


Article

Model for Inverting the Leaf Area Index of Green Plums by Integrating IoT Environmental Monitoring Data and Leaf Relative Content of Chlorophyll Values

Caili Yu ¹, Haiyang Tong ², Daoyi Huang ¹, Jianqiang Lu ^{2,3,4,5,*} , Jiewei Huang ², Dejing Zhou ² and Jiaqi Zheng ²

¹ Center for Intelligent Perception and Internet of Things Research, Shanwei Institute of Technology, Shanwei 516600, China; scauaiteam@gmail.com (C.Y.); huangdaoyi334@gmail.com (D.H.)

² College of Electronic Engineering (College of Artificial Intelligence), South China Agricultural University, Guangzhou 510642, China; thy.scau.edu.cn@stu.scau.edu.cn (H.T.); hjw847197516@stu.scau.edu.cn (J.H.); zhoudejing@stu.scau.edu.cn (D.Z.); 202334610130@stu.scau.edu.cn (J.Z.)

³ Heyuan Sub-Center, Guangdong Provincial Laboratory of Lingnan Modern Agriculture Science and Technology, Heyuan 517000, China

⁴ National Center for International Research on Precision Agricultural Aviation Pesticide Spraying Technology, Guangzhou 510642, China

⁵ Guangdong Provincial Key Laboratory of Agricultural Artificial Intelligence, Guangzhou 510642, China

* Correspondence: lj@scau.edu.cn

Abstract: The quantitative inversion of the leaf area index (LAI) of green plum trees is crucial for orchard field management and yield prediction. The data on the relative content of chlorophyll (SPAD) in leaves and environmental data from orchards show a significant correlation with LAI. Effectively integrating these two data types for LAI inversion is important to explore. This study proposes a multi-source decision fusion LAI inversion model for green plums based on their adjusted determination coefficient (MDF–ADRS). First, three statistical methods—Pearson, Spearman rank, and Kendall rank correlation analyses—were used to measure the linear relationships between variables, and the six environmental factors most highly correlated with LAI were selected from the orchard’s environmental data. Then, using multivariate statistical analysis methods, LAI inversion models based on environmental feature factors (EFs–PM) and SPAD (SPAD–PM) were established. Finally, a weight optimization allocation strategy was employed to achieve a multi-source decision fusion LAI inversion model for green plums. This strategy adaptively allocates weights based on the predictive performance of each data source. Unlike traditional models that rely on fixed weights or a single data source, this approach allows the model to increase the influence of a key data source when its predictive strength is high and reduce noise interference when it is weaker. This dynamic adjustment not only enhances the model’s robustness under varying environmental conditions but also effectively mitigates potential biases when a particular data source becomes temporarily unreliable. Our experimental results show that the MDF–ADRS model achieves an R^2 of 0.88 and an $RMSE$ of 0.39 in the validation set, outperforming other fusion methods. Compared to the EFs–PM and SPAD–PM models, the R^2 increased by 0.19 and 0.26, respectively, and the $RMSE$ decreased by 0.16 and 0.22. This model effectively integrates multiple sources of data from green plum orchards, enabling rapid inversion and improving the accuracy of green plum LAI estimation, providing a technical reference for monitoring the growth and managing the production of green plums.

Keywords: green plum; LAI; SPAD; environmental feature factors; MDF–ADRS



Citation: Yu, C.; Tong, H.; Huang, D.; Lu, J.; Huang, J.; Zhou, D.; Zheng, J. Model for Inverting the Leaf Area Index of Green Plums by Integrating IoT Environmental Monitoring Data and Leaf Relative Content of Chlorophyll Values. *Agriculture* **2024**, *14*, 2076. <https://doi.org/10.3390/agriculture14112076>

Academic Editor: Georgios C. Koubouris

Received: 8 October 2024

Revised: 2 November 2024

Accepted: 11 November 2024

Published: 18 November 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

Green plums, often called sour plums, are a popular stone fruit of the Rosaceae family and have been cultivated in China for over 3000 years [1,2]. Packed with a myriad of vitamins, minerals, and amino acids, green plums enhance digestive health, boost appetite, mitigate fatigue, guard the liver, and delay aging processes [3,4]. Luhe County in Shanwei,

Guangdong Province, known as China's "Green Plum Capital", is a major producer of these economically vital plums [5]. In this region, green plums are primarily used in the creation of plum juice, dried plums, plum wines, and various preserves, alongside their medicinal applications. Furthermore, they are crafted into seasoned sauces that enrich the flavor profiles of numerous dishes. Due to their rich nutritional value and distinct flavor profile, green plums play an integral role in nutrition and health practices. As of 2023, the annual yield of green plums in Luhe County was consistently around 25,000 tons, with peaks of up to 35,000 tons in abundant harvest years. The favorable climatic and soil conditions in Guangdong Province create a strategic advantage for escalating green plum production, prompting local farmers and investors to enhance resource efficiency and expand cultivation areas extensively.

Research has shown that the leaf area index (LAI) is fundamentally connected to a crop's capacity to capture photosynthetically active radiation, thus playing a critical role in photosynthesis, respiration, and transpiration processes. It is also an excellent gauge of growth phases and various abiotic and biotic stresses [6–8]. Son [9] and his team have formulated a method for predicting rice yields using MODIS EVI and LAI data, where they ascertain that incorporating these indices into a quadratic model substantially improves yield simulations within their study's regional focus. Zhuo [10] has shown that integrating LAI into the World Food Studies (WOFOST) model significantly boosts the model's ability to predict regional wheat yields accurately. Additionally, Yamanura [11] has proven that an LAI model derived from NDVI is highly effective for crop vegetation monitoring, stress assessments, pest and disease scouting, and predicting the area and yield of crops, all with notable accuracy. Through the consistent monitoring of the LAI, farmers can tailor their fertilization, irrigation, and pest control approaches to maintain a suitable canopy density in green plum trees, which in turn enhances the health and yield of their fruit [12]. However, the accuracy of LAI measurements can be compromised by weather fluctuations and operator variability. These measurements are dependent on instruments that are sensitive to environmental conditions, where inconsistent sunlight can cause LAI values to vary and strong winds can impact the stability of measurements [13,14]. Variations in operational methods can also lead to deviations in LAI readings, which is particularly problematic in extended studies. Furthermore, LAI measurements are typically conducted point-by-point, meaning they require substantial time and operator experience [12,15,16]. Accurate leaf area index (LAI) data are crucial in agriculture as they help farmers optimize decisions regarding irrigation, fertilization, and pest control. Without these data, farmers may risk overusing resources or missing critical intervention opportunities, which could impact crop yields. In the current cultivation management of plum orchards, inaccurate LAI measurements may lead to misjudgments about the health of the trees, causing farmers to either over-apply or under-apply water and nutrients, directly affecting the yield and quality of their plums. Additionally, delays or improper responses to pest outbreaks could exacerbate damage, further reducing productivity. Given the high economic value of plums in Guangdong, fluctuations in their yield or quality could significantly impact local market prices and profitability. Therefore, there is an urgent need for more reliable and effective methods of measuring the LAI.

Chlorophyll assessments are notably more stable and efficient than measurements of the LAI, as they exhibit minimal variability under different environmental conditions, ensuring consistent results irrespective of weather variations. This technique requires less expertise from operators and is minimally affected by human factors, rendering it suitable for fast and widespread monitoring applications [17–19]. According to research by Dordas [20], photosynthesis plays a vital role in plant growth, including the enhancement of the LAI, and is directly influenced by the chlorophyll concentration in leaves. This relationship underscores that leaf chlorophyll not only depicts vegetative growth but also correlates with the LAI. Presently, IoT technology enables the instantaneous gathering of environmental and crop habitat data in agricultural settings [21–23]. Kinane [24] identified that maximum monthly temperatures and surplus moisture markedly influence the LAI of

managed torch pine forests. Adjustments incorporating environmental factors and local carrying capacities led to a decrease in the root mean square error of 0.0625. Additionally, Karimi [25] observed that remote sensing data for LAI estimation are restricted to clear-sky conditions. Estimating the LAI in forests using traditional multispectral remote sensing also poses significant challenges. Due to these limitations in remote sensing for LAI estimations, Karimi [25] used Gene Expression Programming (GEP) to derive the LAI from weather data.

Accurate LAI monitoring is crucial in plum cultivation management, as it not only supports management decisions throughout the stages of crop growth but also directly impacts yield predictions and resource optimization. However, due to the complex canopy structure of plum trees and their susceptibility to external environmental disturbances, precise LAI measurement remains a significant challenge. The evolution of remote sensing and machine learning technologies has opened new pathways for the monitoring of plant crop growth and yield forecasting [19,26]. Zhang [27] analyzed the relationship between visible and hyperspectral imaging features with winter wheat's LAI, utilizing these correlations for feature selection. They developed multiple linear regression, support vector regression, and random forest regression models using various imaging feature combinations to estimate the LAI and compared the efficacy of single-source versus multi-source data models. Deng [28] employed threshold segmentation to identify the top five vegetation indices correlating with SPAD, LAI, and the canopy chlorophyll content, creating inversion models using PLSR that enhanced the accuracy of physicochemical parameter inversion in winter wheat by mitigating background spectral interference. Chen [29] demonstrated that "Specific spectral transformations", when combined with optimal narrow-band spectral indices and random forest regression, could accurately evaluate small-area LAIs, achieving an R^2 of 0.77 and an $RMSE$ of 0.27. Concurrently, Shao [30] improved the precision of LAI estimation by using deep learning to segment corn ears, significantly enhancing this method's accuracy by removing corn ear images, with results showing an R^2 of 0.816 and an $RMSE$ of 0.399. The aforementioned research primarily focuses on using spectral data for the LAI inversion of crops, emphasizing the internal characteristics of crops while overlooking a comprehensive consideration of their external growth environment. Methodologically, many studies primarily rely on single-source spectral data to construct LAI inversion models, leveraging their high-resolution advantages, with conclusions often focused on the effectiveness of spectral data. However, such methods frequently overlook the impact of environmental factors on the LAI, leading to insufficient model stability when they are confronted with dynamic environmental conditions. Furthermore, the unique growth traits and ecological context of green plums demand specialized monitoring techniques, as their intricate canopy structure and background noise readily lead to remote sensing misinterpretations, with environmental variations in light, soil, and moisture further impacting the reliability of remote sensing signals [13,14,31]. In conclusion, many studies that rely on spectral data assert that spectral characteristics are the primary factors influencing LAI, with limited attention given to how dynamic environmental changes impact the model's long-term effectiveness. Reliance on a single data source in complex and variable conditions may lead to instability in estimation results.

