.0000	2.8284	6.7317
SG + SNV	16.0000	2.8284	6.6766
SG + MN	16.0000	2.8284	7.1582
SG + FD	16.0000	1.0000	4.8223
SG + SD	16.0000	0.5000	6.1215
SG + B	16.0000	0.7071	7.4029
SG + D	16.0000	1.0000	4.5548

In the SVR model, the training set accounts for two-thirds of the dataset, while the test set accounts for one-third. The modeling results are shown in Table 5. In the SVR model, the R^2 of the training set for all nine treatment methods is above 0.90, and the RMSE ranges from 0.27 mg/kg to 1.68 mg/kg. The R_p^2 of the test set is above 0.70, and the RMSEP ranges from 1.80 mg/kg to 2.69 mg/kg. Among them, the SVR model established with the SG + D method shows the best performance. The R^2 of the training set is 0.98, with an RMSE of 0.27 mg/kg. The R_p^2 of the test set is 0.84, which is an improvement of 0.11 compared to the model built with the original spectrum. The RMSEP is 1.80 mg/kg, which is an improvement of 0.89 mg/kg compared to the model established with the original spectrum. The results indicate that the estimation model for cadmium content in lettuce leaves established with the SG + D preprocessing method and SVR approach has effective prediction capabilities.

Table 5. The effectiveness of the SVR models.

Preprocessing Method	Training Set ($n = 90$)		Testing Set ($n = 45$)	
	R^2	RMSE	R_p^2	RMSEP
original spectral	0.99	0.57	0.73	2.69
SG	1.00	0.36	0.84	2.03
SG + MSC	0.99	0.49	0.75	2.49
SG + SNV	0.99	0.47	0.76	2.49
SG + MN	0.99	0.41	0.81	2.22
SG + FD	0.96	1.07	0.82	2.10
SG + SD	0.93	0.64	0.79	2.17
SG + B	0.90	1.68	0.84	1.94
SG + D	0.98	0.27	0.84	1.80

3.6. Model Comparison

The analysis and comparison were conducted by selecting the best combination of the different preprocessing methods and models, as shown in Table 6. Among the four models, the BP-NN (SG + SD) model exhibits the poorest performance on both the training and test sets. Although the SVR (SG + D) model shows the best performance on the training set, its performance on the test set is relatively poor. The PLSR (SG + MSC) and PLSR (SG + SNV) models demonstrate the highest accuracy, with the R_p^2 on the test set being the highest among the four models, indicating a good fit of the models. Furthermore, the RMSEP is the lowest among the four models, suggesting a high prediction accuracy and a small prediction error. Therefore, it can be concluded that the PLSR model with preprocessing using SG + MSC and SG + SNV demonstrates the highest accuracy in estimating lettuce leaf cadmium content. This may be attributed to the fact that PLSR (partial least squares regression) combines the advantages of canonical correlation analysis, principal component analysis, and multiple linear regression analysis, forming a comprehensive multivariate statistical analysis technique. It efficiently harnesses information from all predictor variables to build robust predictive models [42]. Meanwhile, PLSR can effectively suppress or circumvent the problem of multicollinearity among several independent variables, optimizing variable information to achieve maximal collinearity between independent and dependent variables. This feature renders PLSs especially applicable to the analysis of spectral data scenarios in which the quantity of independent variables significantly exceeds that of dependent variables, and there is a noticeable collinearity among the independent variables [43]. Currently, many experts and scholars have qualitatively or quantitatively constructed diagnostic models for crop nutrient parameters using PLSR technology from different perspectives, which align with the results obtained in this study [16,19,42]. Visible–near-infrared spectroscopy technology demonstrates promising prospects in the monitoring of heavy metals. Villatoro-Pulido M et al. [44] believe that NIRS could act as a rapid screening method for assessing the total minerals, iron, sodium, potassium, and zinc contents in the rocket. R. Font et al. [45,46] suggest that visible-to-near-

infrared spectroscopy technology possesses considerable potential in determining the total arsenic (As) in the prostrate amaranth (*Amaranthus blitoides* S. Watson) and in screening for the inorganic arsenic (i-As) content in commercial rice. Combined with this study, it is illustrated that visible–near-infrared spectroscopy technology, when integrated with the PLSR method, is capable of estimating the cadmium content in lettuce leaves.

Table 6. Comparison of different models’ performance.

