Computers and Electronics in Agriculture 214 (2023) 108306

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

CatBoost algorithm for estimating maize above-ground biomass using unmanned aerial vehicle-based multi-source sensor data and SPAD values

Weiguang Zhai [a,b], Changchun Li [b], Shuaipeng Fei [c], Yanghua Liu [d], Fan Ding [a,b], Qian Cheng [a], Zhen Chen [a,*]

[a] Institute of Farmland Irrigation, Chinese Academy of Agricultural Sciences, Xinxiang 453002, China
[b] School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454003, China
[c] College of Land Science and Technology, China Agricultural University, Beijing 100193, China
[d] Piesat Information Technology Co., Ltd, Beijing 100095, China

ARTICLE INFO

Keywords:
Above-ground biomass
Maize
Unmanned aerial vehicle
SPAD values
CatBoost

ABSTRACT

The rapid and accurate estimation of maize above-ground biomass (AGB) is pivotal for precise agricultural management. The rapid evolution of unmanned aerial vehicles (UAVs) and sensor technology has introduced a novel method for obtaining AGB information. Nevertheless, individual sensors may lack comprehensive data, leading to reduced AGB estimation accuracy in certain scenarios. This study collected UAV multi-spectral (MS) and thermal infrared (TIR) data, alongside soil and plant analyzer development (SPAD) values, from maize across multiple growth stages (jointing, trumpet, and big trumpet) during 2022 and 2023. Diverse data fusion programs were devised to explore the potential of combining multi-source sensor data with SPAD values to estimate AGB. The efficacy of CatBoost was evaluated and benchmarked against Support Vector Regression (SVR) and Random Forest Regression (RFR) algorithms. For the entire growth, findings reveal that the fusion of multi-source sensor data (MS + TIR) can mitigate the data insufficiency in single-sensor estimations. The resulting R^2 values range from 0.608 to 0.817. Optimal estimation outcomes were achieved by the fusion of multi-source sensor data with SPAD values (MS + TIR + SPAD), yielding R^2 values ranging from 0.685 to 0.872. For a single growth stage, there are variations in the estimation accuracy across different growth stages. From the jointing stage to the big trumpet stage, the estimation accuracy consistently increases, with the highest accuracy observed during the big trumpet stage, with R^2 ranging from 0.721 to 0.901. Additionally, in alignment with the results for the entire growth stage, the fusion of multi-source sensor data with SPAD values still yields the highest estimation accuracy during different growth stages. In a comparison of different machine learning algorithms, for both the entire growth stage and single growth stages, SVR, RFR, and CatBoost achieved R^2 values ranging from 0.305 to 0.824, 0.368 to 0.881, and 0.451 to 0.901, respectively. Notably, the CatBoost algorithm exhibited heightened estimation accuracy. The fusion of multi-source sensor data with SPAD values combined with the CatBoost algorithm results in accurate and reliable maize AGB estimation accuracy. This high-throughput approach to crop phenotyping is characterized by speed and accuracy and serves as a valuable reference for rapidly acquiring AGB information in this geographical region.

1. Introduction

Above-ground biomass (AGB) is the organic matter accumulated by a crop at a certain time and per unit area. It is an important indicator for crop growth monitoring, nutrient diagnosis and yield estimation (Devia et al., 2019). Timely access to AGB information is of far-reaching significance for agricultural management. The traditional AGB acquisition method is mainly manual measurement, which has high accuracy when applied to small areas. For large areas, the manual method not only requires a lot of human and material resources but also has poor timeliness and is easily affected by human factors. Remote sensing technology has the advantages of low cost and high timeliness, providing the possibility of quickly and accurately obtaining AGB information over large areas. At present, satellite remote sensing technology has the advantages of a wide range and simple image acquisition method, which has incomparable advantages for the acquisition of large-scale AGB

* Corresponding author.
E-mail address: chenzhen@caas.cn (Z. Chen).

https://doi.org/10.1016/j.compag.2023.108306
Received 6 November 2022; Received in revised form 22 August 2023; Accepted 4 October 2023

information. However, satellite remote sensing will be restricted by weather conditions, and there are problems such as low image resolution and long revisit cycles, which is not ideal for the practical effect of precision agriculture (Zhai et al., 2023a). Unmanned aerial vehicle (UAV) remote sensing technology, with its advantages of simple operation, repeatable observation, and access to high spatial and temporal resolution remote sensing images (Jin et al., 2020b). It has shown great advantages in the fine management and regulation of agriculture.

The UAV-based platform mounted sensors showed high accuracy in the estimation of crop phenotypic parameters. There are differences in the data information obtained by different sensors. For example, multispectral (MS) sensors quantitatively estimate crop growth parameters through the interaction of different wavelengths of light with the plant in terms of reflection, absorption, and transmission (Houborg and McCabe, 2016). Thermal infrared (TIR) sensors are unaffected by light and monitor crop growth primarily by measuring crop canopy temperature and the crop response to water stress (Maes and Steppe, 2019). In the estimation of crop phenotypic parameters researches, a single sensor obtains less information about the crop and has a lower estimation accuracy, while the fusion of multi-source sensor data can reflect crop growth information from different aspects, which can effectively improve the estimation accuracy. For example, in the study of winter wheat yield estimation, the fusion of UAV MS, TIR and RGB data has significantly improved the estimation effect compared to using only a single sensor (Fei et al., 2022). The fusion of hyperspectral data with RGB data in the potato AGB estimation study was superior to using hyperspectral data only (Li et al., 2020a). In addition, incorporating auxiliary information such as crop height, crop nitrogen content and soil factor on the basis of UAV remote sensing data can further improve the estimation accuracy of crop phenotypic parameters. Incorporating crop height information based on hyperspectral data further improves the accuracy of wheat AGB estimation (Yue et al., 2017). In addition to remote sensing data, some biochemical indicators such as chlorophyll content are also good indicators of crop growth. Soil and plant analyzer development (SPAD) values are closely related to chlorophyll content and can be used as effective parameters for crop phenotype estimation (Ling et al., 2011). To our knowledge, there are no studies that fusion of UAV multi-source sensor data with SPAD values for maize AGB estimation, which will be explored in this study.