Given the complexities of green plum cultivation and the findings of Dordas and Kinane, which emphasize the crucial impact of environmental factors and relative content of chlorophyll (SPAD) values on the LAI of crops, additional research by Son, Zhuo, and Yamanura highlights the important role of the LAI in pest and disease management and yield prediction for crops. This study employs a multi-source data fusion strategy to combine the internal characteristics of plum trees with external environmental features, introducing a weight optimization method to adjust the importance of each data source. This approach effectively enhances the model's robustness and generalization under variable environmental conditions. Environmental data collected from IoT devices in orchards, including temperature, humidity, and soil temperature and moisture, were used as input variables to develop an LAI inversion model based on environmental features (EFs-PM) using

multivariate statistical techniques. Simultaneously, an LAI inversion model (SPAD–PM) was developed based on collected relative content of chlorophyll (SPAD) data. We then constructed the MDF–ADRS fusion model, which integrates the relative content of chlorophyll (SPAD) and environmental data and dynamically adjusts the importance of each data source through a weight optimization strategy. By optimizing weights to adapt to environmental changes, the MDF–ADRS model improves the robustness and accuracy of LAI inversion, addressing gaps in multi–source data integration and providing a technical reference for monitoring environmentally dependent crops such as plum trees.

Considering the complex environmental conditions influencing green plum cultivation and the crucial role these factors play in crop development, this study uses the relative content of chlorophyll (SPAD) and environmental characteristics as input parameters. By applying multivariate statistical analyses, LAI inversion models based on these environmental inputs and SPAD were developed, and their results were integrated to address the limitations of relying on single–source data in agricultural monitoring. Consequently, this approach introduces a novel, efficient method for green plum LAI estimation that significantly improves field management practices and yield prediction accuracy. The key contributions of this research are as follows:

- (1) This paper aims to identify the key environmental factors affecting the leaf area index (LAI) of green plums, considering various complex conditions. We used statistical methods to evaluate the linear relationships between twelve environmental factors and selected the six with the highest correlation to the green plum LAI for further study.
- (2) Additionally, using these six factors and SPAD values, we constructed two LAI inversion models: one based on environmental factors (EFs–PM) and the other on SPAD (SPAD–PM).
- (3) This study introduces a multi–source decision fusion model based on the adjusted R^2 , utilizing an optimized weight allocation strategy to integrate multi–source data from green plum orchards. Compared to Chen [29], who used spectral transformations with a random forest regression to assess LAIs in small areas, our model improved the R^2 by 0.11. Compared to Shao [30], who quantified LAI estimations using deep learning segmentation methods, our model improved the R^2 by 0.064 and reduced the $RMSE$ by 0.009. The MDF–ADRS fusion model surpasses classic models in both robustness and accuracy, enabling faster and more precise green plum LAI estimation.

2. Materials and Methods

2.1. Study Area and Study Design

This study was conducted in green plum orchards located in Dongkeng Town, Luhe County, Shanwei City, Guangdong Province, China, geographically positioned at 114.03° E and 22.75° N. The field site, situated approximately 240 m above sea level and shown in Figure 1, benefits from a subtropical monsoon climate. The region has an average annual temperature of approximately 21.7 °C and an average annual precipitation of around 1500 mm. The soil in the test area is predominantly loam, which provides good water retention and supplies nutrients to the green plum trees. The terrain consists mainly of gently rolling hills, allowing for good drainage. These conditions create an ideal environment for cultivating subtropical fruits like green plum. The region is primarily focused on agriculture, with a strong emphasis on fruit tree cultivation. Plums, in particular, are highly valued for their historical significance and economic importance. The main variety grown in this area is the Luhe plum, a high–yielding variety prized for its dense canopy and vigorous growth.

We used weather stations and soil temperature and humidity sensors as IoT devices to monitor the environmental characteristics of the green plum orchard. These devices continuously collected real–time environmental data. By integrating these IoT devices into our system, we were able to ensure reliable and continuous data collection, reducing manual errors and inconsistencies. The automation of the monitoring process significantly

enhanced the robustness of the collected data, providing a solid foundation for developing a high-precision LAI inversion model.

Our sensor placement was based on the locations of the selected fruit trees and irrigation variations. If two sample trees share a common area with consistent soil conditions, one sensor is sufficient, as depicted in Figure 2. Sensors are buried as instructed to ensure proper contact and data accuracy. All IoT devices were precalibrated according to the manufacturer's specifications before the experiment, and regular maintenance and recalibration were performed during the experiment in accordance with ISO 17123–8 [32] standards for agricultural monitoring and environmental data collection.

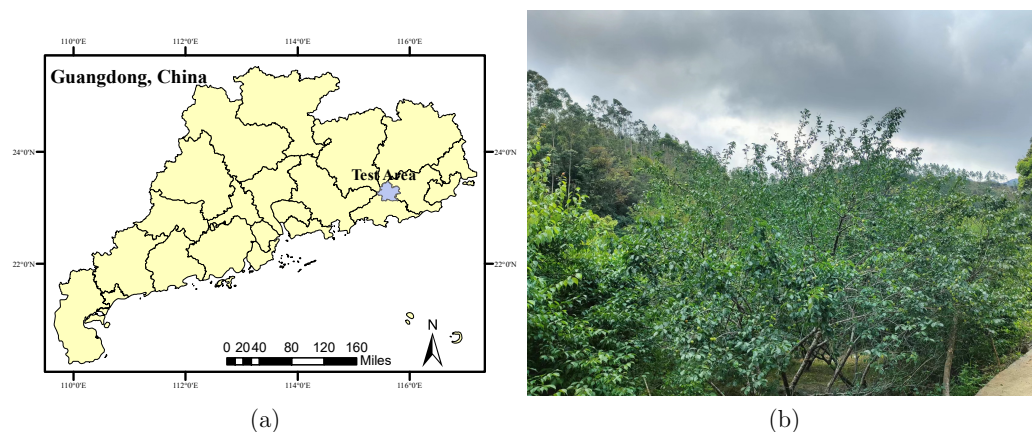


Figure 1. (a) shows the geographical location of the experimental site, and (b) displays the plum trees studied.

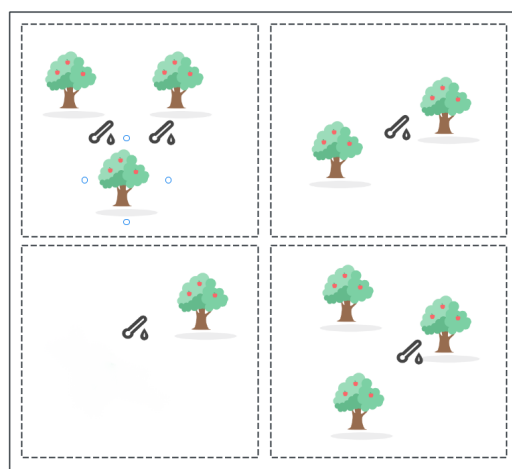


Figure 2. Diagram of the layout of soil temperature and humidity sensors. Sensors should be placed based on differences in soil conditions. If multiple trees share consistent soil conditions, 1 to 2 sensors are sufficient. If conditions vary, sensors should be installed separately.

Fourteen green plum trees were selected for LAI, SPAD, and environmental data measurements, with their positions shown in Figure 3. The circles numbered 1 to 14 represent the sample trees, rectangles mark the weather station, and black dots mark the sensors. These 14 trees are located within the same orchard, where soil and climate conditions are similar. Additionally, the trees are comparable in terms of age and canopy size. Throughout the study, both the control and experimental trees received the same fertilization and pest control treatments. Trees 1 to 12 form the experimental group used to provide representative data, while trees 13 and 14 serve as controls for baseline comparisons.

The methodological framework of this study is illustrated in Figure 4. First, SPAD values were obtained by using the SPAD–502 chlorophyll meter to measure the relative

content of chlorophyll in green plum leaves. Simultaneously, data on environmental characteristics were collected through the orchard’s integrated weather station and sensor systems. The collected SPAD and environmental data were then used to develop two distinct models: the SPAD–PM and the EFs–PM. The outputs of these models were subsequently integrated into the decision fusion model (MDF–ADRS) developed in this study, enabling the accurate inversion of the leaf area index (LAI) for green plums.



Figure 3. A drone aerial image of the experimental area, displaying the locations of the sample green plum trees, meteorological stations, and soil temperature and humidity sensors. Circles numbered 1 to 14 represent the sample green plum trees, the rectangles mark the meteorological stations, and the black dots indicate the placement of the soil temperature and humidity sensors.

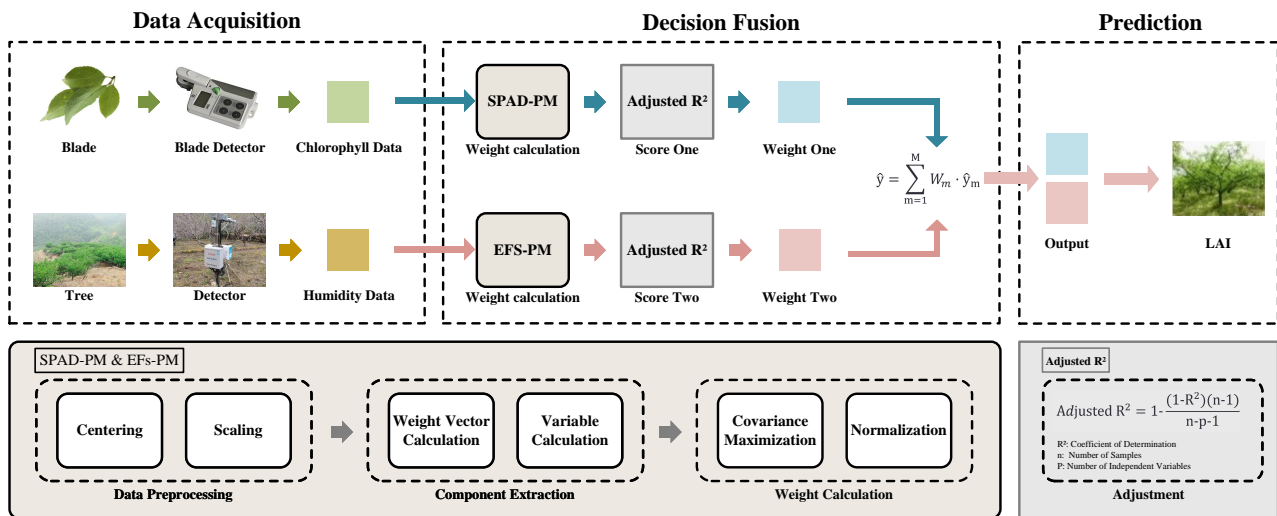


Figure 4. This figure presents the experimental flowchart of the study, which consists of three steps: data acquisition, decision fusion, and Inversion. Data acquisition: the SPAD–502 chlorophyll analyzer is used to collect SPAD data, while environmental characteristic factor data on green plums are gathered through meteorological facilities in the orchard. Decision fusion: This section introduces the MDF–ADRS decision fusion model developed in this study. First, the SPAD–PM and EFs–PM inversion models are constructed using SPAD data and environmental feature factor data, respectively (with the environmental feature factors needing to be screened). The results from the SPAD–PM and EFs–PM models are then combined using a weight reassignment strategy for decision fusion. Inversion: the fused results are used to inverse the LAI of the green plums.

