



Article

Non-Destructive Prediction of Anthocyanin Content of *Rosa chinensis* Petals Using Digital Images and Machine Learning Algorithms

Xiu-Ying Liu ^{1,2,*}, Jun-Ru Yu ^{3,†} and Heng-Nan Deng ¹

- ¹ School of Resources and Environmental Engineering, Mianyang Normal University, Mianyang 621000, China; 15283892633@163.com
- ² College of Agronomy, Henan University of Science and Technology, Luoyang 471023, China
- ³ College of Resources and Environment, Northwest A&F University, Xianyang 712100, China; yujunru@nwafu.edu.cn
- * Correspondence: csfulxy@126.com
- † These authors contributed equally to this work.

Abstract: Anthocyanins are widely found in plants and have significant functions. The accurate detection and quantitative assessment of anthocyanin content are essential to assess its functions. The anthocyanin content in plant tissues is typically quantified by wet chemistry and spectroscopic techniques. However, these methods are time-consuming, labor-intensive, tedious, expensive, destructive, or require expensive equipment. Digital photography is a fast, economical, efficient, reliable, and non-invasive method for estimating plant pigment content. This study examined the anthocyanin content of *Rosa chinensis* petals using digital images, a back-propagation neural network (BPNN), and the random forest (RF) algorithm. The objective was to determine whether using RGB indices and BPNN and RF algorithms to accurately predict the anthocyanin content of *R. chinensis* petals is feasible. The anthocyanin content ranged from 0.832 to 4.549 $\mu\text{mol g}^{-1}$ for 168 samples. Most RGB indices were strongly correlated with the anthocyanin content. The coefficient of determination (R^2) and the ratio of performance to deviation (RPD) of the BPNN and RF models exceeded 0.75 and 2.00, respectively, indicating the high accuracy of both models in predicting the anthocyanin content of *R. chinensis* petals using RGB indices. The RF model had higher R^2 and RPD values, and lower root mean square error (RMSE) and mean absolute error (MAE) values than the BPNN, indicating that it outperformed the BPNN model. This study provides an alternative method for determining the anthocyanin content of flowers.

Keywords: RGB indices; random forest; back-propagation neural network; non-invasive prediction; flower



Citation: Liu, X.-Y.; Yu, J.-R.; Deng, H.-N. Non-Destructive Prediction of Anthocyanin Content of *Rosa chinensis* Petals Using Digital Images and Machine Learning Algorithms. *Horticulturae* **2024**, *10*, 503. <https://doi.org/10.3390/horticulturae10050503>

Academic Editor: Yongfeng Zhou

Received: 6 April 2024
Revised: 11 May 2024
Accepted: 11 May 2024
Published: 13 May 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

Anthocyanins are phenolic water-soluble glycosides or acyl-glycosides of anthocyanidins [1]. They are widely distributed in plants [2]. Anthocyanins are water-soluble flavonoid pigments [3] that accumulate in various organs and are typically stored in vacuoles in the epidermis or mesophyll [2,4]. They contribute to orange-to-blue colors but primarily provide red, purple, or blue hues to leaves, fruits, and flowers [5,6]. The color primarily depends on the anthocyanin type and content [7], pH, co-pigments, and metal ions [2]. Anthocyanins are important secondary plant metabolites produced from the amino acid phenylalanine through the anthocyanin biosynthetic pathway [8]. Their biosynthesis can be affected by biotic or abiotic stresses, such as nutrient deficiency, wounding, pathogens, drought, light, salinity, cold, and ultraviolet (UV) irradiation [7,9,10]. Anthocyanins fulfill essential physiological functions related to adaptation and protection against stresses [5,9]. Accurate detection and quantitative assessment of anthocyanins can provide valuable

information on the physiological responses and adaptation of plants to environmental stresses [11].

Wet chemistry is the most common method for quantifying the anthocyanin content in plant tissues. This method is highly accurate [12] but time-consuming, labor-intensive, tedious, expensive, and destructive [9,13]. Due to advances in spectroscopy and computer analyses, spectroscopic techniques have been widely used to detect the anthocyanin content of plants. Neto et al. [14] predicted the leaf anthocyanin content of *Lactuca sativa* using partial least squares regression (PLSR) and visible and near-infrared (NIR) spectroscopy. Liu et al. [11] used visible and NIR spectroscopy, principal component regression (PCR), PLSR, and a back-propagation neural network (BPNN) to predict the leaf anthocyanin content of *Prunus cerasifera*. These methods are simple, sensitive, non-invasive, and efficient [11], but the equipment is relatively expensive, and the environmental requirements are high [8]. Digital photography has been increasingly used to analyze plant color and quantify the pigment content based on color parameter values extracted from digital images acquired by digital cameras. This method is widely used because it is fast, economical, efficient, reliable, and non-invasive [5,7,15]. For example, Yang et al. [13] calculated six color parameters in both RGB and HIS color space from digital images of *L. sativa* leaves to generate 37 color indices, and then prediction models were developed based on these color indices using curve estimation to predict the anthocyanin content. Del Valle et al. [7] extracted RGB values from digital images to generate 12 color indices, then utilized these indices to construct models to estimate relative anthocyanin concentrations in species with color variations via PSLR. Askey et al. [8] computed color index values from digital images of *Arabidopsis thaliana* leaves across five color spaces, and developed models to predict the anthocyanin content utilizing twenty-two regression models. These studies were all based on digital images and utilized color index values obtained from various color spaces, employing different modeling methods. However, they mainly focus on plant leaves and rarely use machine learning to build models.

Machine learning is a branch of artificial intelligence, widely used to construct estimation models. Mathematical or computer algorithms are employed to train a computational model to solve a problem or perform complex tasks based on input parameters [16]. The algorithm learns to perform tasks based on input data. This method has been used for pattern recognition, classification, and prediction. Machine learning algorithms have high accuracy, automation, and speed, they can be customized and applied at different scales [17], providing excellent performance.

Machine learning algorithms include BPNN, support vector regression (SVR), random forest (RF), extreme learning machine (ELM), and Cubist [18]. BPNN and RF algorithms are two commonly used machine learning algorithms. BPNN is a multilayer feed-forward neural network that corrects errors using a back-propagation algorithm [19]. It has strong nonlinear mapping, self-learning, self-adaptive, and generalization capabilities and high fault tolerance [19,20] and is suitable for regression or classification problems. Therefore, it is the most widely used neural network model and is particularly useful for solving nonlinear problems [21]. RF is a supervised machine learning ensemble algorithm based on the if-then-else rules. It was proposed by Breiman [22] and is known for its robustness, ability to handle high-dimensional data, and resistance to overfitting, noise, and outliers [23]. It is insensitive to collinearity [24] and effective in handling high-dimensional data and covariance among variables [25]. Furthermore, RF provides high prediction accuracy with low computational complexity due to random sampling [26]. Thus, it is a popular machine learning algorithm for classification and prediction.

