remote sensing

MDPI

Article

Enhanced Crop Leaf Area Index Estimation via Random Forest Regression: Bayesian Optimization and Feature Selection Approach

Jun Zhang [1,†], Jinpeng Cheng [1,2,†], Cuiping Liu [1], Qiang Wu [1,2], Shuping Xiong [1], Hao Yang [3], Shenglong Chang [4], Yuanyuan Fu [5], Mohan Yang [1], Shiyu Zhang [1], Guijun Yang [3,6] and Xinming Ma [1,2,*]

[1] College of Agronomy, Henan Agricultural University, Zhengzhou 450046, China; hau_zj121@stu.henau.edu.cn (J.Z.)
[2] Key Laboratory of Regulating and Controlling Crop Growth and Development, Ministry of Education, Zhengzhou 450046, China
[3] Key Laboratory of Quantitative Remote Sensing in Agriculture of Ministry of Agriculture and Rural Affairs, Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China
[4] College of Software, Henan Normal University, Xinxiang 453000, China
[5] College of Information and Management Science, Henan Agricultural University, Zhengzhou 450002, China
[6] College of Geological Engineering and Geomatics, Chang'an University, Xi'an 710054, China
* Correspondence: maxinming@henau.edu.cn
† These authors contributed equally to this work.

Abstract: The Leaf Area Index (LAI) is a crucial structural parameter linked to the photosynthetic capacity and biomass of crops. While integrating machine learning algorithms with spectral variables has improved LAI estimation over large areas, excessive input parameters can lead to data redundancy and reduced generalizability across different crop species. To address these challenges, we propose a novel framework based on Bayesian-Optimized Random Forest Regression (Bayes-RFR) for enhanced LAI estimation. This framework employs a tree model-based feature selection method to identify critical features, reducing redundancy and improving model interpretability. A Gaussian process serves as a prior model to optimize the hyperparameters of the Random Forest Regression. The field experiments conducted over two years on maize and wheat involved collecting LAI, hyperspectral, multispectral, and RGB data. The results indicate that the tree model-based feature selection outperformed the traditional correlation analysis and Recursive Feature Elimination (RFE). The Bayes-RFR model demonstrated a superior validation accuracy compared to the standard Random Forest Regression and Pso-optimized models, with the R^2 values increasing by 27% for the maize hyperspectral data, 12% for the maize multispectral data, and 47% for the wheat hyperspectral data. These findings suggest that the proposed Bayes-RFR framework significantly enhances the stability and predictive capability of LAI estimation across various crop types, offering valuable insights for precision agriculture and crop monitoring.

Keywords: leaf area index; hyperparameter optimization; Bayesian algorithm; random forests

check for updates

Citation: Zhang, J.; Cheng, J.; Liu, C.; Wu, Q.; Xiong, S.; Yang, H.; Chang, S.; Fu, Y.; Yang, M.; Zhang, S.; et al. Enhanced Crop Leaf Area Index Estimation via Random Forest Regression: Bayesian Optimization and Feature Selection Approach. *Remote Sens.* **2024**, *16*, 3917. https://doi.org/10.3390/rs16213917

Academic Editor: David M Johnson

Received: 5 September 2024
Revised: 9 October 2024
Accepted: 16 October 2024
Published: 22 October 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

The Leaf Area Index (LAI), defined as the total one-sided area of leaf tissue per unit ground surface area [1,2], is a crucial parameter for evaluating crop growth, predicting yields, and selecting superior varieties [3,4]. Traditionally, the LAI is measured using either direct sampling or indirect optical methods. Direct methods involve destructive sampling by harvesting vegetation leaves to obtain LAI measurements [5]. In contrast, indirect optical methods estimate the LAI from canopy gap fraction or transmittance using the Beer–Lambert law [6], employing instruments such as plant canopy analyzers or hemispherical

photography to measure light transmission through the canopy [7]. Although canopy analyzers have been widely adopted in recent years, they still cannot achieve high-throughput measurements, limiting their applicability for large-scale field assessments [8].

Empirical models that estimate the LAI based on vegetation indices derived from remotely sensed data have been extensively studied due to their simplicity and applicability [9]. However, key challenges remain in extracting effective nonlinear features from complex data and optimizing estimation models for accurate LAI assessment. Vegetation indices (VIs), calculated using combinations of visible and near-infrared reflectance, can highlight specific features related to the crop LAI, particularly emphasizing the efficacy of red-edge VIs for LAI estimation [10]. VI-based approaches are characterized by their simplicity and speed, but relying solely on single-feature VIs may not fully capture the complexity of the vegetation LAI, potentially limiting the estimation accuracy.

Incorporating multidimensional data features derived from remotely sensed imagery—such as spectral indices, plant height, and vegetation cover—can better characterize canopy or leaf absorption, thereby improving the accuracy of LAI inversion [11]. Previous studies have shown that integrating multi-feature parameters helps overcome the negative effects of VI saturation, enhancing LAI estimation across the entire crop reproductive period [12]. Nonetheless, the inclusion of multidimensional features can lead to feature redundancy, which must be addressed through feature selection techniques to improve the computational efficiency and reduce overfitting [13,14].

Current feature selection methods primarily include filtering and embedded approaches [15]. Filtering methods rely on calculating feature characteristics and selecting the most representative ones without considering the impact of the selected subset on the algorithm's performance [16]. Correlation analysis is a common example of a filtering method [17,18]. In contrast, embedded methods combine the strengths of filtering and wrapper methods, offering greater precision by analyzing the feature interactions and improving the efficiency [19]. These methods perform variable selection during the model training process and are generally specific to a given learning algorithm, meaning the learning step and feature selection cannot be separated [20]. Feature selection methods based on tree models, such as those used in random forests, are typically considered embedded methods.

Besides feature selection, hyperparameter tuning is a critical task in machine learning [21]. Hyperparameters are key factors that affect the prediction accuracy, computational efficiency, and the fitting capability of machine learning algorithms [22,23]. Unlike model parameters learned from data during training, hyperparameters are set prior to training and cannot be automatically estimated from the data. For Random Forest Regression (RFR) models, hyperparameters include the number of decision trees (n_estimators), the maximum depth of each tree (max_depth), and the minimum number of samples required at the leaf nodes (min_samples_leaf) [24]. Selecting appropriate hyperparameter values is essential, as they significantly influence the model's performance and must be determined before model training.

Several studies have sought to improve the performance of RFR models through hyperparameter optimization, yielding better results. For instance, Zhao et al. [25] employed a Particle Swarm Optimization (Pso)-optimized RFR model to predict winter wheat evapotranspiration, achieving an improved prediction accuracy. Lira et al. [26] optimized RFR using a Genetic Algorithm (GA) for nitrogen oxide emission reduction, while Zhu et al. [27] enhanced RFR models by weighting the Kappa coefficients to improve the prediction accuracy. These examples demonstrate that various algorithms can refine RFR models, enhancing their predictive performance. However, differences in datasets and scenarios necessitate selecting tailored optimization algorithms for each unique case. Compared to optimization methods like Pso and GA, Bayesian Optimization often performs better in global optimization problems by leveraging historical data to iteratively propose new hyperparameter sets that are more likely to enhance the model's performance [28,29].

Therefore, this study aims to improve LAI estimation by integrating feature selection and Bayesian hyperparameter optimization of RFR models. We conducted a two-year field experiment on maize and wheat to acquire LAI, hyperspectral, multispectral, and RGB data. The specific objectives were as follows: (1) To construct an RFR model for LAI estimation using the original features and optimize its hyperparameters using Bayesian optimization, evaluating the estimation accuracy before and after optimization. (2) To identify important features for LAI estimation by comparing filtering methods (correlation analysis) and embedded methods (tree model-based feature selection). (3) To evaluate the performance of the coupled Bayesian-optimized RFR model with selected features for LAI estimation.

2. Materials and Methods

2.1. Study Areas and Experimental Setup

This study was conducted at the National Precision Agriculture Research Center in Xiaotangshan (40.17°N, 116.43°E), Changping District, Beijing, China. Field experiments involving maize and wheat were carried out from 2021 to 2022, encompassing a total of 11 cultivars and 6 planting densities (Figure 1). Experiments 1 and 2 focused on maize cultivation. In Experiment 1, five maize varieties—Zhengdan958, Jingjiu16, Tianci19, Jingnuo2008, and Nongke336—were planted at four different densities: 33,000, 45,000, 67,500, and 90,000 plants per hectare, each with four replications. Experiment 1 comprised a total of 80 plots. Experiment 2 included five other maize varieties—Zhengdan958, XianYu335, Jingnong728, Chengdan30, and Jingpin6—planted at densities of 45,000, 67,500, 90,000, and 105,000 plants per hectare, with three replications, totaling 60 plots. Experiments 3 and 4 involved wheat cultivation. These experiments included two wheat varieties—Jinghua11 and Zhongmai1062—with four nitrogen fertilizer treatments and four replications, resulting in 32 plots per experiment. The plant density for wheat was maintained at 37,440 plants per hectare.

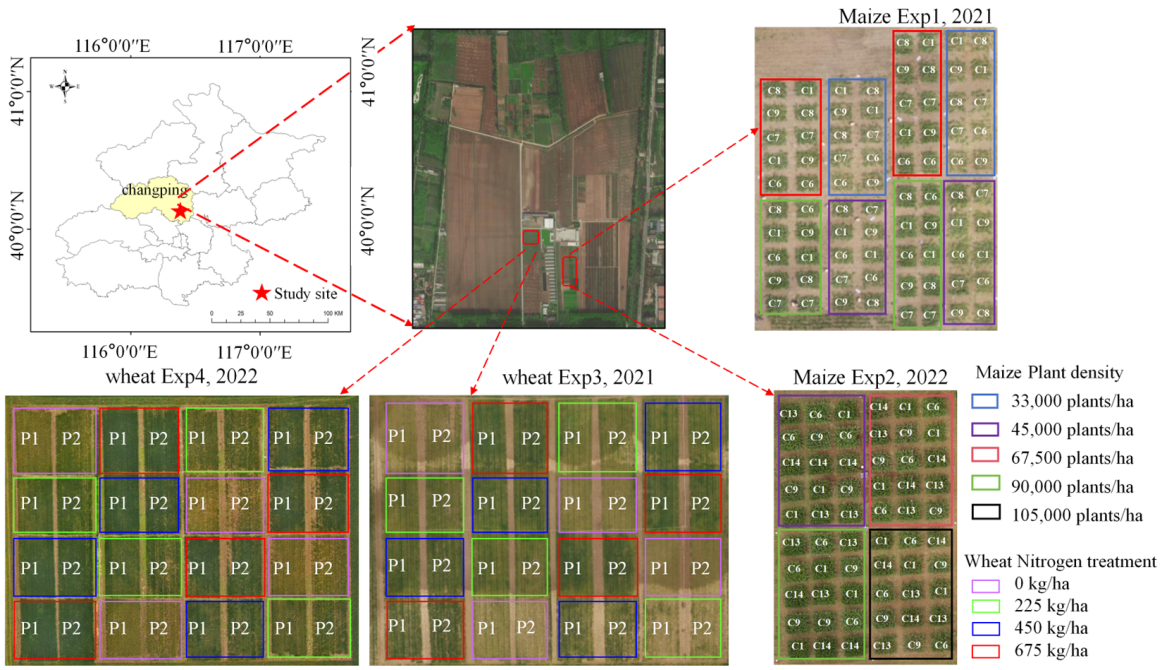


Figure 1. Study area and the distribution of observation plots. Cutivar: C1 (Zhengdan958), C2 (Xianyu335), C3 (Jingnong728), C4 (Chengdan30), C5 (Jingpin6), C6 (ingjiu16), C7 (Tianci19), C8 (Jingnuo2008), C9 (Nongke336), C10 (Jinghua38), C11 (Yufeng506), C12 (Fengnuo), C13 (Jiyuan1), C14 (Jivuan168), P1 (Jinghua11), P2 (Zhongmai1062).

