

Article

Prediction of Live Bulb Weight for Field Vegetables Using Functional Regression Models and Machine Learning Methods

Dahyun Kim ¹, Wanhyun Cho ², Inseop Na ^{3,*}  and Myung Hwan Na ^{2,*}

¹ Department of Mathematics and Statistics, Chonnam National University, Gwangju 61186, Republic of Korea; qhqk7132@jnu.ac.kr

² Department of Statistics, Chonnam National University, Gwangju 61186, Republic of Korea; whcho@chonnam.ac.kr

³ Division of Culture Contents, Chonnam National University, Yeosu 59626, Republic of Korea

* Correspondence: ypencil@hanmail.net (I.N.); nmh@chonnam.ac.kr (M.H.N.)

Abstract: (1) Background: This challenge is exacerbated by the aging of the rural population, leading to a scarcity of available manpower. To address this issue, the automation and mechanization of outdoor vegetable cultivation are imperative. Therefore, developing an automated cultivation platform that reduces labor requirements and improves yield by efficiently performing all the cultivation activities related to field vegetables, particularly onions and garlic, is essential. In this study, we propose methods to identify onion and garlic plants with the best growth status and accurately predict their live bulb weight by regularly photographing their growth status using a multispectral camera mounted on a drone. (2) Methods: This study was conducted in four stages. First, two pilot blocks with a total of 16 experimental units, four horizontals, and four verticals were installed for both onions and garlic. Overall, a total of 32 experimental units were prepared for both onion and garlic. Second, multispectral image data were collected using a multispectral camera repeating a total of seven times for each area in 32 experimental units prepared for both onions and garlic. Simultaneously, growth data and live bulb weight at the corresponding points were recorded manually. Third, correlation analysis was conducted to determine the relationship between various vegetation indexes extracted from multispectral images and the manually measured growth data and live bulb weights. Fourth, based on the vegetation indexes extracted from multispectral images and previously collected growth data, a method to predict the live bulb weight of onions and garlic in real time during the cultivation period, using functional regression models and machine learning methods, was examined. (3) Results: The experimental results revealed that the Functional Concurrence Regression (FCR) model exhibited the most robust prediction performance both when using growth factors and when using vegetation indexes. Following closely, with a slight distinction, Gaussian Process Functional Data Analysis (GPFDA), Random Forest Regression (RFR), and AdaBoost demonstrated the next-best predictive power. However, a Support Vector Machine (SVM) and Deep Neural Network (DNN) displayed comparatively poorer predictive power. Notably, when employing growth factors as explanatory variables, all prediction models exhibited a slightly improved performance compared to that when using vegetation indexes. (4) Discussion: This study explores predicting onion and garlic bulb weights in real-time using multispectral imaging and machine learning, filling a gap in research where previous studies primarily focused on utilizing artificial intelligence and machine learning for productivity enhancement, disease management, and crop monitoring. (5) Conclusions: In this study, we developed an automated method to predict the growth trajectory of onion and garlic bulb weights throughout the growing season by utilizing multispectral images, growth factors, and live bulb weight data, revealing that the FCR model demonstrated the most robust predictive performance among six artificial intelligence models tested.

Keywords: field vegetables; prediction of live bulb weight; multispectral images; various vegetation indices; growth indices; correlation analysis; functional regression models; machine learning methods; performance evaluation measures



Citation: Kim, D.; Cho, W.; Na, I.; Na, M.H. Prediction of Live Bulb Weight for Field Vegetables Using Functional Regression Models and Machine Learning Methods. *Agriculture* **2024**, *14*, 754. <https://doi.org/10.3390/agriculture14050754>

Academic Editors: Francesco Marinello and Gniewko Niedbała

Received: 22 February 2024

Revised: 6 May 2024

Accepted: 10 May 2024

Published: 12 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

Agriculture plays a pivotal role in sustaining the human diet economy. The global population is growing tremendously, leading to a substantial increase in the demand for and supply of food. However, conventional farming methods prove insufficient to meet these escalating requirements. Consequently, there is a pressing need for innovative automated agricultural techniques, and smart agriculture, through the application of artificial intelligence (AI), is emerging as a key solution to the challenges that we face today. Smart agriculture, harnessing the power of AI technologies, such as machine learning, deep learning, and statistical analysis, is crucial in addressing the complexities of modern farming. Various aspects of agricultural processes, including crop selection, crop yield prediction, soil suitability classification, and water management, benefit from these advanced technologies. Machine learning (ML) algorithms contribute to crop selection and management, deep learning techniques enable crop selection and crop production forecasting, and time-series analysis aids in forecasting crop demand, product price, and crop yield. Additionally, ML and deep learning algorithms could be integrated to develop smart agricultural technologies for various tasks, including soil assessment and soil suitability classification. Therefore, the adoption of cutting-edge technologies within AI in agricultural fields holds the potential to assist farmers in producing high-quality crops. The development of the agricultural sector will also contribute considerably to rural development. Within the agricultural sector, diverse technologies, including disease detection, diagnosis, and soil-specific fertilizer recommendations, can be leveraged to improve agricultural practices, such as crop production, real-time monitoring, and harvesting.

Here, we will consider the pivotal role that AI and ML technologies play in revolutionizing smart agriculture in previous research published so far. First, we will examine review papers that describe how artificial intelligence technology is applied to the growth status of field crops, the effects of nitrogen and irrigation, and yield prediction. Eli-Chukwu [1] outlines AI's applications in soil, crop, weed, and disease management, stressing both their potentials and constraints alongside expert system integration for enhanced productivity. Ayed and Hanana [2] showcase AI and ML's efficacy across various agricultural supply chain sectors, indicating a clear upward trend in their adoption to enhance the food industries. Noteworthy contributions by Jagli et al. [3] introduce AI applications such as sensor, robot, and drone utilization for irrigation, weeding, and spraying, collectively aiding in resource conservation, soil fertility preservation, and improved crop productivity. Hossen et al. [4] advocate for AI-driven automation in soil and disease management, crop monitoring, and weed control, highlighting their potential to alleviate agricultural challenges and reduce labor-intensive tasks. Akkem et al. [5] review ML and deep learning's relevance in agriculture, emphasizing their role in soil fertility assessment and crop selection, including time-series analysis and prediction. Oliveira and Silva [6] document the prevalent use of AI technologies, including ML, convolutional neural networks, and IoT, in agriculture, while also outlining the future directions and challenges. Sharma et al. [7] present a systematic review of ML applications in agriculture in their review paper. They focused on a prediction of soil parameters such as organic carbon and moisture content, crop yield prediction, and disease and weed detection in crops and species detection. They also examine ML using computer vision to classify diverse sets of crop images to monitor crop quality and yield assessment.

Second, we will review papers that describe the methods for predicting the raw weight or yield of onions or garlic using statistical models or machine learning techniques. Jeong et al. [8] prepared experimental plots where garlic and onions were grown according to various planting times and fertilizer rates. Then, they collected RGB and multispectral UAV images of onions and garlic at a spatial resolution of less than 1 cm to establish the correlation between the UAV images of garlic and onions and various biophysical parameters. In addition, they estimated the fresh weight of garlic and onions using two spectral indices, vegetation fraction and 3D-based height estimation obtained using an RGB camera, and studied the nutritional status of crops based on NDVI calculated using multispectral

images. Lee et al. [9] assessed the climate change from 1973 to 2017 in Gyeongsangnam-do, South Korea, and analyzed the relationship between weather conditions and onion bulb yield from the 1991/1992 to 2016/2017 growing seasons. Through linear regression analysis, they confirmed that temperature and precipitation were positively correlated with onion bulb yield. Przygocka-Cyna et al. [10] verified the seasonal trend of onion biomass and how its yield varies depending on the dynamics of nitrogen (N) and sulfur (S) uptake based on three years of field surveys. These experimental conditions consisted of three levels of nitrogen and three levels of sulfur absorption, and the total dry weight (TDW), total N uptake (TNU), and total S uptake (TSU) of onions were measured at 10-day intervals. Through experiments, they confirmed that TDW and TNU can be well explained by using an exponential linear model and TSU by using a quadratic growth model during the season. Desta et al. [11] conducted an experiment to evaluate the effects of three harvest stages (60, 80, and 100% top fall), two curing levels (non-cured and cured bulbs), and three storage methods (floor, shelf, and net bag) on the storability of the garlic variety ‘Tseday’ during 2014–2015. They conducted the experiment in a randomized complete block design repeated three times at both sites. The results of their study showed that harvesting at a peak drop rate of 80%, and curing and storing bulbs in racks or mesh bags led to good yields and post-harvest quality and effective storage potential of garlic bulbs under ambient storage conditions. Kim et al. [12] quantify onion and garlic growth parameters using UAV-derived data, developing predictive models for bulb weight estimation. Mwinuka et al. [13] evaluate thermal and multispectral imaging for assessing African eggplant irrigation performance, highlighting the sensitivity of the crop water stress index derived from mobile phone-based thermal images. Salari et al. [14] conducted a study at the Agricultural Research Farm of Kabul University to study the effects of land management practices and sowing dates on onion growth and yield. Different agronomic traits including the number of leaves per plant, leaf length, leaf area per plant, leaf area index, normalized difference vegetation index (NDVI), maturity period, marketable yield, and total yield were studied in these trials. They found that the planting date had a significant effect on onion growth and yield, but land management practices did not have a significant effect on onion growth and maturity. Kim et al. [15] performed a garlic bulb weight estimation analysis reflecting the characteristics of Korean garlic grown in an open field using growth survey data. Additionally, they built a step-by-step model to predict garlic bulb weight based on the fact that factors affecting garlic bulb weight can vary depending on the growth stage. During the analysis process, LASSO regression analysis was used for variable selection and coefficient estimation of garlic bulb weight. Kim and Soon [16] learned the Neural Prophet (NP) lagged time-series model using onion and garlic data from the Korea Rural Economic Institute. And they predicted the average fresh bulb weight of onion and garlic using the learned NP model. The prediction results showed that the average absolute error was within 5%.