However, significant variations in physiological and morphological characteristics occur across different growth stages of crops (Qiao et al., 2022b). These variations pose intricate implications for the estimation of crop phenotypic parameters using remote sensing technology. Each growth stage is accompanied by distinct vegetation structures, canopy densities, and optical properties, which may impact the interpretation and analysis of remote sensing data. For instance, during the initial growth stage of crops, the small and young vegetation might lead to confusion with the background, resulting in errors in remote sensing data. On the other hand, in the mature growth stage of crops, dense vegetation can interfere with remote sensing signals through light transmission and reflection, consequently affecting the accuracy of estimation results (Qiao et al., 2022a). Hence, a thorough investigation into the influence of different growth stages on remote sensing technology for estimating crop phenotypic parameters is crucial. Gaining insights into the reliability, accuracy, and applicability of remote sensing data at different growth stages enhances our ability to utilize remote sensing techniques effectively for monitoring and assessing the growth status of crops.

In recent years, with the rapid development of computer science and artificial intelligence, deep learning models such as Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN) have made significant strides in estimating crop phenotypic parameters (Jin et al., 2020a). As robust deep learning models, ANN and CNN have the ability to learn intricate feature representations from large-scale data and automatically extract patterns relevant to crop phenotypic parameters from images, remote sensing data, or other pertinent data (Yu et al., 2023). However, deep learning models may face challenges in certain scenarios, including high data requirements, substantial computational resource demands, and intricate parameter tuning. To address these issues, machine learning algorithms have demonstrated advantages in crop phenotyping research, including reduced reliance on training time and computational resources. For instance, in constructing crop phenotyping estimation models, individual learning models like Ridge Regression, Least Absolute Shrinkage and Selection Operator, and Support Vector Regression (SVR) have exhibited high predictive accuracy and robustness (Zhai et al., 2023b). Nonetheless, existing research primarily employs single learning model approaches, which are constrained in performance, demanding of substantial datasets, and exhibit poor generalization on small or fluctuating datasets. To enhance the overall modeling performance, ensemble learning is introduced as a machine learning method. Ensemble learning involves training multiple base learners and combining them using specific strategies to yield a more comprehensive strong learner (Feng et al., 2020). Currently, ensemble learning mainly comprises the Bagging framework and the Boosting framework. Among them, Random Forest Regression (RFR) based on the Bagging framework and CatBoost based on the Boosting framework are two commonly used ensemble learning methods. RFR has been successfully applied in crop phenotyping research, yielding excellent estimation accuracy (Zha et al., 2020). On the other hand, the CatBoost algorithm excels in handling categorical features, adaptive feature scaling, missing value treatment, and robustness, thereby enhancing estimation accuracy and reducing implementation complexity. Despite being a novel ensemble learning algorithm with demonstrated utility in hydrological studies (Huang et al., 2019) and structural engineering research (Lee et al., 2021), its application in crop phenotyping research remains relatively limited. However, as a potent gradient boosting framework, CatBoost has not been fully explored in AGB estimation.

In summary, the primary objectives of this study are as follows: (1) To assess the viability of fusing UAV multi-source sensor data with SPAD values for estimating maize AGB. (2) To analyze the influence of different growth stages on the accuracy of maize AGB estimation. (3) To construct a maize AGB estimation model using CatBoost and compare its performance against SVR and RFR models.

2. Materials and methods

2.1. Study area and experimental design

The study was conducted over a two-year period in 2022 and 2023 at the Xinxiang Comprehensive Base of the Chinese Academy of Agricultural Sciences, Xinxiang County, Henan Province, China (113°45′42 ”E, 35°08′05 ”N, Fig. 1). In the 2022 experiment, ten different maize varieties were chosen and cultivated on June 15. Four distinct fertilizer treatments were employed, namely N0: 0 kg/hm^2, N1: 80 kg/hm^2, N2: 120 kg/hm^2, and N3: 160 kg/hm^2. Each fertilizer treatment encompassed three replications of every maize variety, thereby resulting in a total of 120 plots, each measuring 8 m^2 (2 m × 4 m). In the 2023 experiment, the maize cultivating took place on June 17th, and the experimental design remained consistent with that of 2022. The management practices applied in the field were optimized based on local conditions.

2.2. Data acquisition

2.2.1. UAV data acquisition and pre-processing

The DJM210 UAV (SZ DJI Technology Co., Shenzhen, China) was employed as the remote sensing platform, outfitted with both the MS sensors (RedEdge MX, MicaSense Inc., Seattle, USA) and the TIR sensors (ZENMUSE XT, SZ DJI Technology Co., Shenzhen, China) to facilitate the acquisition of multi-sensor data (Fig. 2). Comprehensive information regarding the two sensors is elucidated in Table 1.