Models	Preprocessing Method	Training Set ($n = 90$)		Testing Set ($n = 45$)	
		R^2	RMSE	R_p^2	RMSE _p
PLSR	SG + MSC	0.98	0.68	0.92	1.53
PLSR	SG + SNV	0.98	0.67	0.92	1.53
BP-NN	SG + SD	0.86	1.93	0.88	1.87
SVR	SG + D	0.98	0.27	0.84	1.80

4. Conclusions

In this study, lettuce was selected as the research subject, and an exogenously cadmium-treated potting method was employed to investigate the cadmium content in lettuce leaves and analyze the visible–near-infrared reflectance spectra under cadmium stress. Furthermore, a cadmium content estimation model for lettuce leaves was established. The following conclusions can be drawn:

- Lettuce leaves can effectively accumulate cadmium, and the enrichment effect becomes more pronounced with an increasing cadmium concentration.
- In the visible light range, the spectral reflectance of lettuce leaves increases with the increase in cadmium concentration. In the near-infrared range, under 10 mg/kg and 20 mg/kg cadmium stress, the spectral reflectance of lettuce leaves decreases compared to the CK group.
- Among the PLSR, BP-NN, and SVR models established for estimating lettuce leaf cadmium content, the PLSR model with preprocessing using SG + SNV and SG + MSC demonstrates the highest accuracy. The R^2 values for the training set are both 0.98, while the R_p^2 and RMSE_p values for the testing set are 0.92 and 1.53 mg/kg, respectively.

Therefore, the application of visible–near-infrared spectroscopy technology can offer a theoretical foundation and guidance for ensuring the safe production and quality management of lettuce.

Author Contributions: Conceptualization, L.Z. (Lina Zhou) and L.Z. (Leijinyu Zhou); methodology, L.Z. (Lina Zhou) and L.C.; formal analysis, L.Z. (Leijinyu Zhou) and L.K.; investigation, L.C. and J.L.; resources, L.Z. (Lina Zhou) and L.C.; writing—original draft preparation, L.Z. (Leijinyu Zhou) and H.W.; writing—review and editing, J.L. and T.J.; visualization, L.K. and T.L.; supervision, H.W.; funding acquisition, L.Z. (Lina Zhou). All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Scientific and Technological Research Projects of the Jilin Provincial Department of Education, grant number JJKH20240425KJ, founded by Lina Zhou. And this study was supported by the Science and Technology Development Project of Jilin Province, grant number 20210202051NC, founded by Lina Zhou.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

