Article

Non-Destructive Prediction of Rosmarinic Acid Content in Basil Plants Using a Portable Hyperspectral Imaging System and Ensemble Learning Algorithms

Hyo In Yoon [1,2], Dahye Ryu [1], Jai-Eok Park [1], Ho-Youn Kim [1], Soo Hyun Park [1,*] and Jung-Seok Yang [1,*]

1   Smart Farm Research Center, Korea Institute of Science and Technology (KIST),
    Gangneung 25451, Republic of Korea; yoonplant@korea.kr (H.I.Y.); dahye0507@kist.re.kr (D.R.);
    j-park@kist.re.kr (J.-E.P.); hykim@kist.re.kr (H.-Y.K.)
2   Vegetable Research Division, National Institute of Horticultural and Herbal Science,
    Wanju 55365, Republic of Korea
*   Correspondence: ecoloves@kist.re.kr (S.H.P.); inenviron@kist.re.kr (J.-S.Y.); Tel.: +82-33-650-3702 (J.-S.Y.)

Abstract: Rosmarinic acid (RA) is a phenolic antioxidant naturally occurring in the plants of the Lamiaceae family, including basil (*Ocimum basilicum* L.). Existing analytical methods for determining the RA content in leaves are time-consuming and destructive, posing limitations on quality assessment and control during cultivation. In this study, we aimed to develop non-destructive prediction models for the RA content in basil plants using a portable hyperspectral imaging (HSI) system and machine learning algorithms. The basil plants were grown in a vertical farm module with controlled environments, and the HSI of the whole plant was captured using a portable HSI camera in the range of 400–850 nm. The average spectra were extracted from the segmented regions of the plants. We employed several spectral data pre-processing methods and ensemble learning algorithms, such as Random Forest, AdaBoost, XGBoost, and LightGBM, to develop the RA prediction model and feature selection based on feature importance. The best RA prediction model was the LightGBM model with feature selection by the AdaBoost algorithm and spectral pre-processing through logarithmic transformation and second derivative. This model performed satisfactorily for practical screening with R^2_P = 0.81 and *RMSEP* = 3.92. From in-field HSI data, the developed model successfully estimated and visualized the RA distribution in basil plants growing in the greenhouse. Our findings demonstrate the potential use of a portable HSI system for monitoring and controlling pharmaceutical quality in medicinal plants during cultivation. This non-destructive and rapid method can provide a valuable tool for assessing the quality of RA in basil plants, thereby enhancing the efficiency and accuracy of quality control during the cultivation stage.

Keywords: AdaBoost; feature selection; LightGBM; machine learning; phenolics; Random Forest; XGBoost

1. Introduction

Sweet basil (*Ocimum basilicum* L.) belongs to the Lamiaceae family and is a well-known herb with significant economic value in culinary, cosmetic, industrial, and medicinal/pharmaceutical applications worldwide [1]. Notably, rosmarinic acid (RA), a major phenolic compound in basil, is a caffeic acid ester and exhibits various biological and pharmacological activities, including antioxidant, anti-inflammatory, antiviral, and antitumor effects [2]. These benefits have been linked to reducing oxidative stress and inflammatory symptoms associated with chronic diseases such as cardiovascular and neurodegenerative diseases [3]. Several clinical studies confirmed the health benefits of RA in patients with different medical conditions. One study confirmed the positive effect about pain reduction, in which patients with knee osteoarthritis reduced the number of analgesics by consuming approximately 280 mg of RA daily during 16 weeks as a spearmint tea [4]. In another study,

RA-rich food, namely 0.3–0.6 g lemon balm extract with >6% of RA, improved cognitive performance and had anti-stress effects in a young population [5]. A cream containing RA helped patients with moderate atopic dermatitis to reduce erythema, crusting, edema, and local pruritus [6].

Quality control is essential for the industrial application of RA for therapeutic use. When natural RA is isolated from plants, it may contain non-medicinal parts in addition to the medicinal parts of the herbal plant, which increases the variation in RA content. One of the traditional analytical techniques is high-performance liquid chromatography (HPLC), characterized by higher separation efficiency, less sample consumption, and a wider application range [3]. Kiferle et al. [7] recommended the upscaling of RA production in hydroponically grown sweet basil plants. However, conventional detection methods are accurate but have limitations as they are time-consuming, destructive, and cannot provide real-time feedback [8]. Novel monitoring techniques for RA content from the cultivation field of herbal plants are needed to ensure the quality of natural products.

Hyperspectral imaging (HSI) is a powerful non-destructive tool for analyzing the biophysical characteristics of plants. HSI techniques have been widely used in plant phenotyping for detecting biotic and abiotic stresses, particularly in crops [9,10]. In recent years, HSI has been extended to biochemical estimation, leaf monitoring, and species identification [11]. HSI enables the acquisition of spectral information at high spatial and spectral resolutions. However, extracting useful plant traits from the complex and vast HSI data requires various approaches, including vegetation indices, multiple linear regression, partial least square regression (PLSR), principal component regression, machine learning algorithms, and radiative transfer modeling [12]. Vegetation indices have been developed for the rapid monitoring of phytochemicals and nutrition status, such as the chlorophyll content in corn leaves [13], potassium content in rice leaves [14], and anthocyanin content in grape berries [15]. PLSR is a commonly used regression method for predicting the metabolite content from HSI data. Burnett et al. (2021) provided a comprehensive guide for using PLSR to estimate plant traits from hyperspectral data, covering best practices for data collection and model development [16].

Machine learning techniques, particularly ensemble learning algorithms, have been widely used for developing predictive models for plant status [17,18]. Ensemble learning algorithms combine multiple learners to improve the prediction accuracy and robustness. Two popular ensemble learning methods are bagging and boosting. Bagging (bootstrap aggregation) involves creating multiple base models using the random subsets of training data with the replacement and averaging of the predictions of all base models. Random Forest (RF) is a popular bagging-based algorithm that employs multiple decision trees and random subsets of features at each split [19]. RF algorithms for HSI data have been used to predict the symptom classes of diseases caused by pathogens in sugar beet leaves [20] and vegetation chlorophyll content [12]. Boosting involves creating multiple base models sequentially, where each subsequent base model is trained on the data that the previous models misclassified, and the predictions are combined through the weighted averaging of all base models. Adaptive Boosting (AdaBoost) [21], gradient boosting decision tree, Extreme Gradient Boosting (XGBoost) [22], Light Gradient Boosting Model (LightGBM) [23], and categorical boosting are popular boosting-based algorithms. For example, XGBoost algorithms have demonstrated higher accuracy in classifying the symptoms of barley leaves inoculated with powdery mildew [24] and predicting potassium levels in pepper seedlings than support vector machine algorithms [25].

To enhance the accuracy of HSI-based models, a variety of pre-processing techniques are essential, which include image segmentation, spectral data pre-processing, and wavelength selection. In HSI analysis, image segmentation is employed to identify the region of interest. This can be achieved using various methods, such as histogram-based thresholding [26], rectangular or polygon region selections [16,27], and classification algorithms [28]. Spectral data pre-processing methods are necessary to remove the undesired effects and improve the quality of spectral data by eliminating noise and correcting scattering. Typi-

cally, multiple pre-processing methods are compared and selected, such as normalization, logarithmic transformation, the Savitzky–Golay filter, derivatives, multiplicative scatter correction, and standard normal variate [13,29]. The process to select pre-processing for a given datum is usually difficult and limited to trial-and-error. Combining them can be a strategy to improve model accuracy compared to using them alone [30]. In addition, feature selection is an essential step in reducing the dimensionality of HSI data by selecting only the most informative wavelengths [27]. These pre-processing techniques are usually used in combination [31].