2.2. Data Acquisition

2.2.1. Field Data Collection

The LAI was measured through destructive sampling at key growth stages for both maize and wheat (Table 1). For maize, measurements were taken at the jointing, big flare, and tasseling stages. Wheat measurements were conducted at the jointing, heading, flowering, and filling stages. Detailed sampling information is provided in Table 1. For maize, two to four plants per plot were randomly selected; for wheat, ten plants per plot were chosen. The number of plants and the corresponding ground area for each sampling site were recorded. The leaf area of individual plants was calculated using the aspect factor method, where the sum of the products of leaf length and width was multiplied by a conversion factor of 0.75 [30–32]. The LAI for each plot was then determined based on the number of plants per unit area. The LAI was calculated using the following equation:

$$LAI = 0.75\rho \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} (L_{ij} B_{ij})}{m} \quad (1)$$

where ρ is the number of plants per unit area, m is the number of plants, n is the ith plant, L_{ij} is the leaf length of each leaf, and B_{ij} is the leaf width.

Table 1. Dates, growth stages, and number of data points acquired during the 2021–2022 growing seasons.

No.	Date	Crop Type	Growth Stage	Plots
Experiment 1	13 July	maize	jointing stage	80
	1 August	maize	big flare stage	80
	10 August	maize	tasseling stage	80
	20 August	maize	tasseling stage	79
Experiment 2	21 June	maize	jointing stage	60
	6 July	maize	jointing stage	60
	20 July	maize	big flare stage	60
	6 August	maize	tasseling stage	60
Experiment 3	9 April	wheat	jointing stage	32
	26 April	wheat	heading stage	32
	11, 18 May	wheat	flowering stage	32
	26 May	wheat	filling stage	32
Experiment 4	9 April	wheat	jointing stage	32
	29 April	wheat	heading stage	32
	9 May	wheat	flowering stage	32
	30 May	wheat	filling stage	32

2.2.2. Spectral Data Collection

Multispectral and RGB images were acquired using a DJI PHANTOM 4 Multispectral (P4M) Unmanned Aerial Vehicle (UAV) (SZ DJI Technology Co., Shenzhen, Guangdong, China) equipped with onboard sensors. Flights were conducted between 10:30 and 14:30 under clear, cloudless, and windless conditions. The UAV operated at altitudes ranging from 30 to 50 m, with both horizontal and vertical image overlaps set at 80%, ensuring complete coverage of the experimental plots (Figure 1). The P4M UAV captures images in five spectral bands: blue (450 nm ± 16 nm), green (560 nm ± 16 nm), red (650 nm ± 16 nm), red-edge (730 nm ± 16 nm), and near-infrared (NIR) (840 nm ± 26 nm). Image processing—including stitching, geographic calibration, radiometric calibration, and atmospheric correction—was performed using DJI Terra v2.3.3 software (SZ DJI Technology Co., Shenzhen, Guangdong, China). Terra software integrates user-friendly modules for geographic and radiometric calibration, automatically recognizing calibration target images and applying radiometric calibration using the empirical line method [33,34].

Hyperspectral data were collected using an ASD FieldSpec Pro 2500 canopy spectroradiometer (Analytical Spectral Devices, Boulder, CO, USA), covering a spectral range of 350–2500 nm with a sampling interval of 1 nm. Hyperspectral measurements were taken on the same days as UAV data collection. To minimize measurement errors due to variations in solar angle and temperature, the spectral reflectance of each plot was measured from a height of 1.5 m above the crop canopy, using a 25° field of view, under clear and sunny conditions between 10:00 and 14:00 local time. A $BaSO_4$ panel was used for equipment calibration and baseline reflectance establishment. To enhance data quality and eliminate random noise, the Savitzky–Golay (SG) smoothing algorithm was applied to the raw spectral data [35].

2.2.3. Plant Height and Vegetation Cover Data

Plant height (PH) was extracted from UAV-derived Digital Elevation Models (DEMs) and Digital Orthophoto Maps (DOMs) using ArcGIS v10.8 software (ESRI Inc., Redlands, CA, USA). Accurate acquisition of ground elevation (Digital Terrain Model, DTM) and the upper boundary of the plant canopy is essential for precise PH estimation [36]. The main steps involved were as follows: (1) Processed DEM and DOM files were imported into ArcGIS. (2) Elevation information was assigned to GCPs distributed evenly across the study area. (3) The DEMs were interpolated using the Kriging interpolation method to generate the DTM. (4) The CSM was generated by subtracting the DTM from the DEM. (5) PH was calculated by performing raster calculations on the CSM.

To validate the accuracy of the UAV-derived maize PH, 220 sample measurements across three growth stages were compared with ground truth measurements. The validation results showed a coefficient of determination (R^2) of 0.97 and a relative root mean square error (rRMSE) of 12.17% (Figure 2), indicating high accuracy of the PH extraction method.

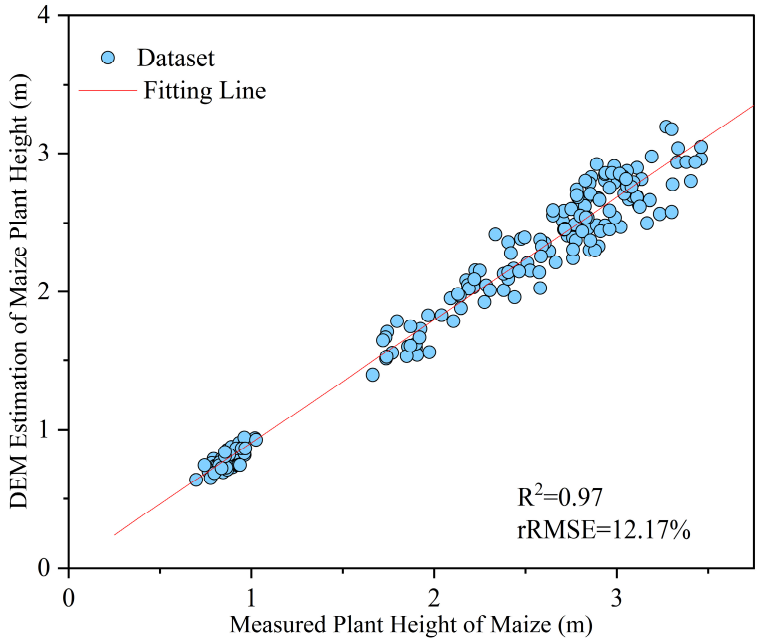


Figure 2. Comparison of UAV-derived maize plant height with ground measurements.

Vegetation fraction (VF), expressing the proportion of vegetative area to total plot area, is a key parameter for estimating crop growth, above-ground biomass (AGB), and yield [37]. The VF was calculated using the following steps: (1) Multispectral images of the crop canopy were extracted and preprocessed. This included applying noise reduction techniques, such as filtering, and ensuring image alignment through registration. (2) Soil background suppression and vegetation element extraction were performed using the Optimized Soil-Adjusted Vegetation Index (OSAVI). Specifically, data from the NIR and red

bands were utilized, and vegetation elements were extracted based on the Otsu thresholding method [38]. (3) The number of vegetation pixels and the total number of pixels within each plot were counted using ArcGIS software. (4) The VF for each plot was calculated using the following equation:

$$VF = \frac{\text{Number of crop pixels in the plot}}{\text{Total number of plot pixels}} \quad (2)$$

2.3. LAI Estimation Method

The LAI estimation was conducted using three datasets: Maize_UAV, Maize_ASD, and Wheat_ASD. The Maize_UAV dataset comprised multispectral vegetation indices, plant height, and vegetation cover extracted from UAV data. The Maize_ASD and Wheat_ASD datasets were derived from canopy spectral reflectance measurements. Initially, data collected in 2021 were used as the training set, while the 2022 data served as the test set. Two feature selection methods were employed: tree model-based feature selection and correlation analysis. For each method, the top five features were selected based on feature importance rankings and absolute correlation coefficients ($|r|$), respectively. A Bayesian-Optimized Random Forest Regression (Bayes-RFR) model was then constructed. A Gaussian process was utilized as the prior model to determine the sampling strategy for hyperparameter optimization. In the training set, the performance of the Bayes-RFR model under different hyperparameter combinations was evaluated using 5-fold cross-validation. The overall workflow is illustrated in Figure 3.

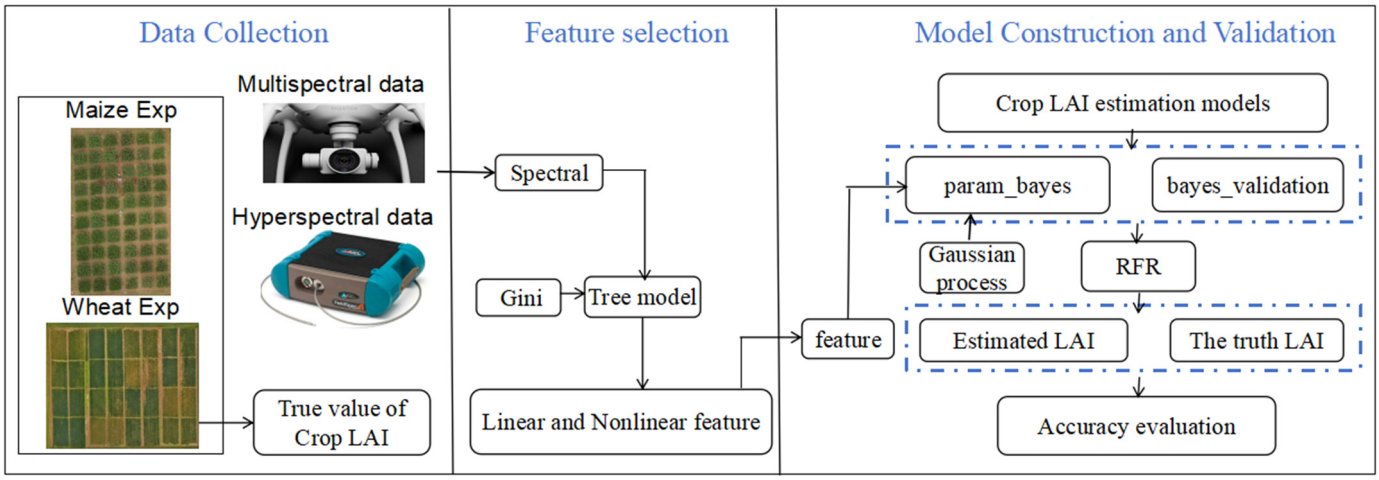


Figure 3. Framework of the LAI estimation process. Note: The dashed box represents the components under this module.

2.3.1. Vegetation Indices

Due to spectral reflectance differences across various crop wavebands, linear or non-linear combinations of reflectance from red (r), green (g), blue (b), red-edge (re), and near-infrared (nir) bands were used to calculate different types of vegetation indices (VIs) [35]. To study LAI estimation using UAV multispectral images and hyperspectral data, a variety of VIs were evaluated based on their stability and applicability, as detailed in Tables 2 and 3. These VIs have been widely used in LAI estimation studies [7].

Tables 2 and 3 present the vegetation indices used in this study along with their calculation formulas. The vegetation indices are categorized into UAV datasets and ASD datasets.

Table 2. Vegetation indices without red-edge bands and their calculation formulas.