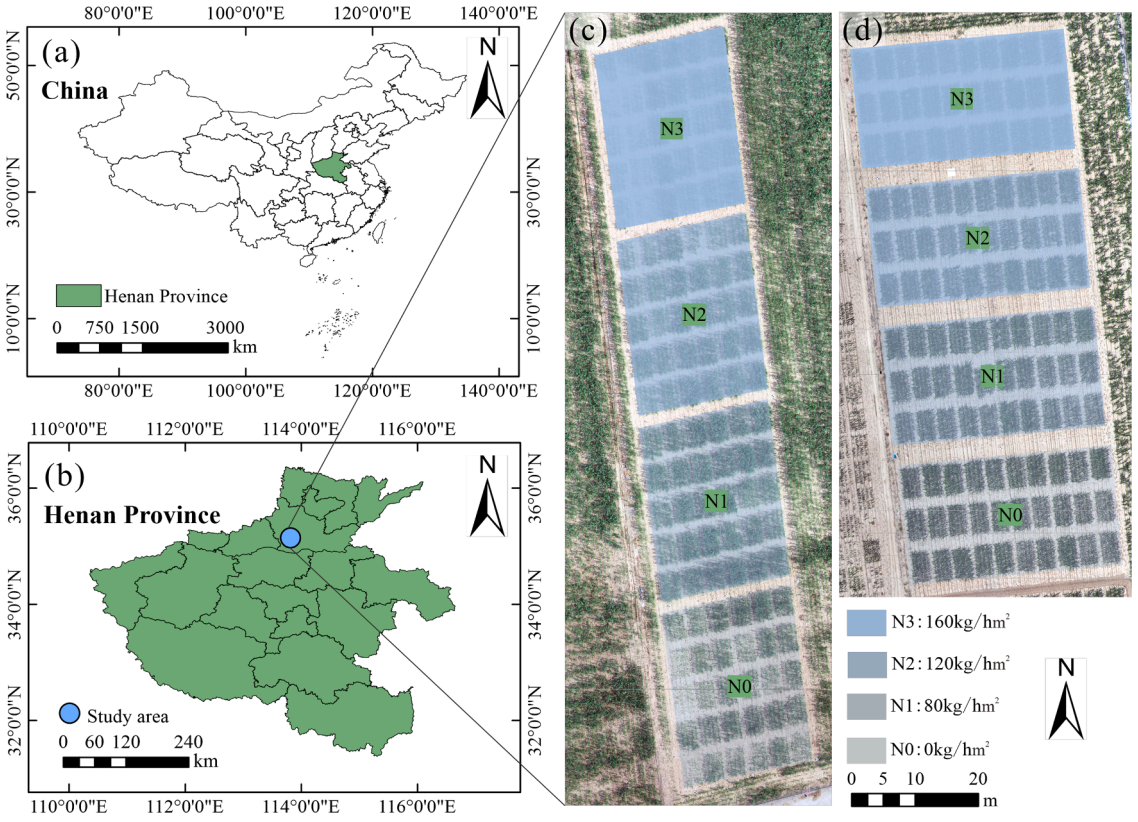

Fig. 1. The geographical location and experimental design of the study. (a) China's border. (b) Henan Province's border. (c) RGB image of the study area in 2022. (d) RGB image of the study area in 2023.

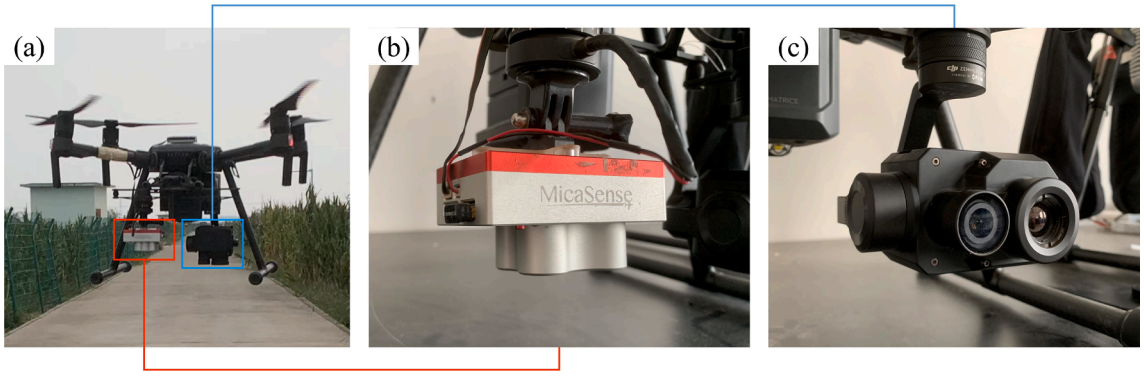

Fig. 2. The UAV and its equipped sensors. (a) DJI M210, (b) RedEdge MX, (c) ZENMUSE XT.

Table 1
RedEdge MX and ZENMUSE XT2 main parameters.

Attribute	MS sensor	TIR sensor
Weight	232 g	629 g
Size	87 × 59 × 45.4 mm	123.7 × 112.6 × 127.1 mm
Spectral band	Red, green, blue, red edge, near infrared	Thermal infrared
Band range	400 ~ 900 nm	7500 ~ 13500 nm
Image resolution	1280 × 960	640 × 512

The aerial missions were executed under optimal, cloud-free conditions between 10:00 and 13:00 BST on different dates. These dates include July 13th, 2022 (jointing stage), July 23rd, 2022 (trumpet stage), and August 2nd, 2022 (big trumpet stage), in addition to July 15th, 2023 (jointing stage), July 23rd, 2023 (trumpet stage), and August 2nd, 2023 (big trumpet stage). DJI GS PRO 2.0.17 software (SZ DJI Technology Co., Shenzhen, China) was used to plan the flight route in the study area. The flight altitude is set to 30 m. The forward and lateral overlaps were set to 85% and 80% respectively. Before and after each flight, calibration board data was collected for post-radiation correction. Stitching and radiometric correction of the images was carried out using Pix4D 4.4.12 software (Pix4D, Lausanne, Switzerland), with the following main steps: Firstly, images and ground control points (GCPs) are imported, and the spatial positioning of GCPs is employed to perform terrain correction on the images. Subsequently, the software conducts feature extraction and matching to establish correlations between images. Based on the matched feature points, a sparse point cloud model is generated to represent the surface structure of the area. Following this, a denser point cloud model is further generated to capture more detailed surface information. Using the dense point cloud, orthorectified images

are generated to eliminate terrain-induced distortions, followed by a color correction to ensure consistent presentation. Ultimately, the software produces orthorectified images in both MS and TIR for the study area, providing high-quality remote sensing data for accurate estimation of maize AGB in this study.

Due to the incomplete coverage of maize, it is necessary to separate the maize plants from the soil. In this study, the separation of plant and soil, along with the extraction of maize canopy features, was conducted using ArcMap 10.8 software (Environmental Systems Research Institute, Inc., Redlands, CA, USA). The soil-adjusted vegetation index was a robust metric for effectively discriminating between maize and soil. In ArcMap 10.8 software, a mask delineating the maize plants was generated from the grayscale image of the soil-adjusted vegetation index. This mask was then applied to the image to eliminate soil-related elements. Finally, seven vegetation indices were extracted from the MS images as canopy spectral features, and the normalized relative canopy temperature was extracted from the TIR images as a canopy thermal feature. The features extracted from both MS and TIR images are listed in Table 2.