Rosa chinensis is a popular flower worldwide. It originated in China and was spread from the Silk Road to Persia, Ceylon, and other countries [27]. Since this flower blooms year-round and produces flowers with diverse colors [28], it has been widely planted and cultivated as an ornamental plant. This flower also has many other values, e.g., cultural [29], medicinal [30,31], and edible [32,33]. These values are attributed to the abundance of an-

thocyanin in the petals, but the anthocyanin content significantly affects these values. Thus, estimating the anthocyanin content of *R. chinensis* petals is essential to assess these values.

This study predicts the anthocyanin content of *R. chinensis* petals using RGB indices and BPNN and RF algorithms. The objective is to investigate the feasibility of using RGB indices and BPNN and RF algorithms to predict the anthocyanin content of *R. chinensis* petals accurately. We hypothesize that RGB indices combined with BPNN and RF can accurately predict the anthocyanin content of *R. chinensis* petals.

2. Materials and Methods

2.1. Plant Materials

A total of 504 petals (3 petals per flower) of *R. chinensis*, ranging from pink to red (Figure 1), were collected on the campus of Henan University of Science and Technology (34.62° N, 112.46° E). Three petals of one flower represented 1 sample, resulting in 168 samples. The petals were immediately sealed in numbered plastic bags and placed in an insulated box with ice cubes to prevent water evaporation. Healthy and homogeneously colored petals without visible symptoms of damage were used for the experiments.

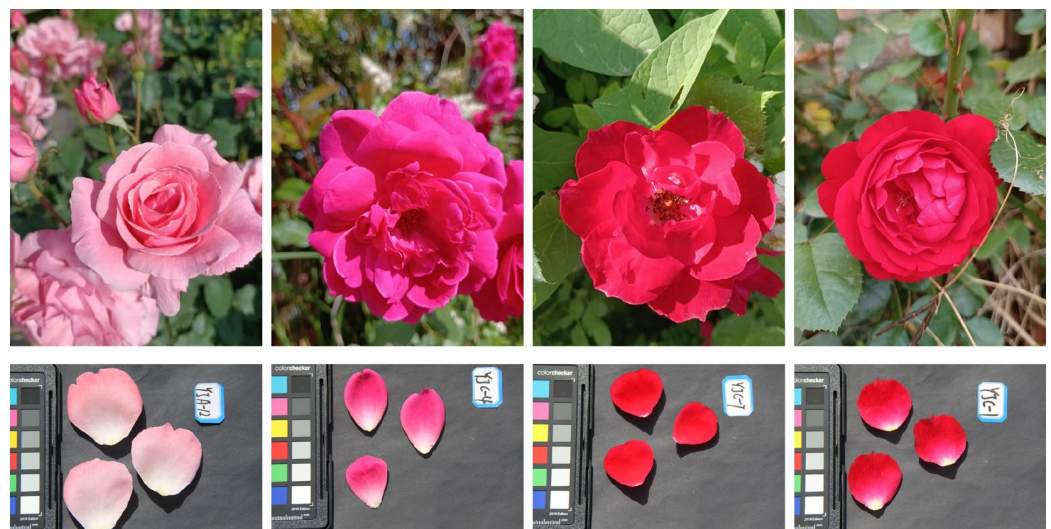


Figure 1. The flowers and petals of *R. chinensis* with different colors.

2.2. Digital Image Acquisition

Images of the samples were captured immediately after petal collection using a digital still color camera (EOS 500D, Canon Inc., Tokyo, Japan) with an EF 100 mm autofocus lens. This camera has a 22.3 × 14.9 mm CMOS sensor (4752 × 3168 pixels). The camera was mounted on a tripod at the nadir position at a constant height of 0.4 m above the top of the petals. Aperture priority mode was selected. The aperture was f/5.6, with ISO 100, white balance fixed at 4900 K, autoexposure, autofocus, and no flash. The sensitivity was manually set to 1600. All samples were photographed with a ColorChecker Passport (X-Rite Inc., Grand Rapids, MI, USA) for standardization for different light conditions. The images were acquired from 11:00 to 14:00 on a sunny day and stored in the Canon raw image (CR2) file format. This format contains unprocessed data that can be linearized using specialized software [7]. A total of 168 images were acquired.

2.3. Image Processing and RGB Index Construction

The “Image Calibration and Analysis Toolbox” [34], a freeware plugin for ImageJ software (version 1.53e) [35], was used for image processing. The method developed by Del Valle et al. [7] was used for image calibration to linearize the RGB values. Nine (three per petal) regions of interest (ROIs) were randomly selected in each sample, the values of the RGB channels were extracted, and the mean RGB values were calculated.

Various RGB indices [7,36,37] were constructed to analyze the colors and estimate the pigment content. In this study, 31 RGB indices (Table 1) were calculated from the mean RGB pixel values.

Table 1. RGB indices used in this study. R, the value of the R channel; G, the value of the G channel; B, the value of the B channel.

Index	Index	Index	Index
R	R/B	$(R - B)/(R + G + B)$	$(G + B - R)/2B$
G	G/B	$(R - G)/(R + G + B)$	$G/B + R$
B	G/R	$(G - B)/(R + G + B)$	$B/G + R$
$(R + B + G)/3$	B/R	$(G + B - R)/2R$	$R/B + G$
$R - B$	B/G	$(R - B)/(R + B)$	$(R + B)/2 - G$
$R - G$	$R/(R + G + B)$	$(R - G)/(R + G)$	$(B + R)/G$
$G - B$	$G/(R + G + B)$	$(G - B)/(G + B)$	$G/((B + R)/2)$
R/G	$B/(R + G + B)$	$(G + B - R)/2G$	

2.4. Anthocyanin Content Measurement

The anthocyanin content of the petals was measured following Xiong et al. [38]. In brief, the petals were chopped, and about 0.20 g of the chopped petals was transferred to 20 mL test tubes, followed by adding 10 mL of 0.1 mol L⁻¹ hydrochloric acid ethanol solution to extract the anthocyanin. The chopped petals were separated from the solvent by filtering, and the extraction was repeated until pale petals were obtained. The absorbance of the extracts was immediately measured with a spectrophotometer. The final anthocyanin content was expressed as a function of the petal amount (i.e., $\mu\text{mol g}^{-1}$).

2.5. Model Construction and Validation

A BPNN consists of three parts: input, hidden, and output layers. The input layer consists of k neurons, where k represents the number of input variables. The hidden layer converts the input into a format compatible with the output layer. The output layer consists of m neurons, where m represents the number of output variables. The accuracy of the BPNN depends on the number of hidden layer neurons, which is usually determined by the number of input and output layers. Equation (1) was used to calculate the number of neurons in the hidden layer. The mean square error (MSE) (Equation (2)) was used to evaluate the performance of the BPNN model. A lower MSE indicates better results [39].

$$q = \sqrt{(k + m) + X} \quad (1)$$

$$\text{MSE} = 1/n \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2)$$

where q is the number of hidden layer nodes, k is the number of input layer nodes, m is the number of output layer nodes, and X is a constant ranging from 1 to 10. MSE is the mean square error, n is the number of samples, y_i is the measured value of the i -th sample, and \hat{y}_i is the predicted value of the i -th sample.