Vegetation Indices	Formula	Platform	Reference
Excess Green (EXG)	$2 \times \mathrm{R} - \mathrm{G} - \mathrm{B}$	UAV	[39]
Excess Green Red (EXGR)	$3 \times \mathrm{G} - 2.4 \times \mathrm{R} - \mathrm{B}$	UAV	[39]
Modified Green–Red Vegetation Index (MGRVI)	$\left(\mathrm{G}^2 - \mathrm{R}^2\right) / \left(\mathrm{G}^2 + \mathrm{B} \times \mathrm{R}\right)$	UAV	[40]
Normalized Green–Red Difference Index (NGRDI)	$\left(R_{\mathrm{green}} - R_{\mathrm{red}}\right) / \left(R_{\mathrm{green}} + R_{\mathrm{red}}\right)$	UAV	[41]
Color Index of Vegetation Extraction (CIVE)	$0.441 \times \mathrm{R} - 0.881 \times \mathrm{G} + 0.385 \times \mathrm{B} + 18.78$	UAV	[40]
Red–Green–Blue Vegetation Index (RGBVI)	$\frac{\mathrm{G}^2 - \mathrm{B} \times \mathrm{R}}{\mathrm{G}^2 + \mathrm{B} \times \mathrm{R}}$	UAV	[40]
Green Leaf Index (GLI)	$\frac{(2 \times \mathrm{G} - \mathrm{R} - \mathrm{B})}{(2 \times \mathrm{G} + \mathrm{R} + \mathrm{B})}$	UAV	[40]
Normalized Difference Index (NDI)	$(\mathrm{R} - \mathrm{G}) / (\mathrm{R} + \mathrm{G} + 0.01)$	UAV	[40]
Green–Red Ratio Index (GRRI)	G / R	UAV	[40]
Difference Vegetation Index (GNDVI)	$\left(R_{\mathrm{nir}} - R_{\mathrm{green}}\right) / \left(R_{\mathrm{nir}} + R_{\mathrm{green}}\right)$	UAV	[42]
Modified Triangular Vegetation Index (MTVI1)	$1.2 \times (1.2 \times (R_{800} - R_{550})) - 2.5 \times (R_{670} - R_{550})$	UAV	[42]
Excess Red (EXR)	$1.4 \times R_{\mathrm{red}} - R_{\mathrm{green}}$	UAV	[39]
Modified Chlorophyll Absorption Ratio Index (MCARI)	$[(R_{702} - R_{671}) - 0.2 \times (R_{702} - R_{549})] \times [R_{702} / R_{671}]$	UAV	[42]
Normalized Difference Vegetation Index (NDVI)	$(R_{\mathrm{nir}} - R_{\mathrm{red}}) / (R_{\mathrm{nir}} + R_{\mathrm{red}})$	UAV	[42]
Enhanced Vegetation Index 2 (EVI2)	$2.5 \times (R_{\mathrm{nir}} - R_{\mathrm{red}}) / (R_{\mathrm{nir}} + 2.4 \times R_{\mathrm{red}} + 1)$	UAV	[43]
Optimized Soil-Adjusted Vegetation Index (OSAVI)	$(R_{800} - R_{670}) / (R_{800} + R_{670} + 0.16)$	UAV	[43]
Difference Vegetation Index (DVI)	$R_{802} - R_{658}$	UAV	[44]
Enhanced Vegetation Index (EVI)	$\frac{(2.5 \times (R_{\mathrm{nir}} - R_{\mathrm{red}}))}{(R_{\mathrm{nir}} + 6 \times R_{\mathrm{red}} - 7.5 \times R_{\mathrm{green}} + 1)}$	UAV	[44]
Soil Adjusted-Vegetation Index (SAVI)	$\frac{1.5 \times (R_{872} - R_{661})}{(R_{872} - R_{661}) + 0.5}$	UAV	[42]
Photochemical reflectance index (PRI)	$(R_{450} - R_{560}) / (R_{450} + R_{560})$	ASD/UAV	[42]
Triangular Vegetation Index (TVI)	$60 \times (R_{800} - R_{550}) - 100 \times (R_{670} - R_{550})$	ASD/UAV	[42]
Normalized Difference Greenness index (NDGI)	$\left(R_{\mathrm{green}} - R_{\mathrm{red}}\right) / \left(R_{\mathrm{green}} + R_{\mathrm{red}}\right)$	ASD	[45]
Structure Insensitive Pigment Index (SIPI)	$(R_{800} - R_{445}) / (R_{800} - R_{680})$	ASD	[45]
Normalized Difference Water Index (NDWI)	$(R_{860} - R_{1240}) / (R_{860} + R_{1240})$	ASD	[45]
Normalized Difference Infrared Index (NDII)	$(R_{850} - R_{1650}) / (R_{850} + R_{1650})$	ASD	[45]
Visible Atmospherically Resistant Index (VARI)	$(R_{559} - R_{661}) / (R_{559} + R_{661} - R_{488})$	ASD	[42]
Green Ratio Vegetation Index (GRVI)	R_{872} / R_{559}	ASD	[42]

Note: In the tables, R, G, and B represent single-channel information from UAV visible RGB images. Italicized *R* represents spectral band reflectance, with subscripts green, red, and nir indicating specific spectral channels of the UAV. For ASD data, vegetation indices involving similar bands are calculated at positions corresponding to the center bands of the UAV's multispectral sensors. Subscript numbers represent single-band information acquired by the ASD.

Table 3. Vegetation indices with red-edge bands and their calculation formulas.

Vegetation Indices	Formula	Platform	Reference
Green Chlorophyll Index (CIGREEN)	$\left(R_{\mathrm{re}} / R_{\mathrm{green}}\right) - 1$	UAV	[44]
Red-edge Chlorophyll Index (CIREDEDGE)	$\left(R_{\mathrm{nir}} / R_{\mathrm{re}}\right) - 1$	UAV	[45]
Normalized Difference Red-Edge Index (NDRE)	$(R_{\mathrm{nir}} - R_{\mathrm{re}}) / (R_{\mathrm{nir}} + R_{\mathrm{re}})$	UAV	[44]
Red-Edge Difference Vegetation Index (REDVI)	$R_{\mathrm{nir}} - R_{\mathrm{re}}$	UAV	[44]
Modified Enhanced Vegetation Index (MEVI)	$\frac{(2.5 \times (R_{\mathrm{nir}} - R_{\mathrm{re}}))}{(R_{\mathrm{nir}} + 6 \times R_{\mathrm{re}} - 7.5 \times R_{\mathrm{green}} + 1)}$	UAV	[44]
Soil-Adjusted Red-Edge Index (SARE)	$1.25 \times ((R_{\mathrm{nir}} - R_{\mathrm{re}}) / (R_{\mathrm{nir}} + R_{\mathrm{re}} + 0.25))$	ASD/UAV	[44]
Green Chlorophyll Index (CIG)	$(R_{730} / R_{560}) - 1$	ASD/UAV	[46]
Chlorophyll Index (CIRE)	$(R_{730} / R_{650}) - 1$	ASD/UAV	[46]
Ratio Vegetation Index (RVI)	R_{840} / R_{730}	ASD/UAV	[42]
Leaf Chlorophyll Index (LCI)	$(R_{850} - R_{710}) / (R_{850} + R_{710})$	ASD	[47]

Note: italicized *R* represents spectral band reflectance.

2.3.2. Feature Selection Methods

The tree model-based feature selection calculates the relative importance of feature variables using Out-of-Bag (OOB) errors in a Random Forest Regression (RFR) model. The main idea is to compute the contribution of each feature to the decision trees in the forest, sum these contributions, and rank the features based on their average importance [48]. The Mean Decrease Gini (MDG) index was used to evaluate the contribution of each feature [49,50]. High correlation among features (e.g., between EVI and EVI2) can lead to redundancy and overfitting. Therefore, extracting features with high importance and rich information optimizes the model's performance.

Correlation analysis is a statistical method used to study the relationship between two or more variables. The Pearson correlation coefficient (R) is commonly used to measure the strength and direction of the linear relationship between two variables, ranging from -1 to 1. An R value close to 1 indicates a strong positive correlation, while a value close to -1 indicates a strong negative correlation. An R value near zero suggests a weak or nonlinear relationship. The Pearson correlation coefficient is calculated using the following formula:

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}} \quad (3)$$

Recursive Feature Elimination (RFE) is another machine learning algorithm used for feature selection. It recursively trains the model and eliminates the least important features based on model weights or coefficients [51]. While RFE is straightforward for linear models, it can be computationally intensive, especially with a large number of features, as it requires model retraining at each iteration.

In this study, the tree model-based feature selection was implemented using the sklearn library in Python, with n_estimators set to 100 and random_state set to 0. Feature importance was extracted using the feature_importances_ method and sorted in descending order. Correlation analysis was performed using the Pearson correlation coefficient method, calculated in Microsoft Excel 2019 (v16.0.11727.20104).

2.3.3. Hyperparameter Optimization of the LAI Estimation Model

Random Forest Regression models, based on decision trees, inherently possess high complexity. The goal of hyperparameter optimization in RFR is to reduce model complexity and prevent overfitting. Key hyperparameters in RFR include the number of decision trees (n_estimators), maximum depth of each tree (max_depth), and the minimum number of samples required at leaf nodes (min_samples_leaf) [52,53]. The default hyperparameter settings in the sklearn library are n_estimators = 100, max_depth = None, and min_samples_leaf = 1. The impact of these hyperparameters on model performance is summarized in Table 4.

The optimal set of hyperparameters requires an iterative process of testing various combinations to identify the best-performing model. In this study, Bayesian Optimization and Pso were employed for hyperparameter tuning. Bayesian Optimization is effective for global optimization problems, as it uses historical data to iteratively refine and propose new hyperparameter sets likely to improve model performance [28,29]. A Gaussian process was used as the surrogate model to guide the search for optimal hyperparameters. Pso algorithms simulate the social behavior of particle swarms to find optimal solutions by updating particle positions and velocities based on individual and collective experiences [54,55]. While Pso is commonly used for optimization problems, Bayesian Optimization may offer better efficiency in certain contexts.

Python's bayes_opt and pyswarm libraries were used to implement Bayesian Optimization and Pso, respectively. Optimization targeted the hyperparameters n_estimators, max_depth, and min_samples_leaf. The objective function was defined as the mean test accuracy (MA) from 5-fold cross-validation within the training data [56]. The 2021 data were used as the training set, and the 2022 data served as the test set.

Table 4. Impact of RFR hyperparameters on model performance.

Hyperparameter	Impact on Model Performance	Upgrading Direction	Degree of Impact
n_estimators	Enhances model stability without significantly affecting individual tree complexity	Increase (↑)	****
max_depth	Controls overfitting; decreasing reduces complexity	Decrease (↓)	***
min_samples_leaf	Increasing reduces complexity by requiring more samples per leaf	Increase (↑)	**

Note: The arrows (↑ or ↓) indicate the direction for hyperparameter adjustment to improve model performance. The number of asterisks (*) represent the relative importance level.

2.3.4. Performance Evaluation

The performance of the LAI estimation models was evaluated using the following statistical metrics: coefficient of determination (R^2), root mean square error (RMSE), mean squared error (MSE), and relative root mean square error (rRMSE). Higher R^2 values indicate a stronger relationship between the estimated and measured LAI values, while lower RMSE and rRMSE values indicate higher precision and stability of the model estimations. The formulas for these metrics are as follows:

$$R^2 = 1 - \frac{\sum_i (y_i - x_i)^2}{\sum_i (y_i - \hat{y})^2} \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (x_i - y_i)^2} \quad (5)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (x_i - y_i)^2 \quad (6)$$

$$rRMSE = \frac{RMSE}{\overline{y}} \times 100\% \quad (7)$$

where $\hat{y}$ is the estimated LAI, y_i is the measured LAI, $\overline{y}$ is the mean of the measured LAI values, n is the number of samples.