2.2.2. SPAD values and AGB acquisition

SPAD values and AGB of maize in each plot were measured simultaneously on the day of the UAV aerial mission. The SPAD values were obtained as follows: Three maize plants were randomly selected from each plot. The upper, middle and lower parts of the top leaves of the plant canopy were measured using a SPAD-502 plus and the average value was taken as the canopy SPAD value of the plant. Then the average of the SPAD values of the three plants was calculated as the SPAD value of the maize canopy in the plot (Fig. 3).

AGB was measured as follows: Two uniformly growing maize plants were randomly selected from each plot as samples and dried in a blast oven until their mass was constant. Then the samples were weighed for dry weight Finally, the AGB per unit area of maize was calculated based on the sample dry weight and population density (Fig. 3).

When observing the SPAD values and AGB for the year 2022, a clear trend can be discerned: with the gradual increase in fertilizer levels, both the SPAD values and AGB exhibit a progressive rise. However, a notable exception is observed in the case of the N3 treatment, corresponding to the highest fertilizer application level, where SPAD values and AGB display a significant decrease. This phenomenon may be the result of the N3 treatment, the plot was previously used for cultivating apple trees until 2022. Apples, being high-nutrient-demanding crops, caused substantial nutrient depletion in the soil, especially without proper nutrient replenishment before maize cultivation. Consequently, during maize cultivation, the soil might have been relatively deficient in nutrients, thereby affecting maize growth and leading to a distinct decrease in both SPAD values and AGB. In the 2023 experiment, utilizing plots with continuous crop cultivation, a trend emerges where increasing fertilizer levels correspond to higher SPAD values and AGB.

Table 2
Two different types of features selected for this study.

Data type	Features	Formula	
MS	Normalized difference vegetation index	(NIR-R)/(NIR + R)	(Tucker 1979)
	Green normalized difference vegetation index	(NIR-G)/(NIR + G)	(Gitelson et al., 1996)
	Red-edge normalized difference vegetation index	(EDGE-R)/(EDGE + R)	(Gitelson et al., 2003)
	Soil adjusted vegetation index	(1 + L)*(NIR-R)/(NIR + R + L)(L = 0.5)	(Huete 1988)
	Enhanced vegetation index	2.5*(NIR-R)/(NIR + 6*R-7.5*B + 1)	(Liu and Huete 1995)
	Ratio vegetation index	NIR/R	(Pearson and Miller 1972)
	Triangle vegetation index	60*(NIR-G)-100(R-G)	(Broge and Leblanc 2001)
TIR	Normalised relative canopy temperature	$(T_{canopy}-T_{min})/(T_{max}-T_{min})$	(Maimaitijiang et al., 2020)

R: red band reflectivity, *G*:green band reflectivity, *B*: blue band reflectivity, *EDGE*: red edge band reflectivity, *NIR*: near infrared band reflectivity, T_{canopy}: mean canopy temperature in that maize plot, T_{max}: maximum canopy temperature measured in all maize plot, T_{min}: minimum canopy temperature measured in all maize plot.

2.3. Regression techniques

2.3.1. SVR

SVR has shown high prediction accuracy and strong stability in crop phenotype studies and is one of the commonly used machine learning algorithms (Bian et al., 2022). The basic principle is to train using a loss function to fit sample data by constructing an optimal decision hyperplane that minimizes the distance between the sample and the hyperplane. In order to achieve optimal performance, machine learning algorithms require parameter adjustments to enhance their effectiveness. In this study, parameter tuning for SVR encompasses the selection of kernel functions, which include linear, polynomial (poly), and radial basis functions (rbf). Determining an appropriate kernel function entails considering the characteristics of the data and the nature of linear and nonlinear relationships. The C parameter, serving as a regularization parameter, is utilized to balance model complexity and tolerance. In this study, the range of C values for adjustment is set from 0.1 to 1, with a step size of 0.01. On the other hand, the gamma parameter is employed to adjust the width of the rbf kernel; a smaller gamma value widens the basis function, while a larger gamma value narrows it. For this study, the range of gamma adjustment is set between 0.1 and 1, with a step size of 0.01. For all machine learning algorithms employed in this study, a grid search approach is utilized to explore various parameter combinations and select the optimal parameter configuration that leads to the best performance.

2.3.2. RFR

RFR is a typical representative algorithm in ensemble learning using bagging as a framework (Breiman 2001). It is characterized by the ability to randomly sample the samples and features in the training set, which can reduce the occurrence of overfitting situations. RFR consists of a combination of multiple decision trees, each tree is randomly sampled and generates corresponding predicted values during the model training process. Eventually, the regression of RFR is completed by integrating the predicted values of all decision trees. RFR has obvious advantages when dealing with large samples and high dimensional data. Parameter adjustments for RFR include n_estimators (the number of decision trees) and max_depth (the maximum depth of decision trees). Increasing n_estimators enhances model diversity, thus increasing robustness and the ability to capture complex relationships within the data. However, an excessive number of decision trees might lead to increased training time and memory consumption, necessitating a balance between efficiency and performance. In this study, the adjustment range for n_estimators is set from 50 to 1000, with a step size of 10. The depth of decision trees determines the model's grasp of data features. Overly deep decision trees can result in overfitting, making the model overly sensitive to training data and leading to poor performance on new data. Constraining max_depth prevents the model from excessively learning noise and details in the training data, thereby enhancing its generalization capability. However, setting max_depth too small might render the model overly simplistic, unable to capture the intricate relationships within the data, resulting in underfitting. In this study, the adjustment range for max_depth is set from 3 to 30, with a step size of 1.