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

References

- Wang, K.R. The Current Situation of Cadmium Pollution in Chinese Farmland and Its Control and Utilization Countermeasures. *J. Agro-Environ. Sci.* **1997**, *16*, 274–278.
- Kumar, V.; Sharma, A.; Kaur, P.; Sidhu, G.P.S.; Bali, A.S.; Bhardwaj, R.; Thukral, A.K.; Cerda, A. Pollution assessment of heavy metals in soils of India and ecological risk assessment: A state-of-the-art. *Chemosphere* **2018**, *216*, 449–462. [[CrossRef](#)] [[PubMed](#)]
- Pandey, C.V.; Bajpai, O.; Singh, N. Energy crops in sustainable phytoremediation. *Renew. Sustain. Energy Rev.* **2016**, *54*, 58–73. [[CrossRef](#)]
- Zhao, F.; Ma, Y.; Zhu, Y.; Tang, Z.; McGrath, S. Soil Contamination in China: Current Status and Mitigation Strategies. *Environ. Sci. Technol.* **2014**, *49*, 750–759. [[CrossRef](#)] [[PubMed](#)]
- Chen, N.; Feng, H.L.; Yang, Y.D.; Chen, P.; Ren, T.B.; Jia, F.F.; Liu, G.S. Establishment of hyperspectral prediction model for cadmium content in flue-cured tobacco leaves. *J. Agric. Resour. Environ.* **2021**, *38*, 570–575.
- Raiesi, F.; Razmkhah, M.; Kiani, S. Salinity stress accelerates the effect of cadmium toxicity on soil N dynamics and cycling: Does joint effect of these stresses matter? *Ecotoxicol. Environ. Saf.* **2018**, *153*, 160–167. [[CrossRef](#)] [[PubMed](#)]
- Xiang, Z.; Xue, C.; Ni, P.; Wei, H.; Liang, S.; Wei, L. Foliar Spraying of Mannose Alleviates Cadmium Stress by Changing the Subcellular Distribution and Chemical Forms of Cadmium in Wheat Root. *Agronomy* **2022**, *12*, 2273.
- Da Silva, E.D.N.; Heerdt, G.; Cidade, M.; Pereira, C.D.; Morgon, N.H.; Cadore, S. Use of in vitro digestion method and theoretical calculations to evaluate the bioaccessibility of Al, Cd, Fe and Zn in lettuce and cole by inductively coupled plasma mass spectrometry. *Microchem. J.* **2015**, *119*, 152–158. [[CrossRef](#)]
- Zhong, L.; Qian, J.W.; Chu, X.Y.; Qian, Z.H.; Wang, M.; Li, J.L. Monitoring heavy metal pollution contamination of wheat soil using hyperspectral remote sensing technology. *Trans. Chin. Soc. Agric. Eng.* **2023**, *39*, 265–270.
- Sun, J.; Hu, S.Q.; Zhou, X.; Zhang, L.; Wu, X.H.; Dai, C.X. Design and Experiment of Portable Non-destructive Tester for Heavy Metal Cadmium Content in Vegetable Leaves. *Trans. Chin. Soc. Agric. Mach.* **2022**, *53*, 195–202+220.
- Tang, X.W.; Yan, P.; Ji, P.H.; Gao, P.C.; Nguyen, T.H.; Tong, Y.A. Cadmium uptake in above-ground parts of lettuce (*Lactuca sativa* L.). *Ecotoxicol. Environ. Saf.* **2016**, *125*, 102–106. [[CrossRef](#)] [[PubMed](#)]
- Wang, X.J.; Wang, W.B.; Yang, L.; Jin, L.; Song, Y.; Jiang, S.J.; Qin, L.L. Transport pathways of cadmium (Cd) and its regulatory mechanisms in plant. *Acta Ecol. Sin.* **2015**, *35*, 7921–7929.
- Nabila, B.; Sangar, K.; Yasir, H.; Asmat, U.M.; Ghaffar, K.A.; Faiza, B.; Peng, J.L.; Qiang, T.L. Effect of foliar application of nanoparticles on growth, physiology, and antioxidant enzyme activities of lettuce (*Lactuca sativa* L.) plants under cadmium toxicity. *Environ. Sci. Pollut. Res. Int.* **2023**, *30*, 99310–99325.