2.2. Data Collection and Dataset Construction

2.2.1. Data Collection

In this study, we conducted comprehensive data collections during two critical growth stages of green plum—the flowering stage (2 February to early March) and the fruit ripening stage (early March to 3 April)—resulting in a total of 13 data collection sessions. LAI values

were obtained using a leaf area index measuring instrument (Figure 5), which features a fish-eye lens with a wide-angle view of 210° , a focal length of 1.05 mm, an illumination level of 0.1 lux, and a pixel size of $5.7 \text{ mm} \times 4.28 \text{ mm}$. The output formats include YUV2, YUYV, and MJPEG. SPAD values were measured using a handheld SPAD-502 chlorophyll meter (Figure 6). Environmental characteristic factors were recorded using the weather stations (Figure 7) and soil temperature and moisture sensors installed throughout the orchard. These enabled the continuous monitoring of ambient temperature, humidity, soil temperature, soil moisture, soil pH, dew point temperature, light intensity, photosynthetically active radiation, air pressure, wind speed, $\text{PM}_{2.5}$ concentration, and CO_2 levels. These data provided essential support for analyzing the environmental conditions affecting the growth of the green plum trees in the experimental area.



Figure 5. The leaf area index (LAI) measuring instrument, with its fish-eye lens facing upwards, is placed horizontally above the crown of the tree during measurement.



Figure 6. Leaves and the SPAD-502 chlorophyll analyzer: the SPAD-502 instrument is used to clamp the leaves and obtain chlorophyll values.



Figure 7. The orchard meteorological station collects environmental data within the orchard and promptly uploads the data to a cloud platform for storage.

The LAI measurements were taken by orchard professionals. During the process, the operators needed to flexibly adjust their strategies based on weather conditions to ensure data accuracy and consistency. On sunny days, strong sunlight may cause excessive glare, so midday should be avoided; measurements should preferably be taken in the morning or evening when the light is more even. To reduce the interference from strong light, the shading device provided with the instrument should be used, and the fish-eye lens of the instrument must be kept perpendicular to the ground and aligned with the central axis of the canopy. The instrument's height should be adjusted according to the actual structure of the canopy, ensuring that the main part of the canopy is covered while avoiding obstruction from lower leaves. On cloudy days, when the light is more even, operators need to pay closer attention to the placement height and angle of the instrument, ensuring that the light

sensor is able to accurately measure the canopy's structure under weaker light conditions. In low-light situations, operators should use supplementary lighting and may need to increase the number of measurement points or extend the measurement time to gather more detailed and comprehensive canopy data. Before each measurement, the instrument must be calibrated, and three measurement points should be set in the east, south, west, and north directions, with one data recording from each point. The final results, processed as an average, can effectively reflect the true LAI value of the canopy, ensuring the data's reliability and repeatability.

During the SPAD measurements, impurities such as dust or stains on the leaves may affect the accuracy of the readings, and leaves damaged by pests or diseases may interfere with SPAD values due to their abnormal color and structure. Therefore, leaves should be carefully inspected before data collection, and any dust or impurities should be gently wiped off with a clean cloth to ensure the measurement area is clean and free of contamination. Leaves affected by pests or diseases should be avoided. Samples are selected from the upper, middle, and lower layers of the crop, with three leaves chosen from each of the east, south, west, and north directions, making a total of 12 leaves. Each leaf is marked, as shown in Figure 8. During the measurement, the SPAD meter must be in complete contact with the leaf surface, avoiding any bending or twisting of the leaf. Each leaf is measured in its upper, middle, and lower parts, avoiding the veins, and the average value of these readings is taken as the chlorophyll content for that leaf. The average chlorophyll content of the 12 leaves is used as the relative content of chlorophyll (SPAD) for that sample. If any of the marked leaves are affected by pests or diseases during the experiment, this must be carefully recorded and explained. Specifically, the severity of the pest or disease impact on the leaf should be noted, and photos should be taken for subsequent analysis. If the diseased area is large or severely affects the color or structure of the leaf, the SPAD data for that leaf will no longer be used in the analysis, and a healthy leaf from the same location will be marked and measured as a replacement.

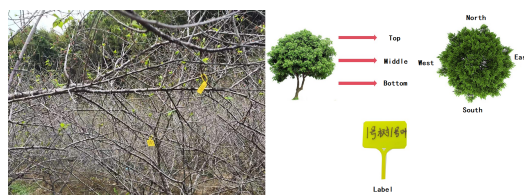


Figure 8. We selected 12 leaves from the upper, middle, and lower levels of the sample, as well as from the east, south, west, and north directions, to measure SPAD values, and used label plates to mark these leaves.

Figure 9 displays the projections of the leaf area index (LAI) values of a sample tree in various directions, with the LAI data are presented in Table 1 and the SPAD data in Table 2. To minimize the uncertainties caused by individual measurements, environmental characteristic data were collected multiple times under different weather conditions and at various time periods to capture the crop's performance under varying environmental states. This approach helps prevent the influence of extreme weather during specific periods from affecting the inversion results. Some of the environmental characteristic data are shown in Table 3. For data from multiple measurements, we applied the moving average smoothing technique. The core idea of this method is to replace individual data points with the average of several data points over a specific period. The window size for the moving average was chosen based on the frequency of data collection to balance noise reduction and data responsiveness. This approach smooths out large short-term fluctuations, highlighting long-term trends and reducing the noise caused by transient weather conditions such as sudden rainfall or gusts of wind, thereby minimizing the impact of environmental variability. This results in more stable and accurate inversion outcomes. All measurement data were meticulously recorded in an Excel file for subsequent analyses.

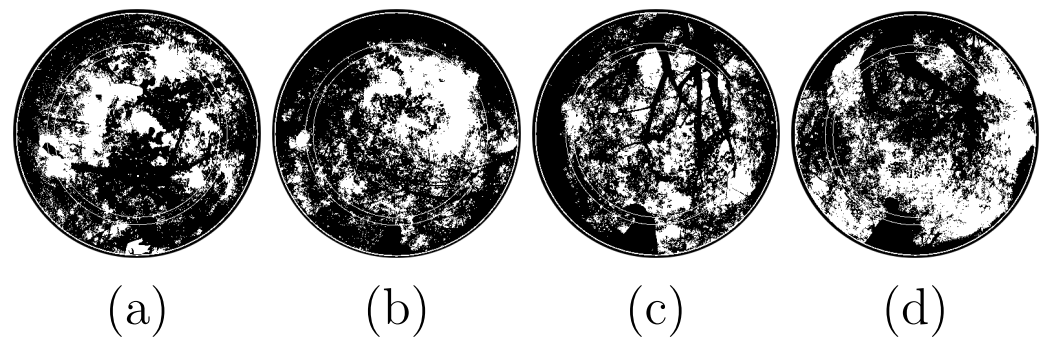


Figure 9. When using leaf area index instruments to measure the LAI from different directions, the projections in the instrument—labeled (a–d)—represent the east, west, south, and north directions, respectively.

Table 1. Table of single-sample leaf area index collection record (1, 2, and 3 represent measurement point 1, measurement point 2, and measurement point 3, respectively).

Sample	East			South			West			North			Ave
	1	2	3	1	2	3	1	2	3	1	2	3	
1	2.41	1.73	1.76	2.4	3.6	2.72	3	2.54	3.16	3.41	2.08	2.3	2.59
2	3.11	3.11	2.2	2.04	2.26	1.9	2.16	2.68	2.41	1.88	2.61	3.31	2.47
3	1.48	1.5	2.61	1.84	1.98	2.18	2.01	2.07	1.6	1.92	2.22	2.07	1.96
4	2.67	2.31	1.89	1.3	1.8	1.8	1.48	2.83	2.02	1.58	1.81	1.33	1.9
5	2.77	2.13	2.16	1.91	1.78	2.55	2.38	3.45	2.59	2.49	1.65	1.65	2.29
6	2.78	2.72	2.85	2.41	2.21	2.72	1.81	1.74	2.72	2.34	2.34	2.77	2.45
7	1.98	2.93	2.55	3.02	2.3	1.95	2.09	2.61	2.88	3.19	2.3	2.75	2.55
8	3.37	1.74	3.02	2.67	2.07	1.96	1.96	2.71	2.68	1.64	1.7	2.34	2.54
9	1.9	2.27	2.3	2.23	2.67	2.89	2.88	2.72	3	3.33	2.62	3.12	2.66
10	2.12	2.78	3.32	2.54	2.89	2.22	1.95	1.79	2.43	2.67	2.24	2.26	2.52
11	3.01	3.1	2.62	2.42	2.3	2	2.18	2.7	2.51	1.92	2.58	3.02	2.53
12	2.82	2.4	2.31	1.95	1.89	2.72	2.4	3.32	2.6	2.5	1.88	1.87	2.39
13	2.48	2.05	2.29	2.64	2.34	2.79	2.42	2.42	2.42	2.41	2.38	2.71	2.45
14	2.49	2.51	2.91	2.37	3.08	2.98	2.54	2.74	2.67	2.44	2.04	2.43	2.6

Table 2. Table of partial collection records of single-sample SPAD (“Up” represents the upper part of the blade, “Mid” represents the middle part of the blade, and “Down” represents the lower part of the blade).

Sample	East			South			West			North		
	Up	Mid	Down	Up	Mid	Down	Up	Mid	Down	Up	Mid	Down
1	28.2	33.3	23.6	29.6	29.9	29.6	32.8	37.8	28.1	31.3	30	33.5
2	29.1	30.6	19.5	28.2	33.8	37	32.6	28.8	33	28.7	28.1	29
3	28.6	32.7	26.3	33.7	30.1	20.9	36.1	23.9	31.3	37.8	31.9	25.3
4	25.2	35.3	32.8	26.8	31.2	30.8	26.3	34.1	27.4	24.6	29.1	31.8
5	30.4	28.8	29.6	36.4	31	36.6	32.4	25.1	23.7	24.1	29.1	23.7
6	39	29.7	31.6	36.3	33.8	35.2	33.7	29.7	29.5	24.5	33	23.7
7	39	37.8	31.7	38.5	32.6	36	36.5	40.2	36.5	35	40	33
8	26.5	30.9	24.3	31.9	32.8	29.5	20.5	19.9	25.1	29.5	28.2	31.9
9	26.7	31.2	29.6	32.3	33.8	28.9	21.2	23.2	25.8	29.7	27.3	32.8
10	27.5	30.4	31.6	32.5	34.4	30.6	25.4	26.7	32.1	28.8	27.4	31.4
11	28.4	29.5	32.7	31.3	34.8	30.2	27.7	30.4	29.8	35.2	34.8	38.3
12	38.3	35.4	34.2	36.5	32.7	35.1	33.3	30.2	29.1	33.4	35.4	35.6
13	27.2	31.3	30.2	32.7	33.9	30.5	21.6	23.9	31.7	29.5	27.6	31.1
14	29.3	33.6	37.5	26.9	43.9	34.5	36.9	36.1	32.6	37	39.2	26

Table 3. Partial environmental data collection (ET represents environmental temperature, EH represents environmental humidity, Soil–T represents soil temperature, Soil–M represents soil moisture, CO₂ represents carbon dioxide concentration, and LI represents light intensity).