2.3.3. CatBoost

CatBoost was first proposed by Yandex in 2017 and is implemented

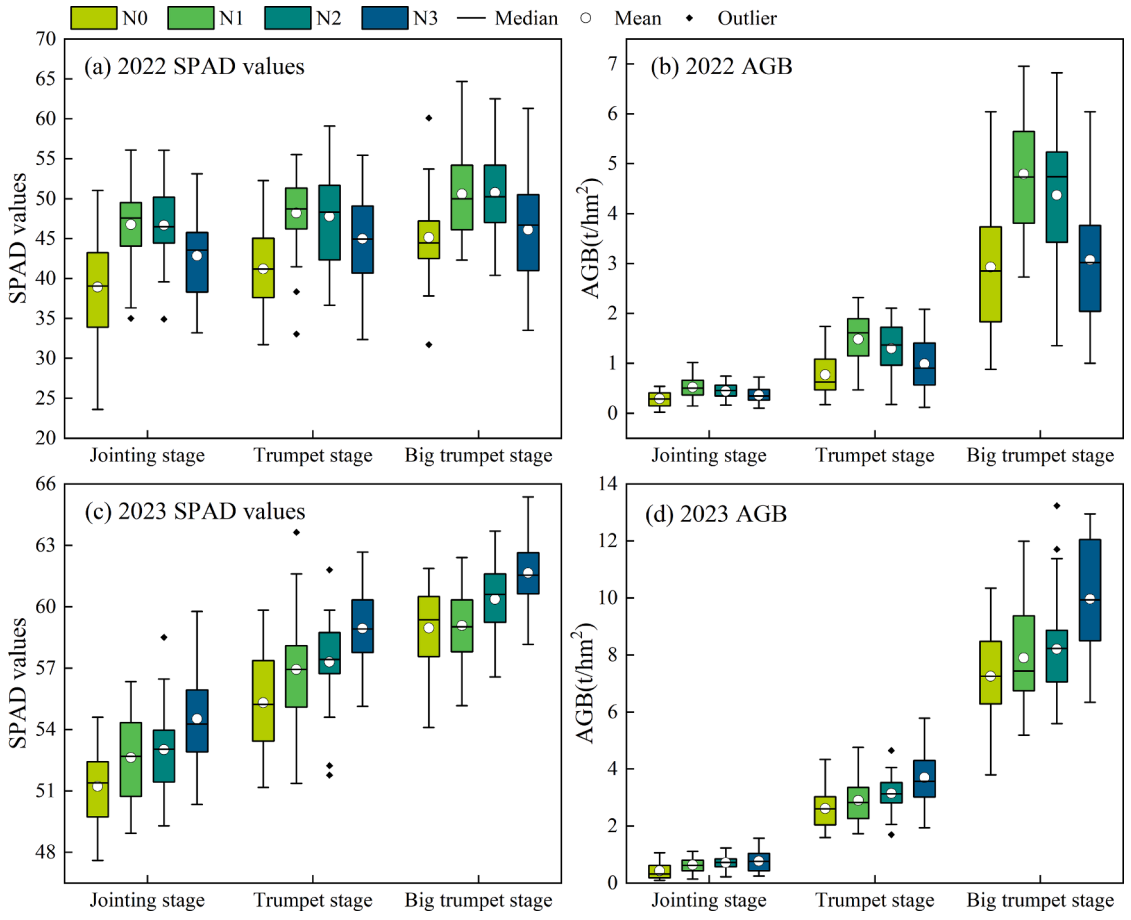

Fig. 3. Maize SPAD values and AGB statistics for four fertilizer treatments at the jointing stage, trumpet stage, and big trumpet stage. (a) 2022 SPAD values, (b) 2022 AGB, (c) 2023 SPAD values, (d) 2023 AGB.

based on the boosting framework (Prokhorenkova et al., 2018), where boosting is implemented as follows (Fig. 4): (1). All base learners use the same training samples and assign the same weights to each training sample. (2). Train the current base learner, and assign higher weights to the samples with larger prediction deviations in the next base learner based on the prediction results of the current base learner. (3). Iterate step 2 until all base learners have been trained. (4). Finally, the final predictions are obtained by weighting the predictions of each base learner. Parameter tuning for CatBoost in this study follows the same approach as that for RFR.

CatBoost is an improved gradient boosting decision tree algorithm. In each iteration of traditional gradient boosting decision tree, the current model gradient is obtained based on the same dataset and the base learner is trained based on that gradient. It leads to biased point-by-point gradient estimation, resulting in overfitting of the final strong learner model. CatBoost uses ordered boosting to optimize the gradient estimation method of gradient boosting decision tree. The ordered boosting method first generates a random permutation σ of [1, n] to sort the original samples and initialize n different models M_1, M_2, ……., M_n, where each M_i is a model trained with only the first i samples in the random permutation. During each iteration step, the unbiased gradient estimate for the jth sample is obtained by model M_{j-1}. The pseudocode is shown in Table 3 (Huang et al., 2019). This approach mitigates the effects of traditional gradient boosting decision tree gradient estimation bias and the model has a higher generalization capability. In addition, unlike the base learner in RFR, CatBoost uses a oblivious tree as the base

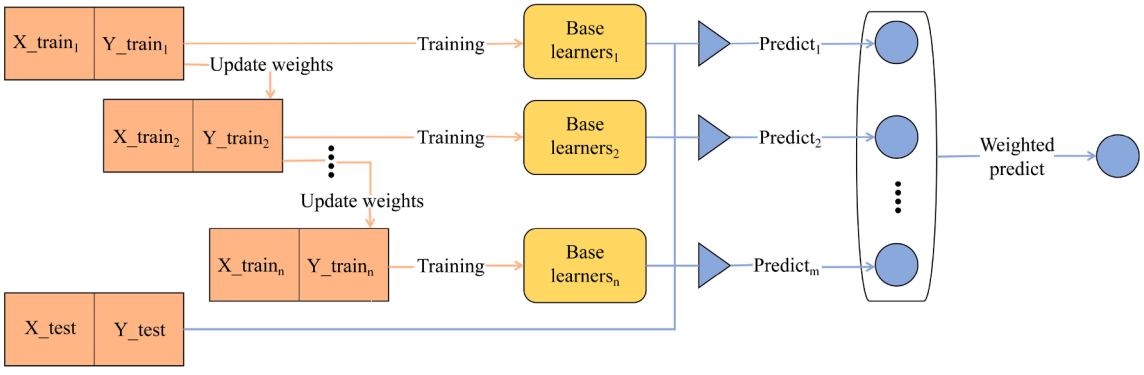

Fig. 4. Boosting implementation process.