3. Results

3.1. LAI of Maize and Wheat at Different Growth Stages

Figure 4 presents box plots illustrating the distribution of the LAI values for maize and wheat at various growth stages. For maize, the LAI increased from the jointing stage to the big flare stage and then stabilized. During the jointing stage, maize leaves grow rapidly due to nutrient absorption, and the leaf area stabilizes during the big flare and tasseling stages. In contrast, wheat exhibited an initial increase in the LAI, followed by a decrease after the flowering stage. The rapid leaf growth in early wheat development leads to a peak in the LAI. After flowering, organic matter translocates to the ears, leading to leaf senescence and a subsequent decline in the LAI during the filling stage.

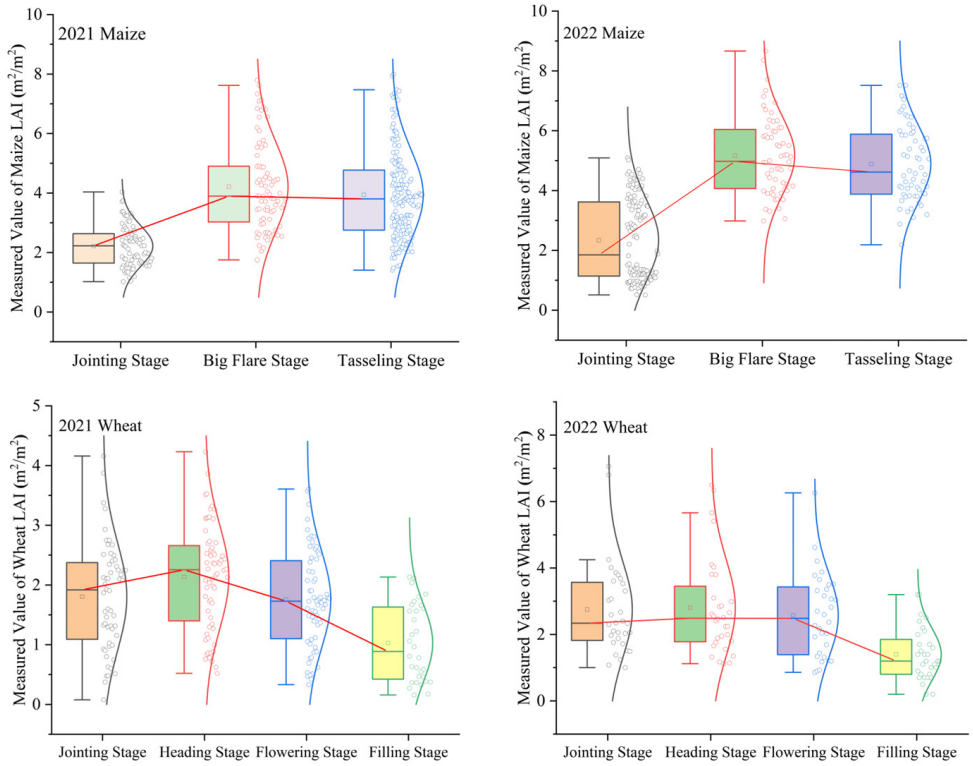


Figure 4. Box plots of measured LAIs for maize and wheat at different growth stages. Note: The different color lines represent the normal distribution of different growth stages.

3.2. Comparison of Optimized LAI Estimation Models

We evaluated the performance of three LAI estimation models—Random Forest Regression (RFR), Particle Swarm Optimization-tuned RFR (Pso-RFR), and Bayesian Optimization-tuned RFR (Bayes-RFR)—using the Maize_ASD, Maize_UAV, and Wheat_ASD datasets. Figure 5 and Table 5 summarize the coefficient of determination (R^2) values and runtime for each model and dataset. The Bayes-RFR model consistently achieved the highest R^2 values across all the datasets. Specifically, for the Maize_ASD dataset, the R^2 values for RFR, Pso-RFR, and Bayes-RFR were 0.28, 0.31, and 0.35, respectively. In the Maize_UAV dataset, the R^2 values were 0.57 for RFR, 0.54 for Pso-RFR, and 0.58 for Bayes-RFR. For the Wheat_ASD dataset, the R^2 values were 0.45 (RFR), 0.45 (Pso-RFR), and 0.53 (Bayes-RFR).

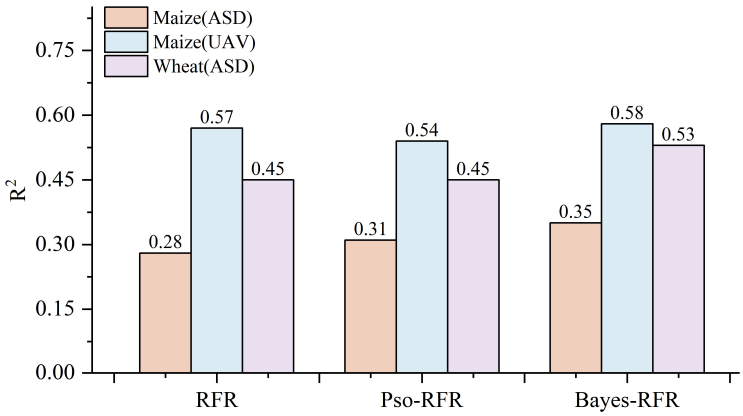


Figure 5. Effect of optimization algorithms on LAI estimation models.

Regarding the runtime, the models utilizing the full feature set generally required more computational time compared to those using selected features. Among the models, RFR had the shortest runtime due to its lack of iterative optimization. Both Bayes-RFR

and Pso-RFR involve iterative processes to identify optimal hyperparameters, leading to longer runtimes. Notably, Bayes-RFR demonstrated shorter runtimes compared to Pso-RFR, indicating a superior computational efficiency during hyperparameter optimization.

Table 5. Runtime for different models and datasets.

Dataset	Model	Runtime with Feature Type	
		Original Features (s)	Selected Features (s)
Maize_ASD	RFR	4.42599 s	4.47022 s
Maize_UAV		3.77637 s	3.65877 s
Wheat_ASD		5.04577 s	3.77398 s
Maize_ASD	Bayes-RFR	187.1543 s	170.3671 s
Maize_UAV		179.2151 s	173.3806 s
Wheat_ASD		189.1506 s	188.6339 s
Maize_ASD	Pso-RFR	271.5143 s	268.6164 s
Maize_UAV		210.4505 s	206.7651 s
Wheat_ASD		243.9103 s	240.5382 s

3.3. Feature Importance Evaluation

We evaluated the feature importance and Pearson correlation coefficients between the feature metrics and LAI using the maize and wheat datasets (Figure 6). In the Maize_ASD dataset, the feature importance method identified NDRE, CIRE, MEVI, MDWI, and VARI as key metrics (Figure 6a,c). In contrast, the correlation analysis highlighted NDRE, CIRE, MEVI, CIG, and NDII as significant features (Figure 6b,d). For the Maize_UAV dataset, the feature importance method highlighted ZG, VF, EXG, EXGR, and EXR, while the correlation analysis identified CIREDEDGE, NDRE, CIGREEN, EVI2, and OSAVI. In the Wheat_ASD dataset, the feature importance analysis revealed CIG, PRI, LCI, SIPI, and NDVI as important features (Figure 6e), whereas the correlation analysis highlighted CIG, MEVI, RVI, CIRE, and PRI (Figure 6f). Notably, the selected features from both methods often included near-infrared bands, such as CIG and PRI. For instance, CIG, derived from the red-edge and near-infrared bands, demonstrates a strong relationship with the LAI. This observation underscores the sensitivity of both the near-infrared and red-edge bands in capturing LAI-related information.

Overall, the features with importance scores greater than 0.1 made significant contributions to LAI estimation. Features with scores between 0.05 and 0.1 provided moderate contributions, while those with scores below 0.05 were less suitable for estimating the LAI. The correlation analysis indicated that features with coefficients between 0.7 and 0.8 were significantly correlated with the LAI, making them effective for estimation. Features with coefficients between 0.5 and 0.6 showed a moderate correlation and were somewhat effective, while those with coefficients between 0.1 and 0.3 were deemed less suitable for LAI estimation.

In this study, the key features were extracted using a tree model-based feature selection method, which effectively eliminated redundant information and fully considered each feature's contribution to the LAI estimation model [57]. By calculating feature importance and accounting for the complex nonlinear relationships between the vegetation indices (VIs) and LAI, our model achieved an improved accuracy. We compared three feature selection methods—correlation analysis, Recursive Feature Elimination (RFE), and tree model-based selection—based on the RFR model. From the estimation of the LAI based on the selected features, the results confirmed previous findings on feature selection (Figure 7), illustrating that machine learning algorithms excel at establishing nonlinear relationships between dependent and independent variables [58]. Furthermore, filtering through a correlation analysis proved to be less effective than machine learning approaches [20,59,60].

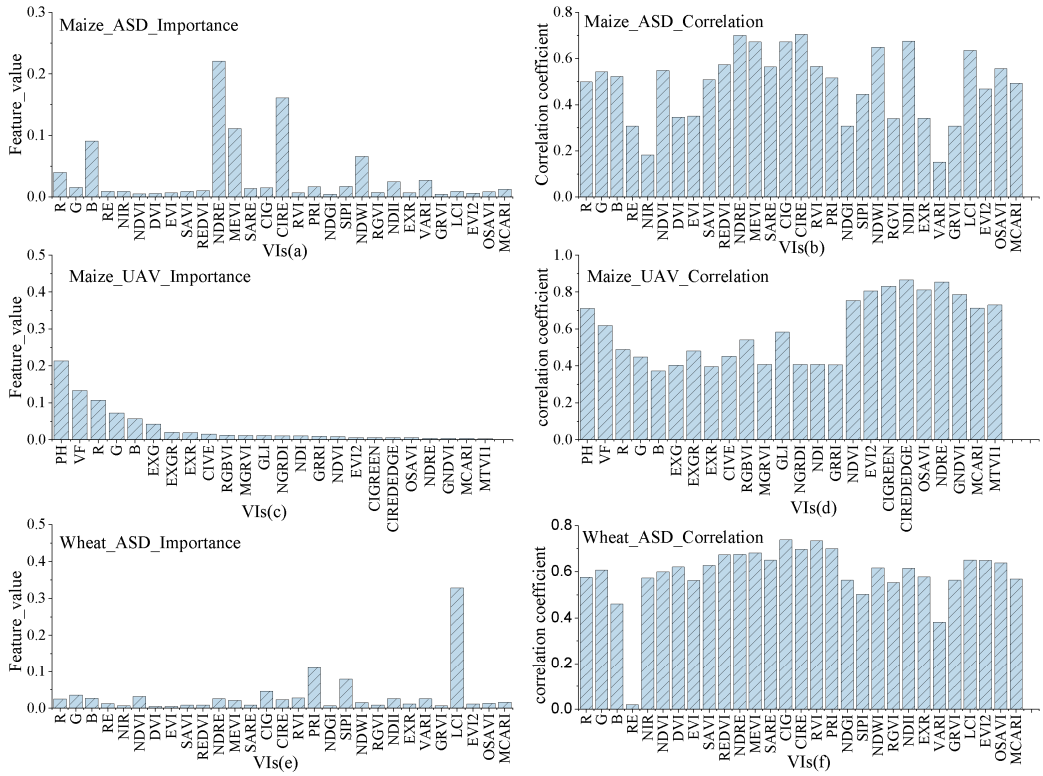


Figure 6. (**a**–**f**) Comparison of feature importance and correlation coefficients across different datasets. Note: (**a**,**c**,**e**) stands for feature importance of Maize_ASD, Maize_UAV, Wheat_ASD; (**b**,**d**,**f**) stands for correlation coefficients of Maize_ASD, Maize_UAV, Wheat_ASD.