ET	EH	Soil–T	Soil–M	CO ₂	LI	Collection Time
18.6	64.9	16.8	36.18	552.9	13,086	20 March 2024-9:00:02
20.3	54.2	16.7	36.08	558.0	9009	20 March 2024-9:30:02
21.9	49.2	16.7	36.06	550.9	11,919	20 March 2024-10:00:00
22.4	48.9	16.6	36.14	544.8	15,644	20 March 2024-10:30:01
24.0	46.5	16.6	36.03	538.8	16,503	20 March 2024-11:00:02
24.0	44.1	16.4	35.94	533.6	13,563	20 March 2024-11:30:02
25.4	44.1	16.3	35.95	524.4	15,171	20 March 2024-12:00:01
25.4	43.7	16.8	35.87	552.9	15,522	20 March 2024-12:30:01
25.2	43.7	16.8	35.77	515.9	16,948	20 March 2024-13:00:01
25.2	43.7	17.0	35.70	512.3	10,526	20 March 2024-13:30:01

2.2.2. Dataset Construction

In this study, we first preprocessed the initial data collected on the LAI, leaf chlorophyll content, and environmental characteristic factors of green plums. Using statistical methods, we thoroughly examined the raw dataset, identifying and removing outliers that deviated significantly from the overall distribution. After this filtering process, we retained a total of 182 data samples, including 156 samples from the experimental group, which served as the foundational dataset for the machine learning algorithm’s modeling. The remaining 26 samples from the control group were used for model validation and comparative analyses. Additionally, to reduce the impact of dimensional differences and seasonal fluctuations across different data sources, we applied standardization and normalization to the dataset. This ensured consistent numerical scales for all features, thereby enhancing the model’s robustness and generalization ability. Table 4 presents the variability and central tendencies of the parameters within the dataset.

Table 4. Environmental data statistics (the maximum, minimum, mean, median, and standard deviation of 12 environmental characteristic factors were statistically analyzed).

Sample	Numbers	Max	Min	Mean	Med	Std
LAI	182	5.175	0.961	2.941	3.151	1.03
SPAD	182	35.14	17.725	25.709	25.383	4.621
ET	182	29.92	11.3	18.364	17.55	4.03
EH	182	93.5	62.82	80.285	80.425	5.242
Soil–T	182	24.25	12.4	18.24	18.26	2.927
Soil–M	182	72	20.37	40.11	40.2	12.867
LI	182	17,902	447	5127	4840	3477
CO ₂	182	557	492	524.6	524	19.496
pH	182	8.26	7.7	8	7.895	0.177
DPT	182	23	8.53	15.207	14.11	4.389
PAR	182	11.73	2.5	5.367	4.68	2.386
AP	182	1016	967	999.1	1001	9.816
WS	182	0.65	0.002	0.164	0.145	0.138
PM _{2.5}	182	4.83	0.01	0.338	0.01	0.964

2.3. LAI Inversion Model Based on Environmental Characteristic Factors (EFs–PM)

2.3.1. Environmental Characteristic Factor Analysis

In this study, we obtained data on 12 environmental characteristic factors of green plums using weather stations and soil temperature and moisture sensors. We evaluated the relationship between these factors and the LAI using statistical methods to assess the linear relationships between variables [33–35]. The Pearson [33] correlation coefficient was employed to measure the linear relationship between two continuous variables. Its value

ranges from -1 to $+1$, where $+1$ indicates a perfect positive linear relationship, -1 indicates a perfect negative linear relationship, and 0 indicates no linear relationship. The calculation formula is provided in Equation (1).

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}} \quad (1)$$

In this formula, n represents the sample size, Σxy is the sum of the products of the corresponding values of the two variables, and Σx and Σy represent the sums of the values of each variable, respectively. Σx^2 and Σy^2 represent the sums of the squares of the values of each variable.

The Spearman rank [34] correlation coefficient, a non-parametric measure, evaluates the correlation between the rankings of two variables. It shares the same range of values as the Pearson coefficient, from -1 to $+1$. Its calculation formula is provided in Equation (2).

$$\rho = 1 - \frac{6\Sigma d_i^2}{n(n^2 - 1)} \quad (2)$$

In this context, d is the rank difference between two variables and n represents the sample size.

The Kendall rank [35] correlation coefficient provides a non-parametric measure of correlation by assessing the agreement and disagreement between pairs of data points. Its formula is provided in Equation (3).

$$\tau = \frac{2(\text{number of concordant pairs}) - 2(\text{number of discordant pairs})}{n(n - 1)} \quad (3)$$

where *number of concordant pairs* is the consistent ranking of the two variables, *number of discordant pairs* is inconsistent ranking of the two variables, and n represents the sample size.

To ensure that environmental characteristic factors (EFs) have a strong explanatory power for the LAI, we first calculated their correlation coefficients using the Pearson, Spearman, and Kendall methods. We then ranked the results from each method and selected the features with the highest correlation coefficients in each analysis. If a feature consistently showed high correlation across multiple methods, it indicated a strong relationship with the LAI. Finally, we integrated the results from all three methods, identifying the top six environmental factors that consistently exhibited high correlation across the analyses. This comprehensive approach enables us to identify the most relevant environmental factors influencing the LAI, enhancing our understanding of how these factors drive changes in plum tree LAI.

2.3.2. EFs–PM Model Construction Method

In this study, we used experimental-group data (from Trees 1 to 12) to develop LAI inversion models by treating environmental feature factors as independent variables and the LAI as the dependent variable. We applied multivariate data analysis algorithms [36], including PLSR [37], SVMR [38], RF [39], MLR [40], and GPR [41]. Subsequently, these models were validated using control–group data (from Trees 13 and 14). By comparing the evaluation metrics of each model, we identified the most effective environmental feature-based LAI inversion model (EFs–PM).

PLSR is a technique that combines principal component analysis and regression analysis which is suitable for handling situations with high multicollinearity among independent variables. It works by projecting both the independent and dependent variables into new subspaces to extract the latent components that best explain their relationship, while simultaneously reducing dimensionality and constructing a regression model to maximize variable correlation. SVMR, based on Support Vector Machines, is a regression method that predicts outcomes by finding a regression hyperplane that maximizes the margin of the support vectors. RF, as an ensemble learning method, builds multiple decision trees for

regression, with the final prediction given by averaging the results of all the trees. MLR assumes a linear relationship between independent and dependent variables, using the least squares method to find the optimal linear mapping between them. GPR is a Bayesian non-parametric regression method that models the functional distribution of independent and dependent variables based on Gaussian processes, providing confidence intervals for these predictions.

Each algorithm has different characteristics and strengths when dealing with complex and high-dimensional data. PLSR efficiently addresses multicollinearity through dimensionality reduction, making it suitable for high-dimensional datasets. SVMR captures the nonlinear relationships between the LAI and environmental variables using kernel functions. RF processes complex nonlinear data by integrating multiple decision trees and has strong noise resistance. MLR serves as a baseline model for evaluating the linear relationship between the LAI and environmental factors. GPR enhances prediction robustness by handling uncertainty. Together, these algorithms provide a comprehensive modeling and analysis framework for LAI inversion.

The coefficient of determination R^2 [42] (Equation (4)) and the root mean square error (RMSE) [43] (Equation (5)) were chosen as the evaluation metrics used to assess the model's inversion accuracy. The R^2 indicates the percentage of variance explained by the model, with larger R^2 values reflecting stronger explanatory power. RMSE, on the other hand, measures the average error between the values predicted by the model and the actual observed values, with smaller RMSE values indicating higher inversion accuracy.

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

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

Let n denote the sample size, y the manually measured values, \hat{y} the predicted values, and \bar{y} the mean value.

To investigate the effect of the correlation analysis on the accuracy of LAI inversion, two models were developed. The first model incorporates the six environmental factors that show the highest correlation with the LAI, whereas the second model uses all the initially collected environmental factors. A comparison of these models' performances was conducted to evaluate the impact of the correlation analysis.

2.4. SPAD-Based LAI Inversion Model for Plum Trees (SPAD-PM)

This study combined measured chlorophyll content (SPAD) data with concurrently measured LAI data and applied multivariate data analysis algorithms, including PLSR, SVMR, RF, MLR, and GPR, to build LAI inversion models. The best-performing model among these five algorithms was used as the SPAD-based LAI inversion model (SPAD-PM) and validated using control-group data.

PLSR improves prediction accuracy by addressing multicollinearity in SPAD data through dimensionality reduction. SVMR captures complex nonlinear relationships between SPAD and the LAI, while RF handles multidimensional nonlinear features with strong noise resistance. MLR assesses the linear relationship between SPAD and the LAI as a baseline comparison model, and GPR enhances model robustness by providing confidence intervals and handling noisy, uncertain data. These algorithms offer diverse modeling approaches to LAI inversions using SPAD data.

2.5. Multi-Source Decision Fusion Model Based on Adjusted Coefficient of Determination (MDF-ADRS)

The two methods described earlier, while capable of performing an LAI inversion to some extent, rely on single-source information and have several limitations. First, the complexity of a plum tree's growing environment cannot be fully captured by a single

information source, leading to reduced inversion accuracy. Second, reliance on a single data source may introduce biases, causing overfitting to specific datasets and impairing performance on new data. Lastly, depending on a single data source complicates data updates and model expansion [44]. Zhao [45] proposed an approach for accurately monitoring LAI and SPAD values in summer maize using multi-source data fusion, demonstrating that multi-source data models are more effective than single-source models. Their random forest model, which incorporated multi-source data fusion, achieved the highest accuracy. Integrating multi-source data allows the LAI inversion model to better adapt to the complex and variable growth conditions of plum trees, thereby providing more precise inversion results.