The framework of the BPNN is shown in Figure 2.

RF is a methodology that uses data as input to make predictions. It generates multiple decision trees by randomly selecting subsets of input variables and subsets of the training data. Each decision tree makes independent predictions, and the final prediction is obtained by averaging or voting the predictions of all trees. The performance of the model is optimized by adjusting two parameters, $n\text{tree}$ and $m\text{try}$. The $n\text{tree}$ denotes the number of decision trees, and $m\text{try}$ is the number of random variables in each data set. The larger the $n\text{tree}$, the better the classification performance of the RF model, but the slower the processing speed. A smaller $m\text{try}$ value may cause overfitting, whereas a larger $m\text{try}$ value increases model diversity and reduces the risk of overfitting. Therefore, $n\text{tree}$ and

mtry must be optimized. The root mean square error (RMSE) (Equation (3)) was used to evaluate the optimization result. A lower RMSE indicates better results. The construction process of RF is shown in Figure 3.

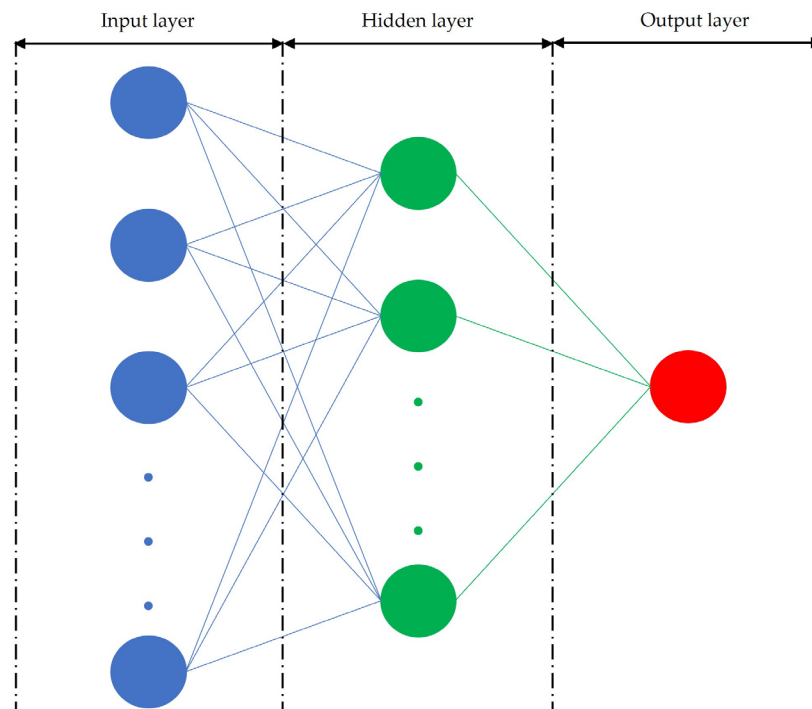


Figure 2. Framework of back-propagation neural network.

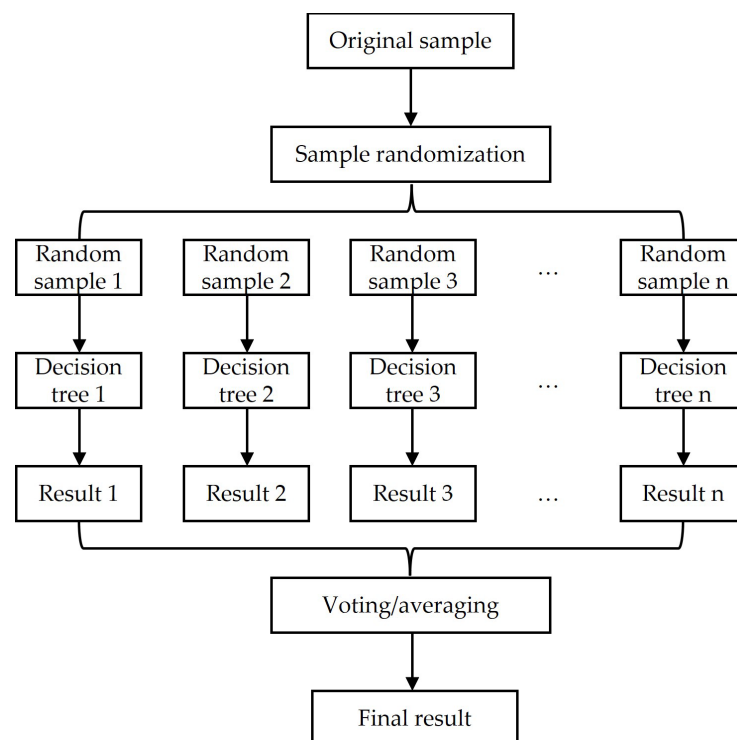


Figure 3. Framework of random forest.

The coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and the ratio of performance to deviation (RPD) were used to evaluate model performance (Equations (3)–(6)). The R^2 represents the proportion of the variance in the

dependent variable that is explained by the independent variable. The closer the R^2 is to 1, and the lower the RMSE and MAE, the better the model's prediction performance. The RPD is the ratio of the measured value's standard deviation to the predicted RMSE. The RPD values were classified as follows: an $RPD < 1.4$ indicated that the model was unreliable, $1.4 \leq RPD < 2.0$ suggested moderate reliability, and $RPD > 2.0$ indicated exceptional prediction ability [8,11,26]. Generally, models with higher R^2 and RPD values and smaller RMSE and MAE values have higher prediction accuracy and stability [40,41].

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (3)$$

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (5)$$

$$RPD = SD / RMSE \quad (6)$$

where \hat{y} is the predicted values; \bar{y} is the mean of the observed values; y is the observed values; n is the number of predicted/observed values with $i = 1, 2, \dots, n$; SD is the standard deviation of the validation or calibration set; and RMSE is the root mean square error of the validation or calibration set.

Pearson correlation analysis was used to determine the relationship between the anthocyanin content and the 31 indices. Collinearity analysis was used to determine the collinearity among the RGB indices using the variance inflation factor (VIF) (Equation (7)).

$$VIF_i = 1 / (1 - r_i^2) \quad (7)$$

where r_i^2 is the R^2 of a regression of regressor x_i on all remaining regressors. When a regressor is orthogonal, $VIF = 1$. VIFs larger than 10 ($r_i^2 > 0.90$) indicate collinearity [42].

The data set was divided into two groups. Two-thirds (112) were randomly selected as the calibration set to construct the models. The remaining one-third (56) was used as the validation set for estimating model prediction accuracy. Pearson correlation analysis and collinearity analysis were performed using SPSS software (version 25.0, IBM Corporation, Armonk, NY, USA). Only significantly correlated RGB indices were used to construct the models. The BPNN and RF models were implemented in MATLAB (version R2016a) and R (version 4.3.2) software, respectively.