In this study, we investigated how vegetation index and growth data derived from multispectral imaging affect the real-time live bulb weight prediction of onions and garlic grown in a field during the cultivation period. To implement this problem, we describe the methods for manually collecting growth data and extracting various vegetation indices from multispectral images. We also used concurrency regression models and various machine learning methods to understand how the weight of raw bulbs increases during the growing season of onions and garlic.

The structure of the paper is as follows. Section 2 describes the procedure that was used to collect multispectral images of onions and garlic grown in each experimental plot using a camera mounted on a drone, and collected data by directly measuring the corresponding growth data and live bulb weight. The methods for extracting various vegetation indexes from multispectral images, which are used to predict the fresh bulb weight of onions and garlic, are presented. Various statistical models and ML methods that can predict the live bulb weight based on the extracted vegetation index and growth data are described. Section 3 describes the experimental procedure for predicting the live bulb

weight using the R-package and compares the performance of various prediction methods. Section 4 discuss the current problem and future works. Section 5 summarizes the research results derived so far.

2. Materials and Methods

2.1. Study Area and Data Collection

First, to obtain the experimental data, onions and garlic were cultivated at the Chive Vegetable Research Institute in Muan-gun, Jeollanam-do. The cultivation period spanned from October 2022 to June 2023, and the final harvest was conducted on 2 June 2023. The experimental field was divided into two blocks, and each block was further divided into a total of 16 experimental plots. Figure 1 illustrates the overall structure of the experimental plot for onions and garlic.

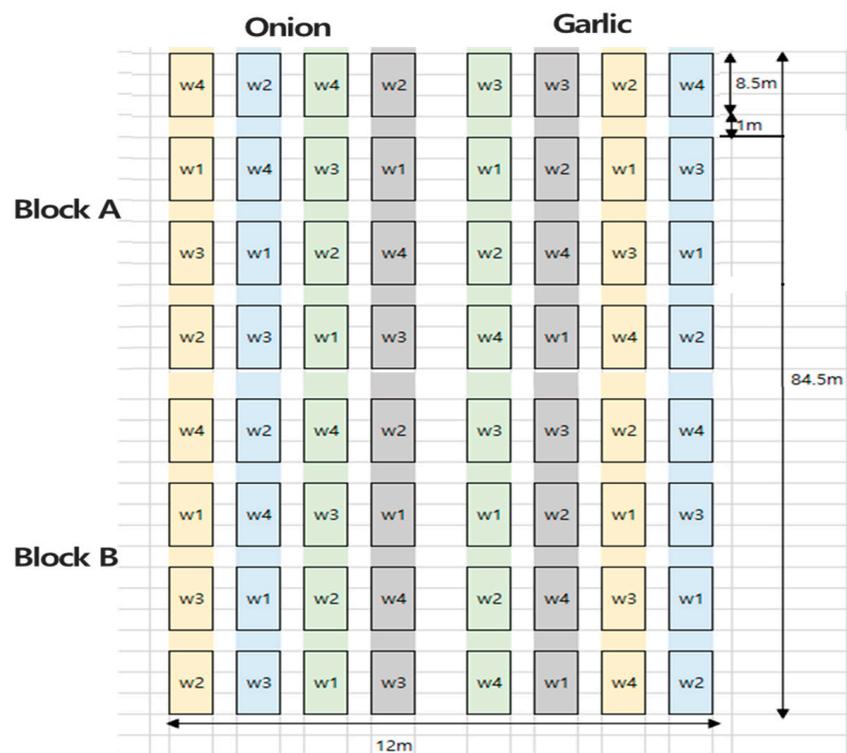


Figure 1. The structure of the experimental plots for onions and garlic.

Second, four growth parameters (plant height, leaf length, leaf width, and the number of leaves), as well as the live bulb weight of onions and garlic were measured at regular intervals. A total of 64 measurements were taken by randomly selecting four points in each of the 16 experimental units, and the same measurement was repeated seven times from 14 March to 30 May. The date on which each measurement was performed, and the measured values of each growth parameter and the live bulb weight at a particular location are given in Table 1.

Third, multispectral images of onions and garlic growing in the experimental area were repeatedly captured seven times using a drone, from 15 March to 31 May. The multispectral images comprised five channels of blue, green, red, red edge, and near-infrared (NIR). Figure 2 presents a sample multi-spectral image and its five distinct channel images.

Table 1. Measurement dates, growth parameter values, and the live bulb weight of onions and garlic during the cultivation period.

Crop	Crop Parameters	Date of Observation					
		3/14	4/3	4/19	5/9	5/23	5/30
Garlic	Plant height	38.25	61.63	86.25	84.25	71.00	62.75
	Leaf length	7	21	27.25	36.25	36	37.75
	Leaf width	10.55	14.225	16.7	14.83	14.6	12.68
	No. of leaves	6.25	7.5	8	5.75	5	4
	Live bulb weight	0.038	0.569	8.533	28.667	44.667	53.000
Onion	Plant height	30.08	48.13	68.75	70.00	58.75	46.33
	Leaf length	3.85	5	9	17.875	18.25	19.67
	Leaf width	9.8	15.38	20.9	19.03	9.63	10.67
	No. of leaves	5.8	7.5	9.5	7.8	5.5	4.7
	Live bulb weight	0.138	1.976	29.633	101	169.33	173.33

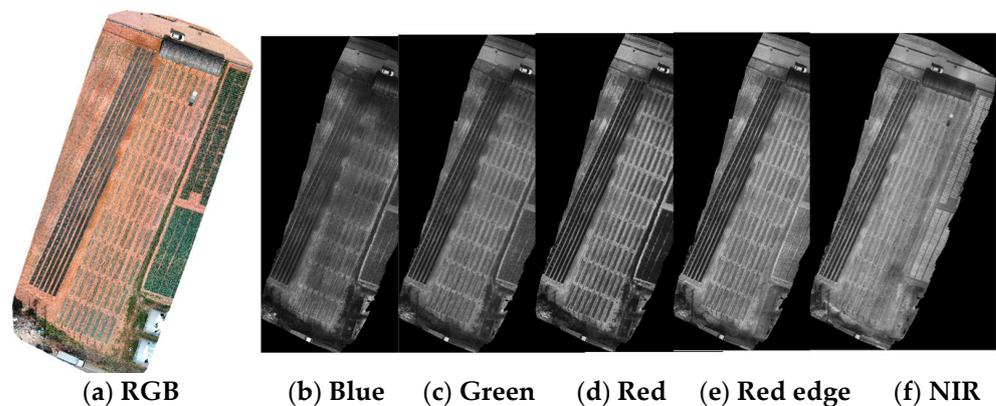


Figure 2. A sample multi-spectral image and its five distinct channel images.

2.2. Extraction of Vegetation Indexes

First, multispectral images were taken using a camera mounted on a drone to monitor the growth status of the onion and garlic crops in the experimental area during the growth period. The multispectral image acquired using the RedEdge-M multispectral camera (Micasense, Seattle, WA, USA) comprised five spectral bands: Blue (475 ± 20 nm), Green (560 ± 20 nm), Red (668 ± 10 nm), Red edge (717 ± 10 nm), and NIR (840 ± 40 nm). The parameters related to drone flight were set with a flight altitude of 20 m, a longitudinal overlap of 80%, and a lateral overlap of 60% for image capture. Multispectral images of the onions and garlic growing in the experimental area were collected through repeated filming six times from 15 March to 31 May. Figure 3 presents a sample multispectral image of the onion and garlic experimental plots.

	Multispectral Image				
	Blue	Green	Red	Red edge	NIR
Onion					
Garlic					

Figure 3. Sample multispectral images of the onion and garlic experimental plots.

Second, the methods for extracting vegetation indexes influencing the live bulb weight from the multispectral images of onions and garlic obtained during the cultivation period were examined [17–21]. For this task, 20 different vegetation indexes estimated from multispectral image bands (R, G, B, Red edge, and NIR) were used. The vegetation indexes were extracted using the numpy and skimage packages in a Python 3.8 environment. Each of the five multispectral images taken for each treatment plot was loaded using the skimage package, and the vegetation index was calculated by applying the relevant formula for each vegetation index using the numpy package. Table 2 specifies the names of the extracted vegetation indexes, their abbreviations, calculation formulas, and related references.

Table 2. Details of the multispectral vegetation indexes evaluated in the present study.

Indices	Acronym	Equation	Reference
Chlorophyll index red	CIred	$\frac{R_n}{R_r} - 1$	Gitelson et al. (2005) [22]
Chlorophyll vegetation index	CVI	$\frac{R_n R_r}{R_g^2}$	Vincini et al. (2008) [23]
Enhanced vegetation index	EVI	$\frac{2.5(R_n - R_r)}{R_n + 6R_r - 7.5R_b + 1}$	Gitelson et al. (2005) [22]
Two-band enhanced vegetation index	EVI2	$\frac{2.5(R_n - R_r)}{R_n + 2.4R_r + 1}$	Jiang et al. (2008) [24]
Green chlorophyll index	GCI	$\frac{R_n}{R_g} - 1$	Gitelson et al. (2005) [22]
Green normalized difference vegetation index	GNDVI	$\frac{R_n - R_g}{R_n + R_g}$	Hunt & Daughtry (2018) [25]
Green ratio vegetation index	GRVI	$\frac{R_n}{R_g}$	Sripada et al. (2006) [26]
Modified excess green	MEXG	$1.262R_g - 0.884R_r - 0.311R_b$	Burgos-Artizzu et al. (2011) [27]
Modified normalized green–red difference	MNGRD	$\frac{R_g^2 - R_r^2}{R_g^2 + R_r^2}$	Bendig et al. (2015) [28]
Normalized difference red edge	NDRE	$\frac{R_n - R_{re}}{R_n + R_{re}}$	Wang et al. (2014) [29]
Normalized difference vegetation index	NDVI	$\frac{R_n - R_r}{R_n + R_r}$	Gitelson et al. (2005) [22]
Normalized green–red difference	NGRD	$\frac{R_g - R_r}{R_g + R_r}$	Hamuda et al. (2016) [30]
Optimized soil adjusted vegetation index	OSAVI	$\frac{R_n - R_r}{R_n + R_r + 0.16}$	Roujeaux et al. (1996) [31]
Pigment-specific normalized vegetation index	PSND	$\frac{R_n - R_b}{R_n + R_b}$	Blackburn (1998) [32]
Renormalized difference vegetation index	RDVI	$\frac{R_n - R_r}{(R_n - R_r)^{0.5}}$	Roujean & Breon (1995) [33]
Red-edge chlorophyll index	RECI	$\frac{R_n}{R_{re}} - 1$	Gitelson et al. (2005) [22]
Ratio vegetation index	RVI	$\frac{R_n}{R_r}$	Gitelson et al. (2005) [22]
Soil adjusted vegetation index	SAVI	$\frac{1.5(R_n - R_r)}{R_n + R_r + 0.5}$	Zhong et al. (2019) [34]
Simplified canopy chlorophyll content index	SCCCI	$\frac{NDRE}{NDVI}$	Raper & Varco (2015) [35]
Triangular greenness index	TGI	$R_g - 0.39R_r - 0.61R_b$	Hunt et al. (2011) [36]