This study introduces a multi-source decision fusion model based on the adjusted coefficient of determination (adjusted R^2) [46]. The model has three components: model output, adjusted R^2 computation, and decision fusion, as shown in Figure 10. The model output section includes two single-source models (SPAD and environmental feature factors) and calculates their respective R^2 values. In the adjusted R^2 computation, the R^2 value is modified to account for the number of independent variables used and sample size. These adjusted R^2 values are then used in the decision fusion process, where each model is weighted based on its adjusted R^2 value. The final inversion result is obtained through weighted averaging, leveraging the complementarity of multiple data sources to enhance accuracy and robustness. The weighted averaging formula used is provided in Equation (6).

$$\hat{y} = \sum_{m=1}^M W_m \cdot \hat{y}_m \tag{6}$$

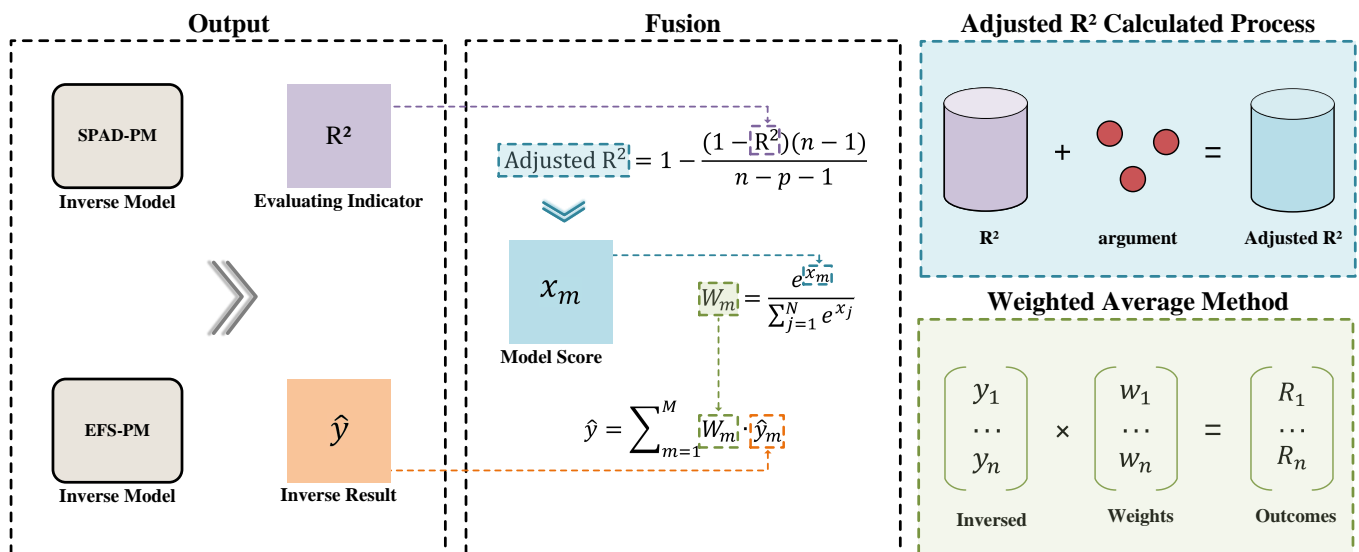


Figure 10. This figure illustrates the process of integrating the output results of two different inversion models (SPAD-PM and EFS-PM) to improve the final inversion accuracy. Both the SPAD-PM and EFS-PM models provide evaluation metrics (R^2) and inversion results (\hat{y}). In the fusion process, their adjusted R^2 value is first calculated, taking into account the number of virtual samples (n) and the variables (p) that affect model evaluation. Next, the score (x_m) for each model is computed, and their weight coefficients (W_m) are standardized. Using a weighted average formula, a weighted sum is applied to obtain the fused inversion result. The weights are adjusted using an exponential function based on model scores to ensure a reasonable weight distribution. Finally, the weighted average method is used to combine multiple inversion results (y_1, y_2, \dots, y_n), outputting the fused inversion results (R_1, R_2, \dots, R_n) and thereby improving the overall performance of the model's inversion.

Here, w_m signifies the weight of the m model, y is the ultimate inversion output, and y_m is the inversion output from the m model. The formula for determining the m model's weight w_m is detailed in Equation (7):

$$W_m = \frac{e^{x_m}}{\sum_{j=1}^N e^{x_j}} \quad (7)$$

In this context, x_m denotes the score assigned to the model, while N indicates the total number of m models. The computation of the adjusted R^2 is detailed in Equation (8):

$$\text{Adjusted-}R^2 = 1 - \frac{(1 - R^2)(n - 1)}{n - p - 1} \quad (8)$$

In this formula, R^2 represents the coefficient of determination, n denotes the total number of samples, and p indicates the number of predictors within the model.

3. Results

This study successfully achieves precise LAI inversion using environmental feature factors and SPAD data. Our findings reveal that our multi-source decision fusion model based on the adjusted coefficient of determination (MDF-ADRS) significantly outperforms the SPAD-PM and EFs-PM single-source models, as well as other traditional model fusion methods, in terms of both inversion precision and accuracy. The detailed experimental results and analyses are provided below.

3.1. Analysis of EFs Results

This study analyzes the relationship between environmental characteristic factors and the LAI using statistical methods used for assessing linear relationships. The correlation coefficients between these 12 environmental factors and the LAI are shown in Figure 11. As depicted in Figure 11a, the Pearson correlation analysis indicates that environmental temperature (ET), soil temperature (Soil-T), soil moisture (Soil-M), and light intensity (LI) exhibit the strongest positive correlations, with coefficients of 0.770, 0.606, 0.731, and 0.637, respectively. In contrast, CO₂ concentration (CO₂) and environmental humidity (EH) show the strongest negative correlations, with coefficients of -0.727 and -0.637, respectively. Figure 11b shows that, under Spearman's rank correlation analysis, the six environmental factors of ET, Soil-T, LI, CO₂, EH, and Soil-M have the highest absolute correlation values with the LAI. ET, Soil-T, Soil-M, and LI exhibit positive correlations, with coefficients of 0.718, 0.661, 0.654, and 0.596, respectively, while CO₂ concentration and environmental humidity show negative correlations, with coefficients of -0.640 and -0.617. Figure 11c reveals that under Kendall's correlation analysis, these same six environmental factors (ET, Soil-T, LI, CO₂, EH, and Soil-M) show the highest absolute correlation values with the LAI. The absolute values of the remaining factors are all below 0.5.

Therefore, by synthesizing the results from the three methods it can be concluded that, among the twelve environmental characteristic factors assessed, environmental temperature (ET), soil temperature (Soil-T), light intensity (LI), CO₂ concentration (CO₂), environmental humidity (EH), and soil moisture (Soil-M) exhibit the strongest correlations with the leaf area index (LAI). These factors are more effective in the regression analysis of LAI inversion and are considered optimal environmental feature factors (OEFs). We assessed the relative importance of the environmental factors in the MDF-ADRS model based on their correlation strength with the LAI, ranking them by the absolute value of their correlation coefficients. The ranking is in the following order: environmental temperature (ET), soil moisture (Soil-M), carbon dioxide concentration (CO₂), soil temperature (Soil-T), environmental humidity (EH), and light intensity (LI). For future research, we plan to use more interpretable methods like Partial Dependence Plots (PDPs) and SHAP values. PDPs will show how individual features impact LAI predictions, while SHAP will quantify

each feature’s contribution, providing a clearer understanding of feature importance and improving model interpretability.

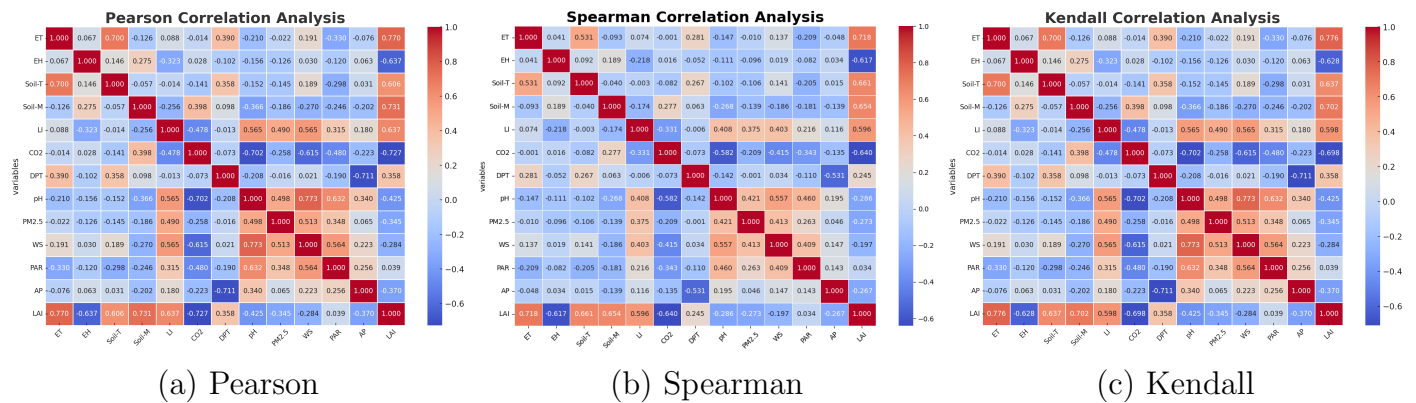


Figure 11. The correlation analysis between the LAI and environmental characteristic factors indicates that a coefficient with a larger absolute value signifies a stronger correlation.

3.2. Analysis of EFs—PM Model Results

We used the experimental-group data as the modeling set and applied multivariate data analysis techniques to construct an LAI inversion model, using the optimal environmental feature factors (OEFs) as predictors and the LAI as the response variable. The constructed model was validated using a validation set composed of control–group data. These data were selected because Trees 13 and 14, which served as control trees, are located in the same orchard as the experimental trees and share similar conditions in terms of their climate, soil, canopy structure, and tree age. This ensures that the validation data are consistent with the experimental data in terms of fundamental conditions. The experimental results are presented in Table 5; the Partial Least Squares Regression (PLSR) [37] model, under optimized conditions, demonstrated a superior inversion accuracy, achieving R^2 values of 0.88 and 0.69 for the modeling and validation sets, respectively, and $RMSE$ values of 0.37 and 0.55, outperforming alternative models. Comparative testing was also performed using all environmental characteristic factors (AEFs) as predictors, with the results presented in Table 6. The analysis of Tables 5 and 6 confirms that the PLSR model maintains a higher inversion accuracy across both optimized and comprehensive environmental conditions. Moreover, the models utilizing optimized features consistently showed an enhanced R^2 and reduced $RMSE$, affirming the beneficial role of optimal feature selection in increasing model accuracy.

Table 5. Evaluation of green plum LAI estimation results based on OEFs.

Algorithm	Modeling		Validation	
	R^2	$RMSE$	R^2	$RMSE$
MLR [40]	0.76	0.51	0.56	0.69
SVMR [38]	0.79	0.47	0.60	0.62
RF [39]	0.70	0.54	0.53	0.75
GPR [41]	0.83	0.39	0.67	0.59
PLSR [37]	0.88	0.37	0.69	0.55

Table 6. Evaluation of green plum LAI estimation results based on AEFs.

Algorithm	Modeling		Validation	
	R^2	$RMSE$	R^2	$RMSE$
MLR [40]	0.68	0.67	0.54	0.74
SVMR [38]	0.71	0.51	0.56	0.71

Table 6. Cont.

Algorithm	Modeling		Validation	
	R^2	RMSE	R^2	RMSE
RF [39]	0.67	0.59	0.48	0.82
GPR [41]	0.73	0.50	0.59	0.67
PLSR [37]	0.82	0.42	0.62	0.64

3.3. Analysis of SPAD–PM Model Results

This research utilized SPAD as the independent variable and measured the LAI as the dependent variable to establish an LAI inversion model based on multivariate data analysis techniques. The performance of this model is summarized in Table 7, which highlights that the Partial Least Squares Regression (PLSR) model displayed the greatest inversion accuracy. The PLSR model achieved an R^2 of 0.73 for the model building set and 0.62 for the validation set, with RMSEs of 0.59 and 0.61, respectively; markedly better than those of competing models. As a result, PLSR was used to construct the SPAD–PM model.