This study aims to develop a non-destructive, rapid, and reliable method for predicting rosmarinic acid content using HSI and machine learning. This approach could fill a significant gap in current practices by providing a faster and more sustainable alternative to traditional methods, which are impractical for real-time monitoring in a cultivation stage. Previous studies that utilized HSI techniques to predict phytochemical content have primarily concentrated on visible pigments like chlorophyll and anthocyanin. However, many health-promoting phytochemicals that contribute to its pharmaceutical properties are not visible, i.e., they are not directly linked to the reflectance spectrum in the visible light region. For practical in-field monitoring applications, the model should be developed using small and affordable equipment such as a portable HSI camera. Therefore, different data analysis techniques should be explored to overcome the limited resolution in a narrow spectral range and to identify invisible indirect correlations. The objective of this study was to develop non-destructive prediction models for RA content in basil plants using a portable HSI camera and machine learning techniques. To achieve this, we evaluated various spectral pre-processing methods and ensemble learning algorithms such as RF, AdaBoost, XGBoost, and LightGBM for feature selection and prediction.

2. Materials and Methods

2.1. Plant Materials and Growth Condition

Basil (*O. basilicum* L.) plants were cultivated using a vertical farming system installed in the Smart U-FARM facility of Korea Institute of Science and Technology (KIST; Gangneung, Republic of Korea) (Figure 1a). The sweet basil seeds were sown on moist rockwool cubes ($25 \times 25 \times 40$ mm, W $\times$ L $\times$ H, respectively, Grodan Co., Roermond, The Netherlands) in a plastic tray using water culture. The seedlings were irradiated with fluorescent lamps (TL5 14 W/865, Philips, Amsterdam, The Netherlands) at a photosynthetic photon flux density (PPFD) of 200 μmol m^{-2} s^{-1} for a 14 h light period and supplied with Hoagland's nutrient solution at an electrical conductivity (EC) of 1.0 dS m^{-1} two weeks after sowing. After five weeks, the 528 seedlings were transplanted into a vertical farming module that utilized a deep flow technique system (44 plants per plot, 12 plots), and they were supplied with the nutrient solution at an EC of 2.0 dS m^{-1} (Figure 1a). The plants were subjected to a full-spectrum light source with a PPFD of 300 μmol m^{-2} s^{-1} for a 14 h light period, and the air temperature was maintained at 25/19 $\pm$ 1 °C (day/night). The CO_2 concentration was maintained at 800 μmol mol^{-1} during the day. At 42, 49, 56, and 63 days after sowing, 48 plants were harvested (4 plants per plot $\times$ 12 plots), resulting in a total of 144 plants were harvested. All leaves from the aerial parts of harvested plants were collected for chemical analysis.

2.2. Determination of RA Content

Basil leaves were freeze-dried at −80 °C for a week under vacuum, followed by fine grinding using an IKA® A11 basic mill (IKA-Werke, Staufen, Germany). To extract the RA, 500 mg of lyophilized basil powder was mixed with 40 mL of 70% ethanol in a falcon tube (50 mL) and sonicated at 60 °C in an ultrasonic water bath (UCP-10, Jeio Tech Co. Ltd., Daejeon, Republic of Korea) for 1 h. The mixture was then centrifuged at 3500 rpm for 15 min, and the supernatant was collected. All supernatant extracts were filtered through a 0.45 μm PTFE syringe filter (Smartpor-ll, Woongki Science Co., Ltd., Seoul, Republic of Korea) and evaporated using a nitrogen concentrator (Allsheng MD 200, Hangzhou

Allsheng Instrument Co., Ltd., Hangzhou, China). The dried crude concentrates were re-dissolved in dimethyl sulfoxide (DMSO), and the resulting solution was filtered through a 0.2 μm Whatman PVDF syringe filter (67791302, Whatman, Maidstone, UK) prior to chromatography analysis.
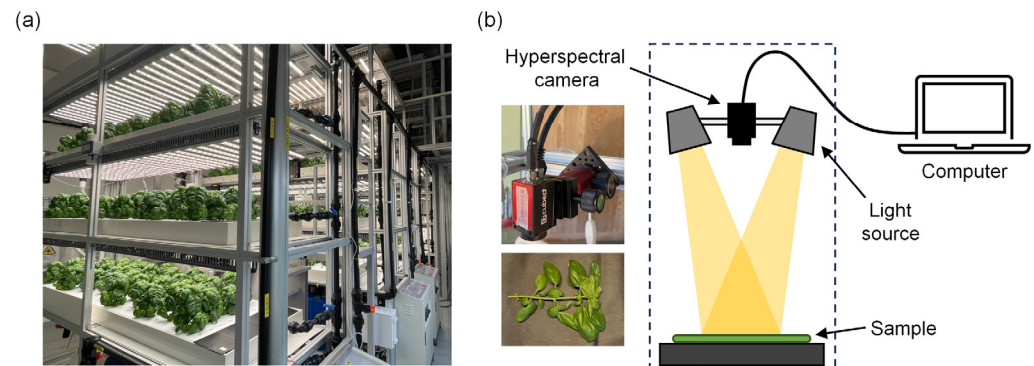

Figure 1. Basil plants cultivated under the growing conditions (**a**) and the schematic diagram of hyperspectral imaging system (**b**) in this study.

The RA content was quantified using an Agilent 1260 Infinity High-Performance Liquid Chromatography (HPLC) system equipped with a diode array detector (DAD). The RA in basil extracts was separated on a YMC Triart C18 column (4.6 × 250 mm, 5 μm; YMC Co., Ltd., Kyoto, Japan). The mobile phases consisted of 0.2% formic acid in water (A) and acetonitrile (B). The gradient program of the mobile phase was used as follows: 0–4 min, 0–20% (B); 4–10 min, 20-37% (B); 10–15 min, 37% (B); 15–18 min, 37–60% (B). The injection volume was 10 μL, and the flow rate was 1 mL/min. The column oven temperature was maintained at 40 °C. The detection of RA was conducted at 330 nm.

A calibration curve was drawn using six concentrations (5–200 μg/mL) of RA standard purchased from Sigma-Aldrich (St. Louis, MO, USA). The RA had a linear regression equation as $Y = 35.223X - 62.422$ with a high correlation coefficient ($R^2 = 0.9995$).

2.3. HSI Data Collection

The hyperspectral images were collected using a portable Ultris 5 hyperspectral imaging camera (Cubert GmbH, Ulm, Germany), which had 51 spectral bands ranging from 450 to 850 nm, a spatial resolution of 290 × 275 pixels, and an integration time of 40 ms. Two 15 W halogen lamps were used as the light source, placed on either side of the camera (Figure 1b). The angle of the light sources was adjusted according to the irradiation distance and measurement area of the camera. For each snapshot, one whole basil plant was placed on a black tray and measured. The raw data were extracted using perClass Mira 3 software (perClass BV, Delft, The Netherlands) and the Spectral library in the Python 3.9 environment.