- Liu, L.; Zhang, W.J.; Wang, W.H.; Zhang, Y.; Li, Q.; Han, D.; Li, J.J.; Lin, L. Estimation models for spectral response and cadmium contents in leaves of *Brassica napus* L. *Chin. J. Oil Crop Sci.* **2019**, *41*, 46–52.
- Gu, Y.W.; Li, S.; Gao, W.; Wei, H. Hyperspectral estimation of the cadmium content in leaves of *Brassica rapa* chinensis based on the spectral parameters. *Acta Ecol. Sin.* **2015**, *35*, 4445–4453.
- Shi, R.J.; Pan, X.Z.; Wang, C.K.; Liu, Y.; Li, Y.L.; Li, Z.T. Prediction of Cadmium Content in the Leaves of Navel Orange in Heavy Metal Contaminated Soil Using VIS-NIR Reflectance Spectroscopy. *Spectrosc. Spectr. Anal.* **2015**, *35*, 3140–3145.
- Abdel-Rahman, M.E.; Mutanga, O.; Odindi, J.; Adam, E.; Odindo, A.; Ismail, R. Estimating Swiss chard foliar macro- and micronutrient concentrations under different irrigation water sources using ground-based hyperspectral data and four partial least squares (PLS)-based (PLS1, PLS2, SPLS1 and SPLS2) regression algorithms. *Comput. Electron. Agric.* **2016**, *132*, 21–33. [[CrossRef](#)]
- Rathod, H.P.; Brackhage, C.; Van der Meer, F.D.; Müller, I.; Noomen, M.F.; Rossiter, D.G.; Dudel, G.E. Spectral changes in the leaves of barley plant due to phytoremediation of metals—Results from a pot study. *Eur. J. Remote Sens.* **2015**, *48*, 283–302. [[CrossRef](#)]
- Li, L.T.; Shen, F.M.; Ma, W.L.; Fan, J.; Li, Y.R.; Liu, H.T. Response Characteristics and Quantitative Monitoring Models Analyzed Using in situ Leaf Hyperspectra under Different Cd Stress Conditions. *Trans. Chin. Soc. Agric. Mach.* **2020**, *51*, 146–155.
- Wang, G.Q.; Deng, S.P.; Fang, Y.H.; Zheng, L.P.; Zhang, Y.; Lin, Y.S. Comparative Study on Soil Environmental Standards for Heavy Metals in China and Other Countries: Cadmium. *J. Ecol. Rural Environ.* **2015**, *31*, 808–821.
- Sun, J.; Zhang, Y.C.; Mao, H.P.; Wu, X.H.; Chen, Y.; Weng, Q.P. Responses Analysis of Lettuce Leaf Pollution in Cadmium Stress Based on Computer Vision. *Trans. Chin. Soc. Agric. Mach.* **2018**, *49*, 166–172.
- Kang, L.; Yuan, J.Q.; Gao, R.; Kong, Q.M.; Jia, Y.J.; Su, Z.B. Early Detection and Identification of Rice Blast Based on Hyperspectral Imaging. *Spectrosc. Spectr. Anal.* **2021**, *41*, 898–902.
- Lei, X.Y.; Hu, Y.; Deng, G.Q.; Li, H.G.; Zhang, Y.L.; Hu, D.; Li, L.H. Effects of Passivator on Cadmium Content and Physiological Characteristics of *Lilium lancifolium*. *Chin. Agric. Sci. Bull.* **2022**, *38*, 66–72.
- Wang, G.; Zhang, F.Y.; Li, M.H.; Wang, J.L.; Wang, Y.B.; Wu, C.W. Air quality monitoring systems based on PLSR algorithm. *Transducer Microsyst. Technol.* **2022**, *41*, 37–40+49.
- Lu, X.H.; Gong, R.X.; Zhang, H.N.; Wang, Q.; Zhang, J.; Xie, R.X. Estimation of Photosynthetic Parameters of *Cinnamomum camphora* in Dwarf Forest Based on UAV Multi-spectral Remote Sensing. *Trans. Chin. Soc. Agric. Mach.* **2023**, *54*, 179–187.
- Zhao, H.H.; Liu, P.J.; Qiao, B.J.; Wu, K.N. The Spatial Distribution and Prediction of Soil Heavy Metals Based on Measured Samples and Multi-Spectral Images in Tai Lake of China. *Land* **2021**, *10*, 1227. [[CrossRef](#)]