3. Results

3.1. Correlation between Anthocyanin Content and Color Indices

The descriptive statistics of the anthocyanin content for the 168 samples are presented in Table 2. The anthocyanin content varied widely, ranging from 0.832 to 4.549 $\mu\text{mol g}^{-1}$ of fresh weight, with a mean value of $2.935 \pm 1.004 \mu\text{mol g}^{-1}$. The calibration set's anthocyanin content ranged from 0.832 to 4.549 $\mu\text{mol g}^{-1}$, with a mean value of $2.924 \pm 1.007 \mu\text{mol g}^{-1}$. The validation set's anthocyanin content ranged from 0.931 to 4.542 $\mu\text{mol g}^{-1}$, with a mean value of $2.958 \pm 1.005 \mu\text{mol g}^{-1}$. The mean of the validation set was slightly higher than the means of the entire set and the calibration set, but the differences were not significant ($p < 0.05$), indicating that all were representative. The coefficient of variation of the entire set was 34%, indicative of moderate variability, which facilitated calibration.

The descriptive statistics of the RGB values for the 168 samples are presented in Table 3. The R, G, and B values ranged from 15.440 to 62.195, 2.842 to 48.647, and 1.856 to 30.307, with means of 33.033 ± 9.952 , 12.125 ± 8.913 , and 9.359 ± 6.026 , respectively.

Table 2. Descriptive statistics of anthocyanin content ($\mu\text{mol g}^{-1}$). SD, the standard deviation; CV, the coefficient of variation.

Data Set	Sample Size	Minimum	Maximum	Average	SD	CV (%)
Entire set	168	0.832	4.549	2.935	1.004	34
Calibration set	112	0.832	4.549	2.924	1.007	34
Validation set	56	0.931	4.542	2.958	1.005	34

Table 3. Descriptive statistics of RGB values. R, the value of the R channel; G, the value of the G channel; B, the value of the B channel; SD, the standard deviation; CV, the coefficient of variation.

Characters	Data Set	Sample Size	Minimum	Maximum	Mean	SD	CV (%)
R	Entire set	168	15.440	62.195	33.033	9.952	30
	Calibration set	112	16.412	61.041	33.482	9.774	29
	Validation set	56	15.440	62.195	32.134	10.328	32
G	Entire set	168	2.842	48.647	12.125	8.913	74
	Calibration set	112	2.842	48.647	12.103	8.630	71
	Validation set	56	3.146	45.227	12.168	9.535	78
B	Entire set	168	1.856	30.307	9.359	6.026	64
	Calibration set	112	1.856	30.307	9.407	6.123	65
	Validation set	56	2.135	27.008	9.264	5.881	63

The Pearson correlation coefficients between the anthocyanin content and the 31 RGB indices are shown in Figure 4. Twenty-eight RGB indices had strong and significant correlations with the anthocyanin content ($p < 0.01$). $R - B$, G/B , and $(G - B)/(R + G + B)$ had positive and non-significant correlations ($p > 0.05$). Among the 28 significantly correlated RGB indices, R , G , B , $(R + B + G)/3$, $G - B$, G/R , B/R , B/G , $G/(R + G + B)$, $B/(R + G + B)$, $(G + B - R)/2R$, $(G + B - R)/2G$, $G/B + R$, $B/G + R$, $R/B + G$, and $G/((B + R)/2)$ were negatively correlated with the anthocyanin content. The remaining 12 RGB indices were positively correlated with the anthocyanin content. The correlations between the anthocyanin content and B , R/G , G/R , $R/(R + G + B)$, $(R - G)/(R + G)$, $(R - B)/(R + G + B)$, $(R - G)/(R + G + B)$, and $(G + B - R)/2R$ were stronger than for the other indices ($|r| \geq 0.800$ for all). The highest correlation coefficient (r) (-0.844) occurred between the anthocyanin content and $(G + B - R)/2R$.

Based on the collinearity analysis, 18 RGB indices were excluded. Nine of the ten predictor RGB indices had VIFs exceeding 10 (Table 4), indicating multicollinearity.

Table 4. Collinearity analysis results of ten predictor RGB indices. R, the value of the R channel; G, the value of the G channel; B, the value of the B channel; VIF, the variance inflation factor.

RGB Indices	VIFs
B	31.66
$R - G$	8.46
$G - B$	48.02
R/B	573.87
G/R	2365.63
B/R	739.06
B/G	377.11
$(G - B)/(G + B)$	1095.59
$(B + R)/G$	366.72
$G/((B + R)/2)$	1426.28

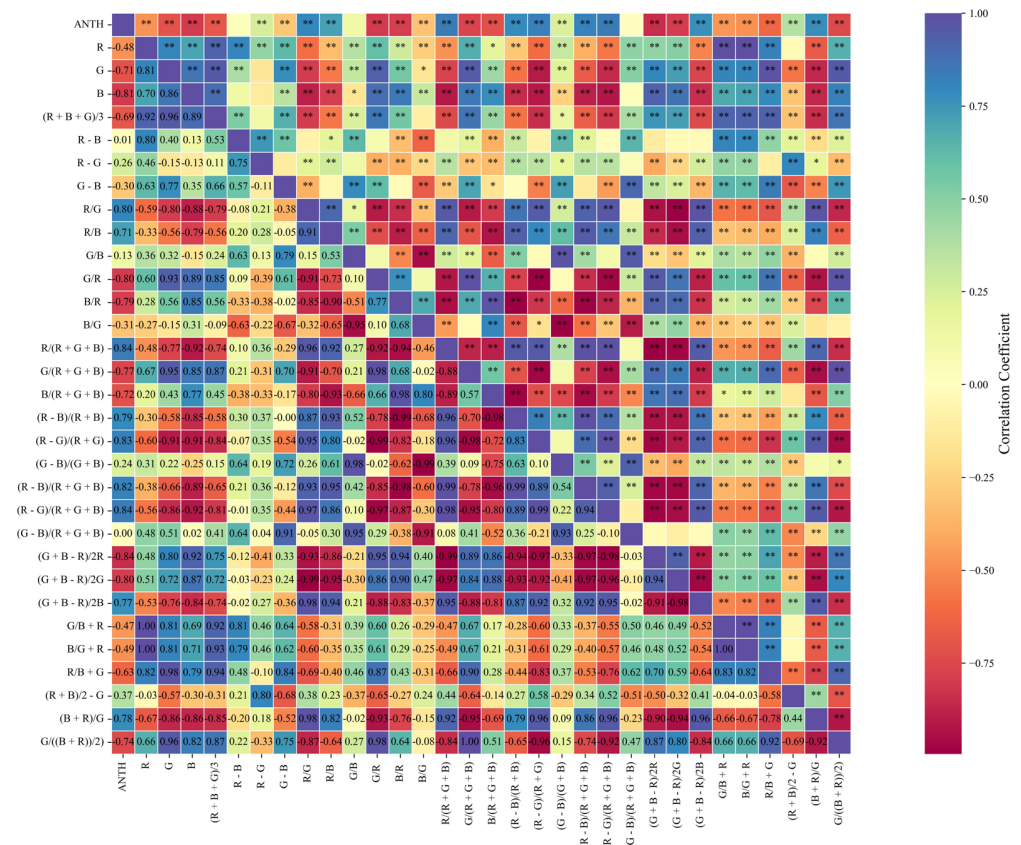


Figure 4. Heat map of Pearson correlation coefficients between anthocyanin content and RGB color indices. ANTH, the anthocyanin content; R, the value of the R channel; G, the value of the G channel; B, the value of the B channel; * and ** are significant at 0.05 and 0.01 levels, respectively.