2.4. Image Segmentation and Spectral Extraction

To segment the basil plants from the background in the hyperspectral cube data, a threshold technique was used to select the regions of interest (ROIs). A normalized band difference (NBD) was calculated using reflectance values at 794 and 666 nm, where $NBD = (R_{794} - R_{666})/(R_{794} + R_{666})$ and R denote the reflectance values at the wavelength in a unit pixel. A threshold was applied to the images to maximize the contrast between the plants and the background, and pixels with NBD > 0.4 were selected as ROI. The average spectral data with 51 bands were obtained for each of the 144 HSI data. The HSI data processing workflow is illustrated in Figure 2.
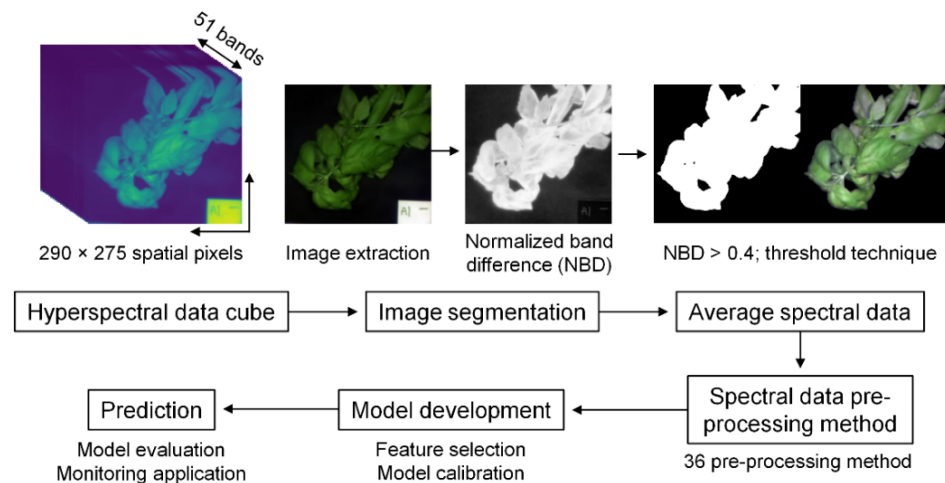
Figure 2. Flowchart illustrating the hyperspectral data processing workflow used for developing the prediction models.

2.5. Spectral Data Pre-Processing Method

Multiple pre-processing methods in HSI analysis could enhance the data quality and extract meaningful features. We classified the various methods into three stages: 1st pre-processing, normalization (Norm), and logarithmic transformation (Log 1/R); 2nd pre-processing, Savitzky–Golay filter (SG filter), first and second derivative after SG-filtering (Der); 3rd pre-processing, multiplicative scatter correction (MSC) and standard normal variate (SNV). First pre-processing helps to minimize baseline differences due to environmental or equipment variation. Log (1/R) is often used as an absorbance spectrum in near-infrared spectroscopy when the medium has both scattering and absorption properties. The transformation helps to normalize the data and improve the linearity of relationships between variables. Second pre-processing reduces the noise and highlights specific spectral features. SG filter is a window-based smoothing operation for spectral data, which involves fitting a polynomial curve of the fixed order within specified points [32]. The SG filter was used for smoothing the data and generating derivative transformations. The first derivative removes an additive/constant baseline offset, while the second derivative removes linear/sloping baseline effects [33]. Although the derivative transform emphasizes the spectral features of the data, it can also promote an emphasis on the level of noise [34]. Third pre-processing can involve advanced corrections or transformations that address complex non-linearities or scatter effects, improving the model's robustness and accuracy. MSC and SNV were used to minimize light scattering effects due to surface heterogeneity [8]. MSC minimizes physical light scattering and reveals the chemical light absorption by fitting a linear model between a reference spectrum and other spectra [35]. The reference spectrum is often chosen as the average of all spectra in the dataset. Typically, the reference spectrum is the average of all spectra in the dataset, which was also the case in this study. SNV also aims to separate the physical and chemical variance [36] in which each spectrum is scaled by subtracting its mean and then dividing by its standard deviation.

We used the 1st, 2nd, and 3rd pre-processing methods alone or in combination, resulting in a total of 36 methods, and details are given in Table S1. The SG filter was implemented with a three-order polynomial fit with five data points using the SciPy package in Python 3.9.

2.6. Feature Selection and Modeling Methods

Feature selection is used to improve the performance of HSI-based model for the following reasons: better accuracy, generalization, computational efficiency, interpretability, less risk of overfitting, and the ease of practical application [33]. We employed an embedded feature selection approach using the same ensemble algorithms as those in our prediction models. Feature importance is a score of how much each feature was used to make key

nodes for constructing a model. Tree-based ensemble algorithms inherently calculate and provide them. We used four ensemble methods for feature importance extraction: Random Forest (RF), Adaptive Boosting (AdaBoost), Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Model (LightGBM). In order to discard irrelevant features, we only selected bands with feature importance greater than 1.25 times the average importance.

After selecting features for each spectral data, the prediction performance was evaluated in prediction models with four algorithms using RF regression, AdaBoost regression, XGBoost regression, and LightGBM regression. The hyperparameters of four algorithms were optimized using the grid-search technique (Table S2). The implementation of model development was programmed based on the scikit-learn, xgboost, and lightgbm packages in Python 3.9.

2.7. Model Calibration and Evaluation

The dataset of the RA content and HSI data were randomly split into a calibration set and a prediction set at a ratio of 8:2. Model performance was evaluated based on the coefficient of determination (R^2) and root mean square error (*RMSE*) using Equations (1) and (2):

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

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

where y_i is the measured value of component analysis, $\hat{y}_i$ is the value predicted by the model, $\overline{y}$ is the mean value of component analysis, and n is the number of samples.

We used 10-fold cross-validation results in the calibration set to determine the spectral pre-processing methods and optimize the hyperparameters for the four algorithms. The final models were determined by evaluation metrics in the prediction set, as a combination of feature selection and prediction algorithms.

2.8. In-Field Application

To confirm the applicability of the developed model for monitoring, we tested it on basil plants growing in an experimental plastic greenhouse at the KIST Gangneung Institute of Natural Products (37.8° N, 128.8° E). The plants were transplanted into pots filled with a horticultural soil substrate on 27 December 2022. HSI data were acquired from the top and side views of plants under growing conditions using four halogen lamps on 6 February 2023 after sunset to exclude the influence of natural light. For ROI selection, pixels with NBD > 0.5 were selected as ROIs. The other conditions of HSI data processing, spectral pre-processing, feature band selection, prediction model, and hyperparameter tuning were the same as the conditions of the final model. The prediction of RA content was applied within the selected ROI using the final model.

2.9. Statistical Analysis

All data analysis and visualization were conducted using Python (Version 3.9). Libraries such as Spectral, NumPy, and Pandas were used for data handling and pre-processing, while SciPy, Scikit-Learn, xgboost, and lightgbm were employed for model development, validation, and evaluation metrics calculation (e.g., R^2, *RMSE*). Matplotlib package was used to visualize these results.

3. Results

3.1. Analysis of RA Content in Basil Plants

The HPLC-DAD analysis was used to quantify the RA content in 144 basil plants, and the results are presented in Table 1. The analytical methods employed in this study are

well-established procedures. To develop the model, we randomly divided the dataset of RA content and HSI data into the calibration set and prediction set at an 8:2 ratio.

Table 1. Descriptive statistics for the rosmarinic acid content in basil plants.