Table 7. Evaluation of green plum LAI estimation results based on SPAD.

Algorithm	Modeling		Validation	
	R^2	RMSE	R^2	RMSE
MLR [40]	0.64	0.75	0.50	0.83
SVMR [38]	0.66	0.73	0.52	0.81
RF [39]	0.60	0.82	0.48	0.87
GPR [41]	0.62	0.77	0.49	0.82
PLSR [37]	0.73	0.59	0.62	0.61

3.4. MDF–ADRS Model Evaluation

This investigation utilized the EFs–PM and SPAD–PM models' outputs as inputs, applying a weighted average fusion method determined through an adjusted coefficient of determination. The MDF–ADRS model, in its validation phase, achieved an R^2 of 0.88 and an RMSE of 0.39, indicating its robust LAI inversion capabilities. To validate the model's enhanced performance, it was compared against classical fusion techniques such as simple averaging, voting, and linear weighing based on Log R^2 . Figure 12 presents a detailed comparison across six models, demonstrating that the MDF–ADRS model, relative to simple averaging, voting, and linear weighing based on Log R^2 , improved the R^2 by 0.17, 0.14, and 0.10 and reduced the RMSE by 0.13, 0.10, and 0.06, respectively. In comparison to the pre-fusion single–source models, the MDF–ADRS model enhanced the R^2 by 0.26 and 0.19 relative to the SPAD–PM and EFs–PM models, respectively, and lowered the RMSE by 0.22 and 0.16.

The comparison between the inversion and measured values in the validation set, as shown in Figure 13, clearly demonstrates that the MDF–ADRS model exhibits a superior fitting performance compared to other models.

The experimental results clearly demonstrate that the MDF–ADRS model has a significant advantage in terms of its LAI inversion accuracy and precision. In terms of data characteristics, the EFs–PM model primarily relies on environmental data, while the SPAD–PM model is based on SPAD data. A single data source alone is insufficient to fully capture the complexity of the LAI. Environmental data reflect macro–level climate conditions, while SPAD data are closely related to the physiological state of vegetation. The MDF–ADRS model enhances LAI inversion accuracy by integrating these two data sources and leveraging their complementary strengths to capture both macro–environmental and micro-vegetation changes. In terms of model structure, the multi–source decision fusion framework of the MDF–ADRS model allows for the simultaneous use of multiple data sources, integrating them through a weight-optimized allocation strategy. Compared

to single-source models (EFs-PM or SPAD-PM), the MDF-ADRS model captures a richer set of features, addressing the limitations of single-source models under extreme environmental conditions. Consequently, the MDF-ADRS model outperforms other models in terms of its overall performance, providing a more reliable and accurate tool for crop growth monitoring in agricultural engineering.

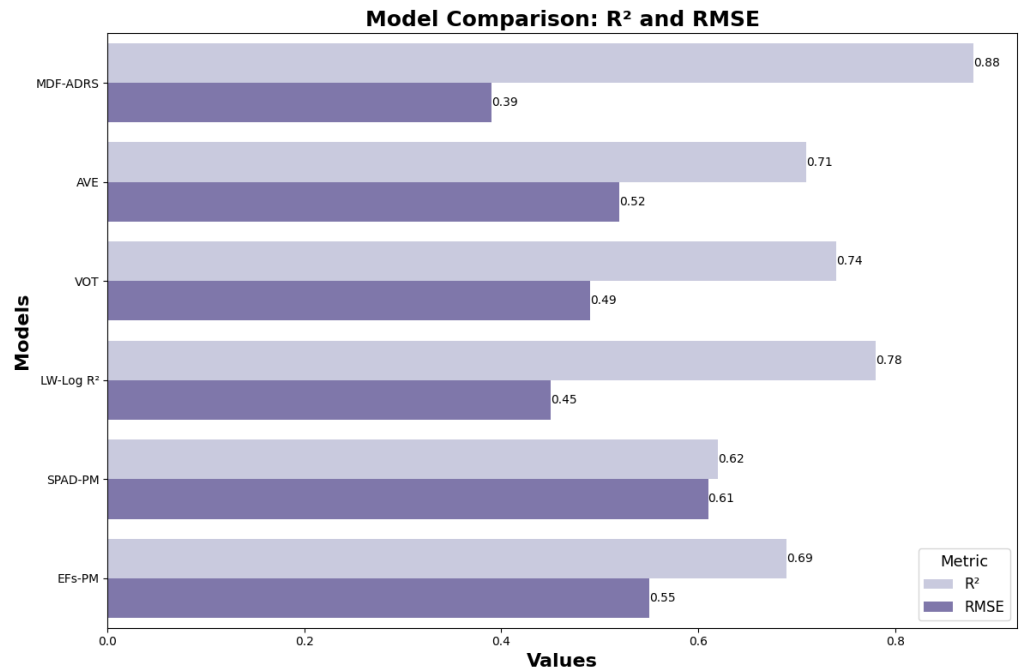


Figure 12. This figure presents the comparative results of our model and the simple average method (AVE), voting method (VOT), and linear weighting method based on Log R² (LW-Log R²).

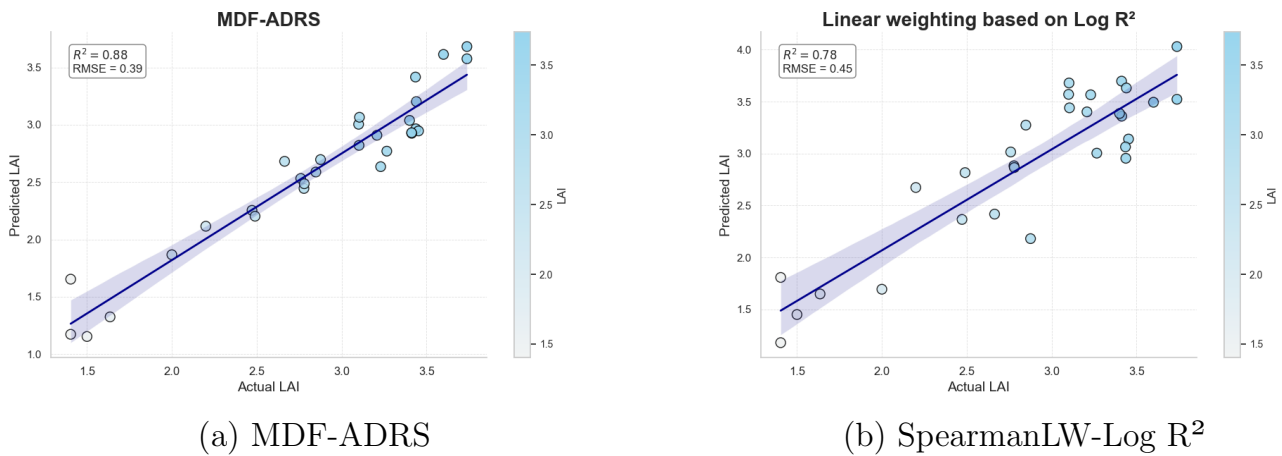


Figure 13. Cont.

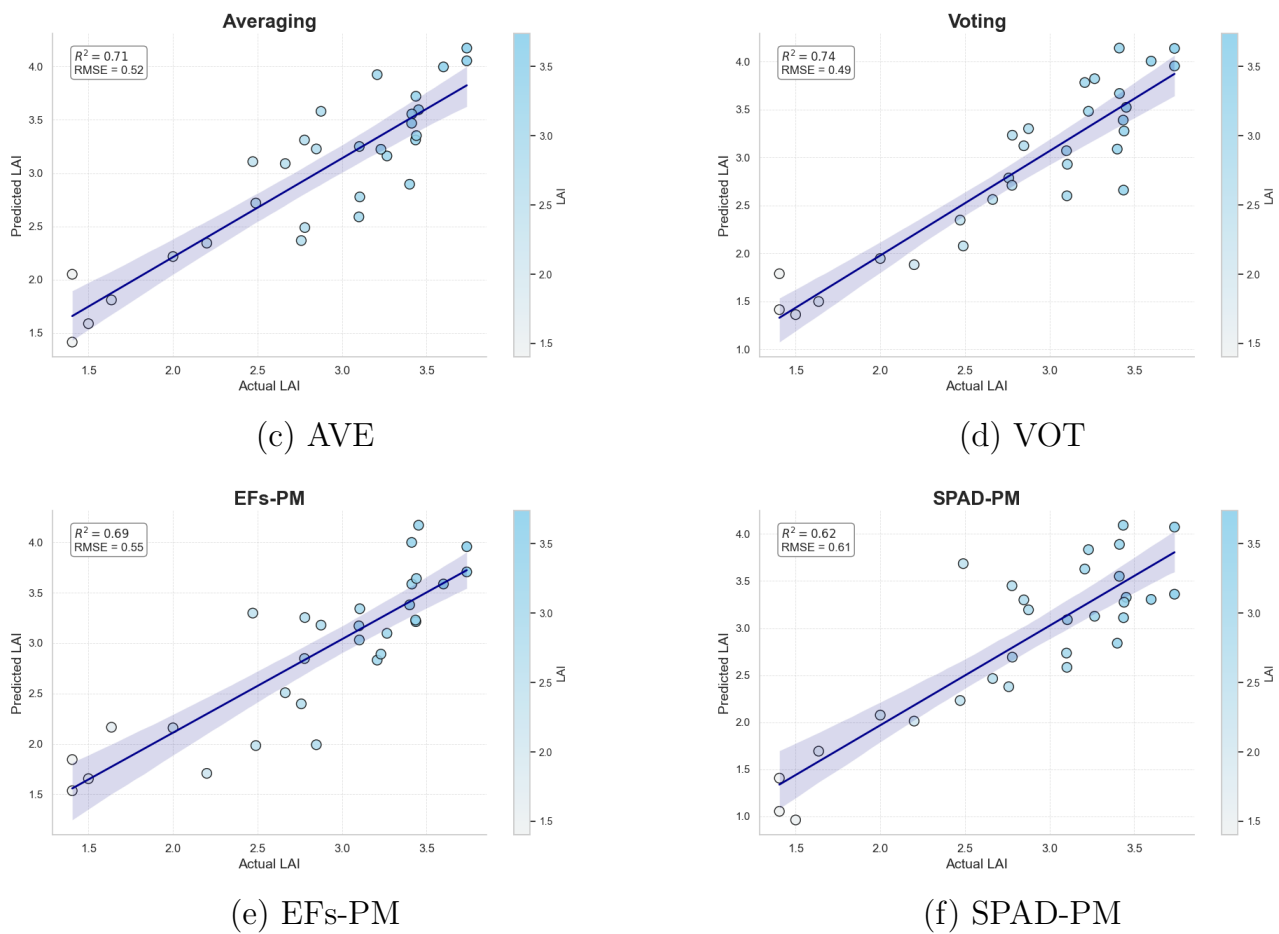


Figure 13. Results of comparison between model inversion values and measured values.