Table 3
Ordered boosting pseudocode.

Algorithm: ordered boosting

Algorithm: ordered boosting

input:$\{(x_k, y_k)\}_{k=1}^{n}$ ordered according to σ, the number of trees I;
$\sigma \leftarrow$ randompermutation of $[1,n]$;
$M_i \leftarrow 0$ for $i = 1,\cdots,\boldsymbol{n}$;
for $t \leftarrow 1$ to I do
  for i $\leftarrow 1$ to I do
    $r_i \leftarrow y_i - M_{\sigma(i)-1}(x_i)$;
  end for
  for $i \leftarrow 1$ to n do
    $\triangle \boldsymbol{M} \leftarrow \boldsymbol{L}\text{earn}\boldsymbol{M}\text{odel}((x_j, r_j) : \sigma(j) \le i)$;
    $M_i \leftarrow M_i + \triangle \boldsymbol{M}$;
  end for
end for
output:M_n

learner, which can increase the model reliability and speed up the computation compared to the normal decision tree in RFR (Tang et al., 2021).

2.4. Modeling evaluation

To begin with, the original data is divided into training and testing sets in a 4:1 ratio (Table 4). Subsequently, a four-fold cross-validation and grid search method are applied to the training set to identify optimal model parameters. The fundamental concept of cross-validation involves further partitioning the training set into four subsets, cyclically utilizing three subsets as training data while the remaining one serves as validation data. For each training subset, different model parameter combinations are attempted using grid search, and the performance of each combination is evaluated using the validation data to determine the best parameter configuration. Once parameter tuning is complete, the entire training set is employed for model training. Finally, an independent testing set is used to predict outcomes with the trained model (Yu et al., 2023). Performance metrics such as coefficient of determination (R^2) and relative root mean square error (rRMSE) are computed on the testing set to assess the accuracy of the model in estimating AGB. The R^2 and rRMSE were calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - \bar{y})^2}$$

$$\text{rRMSE} = \frac{\sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}}}{\bar{y}} \times 100\%$$

where x_i is the measured maize AGB, y_i is the estimated maize AGB, $\bar{y}$ is the mean of the measured maize AGB and n is the sample size of the test set.

3. Results

3.1. AGB estimation for entire growth stage

This study employed SVR, RFR, and CatBoost algorithms to estimate maize AGB throughout its entire growth stage. These algorithms were applied using a diverse set of data sources, which included SPAD values, TIR data, and MS data and different combinations of these data (Table 5, Fig. 5). The experimental results revealed that, for individual sensor data, MS data exhibited the most superior performance in estimating maize AGB. In the SVR, RFR, and CatBoost algorithms, MS data exhibited R^2 values of 0.555, 0.776, and 0.776, with corresponding rRMSE values of 66.54%, 48.86%, and 47.89%. Further investigation into the impact of multi-source sensor data fusion (MS + TIR) on estimation performance was conducted. This combination notably enhanced the estimation accuracy across all machine learning models compared with solely using MS data. In the SVR, RFR, and CatBoost algorithms, this combination achieved higher R^2 values of 0.608, 0.800, and 0.817, along with lower rRMSE values of 62.46%, 45.15%, and

Table 4
Statistics for maize AGB (t/hm^2) of training and testing datasets.

Growth stage	Data set	Number of samples	Maximum value	Minimum value	Average value	Standard deviation
Entire growth stage	Training	576	13.232	0.020	2.866	2.361
	Testing	144	12.176	0.172	3.020	2.389
Jointing stage	Training	192	1.568	0.02	0.527	0.280
	Testing	48	1.232	0.096	0.488	0.290
Trumpet stage	Training	192	5.616	0.116	2.114	1.206
	Testing	48	5.784	0.236	2.103	1.303
Big trumpet stage	Training	192	13.232	0.880	5.972	2.831
	Testing	48	12.816	1.000	6.412	3.173

Table 5
Statistics on the accuracy of AGB estimation for different data combinations, and different algorithms.

Data combination	SVR		RFR		CatBoost	
	R^2	rRMSE (%)	R^2	rRMSE (%)	R^2	rRMSE (%)
SPAD	0.453	73.83	0.458	74.69	0.560	65.83
TIR	0.514	69.95	0.753	49.31	0.725	52.02
MS	0.555	66.54	0.766	48.86	0.776	47.89
TIR + SPAD	0.677	57.07	0.773	47.17	0.788	45.61
MS + TIR	0.608	62.46	0.800	45.15	0.817	43.98
MS + SPAD	0.607	62.67	0.823	41.81	0.836	40.09
MS + TIR + SPAD	0.685	56.22	0.862	36.92	0.872	35.49

43.98%. Subsequently, by fusion of multi-source sensor data with SPAD values (MS + TIR + SPAD), this combination further elevated the performance of the estimation models compared with using MS + TIR. This combined approach not only exhibits the highest R^2 values in the SVR, RFR, and CatBoost models, with values of 0.685, 0.862, and 0.872 respectively, but also achieves the lowest rRMSE values, with values of 56.22%, 36.92%, and 35.49% respectively.

Fig. 6 presents scatter plots depicting the AGB estimations using traditional individual MS data, MS + SPAD, MS + TIR, and MS + TIR + SPAD (Using the CatBoost algorithm as an example). Notably, when the measured AGB exceeds 6 t/hm^2, significant scattering is observed in the scatter plots. This phenomenon can be attributed to the influence of two key factors: spectral data saturation and the limited number of samples with measured AGB surpassing 6 t/hm^2. These factors may contribute to an increase in estimation errors within the models. Especially for samples with high AGB levels, the saturation effect of spectral data might hinder the accurate reflection of measured AGB variations, consequently leading to greater scattering in the estimations. Simultaneously, due to the scarcity of samples with measured AGB exceeding 6 t/hm^2, the model's performance within this range might not be as precise as in other ranges, thereby intensifying estimation fluctuations.