Statistics	Rosmarinic Acid (mg g⁻¹ DW)		
	Total Dataset	**Calibration Set**	**Prediction Set**
Number of samples	144	115	29
Minimum	1.892	1.892	2.991
Maximum	34.29	34.29	33.32
Mean	12.47	12.12	13.93
Standard deviation	7.966	7.624	9.204

3.2. Determination of Spectral Pre-Processing Methods

The spectral data for 144 basil plants were obtained by averaging the HSI data in the ROI, followed by pre-processing either alone or in combination, as shown in Figure 3. Prior to modeling, all algorithms with full spectra were tested using 36 pre-processing methods (Table S1). The best pre-processing methods were determined according to evaluation metrics, R^2 of cross-validation (R^2_{CV}), and RMSE of cross-validation ($RMSECV$). We presented the top five pre-processing methods among 36 methods for each algorithm based on high R^2_{CV} and low $RMSECV$ (Table 2). The pre-processing methods determined for the RF, AdaBoost, and XGBoost algorithms were the combination of Log (1/R), 2nd Der, and SNV (Figure 3j), while for the LightGBM algorithm, the combination of Log (1/R) and 2nd Der (Figure 3i) was chosen. In the XGBoost algorithm with the selected pre-processing method, the R^2 of the calibration (R^2_C) was the largest, and the $RMSEC$ was the smallest among the four algorithms. However, the R^2_{CV} was the smallest at 0.708, indicating that the XGBoost algorithm had an overfitting problem. The AdaBoost algorithm with the selected pre-processing methods had the lowest $RMSECV$ of 3.822.
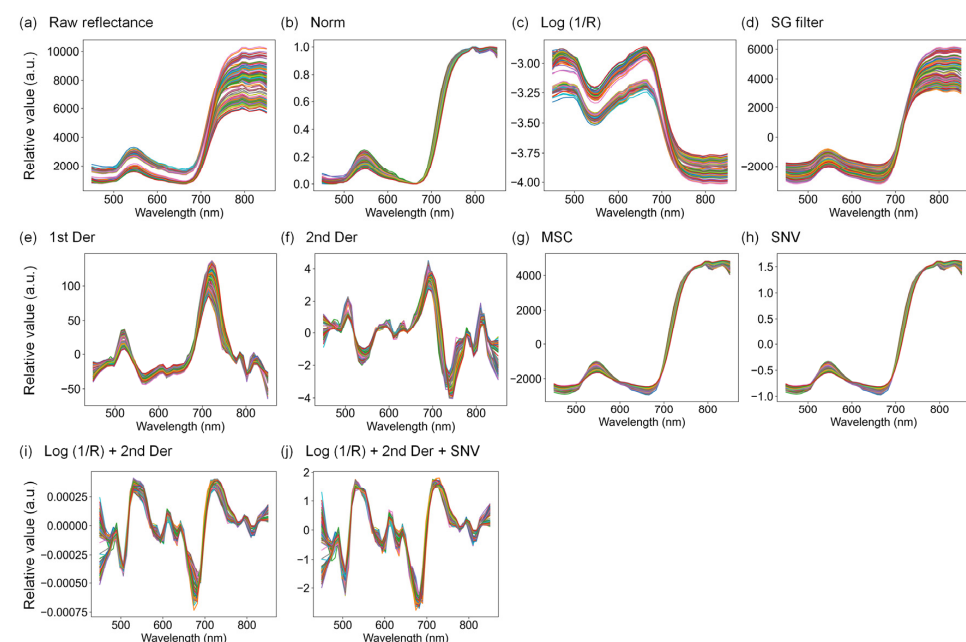

Figure 3. Raw (**a**) and pre-processed (**b**–**j**) spectral data from the basil plants. The spectral pre-processing methods used were normalization (Norm; (**b**)), logarithmic transformation (Log (1/R); (**c**)), Savitzky–Golay filter (SG filter; (**d**)), derivative after SG filter (Der; (**e**,**f**)), multiplicative scatter correction (MSC; (**g**)), and standard normal variate (SNV; (**h**)). For the prediction algorithms, the combinations of Log (1/R) + 2nd Der (**i**) and Log (1/R), 2nd Der, and SNV (**j**) were chosen, as shown in Table 2. Colored lines represent different leaf samples.

Table 2. Performance of four prediction algorithms based on best pre-processing methods for rosmarinic acid content in basil plants.

Prediction Model	Pre-Processing Method	Calibration		Cross-Validation	
		R^2_C	*RMSEC*	R^2_{CV}	*RMSECV*
RF	**Log (1/R) + 2nd Der + SNV**	**0.966**	**1.407**	**0.737**	**3.889**
	Log (1/R) + 2nd Der	0.962	1.485	0.725	3.979
	Log (1/R) + 1st Der + SNV	0.960	1.519	0.725	3.980
	Log (1/R) + 1st Der + MSC	0.963	1.461	0.719	4.021
	Log (1/R) + SNV	0.964	1.434	0.719	4.026
AdaBoost	**Log (1/R) + 2nd Der + SNV**	**0.950**	**1.698**	**0.746**	**3.822**
	Log (1/R) + 2nd Der	0.944	1.803	0.739	3.877
	Log (1/R) + SNV	0.922	2.122	0.713	4.069
	Raw reflectance	0.914	2.229	0.705	4.121
	Log (1/R) + MSC	0.931	1.989	0.705	4.122
XGBoost	**Log (1/R) + 2nd Der + SNV**	**1.000**	**0.001**	**0.708**	**4.099**
	1st Der	1.000	0.001	0.700	4.155
	Log (1/R) + SNV	1.000	0.001	0.697	4.179
	Log (1/R) + SG filter + SNV	1.000	0.001	0.694	4.197
	Log (1/R) + SG filter + MSC	1.000	0.001	0.686	4.254
LightGBM	**Log (1/R) + 2nd Der**	**0.965**	**1.423**	**0.733**	**3.924**
	Log (1/R) + 2nd Der + SNV	0.963	1.466	0.715	4.053
	Log (1/R) + 1st Der + MSC	0.960	1.514	0.712	4.073
	Log (1/R) + 2nd Der + MSC	0.964	1.448	0.711	4.082
	1st Der + MSC	0.957	1.576	0.699	4.165

Only the top five pre-processing methods out of 36 methods for each algorithm are shown. R^2, coefficient of determination; *RMSEC* and *RMSECV*, root mean square error of calibration and cross-validation, respectively. Bold indicates the pre-processing methods determined with the lowest *RMSECV*.

3.3. Selection of Characteristic Wavelength

The two combinations of spectral pre-processing methods, Log (1/R) + 2nd Der and Log (1/R) + 2nd Der + SNV (Figure 3i,j), were used for wavelength selection, denoted by X1 and X2, respectively. The characteristic wavelengths were selected depending on the pre-processed spectral data and selection algorithms (Figure 4). In the spectral data with X1, the feature importance was the highest at 834 and 802 nm among the 10, 8, and 5 bands selected based on RF, AdaBoost, and XGBoost algorithms, respectively. Similarly, the feature importance in the X2 spectra was the highest at 834 nm among the five, seven, and four bands selected based on the RF, AdaBoost, and XGBoost algorithms, respectively. The feature importance based on the LightGBM algorithm was evenly distributed and was the highest at 714–722 nm among 16 bands in the X1 data, and at 578 nm among 14 bands in the X2 data.