3.2. Parameter Optimization Results of BPNN and RF

The 28 RGB indices that were strongly correlated with the anthocyanin content were utilized to establish the prediction models.

The indices were used as input to the BPNN model, and the anthocyanin content was the output variable; $k = 28$ and $m = 1$. The MSE of the model was calculated for different numbers of hidden layer nodes (Table 5). The model with 12 hidden layer nodes had the lowest MSE and the best performance.

Table 5. The training times and errors of BPNN with different hidden layer nodes. BPNN, back-propagation neural network; MSE, the mean square error.

The Number of Hidden Layer Nodes	Training Times	MSE
6	30	0.044
7	30	0.043
8	30	0.034
9	30	0.044
10	30	0.055
11	30	0.036
12	30	0.015
13	30	0.214
14	30	0.032
15	30	0.056

The two parameters ($mtry$ and $mtree$) of the RF model were optimized based on the RMSE. The $mtree$ values ranged from 100 to 500 with an interval of 100, and the range of the

mtry values was 2 to 20 with an interval of 2. The results are shown in Figure 5. The RMSE ($0.235 \mu\text{mol g}^{-1}$) is the lowest when the *ntree* was 100 and *mtry* was 6.

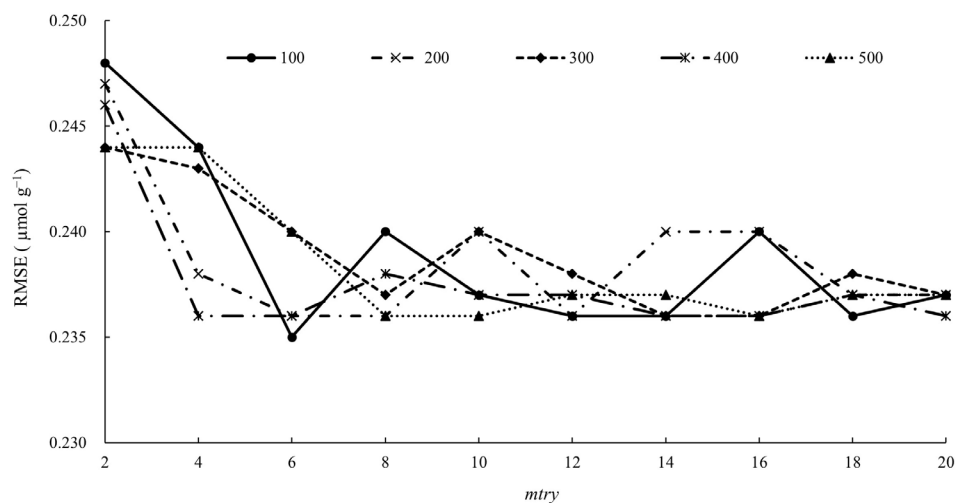


Figure 5. Optimization of RF parameters (*ntree* and *mtry*) using RMSE. RF, random forest; RMSE, the root mean square error.

3.3. Performance of BPNN and RF for Anthocyanin Content Prediction

The results of the BPNN and RF anthocyanin content prediction models are listed in Table 6. The R^2 , RMSE, MAE, and RPD values of the BPNN were 0.784, 0.471, 0.334, and 2.138 for the calibration set and 0.781, 0.475, 0.326 μmol , and 2.116 for the validation set, respectively. The R^2 , RMSE, MAE, and RPD values of the RF were 0.946, 0.235, 0.176, and 4.285 for the calibration set and 0.958, 0.208, 0.152, and 4.832 for the validation set, respectively. The R^2 and RPD values of the RF were 20.66% and 100.42% higher than those of the BPNN for the calibration set, and the RMSE and MAE values were 50.11% and 47.31% lower, respectively. The R^2 and RPD values of the RF were 22.66% and 128.36% higher than those of the BPNN for the validation set, respectively, and the RMSE and MAE values were 56.21% and 53.37% lower, respectively.

Table 6. Performance comparison of BPNN and RF. BPNN, back-propagation neural network; RF, random forest; R^2 , the coefficient of determination; RMSE, root mean square error; MAE, mean absolute error; RPD, the ratio of performance to deviation.

Models	Calibration Set (n = 112)				Validation Set (n = 56)			
	R^2	RMSE	MAE	RPD	R^2	RMSE	MAE	RPD
BPNN	0.784	0.471	0.334	2.138	0.781	0.475	0.326	2.116
RF	0.946	0.235	0.176	4.285	0.958	0.208	0.152	4.832

Table 7 lists the values of some RGB indices and the anthocyanin contents obtained from the BPNN and RF models and wet chemistry. The results indicate that the RF model yielded an anthocyanin content predicted closer to the wet chemistry method than the BPNN.

Figure 6 shows the predicted and measured anthocyanin contents for the validation set. Although both models exhibited similar patterns, the RF model had a better prediction performance (RPD = 4.832). The points were close to the 1:1 line, with fewer outliers than for the BPNN model.

Table 7. The part RGB indices and anthocyanin content obtained by wet chemistry, BPNN, and RF in some samples. R, the value of the R channel; G, the value of the G channel; B, the value of the B channel; BPNN, back-propagation neural network; RF, random forest.