2.3. Prediction of the Live Bulb Weight

A correlation analysis method was designed to determine how the different vegetation indexes and growth data were related to the live bulb weight. Statistical regression models and ML methods capable of accurately predicting the live bulb weight were also examined. Particularly, we considered two functional simultaneous regression model-based methods, namely the Spline Smoothing-based method and Gaussian Process Regression-based method, and four ML methods, including Support Vector Regression, Random Forest Regression, AdaBoost, and Deep Neural Network, were examined.

2.3.1. Correlation Analysis

Generally, a statistical method that analyzes the correlation between two real variables, such as vegetation index and the live bulb weight, is called correlation analysis. Here, a measure of the correlation between the two variables, X and Y , was represented as the correlation coefficient defined as follows:

$$\rho = \frac{\text{Cov}(X, Y)}{SD(X) \times SD(Y)}, \quad (1)$$

where $\text{Cov}(X, Y)$ denotes the covariance of X and Y , and $SD(X)$ and $SD(Y)$ denote the standard deviation of X and Y , respectively. Furthermore, when n observed values for two variables (X, Y) are denoted as $(x_1, y_1), \dots, (x_n, y_n)$, the population correlation coefficient ρ is estimated as the sample correlation coefficient defined as follows.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(x_i - \bar{x})^2} \sqrt{(y_i - \bar{y})^2}}. \quad (2)$$

Next, the problem of testing whether a relationship exists between the two variables was considered. First, the statistical hypothesis for the correlation coefficient is given as follows.

$H_0 : \rho = 0$ (No correlation exists) versus $H_1 : \rho \neq 0$ (Correlation exists).

Second, the test statistic and rejection area at significance level α were defined as follows.

$$|T| = \left| \frac{\sqrt{n-2}}{\sqrt{1-r^2}} r \right| \geq t_{\frac{\alpha}{2}}(n-2), \quad (3)$$

where $t_{\frac{\alpha}{2}}(n-2)$ represents the percentile of the t -distribution with $(n-2)$ degrees of freedom.

2.3.2. Nonparametric Functional Concurrent Regression Model

Generally, function-to-function regression refers to a situation where both independent and dependent variables in a regression model are of a functional nature. Functional concurrent regression is a specific type of function-to-function regression that relates the response function at a specific point to the covariate value at that point and the point itself [24,37–47]. Standard functional concurrent models are linear (a linear combination of the covariates is used), and are often criticized for their linearity assumption and lack of flexibility. To address this issue, a nonparametric functional concurrent regression that models the response function at a specific point using a multivariate non-parametric function of both the point and the covariate value at that point was considered. Such models offer heightened flexibility and predictive accuracy, especially when the underlying relationship is nonlinear.

Here, the nonparametric consistency regression model is expressed as a mathematical formula. The observed data for i th subjects was assumed to be $\{Y_i(t), X_{i1}(t), \dots, X_{iQ}(t)\}$ ($i = 1, \dots, n$), where $Y_i(t)$ represents the functional response for $t = t_1, \dots, t_m$ and $X_{i1}(t), \dots, X_{iQ}(t)$ for $t = t_1, \dots, t_m$ are the Q corresponding functional covariates. The general form of an additive nonlinear functional concurrent model with Q covariates is defined as follows:

$$Y_i(t) = \mu_Y(t) + \sum_{q=1}^Q F_q\{X_{iq}(t), t\} + \epsilon_i(t), i = 1, \dots, n \quad (4)$$

where $\mu(t)$ is an unknown intercept function, and $F_q, q = 1, \dots, Q$ are unknown bivariate smooth functions.

Here, the methods for estimating the unknown intercept function $\mu(t)$ and unknown bivariate smooth function $F_q\{X_{iq}(t), t\}$ given in model (1) were examined. Generally, there

are three methods for estimating these two functions: a Spline Smoothing-based method, a Gaussian Process Regression-based method, and a Kernel Smoothing-based method.

First, an estimation method based on Spline Smoothing was considered. In the modeling process, $\{B_{\mu,d}(t)\}_{d=1}^{K_\mu}$ was considered to be a set of univariate B-spline basis functions defined on $[0, 1]$, where K_μ represents the basis dimensions. Based on these basis functions, $\mu_Y(t) = \sum_{d=1}^{K_\mu} B_{\mu,d}(t)\theta_{\mu,d} = \mathbf{B}_\mu^T(t)\Theta_\mu$, where $\mathbf{B}_\mu(t)$ is the K_μ -dimensional vector of $B_{\mu,d}(t)$'s, and Θ_μ is the vector of unknown parameters $\theta_{\mu,d}$'s. Furthermore, $F_q\{X_{iq}(t), t\}$ was modeled as a bivariate basis expansion using a tensor product of univariate B-spline basis functions. For $q = 1, \dots, Q$, the B-spline basis functions $\{B_{X_q,k}(x)\}_{k=1}^{K_{xq}}$ and $\{B_{T_q,l}(t)\}_{l=1}^{K_{tq}}$ were employed for x and t , respectively, and were used to model $F_q\{X_{iq}(t), t\}$. Thus, $F_q\{X_{iq}(t), t\}$ was represented as follows:

$$F_q\{X_{iq}(t), t\} = \sum_{k=1}^{K_{xq}} \sum_{l=1}^{K_{tq}} \theta_{q,k,l} B_{X_q,k}(x) B_{T_q,l}(t) = \mathbf{Z}_{i,q}^T(t)\Theta_q, \tag{5}$$

where $\mathbf{Z}_{i,q}^T(t)$ is the $K_{xq}K_{tq}$ -dimensional vector of $B_{i,X_q,k}(x)B_{T_q,l}(t)$, and Θ_q is the vector of unknown parameters, $\theta_{i,k,l}$'s. Based on the above expansions, model (1) was expressed as follows:

$$Y_i(t) = \mathbf{B}_\mu^T(t)\Theta_\mu + \sum_{q=1}^Q \mathbf{Z}_{i,q}^T(t)\Theta_q + \epsilon_i(t) \tag{6}$$

In this representation, a large number of basis functions were expected to result in a better but rougher fit, while a small number of basis functions were expected to result in an overly smooth estimate. Inconsistent with the literature, a large number of basis functions were employed to fully capture the complexity of the functions, and the coefficients were penalized to ensure the smoothness of the resulting fit. Consequently, a criterion that incorporates penalty functions for parameter vectors was considered.

Here, Θ_μ and Θ_q , $q = 1, \dots, Q$ were estimated by minimizing a penalized criterion as follows:

$$L(\Theta) = \sum_{i=1}^n \| Y_i(t) - (\mathbf{B}_\mu^T(t)\Theta_\mu + \sum_{q=1}^Q \mathbf{Z}_{i,q}^T(t)\Theta_q) \|^2 + \Theta_\mu^T \mathbb{P}_\mu \Theta_\mu + \sum_{q=1}^Q \Theta_q^T \mathbb{P}_q \Theta_q \tag{7}$$

where \mathbb{P}_μ and \mathbb{P}_q , $q = 1, \dots, Q$ are the penalty matrices for the smoothness of $\mu_Y(t)$ and $F_q\{X_{iq}(t), t\}$, $q = 1, \dots, Q$, respectively, and contain penalty parameters that regularize the trade-off between the goodness of fit and the smoothness of fit. The minimization of $L(\Theta)$ was straightforward. A closed-form expression of the estimator is as follows:

$$\hat{\Theta} = \left(\sum_{i=1}^n \mathbf{Z}_i^T \mathbf{Z}_i + \mathbb{P} \right)^{-1} \sum_{i=1}^n \mathbf{Z}_i^T \mathbb{Y}_i \tag{8}$$

where $\mathbb{Y}_i = [Y_i(t_1), \dots, Y_i(t_m)]^T$ and $\mathbf{Z}_i = [\mathbb{B}_\mu | \mathbb{Z}_{i,1} | \dots | \mathbb{Z}_{i,Q}]$. The penalty parameters can be determined based on Generalized Cross-Validation (GCV) or restricted maximum likelihood (REML).

Finally, the prediction of response trajectory when a new covariate and its evaluation points are given was considered. Considering that a new unknown response $Y_{new}(t)$ is obtained based on new observations $[X_{1,new}(t), \dots, X_{Q,new}(t)]$, then, this new response can be predicted using the following equation:

$$\hat{Y}_{new}(t) = \sum_{d=1}^{K_\mu} B_{\mu,d}(t)\hat{\theta}_{\mu,d} + \sum_{q=1}^Q \sum_{k=1}^{K_{xq}} \sum_{l=1}^{K_{tq}} B_{X_q,k}(X_{q,new}(t)) B_{T_q,l}(t)\hat{\theta}_{q,k,l} \tag{9}$$

where $\hat{\theta}_{\mu,d}$ and $\hat{\theta}_{q,k,l}$ are estimated based on the above formula.