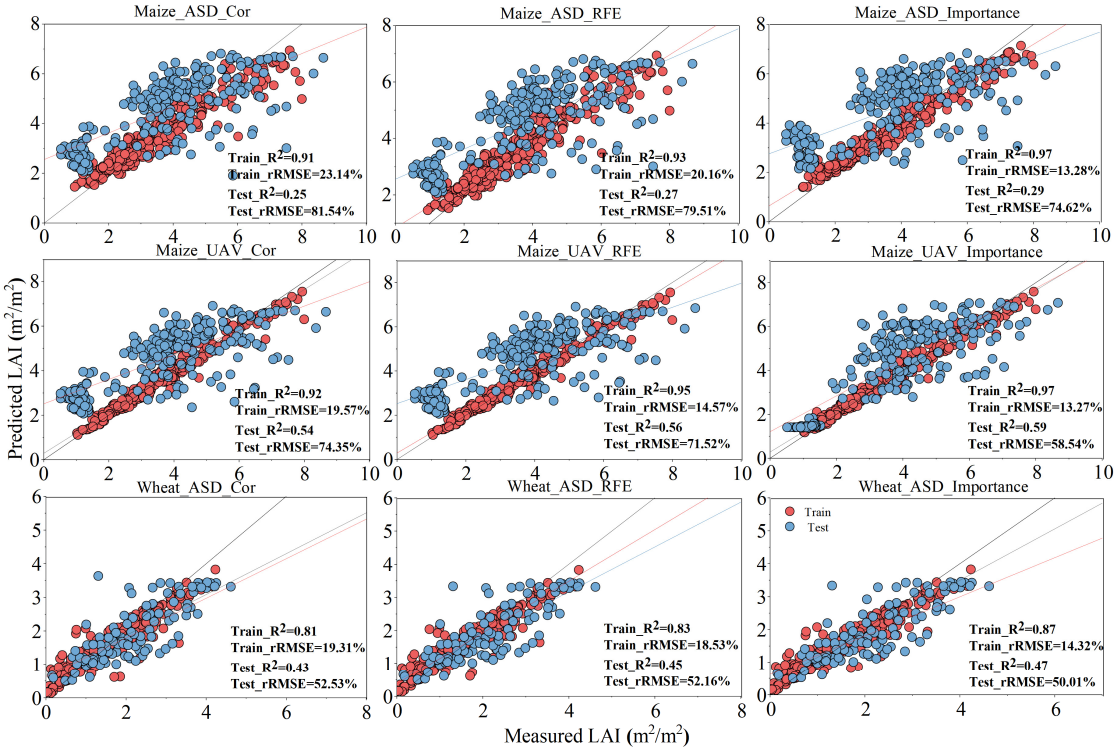


Figure 7. Comparing different feature selection methods in RFR. Note: Cor: correlation analysis, Importance: tree model-based feature.

3.4. LAI Estimation Based on the Bayes-RFR Model

The performance of the Bayes-RFR and Pso-RFR algorithms is influenced by the number of iterations. To achieve optimal model convergence, we selected 100 iterations for both algorithms to search for the best hyperparameter values.

In this study, both the Bayes-RFR and Pso-RFR models were trained using the same number of iterations and datasets (Figure 8), primarily using the mean squared error (MSE) as the evaluation metric to determine the optimal hyperparameters (Table 6). Figure 8 illustrates that both models reached a stable state after 100 iterations across the different datasets. The optimal hyperparameters, determined by minimizing the MSE, are presented in Table 6. Notably, the optimal hyperparameter combinations varied significantly among the models, highlighting the complexities involved in model tuning. We observed that once the number of iterations exceeded 90, the optimization process generally stabilized. For example, when applying the Bayes-RFR model to the Maize_UAV dataset, the optimal hyperparameters were n_estimators = 510, max_depth = 7, and min_samples_leaf = 1, yielding the smallest MSE. Under these optimal hyperparameters, the predicted LAI values closely approximated the actual values, as shown in Figure 9.

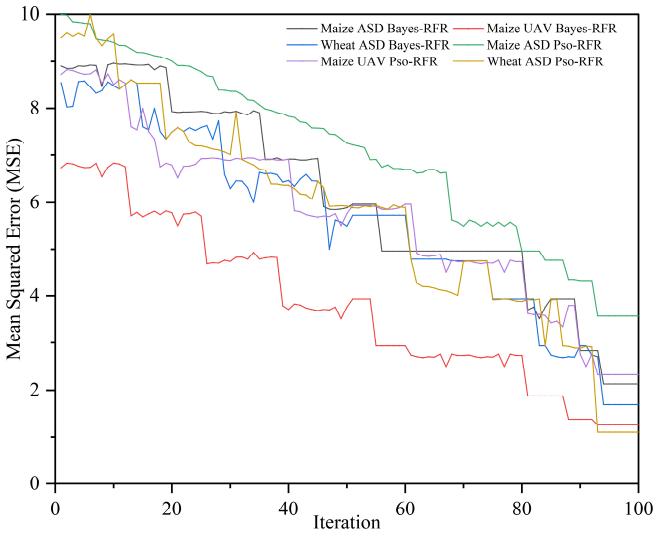


Figure 8. Iteration convergence of the Bayes-RFR and Pso-RFR models across different datasets.

Table 6. Evaluation metrics and optimal hyperparameters for different models and datasets.

Datasets	MSE	Model	Optimal Hyperparameters		
			n_Estimators	max_Depth	min_Samples_Leaf
Maize_ASD	3.875	RFR	140	12	1
Maize_UAV	1.547		503	13	1
Wheat_ASD	1.005		104	13	3
Maize_ASD	2.135	Bayes-RFR	143	11	4
Maize_UAV	1.0696		510	7	1
Wheat_ASD	1.012		106	13	10
Maize_ASD	3.213	Pso-RFR	141	12	1
Maize_UAV	2.346		505	12	1
Wheat_ASD	1.002		103	13	1

Based on the selected features, we used the difference between the predicted and actual LAI values as an evaluation index to assess the fit of the optimized RFR models (represented by the green line in Figure 9). During the parameter optimization process for both the Bayes-RFR and Pso-RFR models, we observed dynamic changes in the iterative optimization as the number of samples varied. Although both models' LAI estimates

closely matched the true values (Figure 9), the Bayes-RFR model demonstrated a superior performance compared to the Pso-RFR model. This efficiency is attributed to the Bayes-RFR model's use of a Gaussian process to represent the prior distribution of the objective function, allowing it to update the estimates based on existing evaluation results.

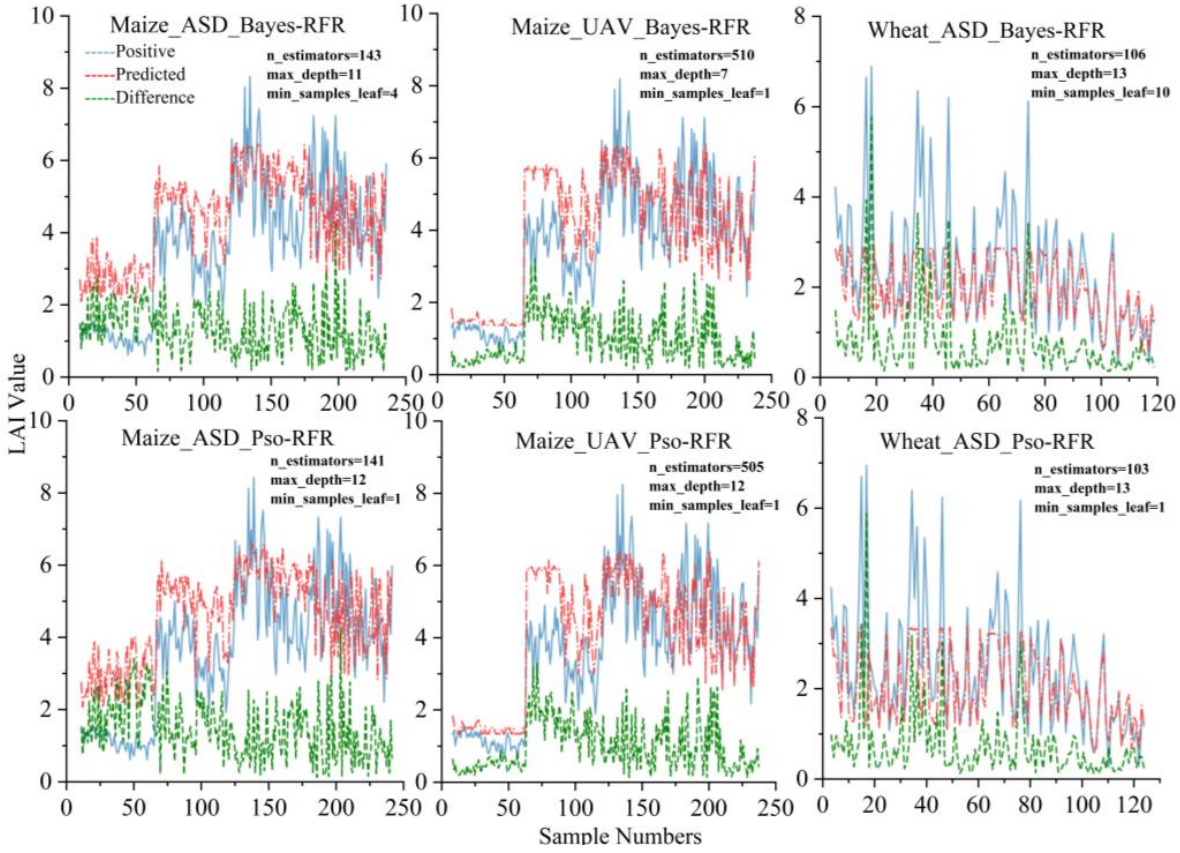


Figure 9. Comparison of actual and predicted LAI values using Bayes-RFR and Pso-RFR models with optimal hyperparameters.

To further compare the Bayes-RFR and standard RFR models, we found that the Bayes-RFR achieved the highest coefficient of determination (R^2) on both the training and test sets across the Maize_ASD, Maize_UAV, and Wheat_ASD datasets (Figures 7 and 10). This suggests that the Bayes-RFR model enhances the stability and accuracy of LAI estimation.

By thoroughly exploring the RFR and Bayes-RFR algorithms, we identified the optimal hyperparameter combinations for each dataset, as shown in Table 6. Compared to the standard RFR, the Bayes-RFR model's hyperparameters were better tuned. Bayesian optimization effectively utilizes historical tuning information, avoiding unnecessary evaluations of sampling points and aiding the algorithm in finding optimal hyperparameter combinations.

As illustrated in Figure 10, under the Maize_ASD dataset, the Bayes-RFR algorithm achieved an R^2 of 0.37, representing a 27% improvement over the RFR algorithm's R^2 of 0.29. For the Maize_UAV dataset, the Bayes-RFR algorithm's R^2 of 0.66 is a 12% improvement compared to the RFR's R^2 of 0.59. In the Wheat_ASD dataset, the Bayes-RFR algorithm attained an R^2 of 0.69, which is 47% higher than the RFR algorithm's R^2 of 0.47. These results indicate that the Bayes-RFR model consistently outperforms the standard RFR model in terms of R^2 values. Although the R^2 of the Bayes-RFR model is lower on the test set compared to the training set, it still surpasses the RFR model, demonstrating a higher predictive accuracy.