Samples	RGB Indices						Anthocyanin Content		
	R	G	B	R/G	R/B	G/R	Wet Chemistry	BPNN	RF
YJA-12	60.282	48.647	30.307	1.239	1.989	0.807	0.832	1.090	1.008
YJA-06	39.221	22.225	23.215	1.765	1.689	0.567	1.044	1.112	1.056
YJB-08	38.667	12.798	12.642	3.021	3.059	0.331	1.130	2.464	1.802
YJC-4	26.744	7.447	5.437	3.591	4.919	0.278	3.428	3.302	3.409
YJC-7	17.211	3.584	2.699	4.802	6.377	0.208	3.767	3.654	3.757
YJC-1	15.440	3.223	2.326	4.791	6.639	0.209	3.770	3.583	3.760

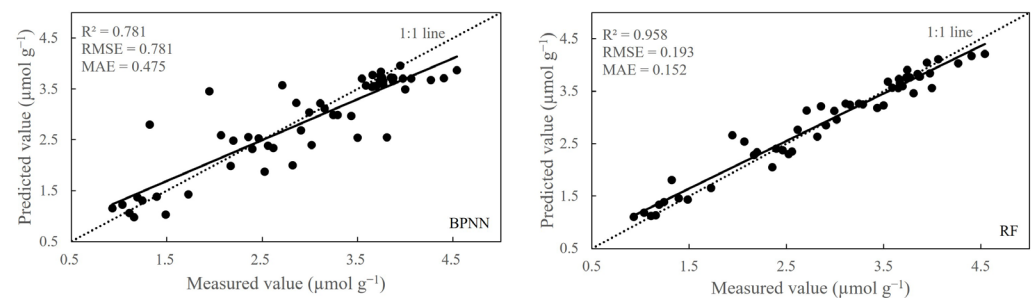


Figure 6. Validation of predicted and measured values of anthocyanins content in *Rosa chinensis* petals. The black line is the regression line, while the dot line is a 1:1 relation line. BPNN, back-propagation neural network; RF, random forest; R^2 , the coefficient of determination; RMSE, root mean square error; MAE, mean absolute error.

4. Discussion

Using the RGB color space is the most common approach to describe a color quantitatively [43]. The RGB value refers to the sum of the three channels (R, G, B) [44], where R, G, and B denote the mean values of the red, green, and blue channels. The digital images can be acquired using a digital camera, smartphone, and scanner. Image processing methods are used to extract the RGB values from digital images and construct RGB indices [45]. Thus, digital photography is a simple, quick, and low-cost method that has been widely used to predict the content of plant pigments. For example, Hassanjililian et al. [46] used RGB indices to predict the leaf chlorophyll content of *Glycine max*. Wood et al. [47] observed that the RGB indices enabled the estimation of the *a*, *b*, and total chlorophyll concentrations of microalgal cultures in situ. Taha et al. [48] demonstrated the feasibility of using RGB indices to estimate the chlorophyll content of lettuce. In this study, the correlations between 28 RGB indices derived from digital camera images and the anthocyanin content were strong, indicating that these indices were suitable for establishing predictive models of the anthocyanin content. The R^2 and RPD values of the BPNN and RF models were greater than 0.75 and 2.00 (Table 5), respectively. An R^2 value higher than 0.7 is indicative of a high-fitting model that explains 70% of the variance [49]. The RPD value exceeding 2.0 indicates that the models had exceptional prediction ability [11,26,50]. These findings imply that predicting the anthocyanin content of *R. chinensis* petals using RGB color indices derived from digital images combined with BPNN and RF is feasible. Both models exhibited excellent robustness and high predictive ability.

Machine learning algorithms and RGB images have been successfully used to predict plant pigment content, such as the anthocyanin content of *A. thaliana* leaves [8] and the chlorophyll content of *G. max* leaves [46]. BPNN and RF are machine learning algorithms. BPNN learns complex nonlinear relationships by iteratively adjusting the weights to minimize the error between the predicted and measured results [51]. RF uses multiple decision trees during training and combines their predictions to improve accuracy [22].

This study utilized BPNN and RF to predict the anthocyanin content of *R. chinense* petals. The RF model had higher R^2 and RPD values and lower RMSE and MAE values than the BPNN for the calibration and validation sets (Table 7). These results indicated that the RF algorithm outperformed the BPNN algorithm, consistent with previous findings. For example, Guo et al. [52] established prediction models for the leaf chlorophyll content of maize and found that the RF algorithm achieved better prediction results than the BPNN algorithm. Yang et al. [53] demonstrated that the RF outperformed the BPNN in predicting the chlorophyll content of trees in coniferous, broad-leaved, and mixed broad-leaved forests and of individual trees. The better performance of RF can be contributed to its insensitivity to multicollinearity, a common problem with RGB indices.

This study has some limitations. First, we used detached petals to enable their analysis under consistent light conditions. Thus, the prediction performance should be assessed for in situ conditions. Second, the petals were obtained from the same location. Future studies should select samples from larger areas. Third, it should be investigated whether images acquired with smartphones could be used with this method instead of cameras since smartphones are ubiquitous. The proposed improvements would improve the application of this method.

5. Conclusions

This study developed models to predict the anthocyanin content of *R. chinensis* petals using RGB indices derived from digital images combined with the BPNN and RF algorithms. Most RGB indices were correlated with the anthocyanin content. The R^2 and RPD values for the two algorithms exceeded 0.75 and 2.0, respectively. The RF model had higher R^2 and RPD values and lower RMSE and MAE values than the BPNN model. This study demonstrates that RGB indices derived from digital camera images can be used to estimate the anthocyanin content of *R. chinensis* petals using BPNN and RF algorithms. The results provide an alternative method for determining the anthocyanin content of flowers.

Author Contributions: Conceptualization, X.-Y.L. and J.-R.Y.; methodology, X.-Y.L., J.-R.Y. and H.-N.D.; software, X.-Y.L. and J.-R.Y.; validation, X.-Y.L., J.-R.Y. and H.-N.D.; formal analysis, X.-Y.L., J.-R.Y. and H.-N.D.; resources, X.-Y.L. and J.-R.Y.; data curation, X.-Y.L., J.-R.Y. and H.-N.D.; writing—original draft preparation, X.-Y.L., J.-R.Y. and H.-N.D.; writing—review and editing, X.-Y.L., J.-R.Y. and H.-N.D.; supervision, X.-Y.L.; project administration, X.-Y.L.; funding acquisition, X.-Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Sichuan Province, grant number 2022NSFSC0145, and the Scientific Research Foundation of Mianyang Normal University, grant number QD2021A27.

Data Availability Statement: The data presented in this study are available within the article.

Acknowledgments: We would like to extend our thanks to Jian-Li Xiong of Mianyang Normal University and Jia-Feng Hu, Lu Wang, and Ze-Jia Chen of Henan University of Science and Technology for their assistance in examining and measuring samples.