Second, an estimation method based on Gaussian Process Regression was explored. A Gaussian process is a set of random variables, where any finite subset follows a multivariate normal distribution. Such a process via $GP\{m(\cdot), C(\cdot, \cdot)\}$, where $m(\cdot)$ and $C(\cdot, \cdot)$ are the mean and covariance functions, respectively. A realization $X(\cdot)$ from such a process is a random function with $E(X(\mathbf{x})) = m(\mathbf{x})$ and $Cov(X(\mathbf{x}), X(\mathbf{x}')) = C(\mathbf{x}, \mathbf{x}')$, where \mathbf{x} and \mathbf{x}' are

points in the domain of the process. Furthermore, for any finite set of points x_1, \dots, x_n , the vector $(X(x_1), \dots, X(x_n))^T$ followed a multivariate normal distribution.

Given a dataset consisting of n data points $\{y_i, \mathbf{x}_i, i = 1, \dots, n\}$, where for each i , \mathbf{x}_i is a Q -dimensional vector of inputs, and y_i is the output, a Gaussian Process Regression model is defined as follows:

$$Y_i = f(\mathbf{x}_i) + \epsilon_i, \tag{10}$$

where $\epsilon_i \sim N(0, \sigma^2)$ is an error term. The unknown function $f(\mathbf{x}_i)$ is a nonlinear function of \mathbf{x}_i . The prior for this function was assumed to correspond to a Gaussian process, i.e., for each i , $f(\mathbf{x}_i)$ followed a multivariate normal distribution with zero mean, and there existed a covariance function $C(\mathbf{x}_i, \mathbf{x}_j) = \text{Cov}(f(\mathbf{x}_i), f(\mathbf{x}_j))$. An example of such a covariance function is as follows:

$$C(\mathbf{x}_i, \mathbf{x}_j) = C(\mathbf{x}_i, \mathbf{x}_j; \boldsymbol{\theta}) = v_0 \exp\left(-\frac{1}{2} \sum_{q=1}^Q w_q (x_{iq} - x_{jq})^2\right) + a_0 + a_1 \sum_{q=1}^Q x_{iq} x_{jq} \tag{11}$$

where $\boldsymbol{\theta} = (w_1, \dots, w_Q, v_0, a_0, a_1)$ denotes the set of unknown parameters. Therefore, $\mathbf{Y} = (y_1, \dots, y_n)$ followed a normal distribution with zero mean and a covariance matrix as follows:

$$\Psi(\boldsymbol{\theta}) = C(\boldsymbol{\theta}) + \sigma_v^2 \mathbf{I} \tag{12}$$

where \mathbf{I} is an identity matrix, $C(\boldsymbol{\theta})$ is an $(n \times n)$ matrix with elements as given in (11), and $\Psi(\boldsymbol{\theta})$ is an $(n \times n)$ matrix.

Shi et al. [45] developed a Gaussian Process Regressions framework for nonparametric concurrent models as follows:

$$Y_i(t) = \mathbf{Z}_i^T \boldsymbol{\beta}(t) + F_i\{X_{i1}(t), \dots, X_{iQ}(t)\} + \epsilon_i(t), \tag{13}$$

where $\boldsymbol{\beta}(t)$ is a vector of unknown coefficient functions, and $F_i\{X_{i1}(t), \dots, X_{iQ}(t)\}$ is the Gaussian process with zero mean and covariance kernel function $C(\mathbf{x}_i, \mathbf{x}_j) \mathbf{x}_i = \mathbf{x}_i(t)$. The model components were estimated using a two-stage approach. In the first stage, each observed response curve was smoothed using a B-spline function: $Y_i(t) = \Phi^T(t) A_i$, where $\Phi(t) = [\Phi_1(t), \dots, \Phi_Q(t)]^T$ are B-spline basis functions, and A_i is a $Q \times 1$ vector of coefficients. These coefficients were estimated by minimizing $\int (Y_i(t) - \Phi^T(t) A_i)^2 dt$ with respect to A_i for each $i = 1, \dots, n$. The unknown coefficient function $\boldsymbol{\beta}(t)$ was also modeled using a B-spline basis function: $\boldsymbol{\beta}(t) = \Phi^T(t) B$, where B is a matrix of unknown coefficients. The matrix B was estimated as $(Z^T Z)^{-1} Z^T A$, where $Z = [Z_1, \dots, Z_n]^T$ and similar calculations were performed for A . In the second stage, the residuals from the first stage were computed as follows:

$$\tilde{F}_i\{X_{i1}(t), \dots, X_{iQ}(t)\} = Y_i(t) - Z_i^T \hat{\boldsymbol{\beta}}(t). \tag{14}$$

Then, $\tilde{F}_i\{X_{i1}(t), \dots, X_{iQ}(t)\}$ was modeled through a Gaussian process with the covariance function $C(\mathbf{x}, \mathbf{x}' : \boldsymbol{\theta})$. The covariance components involved in $C(\cdot)$ were estimated either through Maximum Likelihood Estimation (MLE) or using a Bayesian approach.

Next, the generation of a prediction y^* at a new test point (t^*, \mathbf{x}^*) with $\mathbf{x}^* = \mathbf{x}(t^*)$ was considered. Using Equation (11), the predicted value was given by $\hat{y}^* = Z_i^T \hat{\boldsymbol{\beta}}(t^*) + \hat{F}_i\{\mathbf{x}^*\}$, where $\hat{F}_i\{\mathbf{x}^*\}$ is predicted by its conditional mean $E(F_i\{\mathbf{x}^*\} | D)$ through the Gaussian process with the estimated covariance function $C(\mathbf{x}, \mathbf{x}' : \hat{\boldsymbol{\theta}})$.

2.3.3. Machine Learning Methods

The first method, Support Vector Regression (SVR), was initially proposed by Drucker et al. [48]. It is a variant of support vector machines (SVMs), a popular algorithm for classification tasks. SVR extends an SVM's capabilities to solve regression problems by optimizing an epsilon-insensitive loss function. SVR is an algorithm that aims

to discover the optimal fit for a given dataset. It operates by constructing a hyperplane in high-dimensional feature space to maximize the margin around training data points while minimizing the error on unseen data points. Unlike other regression models that minimize the error across all data points, SVR focuses on minimizing the error within a specific margin.

The second method, Random Forest Regression (RFR), is a supervised learning algorithm that uses an ensemble learning method for regression [49] as shown in Figure 4. The ensemble learning method is a technique that combines predictions from multiple ML algorithms to generate predictions that are more accurate than those generated by a single model.

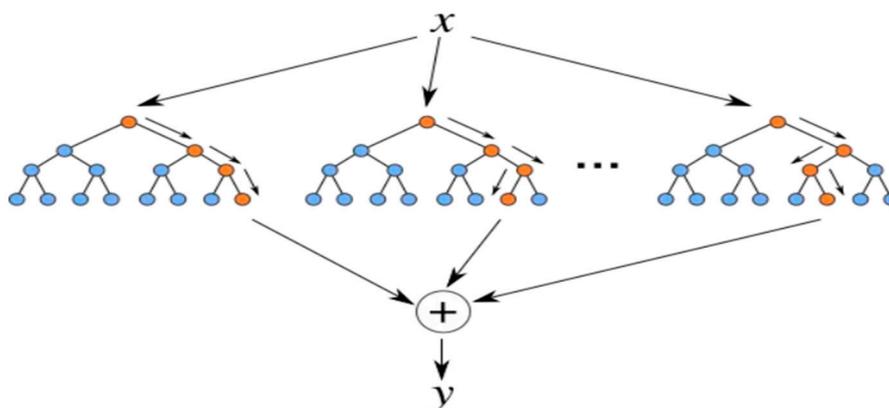


Figure 4. Flow of ensemble learning for regression in Random Forest.

The RFR model is renowned for its robustness and accuracy, excelling across various problem domains, especially those involving features with non-linear relationships. However, it comes with notable drawbacks, including a lack of interpretability, susceptibility to overfitting, and the need to choose the number of trees for inclusion in the model.

The third method, Adaptive Boost (AdaBoost), is an algorithm that sets the initial model as a weak model. Subsequently, it uses weights at each step to sequentially fit a new model that compensates for the weaknesses of the previous model. The final model is created by linearly combining these sequentially fitted models. At every step, AdaBoost increases the weight of data with large errors or misclassification in the previous learning data, while reducing the weight of data with low errors or correct classification. By iteratively adjusting the weights and extracting new training data, AdaBoost emphasizes previously mispredicted data, resulting in a model that predicts them more accurately. Originally introduced for binary classification by Freund and Schapire [50], AdaBoost has been adapted for regression problems. Its success in delivering accurate ensembles and its resistance to overfitting led Breiman to call AdaBoost the “best off-the-shelf classifier in the world” (NIPS Workshop 1996) [51].

The fourth method is based on an Artificial Neural Network (ANN) or a simple traditional neural network, which aims to solve trivial tasks using a straightforward network outline. An ANN is loosely inspired by biological neural networks. It is a collection of layers to perform a specific task. Each layer consists of a collection of nodes that operate together. Typically comprising an input layer, one to two hidden layers, and an output layer, ANNs are suited for solving basic mathematical and computer problems, including fundamental gate structures with their respective truth tables. However, they face challenges when applied to complex tasks like image processing, computer vision, and natural language processing. In contrast, Deep Neural Networks (DNNs) feature a more intricate hidden layer structure, encompassing various layers, such as a convolutional layer, max-pooling layer, dense layer, and other unique layers [52]. These additional layers enhance the model’s ability to understand complex problems and provide optimal solutions. A DNN has more layers (more depth) than an ANN, and each layer adds complexity to the

model and enables the model to process the inputs concisely to provide the ideal solution as the output. DNNs have garnered extremely high traction due to their high efficiency in handling a wide variety of deep-learning projects.

3. Experimental Results

3.1. Graphical Interpretation and Correlation Analysis

First, the relationships between the 20 vegetation indexes and the four growth parameters (plant height, leaf length, leaf width, and the number of leaves) were investigated using graph analysis and correlation analysis. Figure 5 presents scatter plots indicating the relationship between four growth characteristics and the NDVI, a representative vegetation index extracted from the multispectral images of onions and garlic. Figure 5 indicates that plant height, leaf width, and the number of leaves exhibited a highly positive correlation with the NDVI, while the leaf length exhibited a comparatively limited correlation with the NDVI, depending on the duration of cultivation.