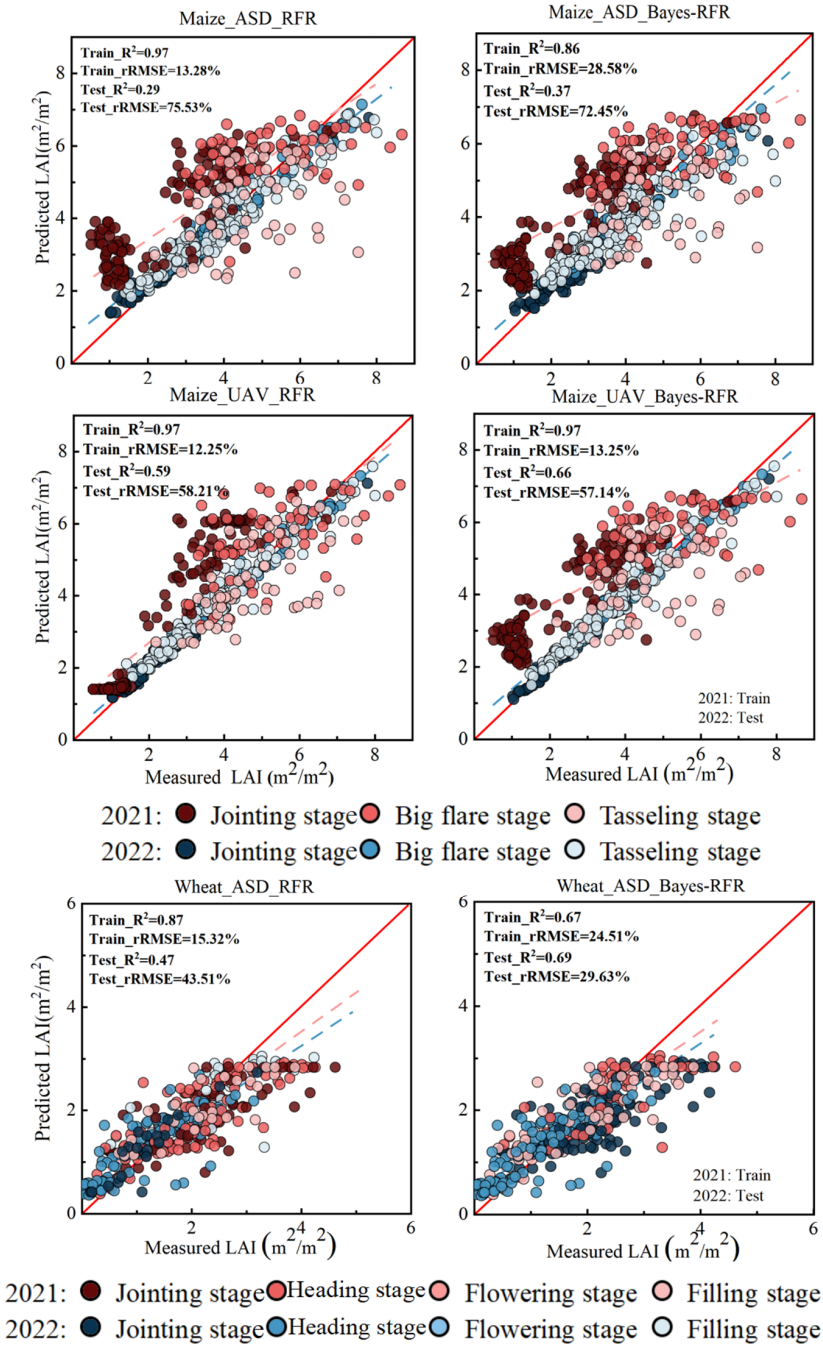


Figure 10. Comparison of actual and predicted LAI values using Bayes-RFR and RFR models with optimal hyperparameters.

The experimental results also revealed a significant difference in accuracy between the Maize_ASD dataset and the other two datasets (Figure 10). The Maize_ASD dataset exhibited much lower accuracies, which can be attributed to variations in crop varieties and planting densities over the two years of the study. It is presumed that the plots with different varieties and planting densities contributed to the lower predictive performance with the Maize_ASD dataset. In contrast, the Wheat_ASD dataset showed much higher accuracies, likely because the same variety and planting density were maintained over the two years. Additionally, in the Maize_UAV dataset, both the plant height and vegetation cover were measured. Under identical coverage conditions, taller plants generally have more leaves, resulting in higher LAI values. A higher vegetation cover typically indicates denser vegetation and larger leaf areas, leading to an increased LAI. Therefore, the inclusion

of structural features in the Maize_UAV dataset contributes to its higher accuracy compared to the Maize_ASD dataset.

4. Discussion

4.1. Influence of Different Features in LAI Estimation

Crop growth exhibits distinct characteristics both horizontally and vertically [61]. Extracting diverse features from different crops provides a comprehensive understanding of their growth status and characteristics. The plant height reflects the crop growth dynamics from a vertical perspective, while visible and multispectral data acquired by UAVs capture spatial distribution information from a horizontal perspective [62].

In this study, we analyzed several factors affecting LAI estimation for maize and wheat, including spectral data, vegetation fraction (VF), plant height, and regression methods. These factors were encapsulated in the Maize_ASD, Maize_UAV, and Wheat_ASD datasets. The Maize_ASD and Wheat_ASD datasets primarily involved vegetation indices extracted from hyperspectral data, whereas the Maize_UAV dataset included multispectral vegetation indices, VF, and plant height. The Maize_UAV dataset improved the LAI prediction accuracy of the Bayes-RFR model compared to the Maize_ASD and Wheat_ASD datasets (Figure 10). This enhancement suggests that multimodal features can reveal crop traits from multiple aspects, enriching the feature dataset and capturing both spectral and structural characteristics [63,64]. The Bayes-RFR model demonstrated a better performance than the standard RFR model across different crop types, indicating its robustness and stability.

Feature-based LAI estimation methods often require extensive feature selection. Traditional statistical models, such as linear regression, struggle to capture the complex relationships between independent variables and LAI. In contrast, machine learning algorithms can explore both linear and nonlinear relationships, determining whether selected features contribute significantly to LAI estimation [59].

The number of selected features (f) is a key parameter influencing the predictive ability of the RFR model. Under the Maize_ASD, Maize_UAV, and Wheat_ASD datasets, we determined that selecting the top five features yielded the best performance (Section 3.3). Since the generalization error of the RFR algorithm cannot be calculated directly, we utilized the Out-of-Bag (OOB) error estimation method proposed by Breiman [65]. By observing changes in the OOB error as f increased, we found that the error decreased initially and then increased after f = 5 (Figure 11). Therefore, selecting five features was optimal for our models.

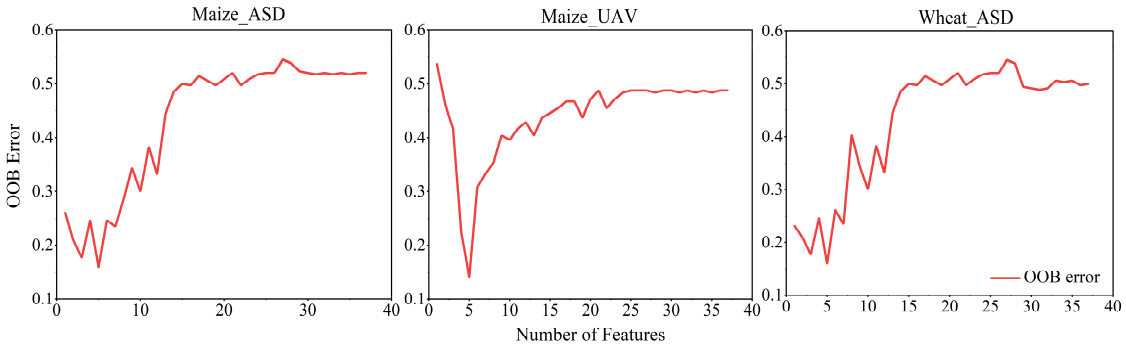


Figure 11. Variation in OOB error with the number of selected features (f) in RFR models.

4.2. Potential of Vegetation Indices in LAI Estimation

We extracted vegetation indices from both hyperspectral and multispectral data to effectively monitor LAI. Spectral data have significant advantages in crop growth monitoring [66–68], as reflectance spectra in the visible range (400–700 nm) are mainly influenced by the chlorophyll content, while the near-infrared range (700–1100 nm) is sensitive to the canopy structure [69]. Vegetation indices combining these spectral ranges can effectively monitor vegetation growth.

Multispectral data derived from UAV imagery have been widely used in crop phenotyping research [8]. Our feature importance ranking experiments for maize (Figure 7) highlight the role of multispectral vegetation indices in assessing phenotypic traits. Multispectral bands, especially those that are near-infrared, provide richer information than RGB bands, improving the relationship between phenotypic traits and vegetation indices [70]. Using a tree-based feature importance analysis, we selected five key vegetation indices, including NDRE and CIRE, which demonstrated strong correlations with the LAI. These indices incorporate critical spectral bands such as red-edge and near-infrared, which are sensitive to the chlorophyll content and canopy structure [71]. Consistent with previous studies, our results confirm that spectral data are closely related to the LAI [8].

4.3. Evaluation of Improved Algorithms for LAI Estimation

We explored the performance of RFR models optimized using Bayesian Optimization (Bayes-RFR) and Particle Swarm Optimization (Pso-RFR) on the ASD and UAV datasets. Figure 5 shows that the Bayes-RFR model outperformed the others in predicting the LAI, achieving the highest accuracy.

Bayesian Optimization employs an active search strategy, using a Gaussian process to model the objective function and guide the search for optimal hyperparameters [28,72]. This approach allows for an efficient exploration of the hyperparameter space, reducing unnecessary evaluations and improving the tuning efficiency. Bayesian Optimization is particularly well suited for problems with low-dimensional hyperparameter spaces due to its fast convergence rate [72]. In contrast, Pso relies on gradient descent methods and may encounter issues such as slow convergence and dimensionality explosion when dealing with high-dimensional parameter spaces [73]. Pso can also struggle with discrete problems and may become trapped in local optima when handling complex functions [54,74]. Therefore, for high-dimensional and complex parameter tuning, the Bayes-RFR model performs better [75]. When using the RFR model for LAI inversion, Bayesian Optimization should be the preferred method for hyperparameter optimization, with Pso serving as a comparative approach. Additionally, to avoid overfitting, outlier detection methods such as the 3σ rule can be employed to reject outliers, enhancing the model stability and prediction accuracy.

Overall, the Bayes-RFR model demonstrated a higher accuracy in LAI estimation (Figure 10). By continuously updating the model based on existing data, Bayes-RFR effectively handles high-dimensional complex function optimization problems, requiring fewer evaluations to determine the optimal solution [72,75]. The Pso-RFR model, being more sensitive to initial conditions, may require more evaluations and is more prone to becoming trapped in local optima [54,74].

To validate the stability of the Bayes-RFR model, we conducted an additional experiment using the 2022 data as the training set and the 2021 data as the test set (Figure 12). The results showed that while the R^2 values decreased and the rRMSE values increased compared to the original training and test sets, the Bayes-RFR model could still predict the 2021 LAI using the 2022 dataset, indicating its robustness.

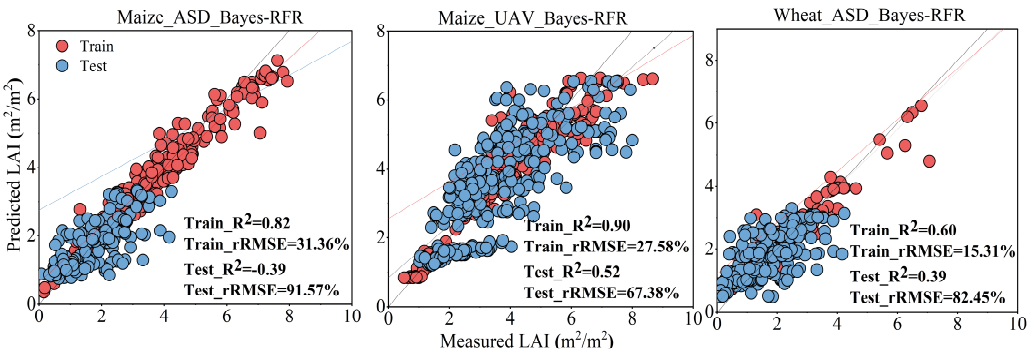


Figure 12. LAI inversion using the Bayes-RFR modeling algorithm.