3.4. Final Prediction Models

In order to improve model accuracy, reduce overfitting, and increase the generalization performance, we fine-tuned hyperparameters such as learning rates, number and depth of trees, and more. Table S2 provides details on the selected hyperparameters for each prediction model and feature selection method used. We developed RF, Adaboost, XGBoost, and LightGBM prediction models with waveband selection or the full band (Table 3). The final model for each prediction algorithm was determined based on the prediction performance, as measured by the R^2 of prediction (R^2_P) and RMSE of prediction (*RMSEP*), as shown in Figure 5. The best model was found to be the LightGBM prediction model with X1 spectral pre-processing and eight bands selected by the AdaBoost method ($R^2_P = 0.812$ and *RMSEP* = 3.924). The eight selected bands were 546, 674, 682, 730, 794, 802, 826, and 834 nm. Both the final RF and XGBoost models were developed using X2 pre-processing, and seven bands were selected by the AdaBoost algorithm ($R^2_P = 0.804$ and 0.796, respectively). The seven selected bands were 506, 562, 682, 706, 802, 826, and 834 nm. The final AdaBoost prediction model used X2 pre-processing, and 14 bands were selected

by the LightGBM algorithm, which were 482, 506, 522, 554, 570, 578, 602, 650, 698, 706, 730, 778, 802, and 826 nm. The performance of the final AdaBoost model was slightly lower than that of the other models, with $R^2{}_P$ = 0.792 and $RMSEP$ = 4.124. While not all feature selections performed better than the full band, the best models utilized selected variables.
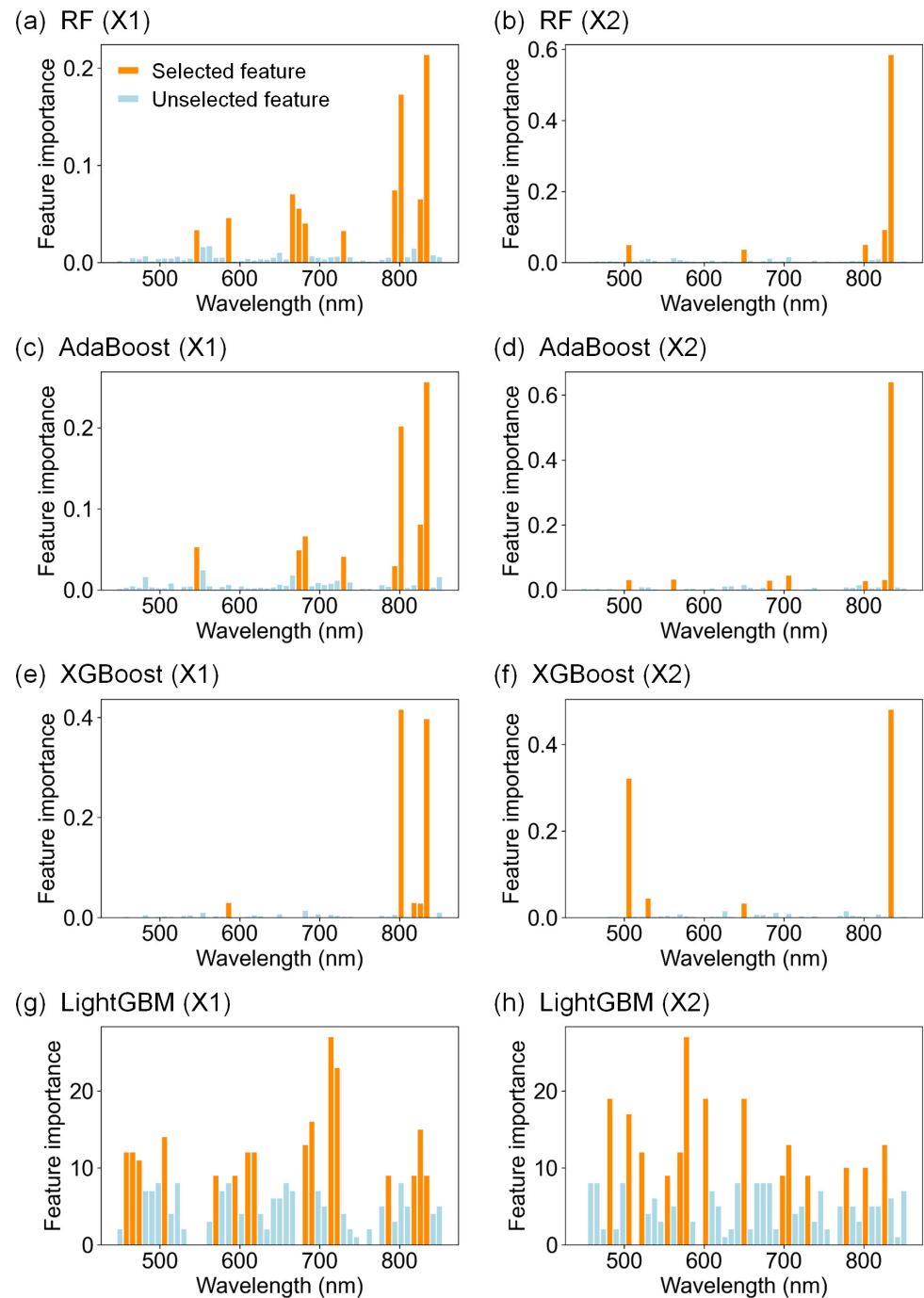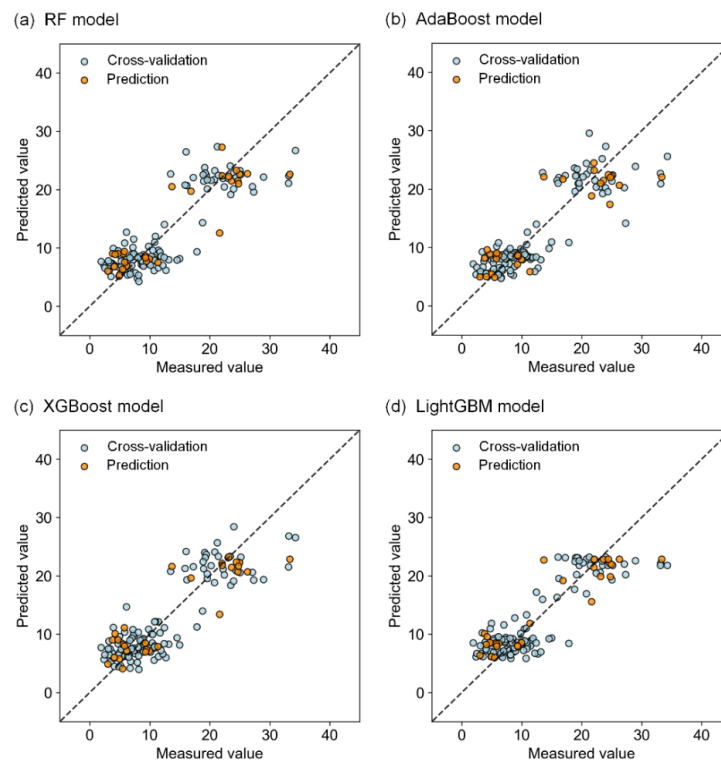

Figure 4. The feature importance of the feature selection algorithms for different spectral pre-processing methods. The feature selection algorithms used were Random Forest (RF) (**a**,**b**), AdaBoost (**c**,**d**), XGBoost (**e**,**f**), and LightGBM (**g**,**h**). The spectral pre-processing methods were Log (1/R) + 2nd Der (X1) (**a**,**c**,**e**,**g**) and Log (1/R) + 2nd Der + SNV (X2) (**b**,**d**,**f**,**h**), as shown in Figure 3i and 3j, respectively. The orange bars represent selected features, and the light blue bars represent unselected features, i.e., not used in the prediction model.

Table 3. Performance of the prediction models according to feature selection algorithms after hyperparameter tuning.