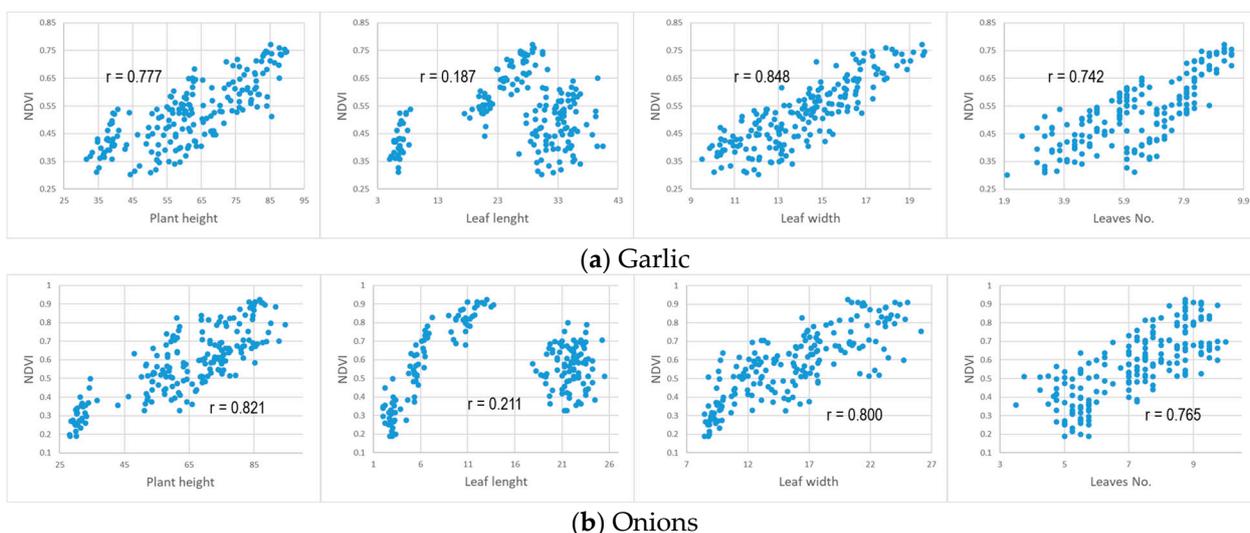


Figure 5. Scatter plots depicting the relationships between four growth characteristics and NDVI.

Next, correlation analysis was conducted to determine the relationship between the 20 vegetation indexes and the four growth parameters. As shown in Table 3, three of the growth parameters, excluding the leaf length, exhibited a very high correlation with the 20 vegetation indexes. Notably, the leaf length exhibited a low correlation with the 20 vegetation indexes. The vegetation indexes that exhibited the highest correlation with the growth parameters were the NDRE, RDVI, PSND, and NDVI, which are commonly used indexes for determining the crop growth status. These findings are consistent with the observations derived from the scatter plots.

Second, graph analysis and correlation analysis were conducted to determine the relationships between the studied vegetation indexes and live bulb weight during the growing season of onions and garlic. Figure 6 visually illustrates the relationship between various vegetation indexes and the live bulb weights of onions and garlic. Figure 6 clearly indicates an overall negative correlation between the live bulb weight and vegetation indexes during the growing period. This observation is consistent with the fact that, as the stems and leaves of onions and garlic mature, they gradually dry out, leading to a decrease in the extracted vegetation indexes.

Table 3. Correlation coefficients for the relationships between four growth parameters and 20 vegetation indexes (Bold number is highest score in the column).

Index	Garlic				Onion			
	Plant Height	Leaf Length	Leaf Width	No. of Leaves	Plant Height	Leaf Length	Leaf Width	No. of Leaves
CIred	0.754	0.179	0.827	0.713	0.611	−0.005	0.663	0.584
CVI	0.768	0.329	0.707	0.472	0.635	−0.096	0.745	0.750
EVI	0.693	0.051	0.837	0.830	0.803	0.150	0.830	0.777
EVI2	0.731	0.097	0.842	0.804	0.820	0.179	0.830	0.788
GCI	0.794	0.220	0.841	0.696	0.666	−0.011	0.732	0.673
GNDVI	0.795	0.207	0.843	0.717	0.781	0.111	0.806	0.787
GRVI	0.794	0.220	0.841	0.696	0.666	−0.011	0.732	0.673
MEXG	0.661	0.136	0.764	0.713	0.827	0.650	0.591	0.519
MNGRD	0.694	0.120	0.812	0.755	0.782	0.249	0.731	0.645
NDRE	0.763	0.166	0.868	0.771	0.733	0.017	0.808	0.780
NDVI	0.777	0.187	0.848	0.742	0.821	0.211	0.800	0.765
NGRD	0.694	0.121	0.811	0.752	0.769	0.235	0.722	0.634
OSAVI	0.753	0.131	0.848	0.785	0.827	0.201	0.821	0.782
PSND	0.859	0.383	0.815	0.567	0.870	0.326	0.782	0.778
RDVI	0.704	0.072	0.825	0.809	0.845	0.241	0.833	0.795
RECI	0.757	0.162	0.864	0.763	0.671	−0.045	0.770	0.714
RVI	0.754	0.179	0.827	0.713	0.611	−0.005	0.663	0.584
SAVI	0.733	0.101	0.842	0.802	0.828	0.196	0.831	0.790
SCCCI	0.552	0.064	0.715	0.693	0.474	−0.297	0.677	0.731
TGI	0.632	0.186	0.670	0.589	0.670	0.810	0.337	0.297

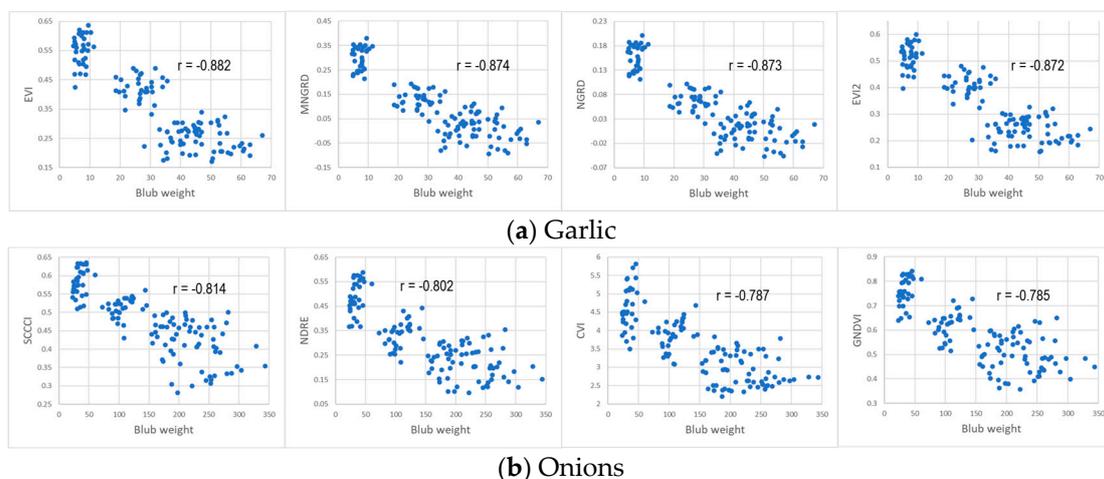


Figure 6. Scatter plots depicting the relationships between live bulb weight and four vegetation indexes.

In addition, we calculated the correlation coefficient to numerically confirm that there is a negative correlation between the live bulb weight and vegetation index, which we identified from the above scatter plot. From Table 4 given below, we can see that most of the vegetation indices have a strong negative correlation with the live bulb weight.

Third, graph analysis and correlation analysis were conducted to determine the relationship between the four growth parameters and the live bulb weights of onions and garlic during the cultivation period. Figure 7 visually illustrates the relationships between the four growth characteristics of onions and garlic and the fresh bulb weight. The results indicate that as onions and garlic grow, the values of the three growth factors, excluding stem length, exhibit a negative correlation with the weight of the fresh bulb. On the other hand, only the leaf length exhibits a positive correlation with the bulb weight.

Table 4. Correlation coefficients for the relationships between live bulb weight and 20 vegetation indexes.

Crop	Index	r	Index	r	Index	r	Index	r
Garlic	EVI	-0.88213	SAVI	-0.86864	RECI	-0.82706	GNDVI	-0.79407
	MNGRD	-0.87429	RDVI	-0.86652	CIred	-0.81859	TGI	-0.73566
	NGRD	-0.87265	OSAVI	-0.85837	RVI	-0.81859	PSND	-0.73122
	EVI2	-0.87170	NDVI	-0.83005	GCI	-0.79714	SCCCI	-0.63974
	MEXG	-0.87026	NDRE	-0.82949	GRVI	-0.79714	CVI	-0.60588
Onion	SCCCI	-0.81404	EVI	-0.78047	NDVI	-0.76211	CIred	-0.64783
	NDRE	-0.80220	OSAVI	-0.77935	PSND	-0.74064	RVI	-0.64783
	CVI	-0.78715	EVI2	-0.77855	GRVI	-0.72242	NGRD	-0.63909
	GNDVI	-0.78523	RDVI	-0.77170	GCI	-0.72242	MEXG	-0.17486
	SAVI	-0.78135	RECI	-0.76769	MNGRD	-0.64898	TGI	0.45755

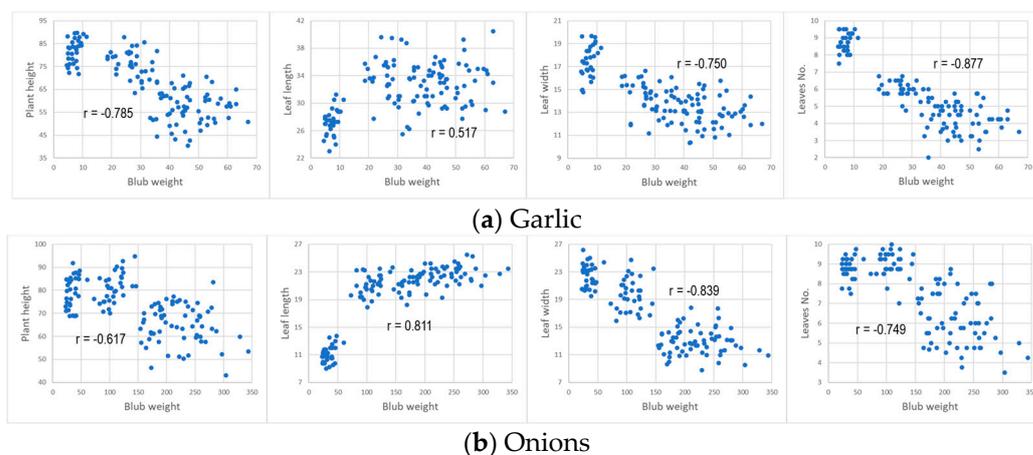


Figure 7. Scatter plots depicting the relationships between live bulb weight and four growth parameters.