27. Wang, X.; Zhang, H.R.; Wu, D.D.; Wang, W.N.; Wang, L.Q.; Luo, S.N.; Yu, D.Y. Fast Quantification of Phosphorus in Crude Soybean Oil by Near-Infrared Spectroscopy. *Food Sci.* **2023**, *44*, 316–322.
28. Huang, Y.Y.; Yao, J. Prediction of rice moisture content based on PSO-GA-BP neural network. *Cereal Feed Ind.* **2023**, 56–60.
29. Yang, C.E.; Su, L.; Feng, W.Z.; Zhou, J.Y.; Wu, H.W.; Yuan, Y.M.; Wang, Q. Identification of *Pleurotus Ostreatus* from Different Producing Areas Based on Mid-Infrared Spectroscopy and Machine Learning. *Spectrosc. Spectr. Anal.* **2023**, *43*, 577–582.
30. Hu, M.W.; Zou, L.; Lu, J.; Yang, Z.Y.; Chen, Y.N.; Xu, Y.Z.; Sun, C.H. Construction of a 5-feature gene model by support vector machine for classifying osteoporosis samples. *Bioengineered* **2021**, *12*, 6821–6830. [\[CrossRef\]](#)
31. Xie, Y.P.; Tong, X.G.; Wang, X.H. Early Monitoring of Rice Koji Disease Based on Hyperspectroscopy. *Trans. Chin. Soc. Agric. Mach.* **2023**, *54*, 288–296.
32. Du, Y.P.; Li, H.J.; Yin, K.L.; Zhai, H. Cadmium accumulation, subcellular distribution, and chemical forms in *Vitis vinifera* cv. Chardonnay grapevine. *Chin. J. Appl. Ecol.* **2012**, *23*, 1607–1612.
33. Yu, P.; Gao, F.; Liu, J.; Liang, Q.; Han, Y.Y.; Wang, J.X.; Jia, Y.H. Effect of Cd on Plant Growth and Its Tolerance Mechanism. *Chin. Agric. Sci. Bull.* **2017**, *33*, 89–95.
34. Newete, S.W.; Erasmus, B.F.; Weiersbye, I.M.; Cho, M.A.; Byrne, M.J. Hyperspectral reflectance features of water hyacinth growing under feeding stresses of *Neochetina* spp. and different heavy metal pollutants. *Int. J. Remote Sens.* **2014**, *35*, 799–817. [\[CrossRef\]](#)
35. Feng, W.; Guo, T.C.; Xie, Y.X.; Wang, Y.H.; Zhu, Y.J.; Wang, C.Y. Spectrum Analytical Technique and Its Applications for the Crop Growth Detection. *Chin. Agric. Sci. Bull.* **2009**, *25*, 182–188.
36. Zhou, L.; Wu, H.; Jing, T.; Li, T.; Li, J.; Kong, L.; Zhou, L. Estimation of Relative Chlorophyll Content in Lettuce (*Lactuca sativa* L.) Leaves under Cadmium Stress Using Visible—Near-Infrared Reflectance and Machine-Learning Models. *Agronomy* **2024**, *14*, 427. [\[CrossRef\]](#)
37. Chen, X.J.; Tao, H.F.; Wu, Y.Z.; Xu, X.M. Effects of Cadmium on metabolism of photosynthetic pigment and photosynthetic system in *Lactuca sativa* L. revealed by physiological and proteomics analysis. *Sci. Hortic.* **2022**, *305*, 111371. [\[CrossRef\]](#)
38. Zhao, J.L.; Zhang, F.; Wan, X.Q.; Xiao, Z. Cadmium tolerance and enrichment characteristics of *Viola prionantha*. *Pratacultural Sci.* **2016**, *33*, 54–60.
39. Liu, S.; Yu, H.Y.; Sui, Y.Y.; Kong, L.J.; Yu, Z.D.; Guo, J.J.; Qiao, J.L. Hyperspectral Data Analysis for Classification of Soybean Leaf Diseases. *Spectrosc. Spectr. Anal.* **2023**, *43*, 1550–1555.
40. Xin, Z.; Jun, S.; Yan, T.; Sheng, C.Q.; Hong, W.X.; Ying, H.Y. A deep learning based regression method on hyperspectral data for rapid prediction of cadmium residue in lettuce leaves. *Chemom. Intell. Lab. Syst.* **2020**, *200*, 103996. [\[CrossRef\]](#)
41. Huang, W.C.; Liu, H.Y.; Zhang, Y.; Mi, R.W.; Tong, C.G.; Xiao, W.; Shuai, B. Railway dangerous goods transportation system risk identification: Comparisons among SVM, PSO-SVM, GA-SVM and GS-SVM. *Appl. Soft Comput. J.* **2021**, *109*, 107541. [\[CrossRef\]](#)
42. Pullanagari, R.R.; Yule, J.I.; Tuohy, P.M.; Hedley, M.J.; Dynes, R.A.; King, W.M. In-field hyperspectral proximal sensing for estimating quality parameters of mixed pasture. *Precis. Agric.* **2012**, *13*, 351–369. [\[CrossRef\]](#)
43. Li, X.; Zhang, Y.; Bao, Y.; Luo, J.; Jin, X.; Xu, X.; Song, X.; Yang, G. Exploring the Best Hyperspectral Features for LAI Estimation Using Partial Least Squares Regression. *Remote Sens.* **2014**, *6*, 6221–6241. [\[CrossRef\]](#)
44. Villatoro-Pulido, M.; Moreno, R.R.; Muñoz-Serrano, A.; Cardeñosa, V.; Amaro, L.M.; Font, R.; Del, R.M. Characterization and prediction by near-infrared reflectance of mineral composition of rocket (*Eruca vesicaria* subsp. *sativa* and *Eruca vesicaria* subsp. *vesicaria*). *J. Sci. Food Agric.* **2012**, *92*, 1331–1340. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Font, R.; Vélez, D.; Río-Celestino, D.M.; Haro-Bailón, D.A.; Montoro, R. Screening Inorganic Arsenic in Rice by Visible and Near-Infrared Spectroscopy. *Microchim. Acta* **2005**, *151*, 231–239. [\[CrossRef\]](#)
46. Font, R.; Río, D.M.; Vélez, D.; Montoro, R.; Haro, D.A. Use of near-infrared spectroscopy for determining the total arsenic content in prostrate amaranth. *Sci. Total Environ.* **2004**, *327*, 93–104. [\[CrossRef\]](#)

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.