Prediction Model	Pre-Processing Method	Feature Selection		Calibration		Cross-Validation		Prediction	
		Method	No. Feature	R^2_C	RMSEC	R^2_{CV}	RMSECV	R^2_P	RMSEP
RF	Log (1/R) + 2nd Der + SNV	Full band	51	0.968	1.368	0.742	3.853	0.790	4.148
		RF	5	0.962	1.486	0.703	4.135	0.770	4.342
		AdaBoost	**7**	**0.968**	**1.364**	**0.732**	**3.927**	**0.804**	**4.003**
		XGBoost	4	0.966	1.399	0.750	3.792	0.787	4.173
		LightGBM	14	0.966	1.405	0.733	3.921	0.788	4.161
AdaBoost	Log (1/R) + 2nd Der + SNV	Full band	51	0.949	1.716	0.750	3.792	0.770	4.335
		RF	5	0.874	2.693	0.724	3.985	0.751	4.516
		AdaBoost	7	0.917	2.193	0.764	3.686	0.766	4.376
		XGBoost	4	0.906	2.330	0.758	3.731	0.758	4.446
		LightGBM	**14**	**0.930**	**2.013**	**0.750**	**3.798**	**0.792**	**4.124**
XGBoost	Log (1/R) + 2nd Der + SNV	Full band	51	0.940	1.851	0.739	3.878	0.768	4.360
		RF	5	0.895	2.455	0.719	4.021	0.773	4.312
		AdaBoost	**7**	**0.983**	**0.997**	**0.744**	**3.842**	**0.796**	**4.082**
		XGBoost	4	0.917	2.187	0.763	3.694	0.761	4.425
		LightGBM	14	0.968	1.350	0.776	3.595	0.752	4.502
LightGBM	Log (1/R) + 2nd Der	Full band	51	0.916	2.199	0.749	3.806	0.801	4.032
		RF	10	0.872	2.718	0.760	3.719	0.789	4.154
		AdaBoost	**8**	**0.828**	**3.151**	**0.744**	**3.837**	**0.812**	**3.924**
		XGBoost	5	0.827	3.159	0.747	3.816	0.791	4.131
		LightGBM	16	0.945	1.776	0.784	3.524	0.750	4.523

R^2, coefficient of determination; *RMSEC*, *RMSECV*, and *RMSEP*, root mean square error of calibration, cross-validation, and prediction, respectively. Bold indicates the best performance based on *RMSEP* for each prediction algorithm.

Figure 5. Performance of the prediction models for rosmarinic acid content in basil plants based on ensemble algorithms. The ensemble algorithms used were Random Forest (RF) (**a**), AdaBoost (**b**), XGBoost (**c**), and LightGBM (**d**), after selecting the spectral pre-processing method and characteristic feature.

3.5. In-Field Application for Monitoring RA Distribution

The HSI-based prediction model can be extended from single-pixel-level predictions to visualize the spatial distribution of the compound content in plants. We used the best LightGBM model to predict the RA distribution of basil plants grown in a greenhouse, as shown in Figure 6. These models, which use a portable HSI camera and ensemble learning algorithms, are fast and non-destructive, making it possible to continuously track the compound distribution in plants.

Figure 6. In-field application for monitoring rosmarinic (RA) acid content in basil plants growing in a greenhouse. Using the same hyperspectral imaging system (**a**), hyperspectral images were acquired after sunset (**b**). Spatial distribution of RA content (**c**) was predicted by the LightGBM model, as shown in Table 3.

4. Discussion

The present study confirmed the potential of the HSI technique as a non-destructive, rapid, and reliable method for predicting RA content. Machine learning algorithms with various data processing techniques such as spectral pre-processing and feature selection were employed to overcome the limited resolution and spectral range of the portable HSI camera. These techniques in multiple steps enable a stable prediction performance under outdoor or variable growing conditions and enhance robustness and practicality.

Multiple spectral pre-processing methods enhance data quality and highlight meaningful features. Overall, the model performance was higher when combining two or three pre-processing methods than when using a single method (Table 2). The Log (1/R) was found to improve the cross-validation performance with derivative formation, MSC, or SNV, regardless of the algorithms used. This method compresses the dynamic range of the data, reducing the influence of extreme values and enhancing the visibility of weaker features [37]. It helps to normalize the data and improve the linearity of relationships between variables. In previous studies, the Log (1/R) transformation is used to predict the anthocyanin content in grape skin [38] and the ABA content in zucchini leaves [39] from the HSI data. Spectral pre-processing is a crucial step in HSI data analysis as it helps to mitigate the effects of undesirable phenomena on spectral measurements. To account for unsteady growing conditions during the in-field measurements of HSI data, proper pre-processing methods were employed. In this study, the second Der transformation was included in the selected pre-processing method for the four models (Table 2). The derivative method can remove baseline offset and highlight the spectral features that differ between samples [33].

Similarly, the second Der combined with MSC and feature selection was chosen to predict the drought-induced components of tea plants based on the HSI data [40]. Despite the second Der method being particularly sensitive to noise, we considered it suitable for monitoring plant functional components in a controlled environment, with small changes in light intensity.

Feature selection techniques in HSI analysis are commonly used to reduce dimensionality, improve the model accuracy, and identify important spectral features. Dimensionality reduction by feature selection can be used to develop new multispectral imaging systems as a practical alternative to overcome the expensive cost of HSI systems [41]. Although the number of full wavebands used in this study was only 51, the model performance of cross-validation and prediction was higher with feature selection than with a full band (Table 3). The feature selection method employed in this study was similar to that used by Luo et al. [42], in which bagging and boosting algorithms were used to select spectral feature bands and predict tea polyphenols as regression models.

The feature importances in the same pre-processed data were generally similar across algorithms, but the pattern in LightGBM was different (Figure 4). In RF algorithms, feature importance is calculated through mean decrease in impurity (MDI), which contributes to a reduction in impurity across all trees [19]. AdaBoost gives a high score to features that help correct misclassification errors from previous rounds [21]. XGBoost uses 'gain', and Light GBM can use 'gain' or 'split' for calculating feature importance, which indicate the quality or frequency of each feature, respectively, for improving the model's accuracy [22,23]. The gain importance could provide a clear view of which features have the most impact on accuracy, but is sensitive to specific features and leads to overemphasis. The split importance is useful for understanding which features are consistently considered useful across different splits, but features that frequently add minimal predictive value may be considered important. In this study, the feature importance in LightGBM was calculated based on a 'split' due to distinguishing it based on 'gain' in XGBoost. Therefore, LightGBM showed a different pattern because it judged how frequently the feature was used, while other algorithms computed those based on quality across splits (Figure 4g,h). Ensemble learning algorithms often rely on decision trees as weak learners, which can handle high-dimensional data but are susceptible to overfitting. Feature selection can also reduce the computational burden of the algorithms, leading to faster training and prediction times.

Hyperparameter tuning is another method that significantly affects the model performance. The hyperparameters of all models in this study were optimized using the grid-search technique (Table S2). The grid-search technique is a commonly used method for hyperparameter tuning that finds the best combination by exhaustively searching over a predefined hyperparameter space. The main parameter for RF is the number of trees, and in general, more trees improve the generalization performance [25] but slow down prediction times. For boosting algorithms, the learning rate is an important hyperparameter that controls the contribution of each tree to the final prediction. For gradient boosting algorithms such as XGBoost and LightGBM, the maximum tree depth should be limited, as deeper trees make the model more complex and prone to overfitting [43].