Additionally, the correlation coefficient was calculated to numerically substantiate the positive correlation observed in the scatter plots depicting the relationships between the live bulb weight and vegetation indexes. The correlation coefficients presented in Table 5 indicate a negative correlation between the live bulb weight and three of the growth factors, excluding the leaf length. On the other hand, among the growth factors studied, only the leaf length exhibited a positive correlation with the live bulb weight. This is consistent with the observations derived from the scatter plots.

Table 5. Hyperparameters of each machine learning algorithm.

DNN		SVR		RFR		AdaBoost	
Hyper Parameters	Value	Hyper Parameters	Value	Hyper Parameters	Value	Hyper Parameters	Value
optimizer	adam	C	10	n_estimators	100	n_estimators	100
learning rate	0.07	epsilon	0.2	criterion	squared_error	loss	linear
loss	mse	kernel	linear	max_depth	8		
epochs	100						
batch_size	8						

3.2. Prediction of the Live Bulb Weight

First, 20 vegetation indexes were used as the input variables and two types of statistical regression models and four types of ML models were considered for predicting the raw weight of onions and garlic. To visually assess the prediction power of these models, two-dimensional scatter plots comparing the actual observed and predicted values were generated. Figure 8 illustrates the performance of the six prediction models. The results

clearly indicate that FCR and RFR have the best prediction ability, followed by AdaBoost and SVR, while Gaussian Process Functional Data Analysis (GPFDA) and DNN have the lowest prediction ability.

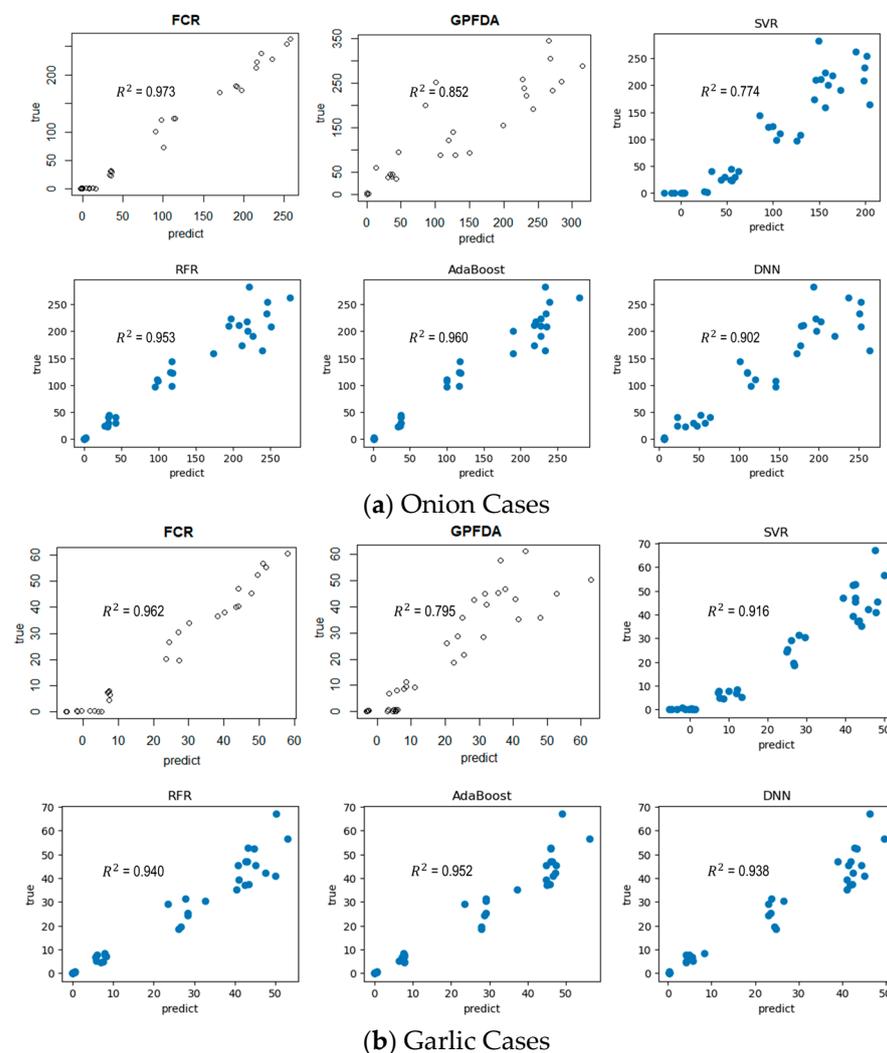


Figure 8. Scatter plots depicting the relationships between the observed and predicted values of live bulb weight based on 20 vegetation indexes.

Second, to predict the fresh bulb weight of onions and garlic, four growth factors (plant height, leaf width, leaf length, and number of leaves) were used as the input variables, and two types of statistical regression models and four types of ML models were considered. Two-dimensional scatter plots comparing the actual observed and predicted values were generated to visually present the prediction power of the models. Figure 9 clearly depicts that, in the case of onions, FCR and GPFDA have the best prediction ability, followed by RFR and AdaBoost, while SVM and DNN have the lowest prediction ability. On the other hand, in the case of garlic, FCR has the best prediction ability, followed by RFR and AdaBoost, while GPFDA, SVM, and DNN have a comparatively poorer prediction ability.

Third, the following four measures were considered to quantitatively evaluate the prediction accuracy of the two types of statistical regression models and four types of machine learning models used to predict the fresh bulb weight of onions and garlic in this paper. These are the coefficient of determination (R^2), root mean square error (RMSE), normalized root mean square error (nRMSE), and mean absolute percentage error (MAPE), respectively.

$$R^2 = 1 - \left(\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \right), \text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - p - 1}}, \text{nRMSE} = 100 \times \frac{\text{RMSE}}{\bar{y}}, \text{MAPE} = \frac{100}{n} \times \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (15)$$

where *SSE* is the sum of square error, *SST* is the total sum of square, and *n* is the number of observations.

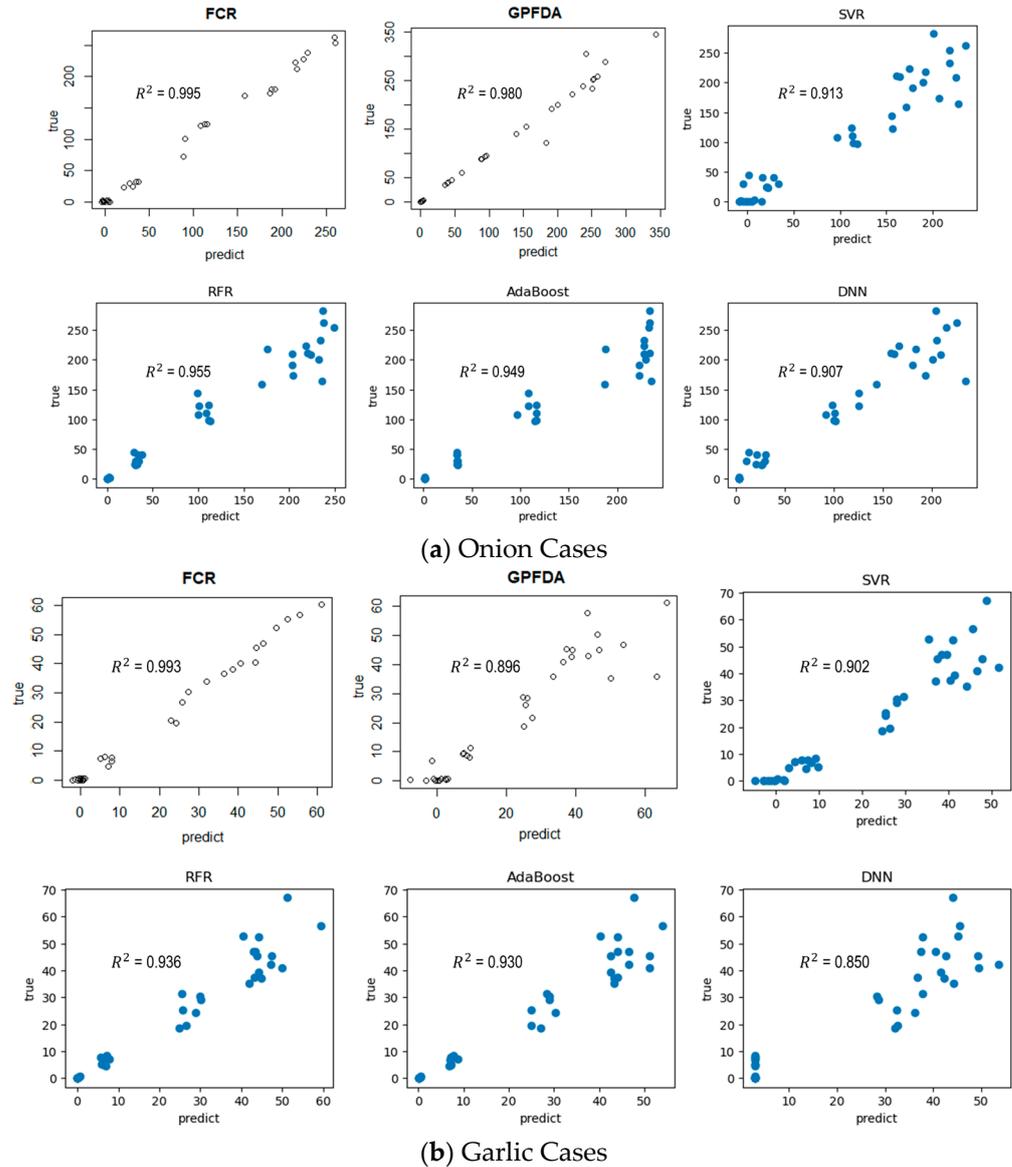


Figure 9. Scatter plots depicting the relationships between the observed and predicted values of live bulb weight based on four growth factors.