4.4. Limitations and Future Prospects

This study utilized two years of data from a single area to estimate the LAI, without including cross-area samples as controls. While the methods demonstrated promising results, their generalizability may be limited when applied to different regions due to variations in environmental conditions and the temporal properties of crops [76,77]. Future research should incorporate crop phenotype data from diverse areas to enhance the accuracy and generalizability of LAI estimation using UAV remote sensing.

5. Conclusions

To optimize Random Forest Regression (RFR) hyperparameters and improve the LAI estimation accuracy, we first employed a tree model-based feature importance selection method to identify critical features for LAI estimation. Subsequently, we constructed a Bayesian-Optimized Random Forest Regression model (Bayes-RFR) to enhance the estimation performance.

The key findings are as follows: (1) The tree model-based feature selection method successfully identified the spectral features strongly correlated with the LAI, reducing the number of features and enhancing the model interpretability. (2) The Bayes-RFR model outperformed both the standard RFR and the Pso-optimized RFR models, significantly improving the LAI estimation accuracy. Under the ASD dataset, the Bayes-RFR model improved the LAI estimation accuracy by 47% for wheat and 27% for maize compared to the standard RFR model. In the Maize_UAV dataset, the Bayes-RFR model increased the LAI estimation accuracy by 12%. (3) The Bayes-RFR model demonstrated robustness and stability across different crop types, suggesting its applicability in diverse agricultural settings. These results demonstrate that integrating effective feature selection with Bayesian hyperparameter optimization can substantially enhance the LAI estimation accuracy. The methodologies developed in this study provide technical support for high-throughput extraction of crop phenotypic traits using UAV remote sensing, contributing to precision agriculture practices.

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

Funding: This work was partially supported by the China Postdoctoral Science Foundation (GZC20230721), Key Technology Research and Development Program of Henan Province (242102111192, 232102111115), the Natural Science Foundation of China (32401705, 42371373), the National Key Research and Development Program of China (2021YFD2000102), the Key R&D Project of Henan Province (241111111500), and the Joint Fund of Science and Technology Research Development Program (Application Research) of Henan Province, China (222103810024).

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Jiang, J.; Weiss, M.; Liu, S.; Baret, F. Effective GAI is best estimated from reflectance observations as compared to GAI and LAI: Demonstration for wheat and maize crops based on 3D radiative transfer simulations. *Field Crops Res.* **2022**, *283*, 108538. [CrossRef]

2. Chen, J.M.; Black, T. Defining leaf area index for non-flat leaves. *Plant Cell Environ.* **1992**, *15*, 421–429. [CrossRef]
3. Liu, Y.; An, L.; Wang, N.; Tang, W.; Liu, M.; Liu, G.; Sun, H.; Li, M.; Ma, Y. Leaf area index estimation under wheat powdery mildew stress by integrating UAV-based spectral, textural and structural features. *Comput. Electron. Agric.* **2023**, *213*, 108169. [CrossRef]
4. Fang, H.; Baret, F.; Plummer, S.; Schaepman-Strub, G. An overview of global leaf area index (LAI): Methods, products, validation, and applications. *Rev. Geophys.* **2019**, *57*, 739–799. [CrossRef]
5. Tooley, E.G.; Nippert, J.B.; Ratajczak, Z. Evaluating methods for measuring the leaf area index of encroaching shrubs in grasslands: From leaves to optical methods, 3-D scanning, and airborne observation. *Agric. For. Meteorol.* **2024**, *349*, 109964. [CrossRef]
6. Yin, T.; Cook, B.D.; Morton, D.C. Three-dimensional estimation of deciduous forest canopy structure and leaf area using multi-directional, leaf-on and leaf-off airborne lidar data. *Agric. For. Meteorol.* **2022**, *314*, 108781. [CrossRef]
7. Yan, G.; Hu, R.; Luo, J.; Weiss, M.; Jiang, H.; Mu, X.; Xie, D.; Zhang, W. Review of indirect optical measurements of leaf area index: Recent advances, challenges, and perspectives. *Agric. For. Meteorol.* **2019**, *265*, 390–411. [CrossRef]
8. Qiao, L.; Zhao, R.; Tang, W.; An, L.; Sun, H.; Li, M.; Wang, N.; Liu, Y.; Liu, G. Estimating maize LAI by exploring deep features of vegetation index map from UAV multispectral images. *Field Crops Res.* **2022**, *289*, 108739. [CrossRef]
9. Shao, M.; Nie, C.; Zhang, A.; Shi, L.; Zha, Y.; Xu, H.; Yang, H.; Yu, X.; Bai, Y.; Liu, S. Quantifying effect of maize tassels on LAI estimation based on multispectral imagery and machine learning methods. *Comput. Electron. Agric.* **2023**, *211*, 108029. [CrossRef]
10. Aasen, H.; Honkavaara, E.; Lucieer, A.; Zarco-Tejada, P.J. Quantitative remote sensing at ultra-high resolution with UAV spectroscopy: A review of sensor technology, measurement procedures, and data correction workflows. *Remote Sens.* **2018**, *10*, 1091. [CrossRef]
11. Chen, Z.; Jia, K.; Wei, X.; Liu, Y.; Zhan, Y.; Xia, M.; Yao, Y.; Zhang, X. Improving leaf area index estimation accuracy of wheat by involving leaf chlorophyll content information. *Comput. Electron. Agric.* **2022**, *196*, 106902. [CrossRef]
12. Qiao, L.; Tang, W.; Gao, D.; Zhao, R.; An, L.; Li, M.; Sun, H.; Song, D. UAV-based chlorophyll content estimation by evaluating vegetation index responses under different crop coverages. *Comput. Electron. Agric.* **2022**, *196*, 106775. [CrossRef]
13. Taşkın, G.; Kaya, H.; Bruzzone, L. Feature selection based on high dimensional model representation for hyperspectral images. *IEEE Trans. Image Process.* **2017**, *26*, 2918–2928. [CrossRef]
14. Yuan, W.; Meng, Y.; Li, Y.; Ji, Z.; Kong, Q.; Gao, R.; Su, Z. Research on rice leaf area index estimation based on fusion of texture and spectral information. *Comput. Electron. Agric.* **2023**, *211*, 108016. [CrossRef]
15. Zhou, Q.; Wang, Q.; Gao, Q.; Yang, M.; Gao, X. Unsupervised Discriminative Feature Selection via Contrastive Graph Learning. *IEEE Trans. Image Process.* **2024**, *33*, 972–986. [CrossRef]
16. Guyon, I.; Gunn, S.; Nikravesh, M.; Zadeh, L.A. *Feature Extraction: Foundations and Applications*; Springer: Berlin/Heidelberg, Germany, 2008; Volume 207.
17. Schober, P.; Boer, C.; Schwarte, L.A. Correlation coefficients: Appropriate use and interpretation. *Anesth. Analg.* **2018**, *126*, 1763–1768. [CrossRef]
18. De Winter, J.C.; Gosling, S.D.; Potter, J. Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psychol. Methods* **2016**, *21*, 273. [CrossRef]
19. Chemmakha, M.; Habibi, O.; Lazaar, M. Improving machine learning models for malware detection using embedded feature selection method. *IFAC-PapersOnLine* **2022**, *55*, 771–776. [CrossRef]
20. Rodriguez-Galiano, V.F.; Luque-Espinar, J.A.; Chica-Olmo, M.; Mendes, M.P. Feature selection approaches for predictive modelling of groundwater nitrate pollution: An evaluation of filters, embedded and wrapper methods. *Sci. Total Environ.* **2018**, *624*, 661–672. [CrossRef]
21. Binder, M.; Moosbauer, J.; Thomas, J.; Bischl, B. Multi-objective hyperparameter tuning and feature selection using filter ensembles. In Proceedings of the 2020 Genetic and Evolutionary Computation Conference, New York, NY, USA, 8–12 July 2020; pp. 471–479.
22. Alshraideh, H.; Del Castillo, E.; Del Val, A.G. Process control via random forest classification of profile signals: An application to a tapping process. *J. Manuf. Process.* **2020**, *58*, 736–748. [CrossRef]
23. Phromphithak, S.; Onsree, T.; Tippayawong, N. Machine learning prediction of cellulose-rich materials from biomass pretreatment with ionic liquid solvents. *Bioresour. Technol.* **2021**, *323*, 124642. [CrossRef]
24. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
25. Zhao, X.; Zhang, L.; Zhu, G.; Cheng, C.; He, J.; Traore, S.; Singh, V.P. Exploring interpretable and non-interpretable machine learning models for estimating winter wheat evapotranspiration using particle swarm optimization with limited climatic data. *Comput. Electron. Agric.* **2023**, *212*, 108140. [CrossRef]
26. Lira, J.O.B.; Riella, H.G.; Padoin, N.; Soares, C. Computational fluid dynamics (CFD), artificial neural network (ANN) and genetic algorithm (GA) as a hybrid method for the analysis and optimization of micro-photocatalytic reactors: NOx abatement as a case study. *Chem. Eng. J.* **2022**, *431*, 133771. [CrossRef]
27. Ying, Z.; Rui, X.; Zheng, R. Optimization of random forest algorithm based on nodes matching cost. *Comput. Eng. Des.* **2020**, *41*, 3106–3111.
28. Shin, S.; Lee, Y.; Kim, M.; Park, J.; Lee, S.; Min, K. Deep neural network model with Bayesian hyperparameter optimization for prediction of NOx at transient conditions in a diesel engine. *Eng. Appl. Artif. Intell.* **2020**, *94*, 103761. [CrossRef]