The final models predicted RA satisfactorily for practical screening with an overall $R^2{}_P$ of approximately 0.8 (Table 3). The model was utilized to generate a chemical distribution map of basil plants based on HSI data obtained under normal growth conditions. However, due to the placement of plants in the back, the RA content was also detected in the black background area (Figure 6). To address this issue, alternative algorithms can only be used to automatically detect plants in front of the camera, improving the accuracy of the chemical distribution map.

The use of HSI techniques in the industry of medicinal plants offers numerous benefits, including the rapid and non-destructive chemical analysis for the online monitoring of quality and authenticity [44]. To ensure the quality assessment of the final product, quality must be assessed at every step of the processing chain [8]. Recently, the novel technique of the early or timely detection of root rot in ginseng caused by *Cylindrocarpon destructans*, a

soil-borne fungus, from the HSI data of aerial parts was reported [28]. In medical cannabis, the changes in cannabinoid composition could be monitored during the drying process based on HSI data [45]. This technique enables real-time quality assurance, which is critical for maintaining consistency in medicinal plant production. Our findings demonstrate the potential use of a portable HSI system for the on-site monitoring of pharmaceutical quality in medicinal plants during cultivation. These applications are non-destructive, rapid, and efficient, supporting quality control and traceability, thereby maintaining the integrity of medicinal products from cultivation to consumer use.

5. Conclusions

In conclusion, our study successfully developed HSI-based prediction models for RA content, which is an important pharmaceutical property in basil plants. We overcame the limitations of the portable HSI camera used in practical applications through the use of various data processing techniques, including 36 spectral pre-processing methods and four well-known ensemble learning algorithms for feature selection and prediction. The developed models exhibited satisfactory performance and were able to generate a distribution map of the RA content in in-field basil plants. These findings highlight the potential of HSI-based monitoring and assessment technology for ensuring and controlling the pharmaceutical quality in medicinal plants at the cultivation stage.

In addition, our study contributes to the existing knowledge on spectral pre-processing and feature selection techniques in HSI analysis. Specifically, we demonstrated the effectiveness of Log(1/R) transformation in normalizing the data and improving the linearity of relationships between variables. We showed that feature selection can be used to reduce the dimensionality of data and improve model accuracy, which may be particularly useful for developing new multispectral imaging systems as a practical alternative to expensive HSI systems.

Moving forward, our study suggests that HSI-based monitoring and assessment technology has the potential to be a valuable tool for the medicinal plant industry, particularly for the online monitoring of quality and authenticity. To maximize productivity and uniformity, it is crucial to control medicinal quality at every step of the processing chain, starting from the plant cultivation stage. We hope that our findings will pave the way for future studies that investigate the feasibility and effectiveness of HSI-based monitoring and assessment technology for a wide range of medicinal plants and pharmaceutical properties.

Author Contributions: All authors have made significant contributions to this research. S.H.P. and J.-S.Y. suggested the experimental design; H.I.Y. conducted hyperspectral data acquisition, analysis, and interpretation; D.R. and H.-Y.K. conducted the chemical analysis; J.-E.P. provided the experimental materials; H.I.Y. wrote the manuscript draft; H.-Y.K., S.H.P. and J.-S.Y. revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

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

References

1. Makri, O.; Kintzios, S. *Ocimum* sp. (Basil): Botany, cultivation, pharmaceutical properties, and biotechnology. *J. Herbs Spices Med. Plants* **2008**, *3*, 123–150. [CrossRef]
2. Guan, H.; Luo, W.; Bao, B.; Cao, Y.; Cheng, F.; Yu, S.; Fan, Q.; Zhang, L.; Wu, Q.; Shan, M.A. Comprehensive review of rosmarinic acid: From phytochemistry to pharmacology and its new insight. *Molecules* **2022**, *27*, 3292. [CrossRef] [PubMed]
3. Hitl, M.; Kladar, N.; Gavarić, N.; Božin, B. Rosmarinic acid–human pharmacokinetics and health benefits. *Planta Medica* **2021**, *87*, 273–282. [CrossRef] [PubMed]
4. Connelly, A.E.; Tucker, A.J.; Tulk, H.; Catapang, M.; Chapman, L.; Sheikh, N.; Yurchenko, S.; Fletcher, R.; Kott, L.S.; Duncan, A.M.; et al. High-rosmarinic acid spearmint tea in the management of knee osteoarthritis symptoms. *J. Med. Food* **2014**, *17*, 1361–1367. [CrossRef] [PubMed]
5. Scholey, A.; Gibbs, A.; Neale, C.; Perry, N.; Ossoukhova, A.; Bilog, V.; Kras, M.; Scholz, C.; Sass, M.; Buchwald-Werner, S. Anti-stress effects of lemon balm-containing foods. *Nutrients* **2014**, *6*, 4805–4821. [CrossRef]
6. Lee, J.; Jung, E.; Koh, J.; Kim, Y.S.; Park, D. Effect of rosmarinic acid on atopic dermatitis. *J. Dermatol.* **2008**, *35*, 768–771. [CrossRef]
7. Kiferle, C.; Lucchesini, M.; Mensuali-Sodi, A.; Maggini, R.; Raffaelli, A.; Pardossi, A. Rosmarinic acid content in basil plants grown in vitro and in hydroponics. *Cent. Eur. J. Biol.* **2011**, *6*, 946–957. [CrossRef]
8. Amigo, J.M.; Martí, I.; Gowen, A. Hyperspectral imaging and chemometrics: A perfect combination for the analysis of food structure, composition and quality. *Data Handl. Sci. Technol.* **2013**, *28*, 343–370. [CrossRef]
9. Behmann, J.; Steinrücken, J.; Plümer, L. Detection of early plant stress responses in hyperspectral images. *ISPRS J. Photogramm. Remote Sens.* **2014**, *93*, 98–111. [CrossRef]
10. Humplík, J.F.; Lazár, D.; Husičková, A.; Spíchal, L. Automated phenotyping of plant shoots using imaging methods for analysis of plant stress responses—A review. *Plant Methods* **2015**, *11*, 29. [CrossRef]
11. Mishra, P.; Lohumi, S.; Ahmad Khan, H.; Nordon, A. Close-range hyperspectral imaging of whole plants for digital phenotyping: Recent applications and illumination correction approaches. *Comput. Electron. Agric.* **2020**, *178*, 105780. [CrossRef]
12. Lu, B.; He, Y. Evaluating empirical regression, machine learning, and radiative transfer modelling for estimating vegetation chlorophyll content using bi-seasonal hyperspectral images. *Remote Sens.* **2019**, *11*, 1979. [CrossRef]
13. Wu, C.; Niu, Z.; Tang, Q.; Huang, W. Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. *Agric. For. Meteorol.* **2008**, *148*, 1230–1241. [CrossRef]
14. Lu, J.; Yang, T.; Su, X.; Qi, H.; Yao, X.; Cheng, T.; Zhu, Y.; Cao, W.; Tian, Y. Monitoring leaf potassium content using hyperspectral vegetation indices in rice leaves. *Precis. Agric.* **2020**, *21*, 324–348. [CrossRef]
15. Gomes, V.; Rendall, R.; Reis, M.S.; Mendes-Ferreira, A.; Melo-Pinto, P. Determination of Sugar, pH, and Anthocyanin Contents in Port Wine Grape Berries through Hyperspectral Imaging: An Extensive Comparison of Linear and Non-Linear Predictive Methods. *Appl. Sci.* **2021**, *11*, 10319. [CrossRef]
16. Burnett, A.C.; Anderson, J.; Davidson, K.J.; Ely, K.S.; Lamour, J.; Li, Q.; Morrison, B.D.; Yang, D.; Rogers, A.; Serbin, S.P. A best-practice guide to predicting plant traits from leaf-level hyperspectral data using partial least squares regression. *J. Exp. Bot.* **2021**, *72*, 6175–6189. [CrossRef]
17. Zhang, Y.; Liu, J.; Shen, W. A review of ensemble learning algorithms used in remote sensing applications. *Appl. Sci.* **2022**, *12*, 8654. [CrossRef]
18. Saha, D.; Manickavasagan, A. Machine learning techniques for analysis of hyperspectral images to determine quality of food products: A review. *Curr. Res. Food Sci.* **2021**, *4*, 28–44. [CrossRef]
19. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
20. Brugger, A.; Yamati, F.I.; Barreto, A.; Paulus, S.; Schramowsk, P.; Kersting, K.; Steiner, U.; Neugart, S.; Mahlein, A.K. Hyperspectral imaging in the UV range allows for differentiation of sugar beet diseases based on changes in secondary plant metabolites. *Phytopathology* **2023**, *113*, 44–54. [CrossRef]
21. Freund, Y.; Schapire, R.E. A short introduction to boosting. *J. Jpn. Soc. Artif. Intell.* **1999**, *14*, 771–780.
22. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
23. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. LightGBM: A highly efficient gradient boosting decision tree. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 3147–3154.
24. Brugger, A.; Schramowski, P.; Paulus, S.; Steiner, U.; Kersting, K.; Mahlein, A.K. Spectral signatures in the UV range can be combined with secondary plant metabolites by deep learning to characterize barley–powdery mildew interaction. *Plant Pathol.* **2021**, *70*, 1572–1582. [CrossRef]
25. Weksler, S.; Rozenstein, O.; Haish, N.; Moshelion, M.; Wallach, R.; Ben-Dor, E. Detection of potassium deficiency and momentary transpiration rate estimation at early growth stages using proximal hyperspectral imaging and extreme gradient boosting. *Sensors* **2021**, *21*, 958. [CrossRef] [PubMed]
26. Lu, Y.; Young, S.; Linder, E.; Whipker, B.; Suchoff, D. Hyperspectral imaging with machine learning to differentiate cultivars, growth stages, flowers, and leaves of industrial hemp (*Cannabis sativa* L.). *Front. Plant Sci.* **2022**, *12*, 810113. [CrossRef] [PubMed]
27. Yuan, Z.; Ye, Y.; Wei, L.; Yang, X.; Huang, C. Study on the optimization of hyperspectral characteristic bands combined with monitoring and visualization of pepper leaf SPAD value. *Sensors* **2022**, *22*, 183. [CrossRef] [PubMed]