Fourth, the onion and garlic fresh bulb weight prediction performance of the six prediction models was quantitatively analyzed using four indicators. Hyperparameters for each machine learning algorithm are listed in Table 5. For the training and testing, we randomly shuffled the entire dataset and utilized 80% as training data, 20% as test data

Table 6 shows the four measurement values corresponding to the six prediction models used for predicting the live bulb weight of onions and garlic when 20 vegetation indexes were used as input variables. These indexes were obtained from images captured over multiple dates and calculated from those images. For each date of image capture, 20 vegetation indices were calculated, resulting in 20 vegetation values measured on the

same date. The results clearly indicate that the FCR model has the best prediction power, the RFR and AdaBoost models have the second-best prediction power, and the SVR and DNN models have the third-best prediction power, while the GPFDA model has the poorest prediction power.

Table 6. Predictions of live bulb weight using 20 vegetation indexes.

Crop	Model	R^2	RMSE	nRMSE	MAPE
Garlic	FCR	0.962	3.233	15.108	13.283
	GPFDA	0.795	10.383	48.856	39.654
	SVR	0.916	5.750	25.182	19.688
	RFR	0.940	4.868	21.610	14.757
	AdaBoost	0.952	4.390	19.487	12.927
	DNN	0.938	4.989	22.145	16.217
Onion	FCR	0.973	11.413	12.495	9.428
	GPFDA	0.852	52.764	49.319	39.490
	SVR	0.774	38.498	37.006	27.215
	RFR	0.953	21.161	20.341	12.310
	AdaBoost	0.960	19.600	18.840	11.858
	DNN	0.902	29.372	28.233	19.438

Table 7 shows the four measurement values corresponding to the six prediction models used for predicting the live bulb weight of onions and garlic when four growth factors were used as the input variables. The results indicate that the FCR model has the best prediction power, the RFR and AdaBoost models have the second-best prediction power, and the SVR and GPFDA models have the third-best prediction power, while the DNN model has the poorest prediction power.

Table 7. Predictions of live bulb weight using growth factor items.

Crop	Model	R^2	RMSE	nRMSE	MAPE
Garlic	FCR	0.993	2.024	9.459	7.744
	GPFDA	0.896	6.838	32.176	20.253
	SVR	0.902	6.150	27.302	18.918
	RFR	0.936	5.237	23.247	15.350
	AdaBoost	0.930	5.358	23.784	15.219
	DNN	0.850	7.475	33.182	24.404
Onion	FCR	0.995	7.782	8.520	7.157
	GPFDA	0.980	15.116	14.129	4.111
	SVR	0.913	27.281	26.223	19.043
	RFR	0.955	20.074	19.295	11.764
	AdaBoost	0.949	21.842	20.995	14.548
	DNN	0.907	26.810	25.771	16.809

Together, the results of the two types of analyses revealed that using the four growth factors rather than the 20 vegetation indexes as the input variables resulted in improved live bulb weight prediction for both onions and garlic. Moreover, the prediction power of the six prediction models maintained a consistent ranking, irrespective of whether growth factors or vegetation indexes were used as the input variables.

4. Discussion

Previous studies have primarily utilized artificial intelligence and machine learning technologies for productivity enhancement, disease management, and crop monitoring [1–7]. Additionally, attempts have been made to use UAVs and spectral images to acquire various biological parameters [17–21]. However, not much research has been conducted so far on predicting the weight of onion and garlic bulbs in real time.

This study was conducted in four stages. First, two pilot blocks each were selected for both onions and garlic, with each block comprising a total of 16 experimental units, four horizontals and four verticals. Thus, a total of 32 experimental units were prepared for both onions and garlic. Second, image data were collected using a multispectral camera, which was repeated a total of seven times for each area in the 32 experimental units prepared for both onions and garlic. Simultaneously, growth data and the live bulb weight at the corresponding points were recorded manually. Third, correlation analysis was conducted to determine the relationship between various vegetation indexes extracted from multispectral images and the manually measured growth data and live bulb weights. Fourth, based on the vegetation index extracted from multispectral images and previously collected growth data, we proposed a method to predict the live bulb weight of onions and garlic in real time during the cultivation period using a functional concurrence regression model and ML techniques.

The academic value and applicability of our research results described in this paper are as follows. First, our findings can inform agricultural practices aimed at optimizing the cultivation of onions and garlic. By understanding the factors influencing the weight of fresh bulbs during the growth period, farmers can adjust their cultivation methods, such as irrigation, fertilization, and pest control, to enhance the yield and quality. Second, our study contributes to the advancement of machine learning applications in agriculture. The development of predictive models for onion and garlic bulb weight estimation opens avenues for the creation of decision support systems for farmers. These systems can provide real-time recommendations based on environmental conditions and crop characteristics, aiding in more precise and efficient farming practices. Third, our research results can be applied to various fields beyond agriculture. The methodologies and techniques employed, such as image analysis and machine learning algorithms, can be adapted for the monitoring and managing of various other crops and plant species. This broader application potential underscores the significance of our research in advancing technology-driven solutions for sustainable agriculture and food security.

5. Conclusions

In this study, we introduced an automated method to predict the growth trajectory of onion and garlic bulb weights during the growing season. To establish this method, data, including multispectral images, growth factors, and the live bulb weight, were collected at consistent intervals from planting to harvest. Subsequently, six artificial intelligence models were employed to predict the live bulb weight based on 20 vegetation indexes derived from multispectral images, pre-collected growth factors, and live bulb weight data. The experimental results revealed that the FCR model showed the most robust prediction performance both when using four growth factors and when using 20 vegetation indexes. Following closely, with a slight distinction, GPFDA, RFR, and AdaBoost exhibited the next-best predictive power. However, SVM and DNN displayed comparatively poorer predictive power. Notably, when using the four growth factors as explanatory variables, all the predictive models exhibited improved prediction performance compared to that when using the 20 vegetation indexes.

Author Contributions: Study conception and design: M.H.N., W.C. and I.N.; data collection: D.K. and M.H.N.; software, D.K.; analysis and interpretation of results: I.N., W.C. and D.K.; manuscript draft preparation: W.C., D.K. and I.N. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the research funding from the Rural Development Administration (RDA) (RS-2022-RD010424).

Institutional Review Board Statement: Ethical review and approval were waived for this study due to the data collection being conducted through observational methods without direct interaction with human subjects. Therefore, no separate ethical review or approval process was deemed necessary.

Data Availability Statement: The datasets generated or analyzed during the study are available from the corresponding author upon reasonable request; however, some data are unavailable due to commercial restrictions.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