4. Discussion

This paper presents the development of a cutting-edge multi-source decision fusion model (MDF-ADRS) based on the adjusted coefficient of determination for inverting the leaf area index (LAI) of green plums. This model integrates both environmental characteristic factors and SPAD values to achieve precise LAI estimates. Demonstrating exceptional accuracy, with an R^2 of 0.88 and an RMSE of 0.39 in the validation set, the MDF-ADRS model significantly outperforms established single-source models (EFs-PM and SPAD-PM) as well as other fusion techniques. This confirms the model's effectiveness in utilizing multi-source data for LAI inversion and aligns with Zhao's [45] findings on the superior performance of multi-source models in agricultural applications. This study proposes an innovative multi-source data fusion approach that combines leaf SPAD values with IoT-derived environmental factors. Compared to traditional single-source LAI estimation methods, this approach captures the dynamic changes in the LAI more comprehensively and significantly improves predictive accuracy. For example, Wei [47] used Sentinel-2 imagery and PROSAIL model parameter calibration to invert the maize canopy LAI, relying mainly on remote sensing data. In contrast, our study integrates IoT-based environmental data, providing more detailed ground-level observations and enhancing the robustness of the model. Shi [48] combined airborne LiDAR waveforms with Sentinel-2 imagery to estimate the LAI using a physical model. While effective, their method may have limitations in handling multi-source data. Our research, however, employs an adjusted R^2 -based multi-source decision fusion model, optimizing the weighting of various data sources and significantly boosting both model generalization and prediction accuracy.

This study overcomes the limitations of single-source data approaches, excelling particularly under complex environmental conditions. By integrating IoT environmental data with SPAD values, our MDF-ADRS model effectively captures critical growth factors, delivering more accurate and robust LAI estimates. The model employs a weight optimization strategy to seamlessly integrate diverse data inputs, mitigating the biases of individual data sources. The application of the adjusted R^2 further enhances the model's generalization capability, ensuring high predictive accuracy across different environments. Compared to traditional methods, our fusion strategy effectively resolves issues of data redundancy and conflict, with a clear advantage demonstrated in its results.

4.1. Limitations and Future Directions

Despite the significant progress made in this study by integrating multi-source data, certain limitations remain. Firstly, the model's effectiveness relies heavily on the extensive coverage and reliability of IoT sensor networks. If the sensors' data are inaccurate, the model's performance could be compromised. Secondly, the model's dependence on multi-source data faces additional limitations, especially under rapidly changing environmental conditions. Multi-source data typically originate from different sensor systems, leading to potential discrepancies in data collection timings and spatial resolution. In conditions of rapid environmental change, inconsistent data synchronization can result in input biases, affecting the accuracy of the model's predictions. Additionally, the quality of different data sources may vary, particularly when sensors age, maintenance is delayed, or environmental interference increases, potentially introducing noise, missing values, or errors that could impact model stability. Future research will consider incorporating more advanced data preprocessing techniques, along with data quality monitoring and anomaly detection mechanisms, to reduce the interference from data noise. At the same time, we plan to use methods like transfer learning to enhance the model's adaptability, ensuring it maintains its high performance across different environments.

Moreover, the model's applicability across different plum varieties is another important issue. This study focused on the key growth stages of the Luhe plum variety, which has a short growth cycle and thrives in mild, humid climates. Different plum varieties exhibit variations in their canopy structure, leaf morphology, and environmental adaptability, making it difficult for a single model to be generalized to all plum types. Future research should incorporate data from multiple plum varieties during model training, or design a modular model that adjusts predictions based on specific varietal characteristics, to improve the breadth of the model's applicability.

Lastly, although this study did not observe significant overfitting or underfitting issues during model development, these risks should not be overlooked. As the model integrates a greater variety of data sources, its structure will become more complex. Therefore, in future studies, we plan to use cross-validation, regularization techniques, and data augmentation methods to mitigate the risk of overfitting. Additionally, we will explore advanced algorithms such as deep learning or ensemble learning to better address underfitting and improve the model's adaptability to complex data.

In the validation phase, especially when using control-group data, challenges and biases emerged, such as slight environmental differences between the control and experimental trees, relatively smaller sample sizes for the control data, and physiological variations between individual trees—all of which could introduce bias into the results. To mitigate these effects, we preprocessed the data (e.g., through standardization) and carefully analyzed individual trees' physiological differences. In future studies, expanding the sample size and incorporating additional multi-scenario data will further reduce validation biases and enhance the model's generalization across diverse environments and crop conditions.

4.2. The Application and Challenges of Models in Agricultural Practice

The proposed MDF–ADRS model demonstrates high accuracy and reliability in leaf area index (LAI) prediction. However, to effectively integrate it into routine agricultural practices, several measures are required to ensure the model's applicability and usability in real–world settings. The key steps for model deployment and the associated potential challenges are as follows:

- (1) The initial steps involve setting up the necessary infrastructure and deploying data collection systems within agricultural production areas, including systems for gathering environmental, SPAD, and LAI data. This will ensure that the MDF–ADRS model receives high–quality, real–time data support.
- (2) After data collection, a data management and processing platform is needed. This platform will clean, calibrate, and standardize the collected data, ensuring consistency and accuracy.
- (3) Farmers or agricultural technicians will undergo training to learn how to operate the model, interpret its outputs, and translate the model's recommendations into practical agricultural management decisions.

During this transition, farmers may face challenges related to equipment investments, technical training, and data management. To support a smooth integration, we plan to provide technical support, cost–control solutions, and user–friendly platform development to help farmers overcome these challenges.

The MDF–ADRS model has broad potential beyond LAI inversion estimations for green plum trees and can be applied to other crops and agricultural settings. By efficiently integrating multi–source data with a weight optimization strategy, it will be suitable for various crop management needs. For field crops like rice and wheat, which are heavily influenced by environmental factors and SPAD values, the model can accurately monitor growth and optimize irrigation and fertilization, improving yield and efficiency. For fruit crops like grapes and apples, where the LAI is key to assessing canopy health and predicting yields, the model can precisely evaluate tree health by combining environmental data with SPAD values. This is especially beneficial for high–value crops, where precision management boosts yield and quality.

In conclusion, the MDF–ADRS model represents a significant breakthrough in non–destructive LAI estimation for green plums, offering valuable tools for precision agriculture and improving monitoring and management practices across large–scale farming operations. Its successful application exemplifies the profound impact of integrating multi–source data on the accuracy and efficiency of agricultural monitoring, setting a new standard for the industry.

5. Conclusions

This study proposed a multi–source data fusion model for leaf area index (LAI) inversion, significantly improving the prediction accuracy of the LAI through the integration of environmental and SPAD data for plum trees. The main findings of the study are as follows:

- (1) In the analysis of key environmental factors, the six critical factors most correlated with plum tree LAIs were selected by comprehensively evaluating linear relationships using statistical methods, including Pearson, Spearman, and Kendall correlation analyses. These factors include environmental temperature (ET), soil temperature (Soil–T), environmental humidity (EH), soil moisture (Soil–M), carbon dioxide concentration (CO₂), and light intensity (LI). These factors significantly influence the LAI at different growth stages and are fundamental to constructing high–precision models.
- (2) During model construction, two separate models were developed: the Environmental Features–based Model (EFs–PM) and the SPAD–based Model (SPAD–PM). The EFs–PM model used the selected key environmental factors and the PLSR algorithm to capture macro–environmental effects on crop growth. The EFs–PM model achieved an R^2 of 0.69 and a root mean square error (RMSE) of 0.55 on the validation

set. The SPAD–PM model, on the other hand, used SPAD data and PLSR to handle the complex relationship between SPAD and the LAI, achieving an R^2 of 0.62 and an RMSE of 0.61 on the validation set.

- (3) This study introduced a multi–source decision fusion model based on the adjusted R^2 (MDF–ADRS), which effectively combined the strengths of both the EFs–PM and SPAD–PM models. By incorporating a weight optimization allocation method, the model flexibly adjusts the weights of environmental features and SPAD data for LAI prediction based on different environmental conditions, significantly improving its accuracy and robustness. The MDF–ADRS model achieved an R^2 of 0.88 and an RMSE of 0.39 on the validation set, demonstrating its superior performance under complex environmental conditions compared to that of single–source models.

In future research, adding multispectral or hyperspectral image data is expected to significantly enhance the prediction accuracy and applicability of LAI inversion models. Specifically, key spectral features related to the LAI, such as vegetation indices and reflectance information from specific bands, can be extracted from multispectral or hyperspectral data. By proposing novel fusion techniques that can integrate these different dimensions of data, future research will enable more precise agricultural management and improve real–time crop growth monitoring through dynamic model adjustments. However, challenges such as data processing, equipment costs, and practical implementation need to be addressed in real–world applications. With proper technology choices and effective training support, this direction promises to provide stronger technical tools and management solutions for modern agriculture.

Author Contributions: Conceptualization, C.Y. and H.T.; Methodology, C.Y., H.T., and D.H.; Experiment–conception and design, C.Y., H.T., D.H., and J.L.; Experiment–execution, C.Y., H.T., D.H., J.L., and J.H.; Writing–original draft preparation, D.H., J.L., J.H., D.Z., and J.Z.; Writing–review and editing, H.T., J.H., D.Z., and J.Z.; Funding acquisition, C.Y. and J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by: Open Project Program of Guangdong Provincial Key Laboratory of Agricultural Artificial Intelligence (GDKL-AAI-2023007); Open Fund of the State Key Laboratory of Agricultural Equipment Technology, South China Agricultural University (SKLAET-202412); The National Natural Science Foundation of China (42061046); The Special Projects in Key Fields of Ordinary Universities in Guangdong Province (2021ZDZX4111); The Science and Technology Planning Project of Guangdong Province (2021B1212040009).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data used in this study are available from the corresponding author upon reasonable request.