29. Lv, F.; Yu, J.; Zhang, J.; Yu, P.; Tong, D.-W.; Wu, B.-P. A novel stacking-based ensemble learning model for drilling efficiency prediction in earth-rock excavation. *J. Zhejiang Univ.-SCIENCE A* **2022**, *23*, 1027–1046. [CrossRef]
30. Cheng, Q.; Ding, F.; Xu, H.; Guo, S.; Li, Z.; Chen, Z. Quantifying corn LAI using machine learning and UAV multispectral imaging. *Precis. Agric.* **2024**, *25*, 1777–1799. [CrossRef]
31. Liu, F.; Gao, M.; Zhang, H.; Yuan, H.; Hu, B.; Zong, R.; Zhang, M.; Ma, Y.; Li, Q. Synergistic impact of various straw-return methods and irrigation regimes on winter wheat physiological growth and yield. *Field Crops Res.* **2024**, *316*, 109516. [CrossRef]
32. Jiao, F.; Hong, S.; Zhang, Q.; Li, M.; Shi, R.; Ma, Y.; Li, Q. Subsoiling before winter wheat cultivation increases photosynthetic characteristics and leaf water-use efficiency of summer maize in a double-cropping system. *Arch. Agron. Soil Sci.* **2023**, *69*, 847–860. [CrossRef]
33. Daniels, L.; Eeckhout, E.; Wieme, J.; Dejaegher, Y.; Audenaert, K.; Maes, W.H. Identifying the optimal radiometric calibration method for UAV-based multispectral imaging. *Remote Sens.* **2023**, *15*, 2909. [CrossRef]
34. Zhu, H.; Huang, Y.; An, Z.; Zhang, H.; Han, Y.; Zhao, Z.; Li, F.; Zhang, C.; Hou, C. Assessing radiometric calibration methods for multispectral UAV imagery and the influence of illumination, flight altitude and flight time on reflectance, vegetation index and inversion of winter wheat AGB and LAI. *Comput. Electron. Agric.* **2024**, *219*, 108821. [CrossRef]
35. Savitzky, A.; Golay, M.J. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. [CrossRef]
36. Niu, Y.; Han, W.; Zhang, H.; Zhang, L.; Chen, H. Estimating maize plant height using a crop surface model constructed from UAV RGB images. *Biosyst. Eng.* **2024**, *241*, 56–67. [CrossRef]
37. Shu, M.; Li, Q.; Ghafoor, A.; Zhu, J.; Li, B.; Ma, Y. Using the plant height and canopy coverage to estimation maize aboveground biomass with UAV digital images. *Eur. J. Agron.* **2023**, *151*, 126957. [CrossRef]
38. Almeida-Ñauñay, A.F.; Tarquis, A.M.; López-Herrera, J.; Pérez-Martín, E.; Pancorbo, J.L.; Raya-Sereno, M.D.; Quemada, M. Optimization of soil background removal to improve the prediction of wheat traits with UAV imagery. *Comput. Electron. Agric.* **2023**, *205*, 107559. [CrossRef]
39. Xue, J.; Su, B. Significant remote sensing vegetation indices: A review of developments and applications. *J. Sens.* **2017**, *2017*, 1353691. [CrossRef]
40. Liu, S.B.; Jin, X.L.; Bai, Y.; Wu, W.B.; Cui, N.B.; Cheng, M.H.; Liu, Y.D.; Meng, L.; Jia, X.; Nie, C.W.; et al. UAV multispectral images for accurate estimation of the maize LAI considering the effect of soil background. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *121*, 103383. [CrossRef]
41. Imran, A.; Khan, K.; Ali, N.; Ahmad, N.; Ali, A.; Shah, K. Narrow band based and broadband derived vegetation indices using Sentinel-2 Imagery to estimate vegetation biomass. *Glob. J. Environ. Sci. Manag.* **2020**, *6*, 97–108.
42. Li, C.; Li, X.; Meng, X.; Xiao, Z.; Wu, X.; Wang, X.; Ren, L.; Li, Y.; Zhao, C.; Yang, C. Hyperspectral Estimation of Nitrogen Content in Wheat Based on Fractional Difference and Continuous Wavelet Transform. *Agriculture* **2023**, *13*, 1017. [CrossRef]
43. Yue, J.; Yang, G.; Tian, Q.; Feng, H.; Xu, K.; Zhou, C. Estimate of winter-wheat above-ground biomass based on UAV ultrahigh-ground-resolution image textures and vegetation indices. *ISPRS J. Photogramm. Remote Sens.* **2019**, *150*, 226–244. [CrossRef]
44. Li, W.; Li, D.; Liu, S.; Baret, F.; Ma, Z.; He, C.; Warner, T.A.; Guo, C.; Cheng, T.; Zhu, Y. RSARE: A physically-based vegetation index for estimating wheat green LAI to mitigate the impact of leaf chlorophyll content and residue-soil background. *ISPRS J. Photogramm. Remote Sens.* **2023**, *200*, 138–152. [CrossRef]
45. Li, H.; Zhao, C.; Yang, G.; Feng, H. Variations in crop variables within wheat canopies and responses of canopy spectral characteristics and derived vegetation indices to different vertical leaf layers and spikes. *Remote Sens. Environ.* **2015**, *169*, 358–374. [CrossRef]
46. Wen, W.; Timmermans, J.; Chen, Q.; van Bodegom, P.M. A review of remote sensing challenges for food security with respect to salinity and drought threats. *Remote Sens.* **2020**, *13*, 6. [CrossRef]
47. Narmilan, A.; Gonzalez, F.; Salgadoe, A.S.A.; Kumarasiri, U.W.L.M.; Weerasinghe, H.A.S.; Kulasekara, B.R. Predicting canopy chlorophyll content in sugarcane crops using machine learning algorithms and spectral vegetation indices derived from UAV multispectral imagery. *Remote Sens.* **2022**, *14*, 1140. [CrossRef]
48. Raschka, S. *Python Machine Learning*; Packt publishing Ltd.: Birmingham, UK, 2015.
49. Chen, H.; Du, M.; Zhang, Y.; Yang, C. Research on Disease Prediction Method Based on R-Lookahead-LSTM. *Comput. Intell. Neurosci.* **2022**, *2022*, 8431912. [CrossRef]
50. Dunne, R.; Reguant, R.; Ramarao-Milne, P.; Szul, P.; Sng, L.M.; Lundberg, M.; Twine, N.A.; Bauer, D.C. Thresholding Gini variable importance with a single-trained random forest: An empirical Bayes approach. *Comput. Struct. Biotechnol. J.* **2023**, *21*, 4354–4360. [CrossRef]
51. Darst, B.F.; Malecki, K.C.; Engelman, C.D. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. *BMC Genet.* **2018**, *19*, 65. [CrossRef]
52. Chung, H.-T.; Tsai, C.-C.; Jen, K.-K.; Huang, Y.-S.; Ferng, Y.-C.; Lo, C.-Y.; Chen, T.-W.; Chang, K.-H.; Yeh, A.-C. Optimization of process parameters of selective laser melted nickel-based superalloy for densification by random forest regression algorithm and response surface methodology. *Results Eng.* **2024**, *22*, 102182. [CrossRef]
53. Signori-Iamin, G.; Santos, A.F.; Mazega, A.; Corazza, M.L.; Aguado, R.J.; Delgado-Aguilar, M. Bayesian-optimized random forest prediction of key properties of micro-/nanofibrillated cellulose from different woody and non-woody feedstocks. *Ind. Crops Prod.* **2023**, *206*, 117719. [CrossRef]

54. Zeng, J.; Roy, B.; Kumar, D.; Mohammed, A.S.; Armaghani, D.J.; Zhou, J.; Mohamad, E.T. Proposing several hybrid PSO-extreme learning machine techniques to predict TBM performance. *Eng. Comput.* **2022**, *38*, 3811–3827. [CrossRef]
55. Wang, Y.; Wang, K.; Zhang, M.; Gu, T.; Zhang, H. Reliability-enhanced surrogate-assisted particle swarm optimization for feature selection and hyperparameter optimization in landslide displacement prediction. *Complex Intell. Syst.* **2023**, *9*, 5417–5447. [CrossRef]
56. Meckesheimer, M.; Barton, R.R.; Simpson, T.W.; Booker, A.J. Computationally inexpensive metamodel assessment strategies. In Proceedings of the International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Pittsburgh, PA, USA, 9–12 September 2001; pp. 191–201.
57. Tuna, S.; Korkmaz Özay, E.; Tunga, B.; Gürvit, E.; Tunga, M.A. An efficient feature extraction approach for hyperspectral images using Wavelet High Dimensional Model Representation. *Int. J. Remote Sens.* **2022**, *43*, 6899–6920. [CrossRef]
58. Luo, Q.; Li, Z.; Huang, Z.; Abulaiti, Y.; Yang, Q.; Yu, S. Retrieval of Mangrove leaf area index and its response to typhoon based on WorldView-3 image. *Remote Sens. Appl. Soc. Environ.* **2023**, *30*, 100931. [CrossRef]
59. Chen, Z.; Jia, K.; Xiao, C.; Wei, D.; Zhao, X.; Lan, J.; Wei, X.; Yao, Y.; Wang, B.; Sun, Y. Leaf area index estimation algorithm for GF-5 hyperspectral data based on different feature selection and machine learning methods. *Remote Sens.* **2020**, *12*, 2110. [CrossRef]
60. Barbieri, M.C.; Grisci, B.I.; Dorn, M. Analysis and comparison of feature selection methods towards performance and stability. *Expert Syst. Appl.* **2024**, *249*, 123667. [CrossRef]
61. Dzvene, A.R.; Tesfuhuney, W.A.; Walker, S.; Ceronio, G. Planting time and stand density effect on radiation interception and use efficiency of maize and sunn hemp intercropping in semi-arid South Africa. *Agric. For. Meteorol.* **2023**, *341*, 109690. [CrossRef]
62. Tian, J.-Y.; Wang, B.; Zhang, Z.-M.; Lin, L.-X. Application of spectral diversity in plant diversity monitoring and assessment. *Chin. J. Plant Ecol.* **2022**, *46*, 1129. [CrossRef]
63. Adeluyi, O.; Harris, A.; Foster, T.; Clay, G.D. Exploiting centimetre resolution of drone-mounted sensors for estimating mid-late season above ground biomass in rice. *Eur. J. Agron.* **2022**, *132*, 126411. [CrossRef]
64. Maimaitijiang, M.; Sagan, V.; Sidike, P.; Hartling, S.; Esposito, F.; Fritschi, F.B. Soybean yield prediction from UAV using multimodal data fusion and deep learning. *Remote Sens. Environ.* **2020**, *237*, 111599. [CrossRef]
65. Smith, H.; Biggs, P.J.; French, N.P.; Smith, A.; Marshall, J. Lost in the forest. *BioRxiv* **2022**. [CrossRef]
66. Stanton, C.; Starek, M.J.; Elliott, N.; Brewer, M.; Maeda, M.M.; Chu, T. Unmanned aircraft system-derived crop height and normalized difference vegetation index metrics for sorghum yield and aphid stress assessment. *J. Appl. Remote Sens.* **2017**, *11*, 026035. [CrossRef]
67. Tilly, N.; Aasen, H.; Bareth, G. Fusion of plant height and vegetation indices for the estimation of barley biomass. *Remote Sens.* **2015**, *7*, 11449–11480. [CrossRef]
68. Weiss, M.; Jacob, F.; Duveiller, G. Remote sensing for agricultural applications: A meta-review. *Remote Sens. Environ.* **2020**, *236*, 111402. [CrossRef]
69. Li, X.; Zhang, Y.; Bao, Y.; Luo, J.; Jin, X.; Xu, X.; Song, X.; Yang, G. Exploring the best hyperspectral features for LAI estimation using partial least squares regression. *Remote Sens.* **2014**, *6*, 6221–6241. [CrossRef]
70. Xu, D.; An, D.; Guo, X. The impact of non-photosynthetic vegetation on LAI estimation by NDVI in mixed grassland. *Remote Sens.* **2020**, *12*, 1979. [CrossRef]
71. Li, D.; Chen, J.M.; Yu, W.; Zheng, H.; Yao, X.; Zhu, Y.; Cao, W.; Cheng, T. A chlorophyll-constrained semi-empirical model for estimating leaf area index using a red-edge vegetation index. *Comput. Electron. Agric.* **2024**, *220*, 108891. [CrossRef]
72. Frazier, P.I. A tutorial on Bayesian optimization. *arXiv* **2018**, arXiv:1807.02811.
73. Wang, D.; Tan, D.; Liu, L. Particle swarm optimization algorithm: An overview. *Soft Comput.* **2018**, *22*, 387–408. [CrossRef]
74. Byeon, H. Comparing the accuracy and developed models for predicting the confrontation naming of the elderly in South Korea using weighted random forest, random forest, and support vector regression. *Int. J. Adv. Comput. Sci. Appl.* **2021**, *12*, 326–331. [CrossRef]
75. Pranckevičius, T.; Marcinkevičius, V. Comparison of naive bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification. *Balt. J. Mod. Comput.* **2017**, *5*, 221. [CrossRef]
76. Wang, Y.; Feng, L.; Zhang, Z.; Tian, F. An unsupervised domain adaptation deep learning method for spatial and temporal transferable crop type mapping using Sentinel-2 imagery. *ISPRS J. Photogramm. Remote Sens.* **2023**, *199*, 102–117. [CrossRef]
77. Mohammadi, S.; Belgiu, M.; Stein, A. A source-free unsupervised domain adaptation method for cross-regional and cross-time crop mapping from satellite image time series. *Remote Sens. Environ.* **2024**, *314*, 114385. [CrossRef]

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