- Eli-Chukwu, N.C. Applications of Artificial Intelligence in Agriculture: A Review. *Eng. Technol. Appl. Sci. Res.* **2019**, *9*, 4377–4383. [[CrossRef](#)]
- Ayed, R.B.; Hanana, M. Artificial Intelligence to Improve the Food and Agriculture Sector. *J. Food Qual.* **2021**, *2021*, 5584754.
- Jagli, D.; Purohit, S.; Agale, T.; Kahar, D. Smart Farming Using Artificial Intelligence. In Proceedings of the ACM-2022: Algorithms Computing and Mathematics Conference, Chennai, India, 29–30 August 2022.
- Hossen, M.I.; Fahad, N.; Sarkar, M.R.; Rabbi, M.R. Artificial Intelligence in Agriculture: A Systematic Literature Review. *Turk. J. Comput. Math. Educ.* **2023**, *14*, 137–146.
- Akkem, Y.; Biswas, S.K.; Varanasi, A. Smart farming using artificial intelligence: A review. *Eng. Appl. Artif. Intell.* **2023**, *120*, 105899. [[CrossRef](#)]
- Oliveira, R.C.; Souza e Silva, R.D. Artificial Intelligence in Agriculture: Benefits, Challenges, and Trends. *Appl. Sci.* **2023**, *13*, 7405. [[CrossRef](#)]
- Sharma, A.; Jain, A.; Gupta, P.; Chowdary, V. Machine Learning Applications for Precision Agriculture: A Comprehensive Review. *IEEE Access* **2021**, *9*, 4843–4873. [[CrossRef](#)]
- Jeong, S.-J.; Kim, D.-W.; Yun, H.; Cho, W.-J.; Kwon, Y.-S.; Kim, H.-J. Monitoring the growth status variability in Onion (*Allium cepa*) and Garlic (*Allium sativum*) with RGB and multi-spectral UAV remote sensing imagery. In Proceedings of the 7th Asian-Australasian Conference on Precision Agriculture, Hamilton, New Zealand, 16–18 October 2017.
- Lee, J.; Min, B.; Yoon, S.; Lee, M.; Kim, H.; Hong, K. A multiple-regression model of bulb onion yield in response to meteorological conditions in Gyeongsangnam province, Republic of Korea. *Acta Hortic.* **2019**, *1251*, 81–90. [[CrossRef](#)]
- Przygocka-Cyna, K.; Barłóg, P.; Grzebisz, W.; Spizewski, T. Onion (*Allium cepa* L.) Yield and Growth Dynamics Response to In-Season Patterns of Nitrogen and Sulfur Uptake. *Agronomy* **2020**, *10*, 1146. [[CrossRef](#)]
- Desta, B.; Woldetsadik, K.; Ali, E.N. Effect of Harvesting Time, Curing and Storage Methods on Storability of Garlic Bulbs. *Open Biotechnol. J.* **2021**, *15*, 36–45. [[CrossRef](#)]
- Kim, D.-W.; Jeong, S.-J.; Lee, W.S.; Yun, H.; Chung, Y.S.; Kwon, Y.-S.; Kim, H.-J. Growth monitoring of field-grown onion and garlic by CIE L*a*b* color space and region-based crop segmentation of UAV RGB images. *Precis. Agric.* **2023**, *24*, 1982–2001. [[CrossRef](#)]
- Mwinuka, P.R.; Mbilinyi, B.P.; Mbungu, W.B.; Mourice, S.K.; Mahoo, H.F.; Schmitter, P. The feasibility of hand-held thermal and UAV-based multispectral imaging for canopy water status assessment and yield prediction of irrigated African eggplant (*Solanum aethiopicum* L.). *Agric. Water Manag.* **2021**, *245*, 106584. [[CrossRef](#)]
- Salari, H.; Antil, R.S.; Saharawat, Y.S. Responses of onion growth and yield to different planting dates and land management practices. *Agron. Res.* **2021**, *19*, 914–1928.
- Kim, J.; Lee, H.; Suh, T. Analysis of Predictions of Garlic Bulb Weight using LASSO Regression Model. *Hortic. Sci. Technol.* **2023**, *41*, 437–447.
- Kim, W.; Soon, B.M. Advancing Agricultural Predictions: A Deep Learning Approach to Estimating Bulb Weight Using Neural Prophet Model. *Agronomy* **2023**, *13*, 1362. [[CrossRef](#)]
- Xue, J.; Su, B. Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *Hindawi J. Sens.* **2017**, *2017*, 1353691. [[CrossRef](#)]
- Gong, Y.; Duan, B.; Fang, S.; Zhu, R.; Wu, X.; Ma, Y.; Peng, Y. Remote estimation of rapeseed yield with unmanned aerial vehicle (UAV) imaging and spectral mixture analysis. *Plant Methods* **2018**, *14*, 70. [[CrossRef](#)] [[PubMed](#)]
- García-Martínez, H.; Flores-Magdaleno, H.; Ascencio-Hernández, R.; Khalil-Gardezi, A.; Tijerina-Chávez, L.; Mancilla-Villa, O.R.; Vázquez-Peña, M.A. Corn Grain Yield Estimation from Vegetation Indices, Canopy Cover, Plant Density, and a Neural Network Using Multispectral and RGB Images Acquired with Unmanned Aerial Vehicles. *Agriculture* **2020**, *10*, 277. [[CrossRef](#)]
- da Silva, E.E.; Baio, F.H.R.; Teodoro, L.P.R.; da Silva Junior, C.A.; Borges, R.S.; Teodoro, P.E. UAV-multispectral and vegetation indices in soybean grain yield prediction based on in situ observation. *Remote Sens. Appl. Soc. Environ.* **2020**, *18*, 100318. [[CrossRef](#)]
- Wan, L.; Cen, H.; Zhu, J.; Zhang, J.; Zhu, Y.; Sun, D.; Du, X.; Zhai, L.; Weng, H.; Li, Y.; et al. Grain yield prediction of rice using multi-temporal UAV-based RGB and multispectral images and model transfer—A case study of small farmlands in the South of China. *Agric. For. Meteorol.* **2020**, *291*, 108096. [[CrossRef](#)]
- Gitelson, A.A.; Vina, A.; Ciganda, V.; Rundquist, D.C.; Arkebauer, T.J. Remote estimation of canopy chlorophyll content in crops. *Geophys. Res. Lett.* **2005**, *32*. [[CrossRef](#)]
- Vincini, M.; Frazzi, E.; D’Alessio, P. A broad-band leaf chlorophyll vegetation index at the canopy scale. *Precis. Agric.* **2008**, *9*, 303–319. [[CrossRef](#)]
- Jiang, C.-R.; Wang, J.-L. Functional Single Index Models for Longitudinal Data. *Ann. Stat.* **2011**, *39*, 362–388. [[CrossRef](#)]

25. Hunt, E.R.; Daughtry, C.S. What good are unmanned aircraft systems for agricultural remote sensing and precision agriculture? *Int. J. Remote Sens.* **2018**, *39*, 5345–5376. [[CrossRef](#)]
26. Sripada, R.P.; Heiniger, R.W.; White, J.G.; Weisz, R. Aerial Color Infrared Photography for Determining Early In-Season Nitrogen Requirements in Corn. *Agron. J.* **2006**, *98*, 968–977. [[CrossRef](#)]
27. Burgos-Artizzu, X.P.; Ribeiro, A.; Guijarro, M.; Pajares, G. Real-time image processing for crop/weed discrimination in maize fields. *Comput. Electron. Agric.* **2011**, *75*, 337–346. [[CrossRef](#)]
28. Bendig, J.; Yu, K.; Aasen, H.; Bolten, A.; Bennertz, S.; Broscheit, J.; Gnyp, M.L.; Bareth, G. Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *39*, 79–87. [[CrossRef](#)]
29. Wang, C.; Gao, Y.Y.; Pop, I.M.; Vool, U.; Axline, C.; Brecht, T.; Heeres, R.W.; Frunzio, L.; Devoret, M.H.; Catelani, G.; et al. Measurement and control of quasiparticle dynamics in a superconducting qubit. *Nat. Commun.* **2014**, *5*, 5836. [[CrossRef](#)] [[PubMed](#)]
30. Hamuda, E.; Glavin, M.; Jones, E. A survey of image processing techniques for plant extraction and segmentation in the field. *Comput. Electron. Agric.* **2016**, *125*, 184–199. [[CrossRef](#)]
31. Rondeaux, G.; Steven, M.; Baret, F. Optimization of soil-adjusted vegetation indices. *Remote Sens. Environ.* **1996**, *55*, 95–107. [[CrossRef](#)]
32. Black, P.; William, D. Assessment and classroom learning. *Assess. Educ.* **1998**, *5*, 7–74. [[CrossRef](#)]
33. Roujean, J.-L.; Breon, F.-M. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sens. Environ.* **1995**, *51*, 375–384. [[CrossRef](#)]
34. Zhong, L.; Zhang, Y.; Duan, C.A.; Deng, J.; Pan, J.; Xu, N.L. Causal contributions of parietal cortex to perceptual decision-making during stimulus categorization. *Nat. Neurosci.* **2019**, *22*, 963–973. [[CrossRef](#)] [[PubMed](#)]
35. Raper, T.B.; Varco, J.J. Canopy-scale wavelength and vegetative index sensitivities to cotton growth parameters and nitrogen status. *Precis. Agric.* **2015**, *16*, 62–76. [[CrossRef](#)]
36. Hunt, T.E.; Clark-Carter, D.; Sheffield, D. The Development and Part Validation of a U.K. Scale for Mathematics Anxiety. *J. Psychoeduc. Assess.* **2011**, *29*, 455–466. [[CrossRef](#)]
37. Morris, J.S. Functional Regression. *Annu. Rev. Stat. Its Appl.* **2015**, *2*, 321–359. [[CrossRef](#)]
38. Maity, A. Nonparametric functional concurrent regression models. *Adv. Rev.* **2019**, *9*, e1394. [[CrossRef](#)]
39. Leroux, A.; Xiao, L.; Crainiceanu, C.; Checkley, W. Dynamic prediction in functional concurrent regression with an application to child growth. *Stat. Med.* **2018**, *37*, 1376–1388. [[CrossRef](#)]
40. Janet, S.; Kim, J.S.; Arnab Maity, A.; Ana-Maria Staicu, A.-M. Additive nonlinear functional concurrent model. *Stat. Its Interface* **2018**, *11*, 669–685.
41. Bhattacharjee, S.; Müller, H.-G. Concurrent Object Regression. *Electron. J. Stat.* **2022**, *16*, 4031–4089. [[CrossRef](#)]
42. Pan, R.; Wang, Z.; Wu, Y. Detection of Interaction Effects in a Nonparametric Concurrent Regression Model. *Entropy* **2023**, *25*, 1327. [[CrossRef](#)]
43. Ghosal, R.; Maity, A. Variable selection in nonparametric functional concurrent regression. *Can. J. Stat.* **2022**, *50*, 142–161. [[CrossRef](#)]
44. Shi, J.Q.; Wang, B.; Murry-Smith, R.; Titterton, D.M. Gaussian Process Functional Regression Modelling for Batch Data. *Biometrics* **2007**, *63*, 714–723. [[CrossRef](#)] [[PubMed](#)]
45. Shi, J.Q.; Murry-Smith, R.; Titterton, D.M. Hierarchical Gaussian process mixtures for regression. *Stat. Comput.* **2005**, *15*, 31–41. [[CrossRef](#)]
46. Sun, Y.; Fang, X. Robust estimators of functional single index models for longitudinal data. In *Communications in Statistics-Theory and Methods*; Taylor & Francis: Abingdon, UK, 2023; pp. 1–15.
47. Konzon, E.; Cheng, Y.; Shi, J.Q. Gaussian Process for Functional Data Analysis: The GPFDA Package for R. *arXiv* **2021**, arXiv:2102.00249, 1–24.
48. Drucker, H.; Burges, C.J.C.; Kaufman, L.; Smola, A.; Vapnik, V. Support vector regression machines. In *Advances in Neural Information Processing Systems 9 (NIPS'96), Proceedings of the 9th International Conference on Neural Information Processing Systems, Denver CO, USA, 3–5 December 1996*; MIT Press: Cambridge, MA, USA, 1996; pp. 155–161.
49. Borup, D.; Christensen, B.J.; Mühlbach, N.S.; Nielsen, M.S. Targeting predictors in random forest regression. *Int. J. Forecast.* **2023**, *39*, 841–868. [[CrossRef](#)]
50. Freund, Y.; Schapire, R.E. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [[CrossRef](#)]
51. Ferrario, A.; Hämmerli, R. On Boosting: Theory and Applications. ETH Zurich Research Collection 2019; pp. 1–40. Available online: <https://ssrn.com/abstract=3402687> (accessed on 9 May 2024).
52. Iqbal, H.; Sarker, I.H. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Comput. Sci.* **2021**, *2*, 420.

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.