Fig. 7 further illustrates the results of a statistical involving the absolute value of residuals between the estimated AGB and the measured AGB for exceeding 6 t/hm^2. It is noteworthy that the employment of MS + SPAD, MS + TIR, and MS + TIR + SPAD yields a lower absolute value of residuals in comparison to using solely individual MS data. This observation robustly validates the potential of multi-source data fusion in overcoming the errors stemming from MS data saturation and insufficient data volume. Furthermore, it accentuates the fact that the fusion of multi-source data can significantly enhance the stability and precision of estimation models.

3.2. AGB estimation for different growth stage

In addition to exploring the entire growth stage for AGB estimation, this study also delved into the impact of different growth stages on AGB estimation. The estimated AGB results for different growth stages of maize are presented in Table 6 and Fig. 8. The results revealed varying accuracy in the estimation of AGB for different growth stages. For the jointing stage, the R^2 values for AGB estimation ranged from 0.305 to 0.726, with rRMSE ranging from 29.68% to 49.84%. For the trumpet stage, the R^2 values improved to a range from 0.452 to 0.866, with rRMSE values reducing from 26.65% to 49.12%. Notably, the highest accuracy was achieved for the big trumpet stage, where the R^2 values further increased to a range from 0.721 to 0.901, accompanied by rRMSE values of 18.72% to 34.97%. This progression underscores the influence of the different growth stages on the accuracy of AGB estimation. As the maize growth stage progresses, the estimation accuracy shows an increasing trend, reaching its highest during the big trumpet stage. The study further delved into the impact of fusing multi-source sensor data with SPAD values on maize AGB estimation across different growth stages, as illustrated in Fig. 9. It was observed that the fusion of multi-source data continued to enhance the accuracy of AGB estimation during different growth stages. This improvement was consistent with the trends observed in the entire growth stage (Table 5 and Fig. 5). These findings underscore the robustness of the multi-source sensor data and SPAD value fusion technique in enhancing AGB estimation accuracy, not only over the entire growth stage but also during each single growth stage.

3.3. AGB estimation for different machine learning algorithms

Fig. 10, combining Tables 5 and 6, illustrates the accuracy of different machine learning algorithms in estimating mazie AGB across the entire growth stage as well as during single growth stages. The result of the study revealed distinct performances of the employed machine learning algorithms in estimating maize AGB. The SVR algorithm exhibited R^2 ranged from 0.305 to 0.824, with rRMSE ranging from 25.18% to 73.83%. For the RFR algorithm, the R^2 ranges from 0.368 to 0.881, with rRMSE ranging from 19.76% to 74.69%. For the CatBoost

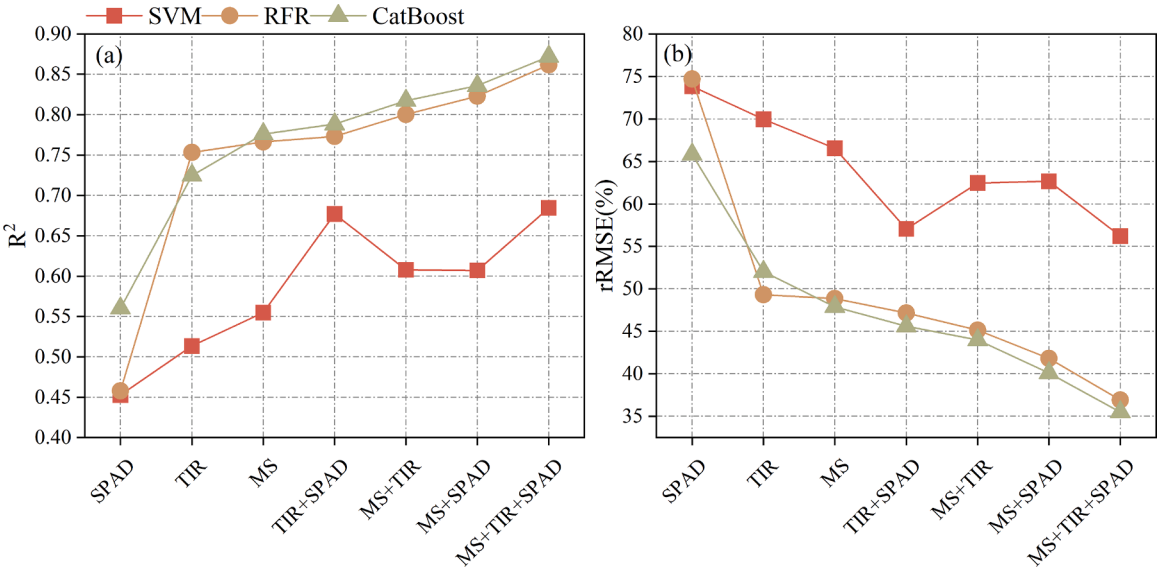

Fig. 5. Accuracy of AGB estimation for different data combinations, and different algorithms. (a) R^2, (b) rRMSE.