28. Park, E.; Kim, Y.S.; Faqeerzada, M.A.; Kim, M.S.; Baek, I.; Cho, B.K. Hyperspectral reflectance imaging for nondestructive evaluation of root rot in Korean ginseng (*Panax ginseng* Meyer). *Front. Plant Sci.* **2023**, *14*, 1109060. [CrossRef]

29. Caporaso, N.; Whitworth, M.B.; Fowler, M.S.; Fisk, I.D. Hyperspectral imaging for non-destructive prediction of fermentation index, polyphenol content and antioxidant activity in single cocoa beans. *Food Chem.* **2018**, *258*, 343–351. [CrossRef]

30. Bian, X.; Wang, K.; Tan, E.; Diwu, P.; Zhang, F.; Guo, Y. A selective ensemble preprocessing strategy for near-infrared spectral quantitative analysis of complex samples. *Chemom. Intell. Lab. Syst.* **2020**, *197*, 103916. [CrossRef]

31. Liu, Y.; Wang, Q.; Gao, X.; Xie, A. Total phenolic content prediction in Flos Lonicerae using hyperspectral imaging combined with wavelengths selection methods. *J. Food Process Eng.* **2019**, *42*, 13224. [CrossRef]

32. Savitzky, A.; Golay, M.J. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. [CrossRef]

33. Vidal, M.; Amigo, J.M. Pre-processing of hyperspectral images. Essential steps before image analysis. *Chemom. Intell. Lab. Syst.* **2012**, *117*, 138–148. [CrossRef]

34. Eshkabilov, S.; Lee, A.; Sun, X.; Lee, C.W.; Simsek, H. Hyperspectral imaging techniques for rapid detection of nutrient content of hydroponically grown lettuce cultivars. *Comput. Electron. Agric.* **2021**, *181*, 105968. [CrossRef]

35. Geladi, P.; MacDougall, D.; Martens, H. Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Appl. Spectrosc.* **1985**, *39*, 491–500. [CrossRef]

36. Barnes, R.J.; Dhanoa, M.S.; Lister, S.J. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc.* **1989**, *43*, 772–777. [CrossRef]

37. Zhao, T.; Nakano, A.; Iwaski, Y.; Umeda, H. Application of hyperspectral imaging for assessment of tomato leaf water status in plant factories. *Appl. Sci.* **2020**, *10*, 4665. [CrossRef]

38. Diago, M.P.; Fernandez-Novales, J.; Fernandes, A.M.; Melo-Pinto, P.; Tardaguila, J. Use of visible and short-wave near-infrared hyperspectral imaging to fingerprint anthocyanins in intact grape berries. *J. Agric. Food Chem.* **2016**, *64*, 7658–7666. [CrossRef]

39. Burnett, A.C.; Serbin, S.P.; Davidson, K.J.; Ely, K.S.; Rogers, A. Detection of the metabolic response to drought stress using hyperspectral reflectance. *J. Exp. Bot.* **2021**, *72*, 6474–6489. [CrossRef]

40. Chen, S.; Gao, Y.; Fan, K.; Shi, Y.; Luo, D.; Shen, J.; Ding, Z.; Wang, Y. Prediction of drought-induced components and evaluation of drought damage of tea plants based on hyperspectral imaging. *Front. Plant Sci.* **2021**, *12*, 695102. [CrossRef]

41. Gutiérrez, S.; Wendel, A.; Underwood, J. Spectral filter design based on in-field hyperspectral imaging and machine learning for mango ripeness estimation. *Comput. Electron. Agric.* **2019**, *164*, 104890. [CrossRef]

42. Luo, X.; Xu, L.; Huang, P.; Wang, Y.; Liu, J.; Hu, Y.; Wang, P.; Kang, Z. Nondestructive testing model of tea polyphenols based on hyperspectral technology combined with chemometric methods. *Agriculture* **2021**, *11*, 673. [CrossRef]

43. Jafarzadeh, H.; Mahdianpari, M.; Gill, E.; Mohammadimanesh, F.; Homayouni, S. Bagging and boosting ensemble classifiers for classification of multispectral, hyperspectral and polSAR data: A comparative evaluation. *Remote Sens.* **2021**, *13*, 4405. [CrossRef]

44. Kiani, S.; van Ruth, S.M.; Minaei, S.; Varnamkhastid, M.G. Hyperspectral imaging, a non-destructive technique in medicinal and aromatic plant products industry: Current status and potential future applications. *Comput. Electron. Agric.* **2018**, *152*, 9–18. [CrossRef]

45. Yoon, H.I.; Lee, S.H.; Ryu, D.; Choi, H.; Park, S.H.; Jung, J.H.; Kim, H.-Y.; Yang, J.-S. Non-destructive assessment of cannabis quality during drying process using hyperspectral imaging and machine learning. *Front. Plant Sci.* **2024**, *15*, 1365298. [CrossRef] [PubMed]