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

References

1. Gong, X.P.; Tang, Y.; Song, Y.Y.; Du, G.; Li, J. Comprehensive review of phytochemical constituents, pharmacological properties, and clinical applications of *Prunus mume*. *Front. Pharmacol.* **2021**, *12*, 679378. [[CrossRef](#)] [[PubMed](#)]
2. Wang, J.; Wang, J.; Wu, C.; Huang, J.; Zhou, R.; Jin, Y. Characteristics and valorization potential of fermentation waste of greengage (*Prunus mume*). *Appl. Sci.* **2021**, *11*, 8296. [[CrossRef](#)]
3. Malavolta, M.; Mocchegiani, E. (Eds.). *Trace Elements and Minerals in Health and Longevity*; Springer International Publishing: Cham, Switzerland, 2018; Volume 8.
4. Shen, L.; Wang, H.; Liu, Y.; Liu, Y.; Zhang, X.; Fei, Y. Prediction of soluble solids content in green plum by using a sparse autoencoder. *Appl. Sci.* **2020**, *10*, 3769. [[CrossRef](#)]
5. Xu, C. Greengage mau: Green branches full of fruit, growers' income makes you happy. *Shanwei J.* **2023**, *3*,. [[CrossRef](#)]
6. Yamaguchi, H.; Yasutake, D.; Hirota, T.; Nomura, K. Nondestructive Measurement Method of Leaf Area Index Using Near-infrared Radiation and Photosynthetically Active Radiation Transmitted through a Leafy Vegetable Canopy. *HortScience* **2022**, *58*, 16–22. [[CrossRef](#)]
7. Weraduwege, S.M.; Chen, J.; Anozie, F.C.; Morales, A.; Weise, S.E.; Sharkey, T.D. The relationship between leaf area growth and biomass accumulation in *Arabidopsis thaliana*. *Front. Plant Sci.* **2015**, *6*, 167. [[CrossRef](#)]

8. Wu, J.; Wang, J.; Hui, W.; Zhao, F.; Wang, P.; Su, C.; Gong, W. Physiology of plant responses to water stress and related genes: A review. *Forests* **2022**, *13*, 324. [[CrossRef](#)]
9. Son, N.T.; Chen, C.F.; Chen, C.R.; Chang, L.Y.; Duc, H.N.; Nguyen, L.D. Prediction of rice crop yield using MODIS EVI–LAI data in the Mekong Delta, Vietnam. *Int. J. Remote Sens.* **2013**, *34*, 7275–7292. [[CrossRef](#)]
10. Zhuo, W.; Fang, S.; Gao, X.; Wang, L.; Wu, D.; Fu, S.; Wu, Q.; Huang, J. Crop yield prediction using MODIS LAI, TIGGE weather forecasts and WOFOST model: A case study for winter wheat in Hebei, China during 2009–2013. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *106*, 102668. [[CrossRef](#)]
11. Yamanura, R.M.K.; Patil, B. NDVI derived LAI model: A novel tool for crop monitoring. *J. Crop Weed.* **2021**, *17*, 49–54. [[CrossRef](#)]
12. Pokovai, K.; Fodor, N. Adjusting ceptometer data to improve leaf area index measurements. *Agronomy* **2019**, *9*, 866. [[CrossRef](#)]
13. Mudi, S.; Paramanik, S.; Behera, M.D.; Prakash, A.J.; Deep, N.R.; Kale, M.P.; Kumar, S.; Sharma, N.; Pradhan, P.; Chavan, M.; et al. Moderate resolution LAI prediction using Sentinel–2 satellite data and indirect field measurements in Sikkim Himalaya. *Environ. Monit. Assess.* **2022**, *194*, 897. [[CrossRef](#)] [[PubMed](#)]
14. Jia, D.; Lei, S.; Numata, I.; Tian, L. Accuracy assessment and impact factor analysis of GEDI leaf area index product in temperate forest. *Remote Sens.* **2023**, *15*, 1535. [[CrossRef](#)]
15. Frazer, G.W.; Trofymow, J.A.; Lertzman, K.P. *A Method for Estimating Canopy Openness, Effective Leaf Area Index, and Photosynthetically Active Photon Flux Density Using Hemispherical Photography and Computerized Image Analysis Techniques*; Pacific Forestry Centre: Victoria, BC, Canada, 1997; Volume 373.
16. Zhang, J.; Wang, J.; Khromykh, V.; Li, J. Global leaf area index research over the past 75 years: A comprehensive review and bibliometric analysis. *Sustainability* **2023**, *15*, 3072. [[CrossRef](#)]
17. Lee, H.; Cho, H. Chlorophyll fluorescence imaging for environmental stress diagnosis in crops. *Sensors* **2024**, *24*, 1442. [[CrossRef](#)]
18. Li, W.; Pan, K.; Liu, W.; Xiao, W.; Ni, S.; Shi, P.; Chen, X. Monitoring maize canopy chlorophyll content throughout the growth stages based on UAV MS and RGB feature fusion. *Agriculture* **2024**, *14*, 1265. [[CrossRef](#)]
19. Haboudane, D.; Miller, J.R.; Tremblay, N.; Zarco-Tejada, P.J.; Dextraze, L. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sens. Environ.* **2002**, *81*, 416–426. [[CrossRef](#)]
20. Dordas, C.A.; Sioulas, C. Safflower yield, chlorophyll content, photosynthesis, and water use efficiency response to nitrogen fertilization under rainfed conditions. *Ind. Crop. Prod.* **2008**, *27*, 75–85. [[CrossRef](#)]
21. Feng, Y.; Sui, X.; Tang, J.; Liu, R.; Ling, X.; Liang, W.; Wei, X. Responses of belowground fine root biomass and morphology in *Robinia pseudoacacia* L. plantations to aboveground environmental factors. *Glob. Ecol. Conserv.* **2024**, *50*, e02863. [[CrossRef](#)]
22. Tao, T.Y.; He, F.; Fan, H.Y. Research progress and comprehensive control of litchi downy mildew. *Chin. Trop. Agric.* **2008**, *2008*, 51–52.
23. Xu, D.; Deng, Y.; Xi, P.; Zhu, Z.; Kong, X.; Wan, L.; Situ, J.; Li, M.; Gao, L.; Jiang, Z. Biological activity of pterostilbene against *Peronophythora litchii*, the litchi downy blight pathogen. *Postharvest Biol. Technol.* **2018**, *144*, 29–35. [[CrossRef](#)]
24. Kinane, S.M.; Montes, C.R.; Zapata, M.; Bullock, B.P.; Cook, R.L.; Mishra, D.R. Influence of environmental variables on leaf area index in loblolly pine plantations. *For. Ecol. Manag.* **2022**, *523*, 120445. [[CrossRef](#)]
25. Karimi, S.; Nazemi, A.H.; Sadraddini, A.A.; Xu, T.R.; Bateni, S.M.; Fard, A.F. Estimation of forest leaf area index using meteorological data: Assessment of heuristic models. *J. Environ. Inform.* **2020**, *36*, 119–132. [[CrossRef](#)]
26. Gao, F.; Zhang, X. Mapping crop phenology in near real–time using satellite remote sensing: Challenges and opportunities. *J. Remote Sens.* **2021**, *2021*, 8379391. [[CrossRef](#)]
27. Zhang, D.; Han, X.; Lin, F.; Du, S.; Zhang, G.; Hong, Q. Estimation of winter wheat LAI based on multi–source UAV image feature fusion. *Trans. Chin. Soc. Agric. Eng.* **2022**, *38*, 171–179.
28. Deng, S.Q.; Zhao, Y.; Bai, X.Y.; Li, X.; Sun, Z.D.; Liang, J.; Sun, Z.H.; Cheng, S. Inversion of chlorophyll and leaf area index of winter wheat based on UAV image segmentation. *Trans. Chin. Soc. Agric. Eng.* **2022**, *38*, 136–145.
29. Chen, R.; Feng, H.; Yang, F.; Li, C.; Yang, G.; Pei, H.; Pan, L.; Chen, P. Estimation of leaf area index of winter wheat based on UAV hyperspectral remote sensing. *Trans. Chin. Soc. Agric. Eng.* **2020**, *36*, 40–49.
30. Shao, M.; Nie, C.; Zhang, A.; Shi, L.; Zha, Y.; Xu, H.; Yang, H.; Yu, X.; Bai, Y.; Liu, S.; et al. Quantifying effect of maize tassels on LAI estimation based on multispectral imagery and machine learning methods. *Comput. Electron. Agric.* **2023**, *211*, 108029. [[CrossRef](#)]
31. Guo, Q.; Su, Y.; Hu, T.; Guan, H.; Jin, S.; Zhang, J.; Coops, N.C. Lidar boosts 3D ecological observations and modelings: A review and perspective. *IEEE Geosci. Remote Sens. Mag.* **2020**, *9*, 232–257. [[CrossRef](#)]
32. Garrido-Carretero, M.S.; Borque-Arancón, M.J.; Ruiz-Armenteros, A.M.; Moreno-Guerrero, R.; Gil-Cruz, A.J. Low-cost GNSS receiver in RTK positioning under the standard ISO-17123-8: A feasible option in geomatics. *Measurement* **2019**, *137*, 168–178. [[CrossRef](#)]
33. Cohen, I.; Huang, Y.; Chen, J.; Benesty, J.; Benesty, J. Pearson correlation coefficient. In *Noise Reduction in Speech Processing*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 1–4.
34. De Winter, J.C.F.; 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](#)] [[PubMed](#)]
35. Abdi, H. The Kendall rank correlation coefficient. In *Encyclopedia of Measurement and Statistics*; Sage Publications: Thousand Oaks, CA, USA, 2007; Volume 2, pp. 508–510.

36. Murtagh, F.; Heck, A. *Multivariate Data Analysis*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012; Volume 131.
37. Abdi, H.; Williams, L.J. Partial least squares methods: Partial least squares correlation and partial least square regression. In *Computational Toxicology: Volume II*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 549–579.
38. Awad, M.; Khanna, R.; Awad, M.; Khanna, R. Support vector regression. In *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*; Apress: New York, NY, USA, 2015; pp. 67–80.
39. Rigatti, S.J. Random forest. *J. Insur. Med.* **2017**, *47*, 31–39. [[CrossRef](#)] [[PubMed](#)]
40. Tranmer, M.; Elliot, M. Multiple linear regression. *Cathie Marsh Cent. Census Surv. Res. (CCSR)* **2008**, *5*, 1–57.
41. Williams, C.; Rasmussen, C. Gaussian processes for regression. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 1995; Volume 8.
42. Nakagawa, S.; Schielzeth, H. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods Ecol. Evol.* **2013**, *4*, 133–142. [[CrossRef](#)]
43. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE). *Geosci. Model Dev. Discuss.* **2014**, *7*, 1525–1534.
44. Ma, J.; Chen, P.; Wang, L. A comparison of different data fusion strategies' effects on maize leaf area index prediction using multisource data from unmanned aerial vehicles (UAVs). *Drones* **2023**, *7*, 605. [[CrossRef](#)]
45. Zhao, J.; Pan, F.; Xiao, X.; Hu, L.; Wang, X.; Yan, Y.; Lan, Y. Summer maize growth estimation based on near-surface multi-source data. *Agronomy* **2023**, *13*, 532. [[CrossRef](#)]
46. Akossou, A.Y.J.; Palm, R. Impact of data structure on the estimators R-square and adjusted R-square in linear regression. *Int. J. Math. Comput.* **2013**, *20*, 84–93.
47. Su, W.; Wu, J.-Y.; Wang, X.-S.; Xie, Z.-X.; Zhang, Y.; Tao, W.-C.; Jian, T. Retrieving corn canopy leaf area index based on Sentinel-2 image and PROSAIL model parameter calibration. *Spectrosc. Spectr. Anal.* **2021**, *41*, 1891–1897.
48. Shi, Z.; Shi, S.; Gong, W.; Xu, L.; Wang, B.; Sun, J.; Chen, B.; Xu, Q. LAI estimation based on physical model combining airborne LiDAR waveform and Sentinel-2 imagery. *Front. Plant Sci.* **2023**, *14*, 1237988. [[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.