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

References

1. Alappat, B.; Alappat, J. Anthocyanin pigments: Beyond aesthetics. *Molecules* **2020**, *25*, 5500. [[CrossRef](#)]
2. Tanaka, Y.; Sasaki, N.; Ohmiya, A. Biosynthesis of plant pigments: Anthocyanins, betalains and carotenoids. *Plant J.* **2008**, *54*, 733–749. [[CrossRef](#)]
3. Strack, D. Plant biochemistry. In *Phenolic Metabolism*; Harborne, J.B., Dey, P.M., Eds.; Academic Press: London, UK, 1997; Volume 387–416.
4. Gould, K.; Davies, K.M.; Winefield, C. *Anthocyanins: Biosynthesis, Functions, and Applications*; Springer: New York, NY, USA, 2008.
5. Kim, C.; van Iersel, M.W. Image-based phenotyping to estimate anthocyanin concentrations in lettuce. *Front. Plant Sci.* **2023**, *14*, 1155722. [[CrossRef](#)]
6. Deepa, P.; Hong, M.; Sowndhararajan, K.; Kim, S. A review of the role of an anthocyanin, cyanidin-3-o-beta-glucoside in obesity-related complications. *Plants* **2023**, *12*, 3889. [[CrossRef](#)]

7. Del Valle, J.C.; Gallardo López, A.; Buide, M.L.; Whittall, J.B.; Narbona, E. Digital photography provides a fast, reliable, and noninvasive method to estimate anthocyanin pigment concentration in reproductive and vegetative plant tissues. *Ecol. Evol.* **2018**, *8*, 3064–3076. [[CrossRef](#)]
8. Askey, B.C.; Dai, R.; Lee, W.S.; Kim, J. A noninvasive, machine learning-based method for monitoring anthocyanin accumulation in plants using digital color imaging. *Appl. Plant Sci.* **2019**, *7*, e11301. [[CrossRef](#)]
9. Gitelson, A.A.; Chivkunova, O.B.; Merzlyak, M.N. Nondestructive estimation of anthocyanins and chlorophylls in anthocyanic leaves. *Am. J. Bot.* **2009**, *96*, 1861–1868. [[CrossRef](#)]
10. Shan, X.; Zhang, Y.; Peng, W.; Wang, Z.; Xie, D. Molecular mechanism for Jasmonate-induction of anthocyanin accumulation in *arabidopsis*. *J. Exp. Bot.* **2009**, *60*, 3849–3860. [[CrossRef](#)]
11. Liu, X.; Liu, C.; Shi, Z.; Chang, Q. Comparison of prediction power of three multivariate calibrations for estimation of leaf anthocyanin content with visible spectroscopy in *Prunus cerasifera*. *PeerJ* **2019**, *7*, e7997. [[CrossRef](#)]
12. Lee, J.; Rennaker, C.; Wrolstad, R.E. Correlation of two anthocyanin quantification methods: HPLC and spectrophotometric methods. *Food Chem.* **2008**, *110*, 782–786. [[CrossRef](#)]
13. Yang, X.; Zhang, J.; Guo, D.; Xiong, X.; Chang, L.; Niu, Q.; Huang, D. Measuring and evaluating anthocyanin in lettuce leaf based on color information. *IFAC-PapersOnLine* **2016**, *49*, 96–99. [[CrossRef](#)]
14. Neto, A.J.S.; Moura, L.D.O.; Lopes, D.D.C.; Carlos, L.D.A.; Martins, L.M.; Ferraz, L.D.C.L. Non-destructive prediction of pigment content in lettuce based on visible–NIR spectroscopy. *J. Sci. Food Agric.* **2017**, *97*, 2015–2022. [[CrossRef](#)]
15. Costa, C.; Schurr, U.; Loreto, F.; Menesatti, P.; Carpentier, S. Plant phenotyping research trends, a science mapping approach. *Front. Plant Sci.* **2019**, *9*, 1933. [[CrossRef](#)]
16. Russell, S.; Norvig, P. *Artificial Intelligence: A Modern Approach*, 4th ed.; Pearson Education, Inc.: London, UK, 2020; pp. 1–1069.
17. Liu, Y.; Li, J.; Liu, C.; Wei, J. Evaluation of cultivated land quality using attention mechanism-back propagation neural network. *PeerJ Comput. Sci.* **2022**, *8*, e948. [[CrossRef](#)]
18. Wei, G.; Li, Y.; Zhang, Z.; Chen, Y.; Chen, J.; Yao, Z.; Lao, C.; Chen, H. Estimation of soil salt content by combining UAV-borne multispectral sensor and machine learning algorithms. *PeerJ* **2020**, *8*, e9087. [[CrossRef](#)]
19. Fu, J.; Chang, Y.; Huang, B. Prediction and sensitivity analysis of CO₂ capture by amine solvent scrubbing technique based on BP neural network. *Front. Bioeng. Biotechnol.* **2022**, *10*, 907904. [[CrossRef](#)]
20. Lyu, Z.; Yu, Y.; Samali, B.; Rashidi, M.; Mohammadi, M.; Nguyen, T.N.; Nguyen, A. Back-propagation neural network optimized by k-fold cross-validation for prediction of torsional strength of reinforced concrete beam. *Materials* **2022**, *15*, 1477. [[CrossRef](#)]
21. Montanaro, G.; Petrozza, A.; Rustioni, L.; Cellini, F.; Nuzzo, V. Phenotyping key fruit quality traits in olive using RGB images and back propagation neural networks. *Plant Phenomics* **2023**, *5*, 61. [[CrossRef](#)]
22. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
23. Hu, X.; Zhang, J.; Xue, W.; Zhou, L.; Che, Y.; Han, T. Estimation of the near-surface ozone concentration with full spatiotemporal coverage across the Beijing-Tianjin-Hebei region based on extreme gradient boosting combined with a WRF-chem model. *Atmosphere* **2022**, *13*, 632. [[CrossRef](#)]
24. Wang, L.; Zhou, X.; Zhu, X.; Dong, Z.; Guo, W. Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. *Crop J.* **2016**, *4*, 212–219. [[CrossRef](#)]
25. Sun, X.; Yang, Z.; Su, P.; Wei, K.; Wang, Z.; Yang, C.; Wang, C.; Qin, M.; Xiao, L.; Yang, W.; et al. Non-destructive monitoring of maize lai by fusing UAV spectral and textural features. *Front. Plant Sci.* **2023**, *14*, 1158837. [[CrossRef](#)]
26. Zou, M.; Liu, Y.; Fu, M.; Li, C.; Zhou, Z.; Meng, H.; Xing, E.; Ren, Y. Combining spectral and texture feature of UAV image with plant height to improve lai estimation of winter wheat at jointing stage. *Front. Plant Sci.* **2024**, *14*, 1272049. [[CrossRef](#)]
27. Li, Y.; Pu, M.; Cui, Y.; Gu, J.; Chen, X.; Wang, L.; Wu, H.; Yang, Y.; Wang, C. Research on the isolation and identification of black spot disease of *rosa chinensis* in Kunming, China. *Sci. Rep.* **2023**, *13*, 8299. [[CrossRef](#)]
28. Han, Y.; Yu, J.; Zhao, T.; Cheng, T.; Wang, J.; Yang, W.; Pan, H.; Zhang, Q. Dissecting the genome-wide evolution and function of r2r3-myb transcription factor family in *Rosa chinensis*. *Genes* **2019**, *10*, 823. [[CrossRef](#)]
29. Yu, R.; Xiong, Z.; Zhu, X.; Feng, P.; Hu, Z.; Fang, R.; Zhang, Y.; Liu, Q. Rcsp11-rctaf15b regulates the flowering time of rose (*Rosa chinensis*). *Horti. Res.* **2023**, *10*, uhad083. [[CrossRef](#)] [[PubMed](#)]
30. Luo, Y.; Wang, H.; Li, X.; He, T.; Wang, D.; Wang, W.; Jia, W.; Lin, Z.; Chen, S. One injection to profile the chemical composition and dual-antioxidation activities of *Rosa chinensis* jacq. *J. Chromatogr. A* **2020**, *1613*, 460663. [[CrossRef](#)]
31. Cai, Y.; Xing, J.; Sun, M.; Zhan, Z.; Corke, H. Phenolic antioxidants (hydrolyzable tannins, flavonols, and anthocyanins) identified by LC-ESI-MS and MALDI-QIT-TOF MS from *Rosa chinensis* flowers. *J. Agric. Food Chem.* **2005**, *53*, 9940–9948. [[CrossRef](#)]
32. Cui, W.; Du, X.; Zhong, M.; Fang, W.; Suo, Z.; Wang, D.; Dong, X.; Jiang, X.; Hu, J. Complex and reticulate origin of edible roses (*Rosa rosaceae*) in China. *Horti. Res.* **2022**, *9*, uhab051. [[CrossRef](#)]
33. Quan, W.; Jin, J.; Qian, C.; Li, C.; Zhou, H. Characterization of volatiles in flowers from four *rosa chinensis* cultivars by hs-spme-gc × gc-qtofms. *Front. Plant Sci.* **2023**, *14*, 1060747. [[CrossRef](#)]
34. Troscianko, J.; Stevens, M. Image calibration and analysis toolbox—A free software suite for objectively measuring reflectance, colour and pattern. *Methods Ecol. Evol.* **2015**, *6*, 1320–1331. [[CrossRef](#)] [[PubMed](#)]
35. Schneider, C.A.; Rasband, W.S.; Eliceiri, K.W. Nih image to imagej: 25 years of image analysis. *Nat. Methods* **2012**, *9*, 671–675. [[CrossRef](#)] [[PubMed](#)]

36. He, Y.; Deng, L.; Mao, Z.; Sun, J. Remote sensing estimation of canopy SPAD value for maize based on digital camera. *Sci. Agric. Sin.* **2018**, *51*, 2886–2897. [[CrossRef](#)]
37. Mizunuma, T.; Mencuccini, M.; Wingate, L.; Ogée, J.; Nichol, C.; Grace, J.; Muller Landau, H.; Muller Landau, H. Sensitivity of colour indices for discriminating leaf colours from digital photographs. *Methods Ecol. Evol.* **2014**, *5*, 1078–1085. [[CrossRef](#)]
38. Xiong, Q.E.; Ye, Z.; Yang, S.M.; Wang, X.Y.; Li, F.A.; Li, X.L.; Liu, F.; Ni, S. *Plant Physiology Experiment Course*; Sichuan Science & Technology Publishing House: Chengdu, China, 2003.
39. Karsoliya, S. Approximating number of hidden layer neurons in multiple hidden layer BPNN architecture. *Int. J. Eng. Trends Technol.* **2012**, *3*, 714–717.
40. Yu, X.; Chang, C.; Song, J.; Zhuge, Y.; Wang, A. Precise monitoring of soil salinity in China's yellow river delta using UAV-borne multispectral imagery and a soil salinity retrieval index. *Sensors* **2022**, *22*, 546. [[CrossRef](#)]
41. Li, Y.; Huang, S.; Miao, L.; Wu, Z. Simulation analysis of carbon peak path in China from a multi-scenario perspective: Evidence from random forest and back propagation neural network models. *Environ. Sci. Pollut. Res.* **2023**, *30*, 46711–46726. [[CrossRef](#)] [[PubMed](#)]
42. Hair, J.F., Jr.; Anderson, R.E.; Tatham, R.L.; Black, W.C. *Multivariate Data Analysis*, 3rd ed.; Macmillan Publishing Company: New York, NY, USA, 1995.
43. Firdaus, M.L.; Parlindungan, D.; Sundaryono, A.; Farid, M.; Rahmidar, L.; Maidartati, M.; Amir, H. Development of low-cost spectrophotometry laboratory practice based on the digital image for analytical chemistry subject. In *The 3rd Asian Education Symposium (AES 2018)*; Atlantis Press: Bandung, Indonesia, 2018.
44. Salimi, M.; Sun, B.R.; Tabunag, J.S.; Li, J.; Yu, H. A mobile analytical device for on-site quantitation of anthocyanins in fruit beverages. *Micromachines* **2021**, *12*, 246. [[CrossRef](#)] [[PubMed](#)]
45. Singh, A.; Ganapathysubramanian, B.; Singh, A.K.; Sarkar, S. Machine learning for high-throughput stress phenotyping in plants. *Trends Plant Sci.* **2016**, *21*, 110–124. [[CrossRef](#)]
46. Hassanijalilian, O.; Igathinathane, C.; Doetkott, C.; Bajwa, S.; Nowatzki, J.; Haji Esmaeili, S.A. Chlorophyll estimation in soybean leaves infield with smartphone digital imaging and machine learning. *Comput. Electron. Agric.* **2020**, *174*, 105433. [[CrossRef](#)]
47. Wood, N.J.; Baker, A.; Quinell, R.J.; Camargo-Valero, M.A. A simple and non-destructive method for chlorophyll quantification of chlamydomonas cultures using digital image analysis. *Front. Bioeng. Biotech.* **2020**, *8*, 746. [[CrossRef](#)] [[PubMed](#)]
48. Taha, M.F.; Mao, H.; Wang, Y.; Elmanawy, A.I.; Elmasry, G.; Wu, L.; Memon, M.S.; Niu, Z.; Huang, T.; Qiu, Z. High-throughput analysis of leaf chlorophyll content in aquaponically grown lettuce using hyperspectral reflectance and RGB images. *Plants* **2024**, *13*, 392. [[CrossRef](#)] [[PubMed](#)]
49. Frost, J. *Regression Analysis: An Intuitive Guide for Using and Interpreting Linear Models*; Statistics by Jim Publishing: State College, PA, USA, 2019.
50. Chang, C.; Laird, D.A.; Mausbach, M.J.; Hurburgh, C.R. Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. *Soil Sci. Soc. Am. J.* **2001**, *65*, 480–490. [[CrossRef](#)]
51. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [[CrossRef](#)]
52. Guo, Y.; Yin, G.; Sun, H.; Wang, H.; Chen, S.; Senthilnath, J.; Wang, J.; Fu, Y. Scaling effects on chlorophyll content estimations with rgb camera mounted on a UAV platform using machine-learning methods. *Sensors* **2020**, *20*, 5130. [[CrossRef](#)] [[PubMed](#)]
53. Yang, T.; Yu, Y.; Yang, X.G.; Du, H.X. UAV hyperspectral combined with lidar to estimate chlorophyll content at the stand and individual tree scales. *Chin. J. Appl. Ecol.* **2023**, *34*, 2101–2112. [[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.