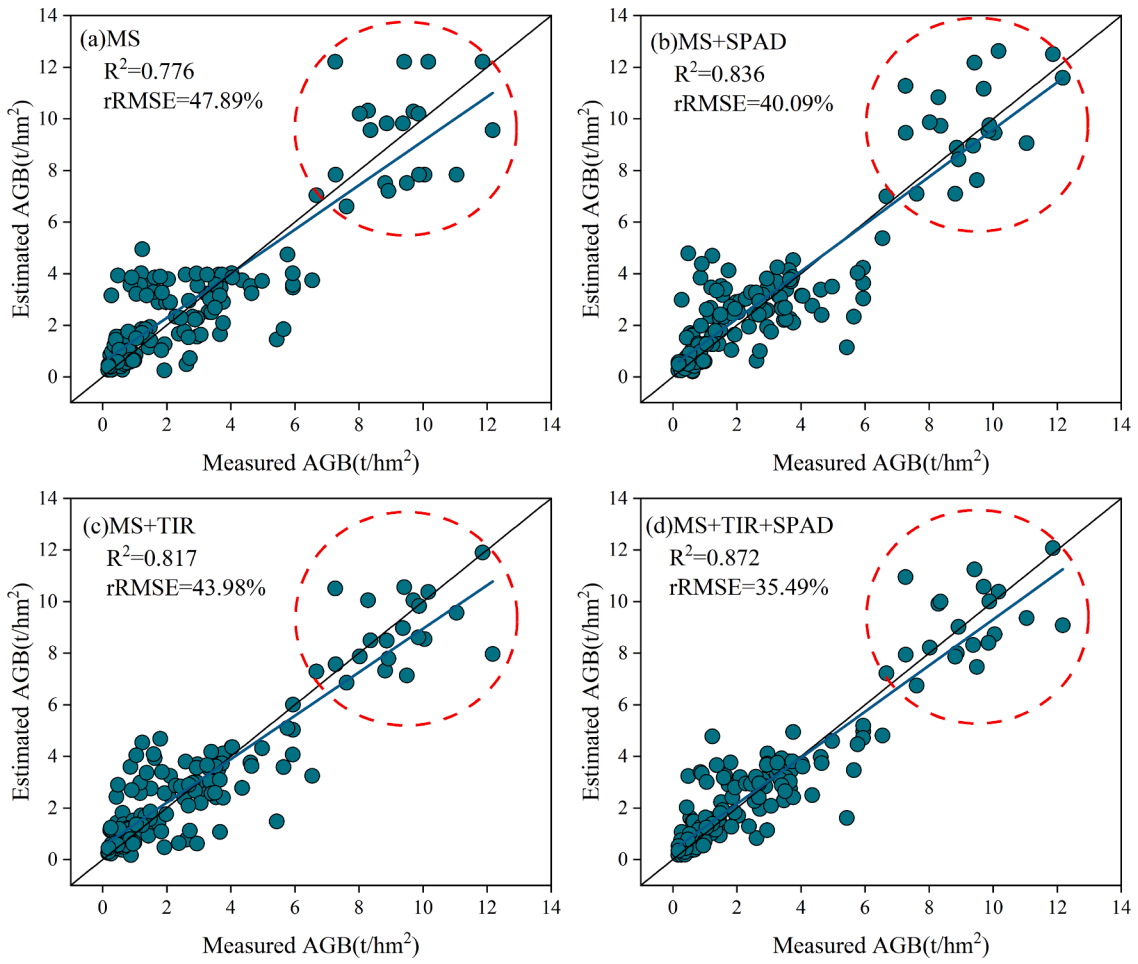

Fig. 6. Scatter plot of AGB estimation using CatBoost algorithm. (a) MS, (b) MS + SPAD, (c) MS + TIR, (d) MS + TIR + SPAD.

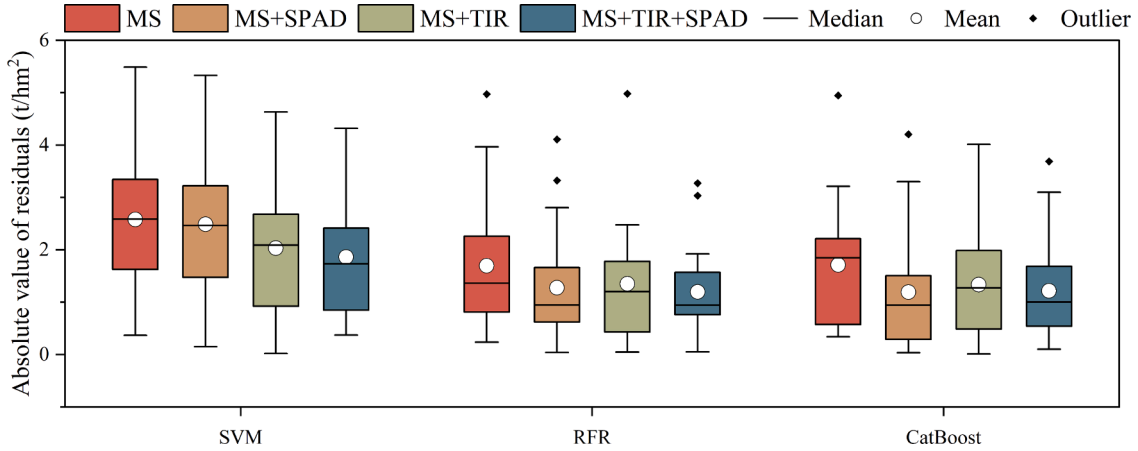

Fig. 7. Statistics of absolute values of residuals between estimated AGB and measured AGB in exceeding 6 t/hm^2.

algorithm, the R^2 ranges from 0.451 to 0.901, with rRMSE ranging from 18.72% to 65.83%. The results underscore the CatBoost algorithm exhibited superior performance compared to both RFR and SVR, indicating its ability to produce more accurate and precise AGB estimates.

3.4. Spatial and temporal distribution of AGB

By using the fusion of multi-sensor data with SPAD values (MS + TIR + SPAD) combined with the CatBoost algorithm, this study achieved the optimal estimation model for maize AGB. Fig. 11 displays the statistical of estimation errors for this model across different growth stages of maize in the years 2022 and 2023. In 2022, during the jointing, trumpet, and big trumpet stages, the estimation errors were primarily distributed within the ranges of ±0.2 t/hm^2, ±0.4 t/hm^2, and ±1.0 t/hm^2, respectively. Similarly, in 2023, during the same growth stages, the estimation errors were primarily distributed within the ranges of ±0.4 t/hm^2, ±1.0 t/hm^2, and ±2.0 t/hm^2, respectively. It is noteworthy that the estimation errors varied across different growth stages and years, primarily

Table 6
Comparison of AGB estimation accuracy for different maize growth stages and algorithms.

Growth stage	Data combination	SVR		RFR		CatBoost	
		R^2	rRMSE(%)	R^2	rRMSE(%)	R^2	rRMSE(%)
Jointing stage	SPAD	0.305	49.84	0.368	43.28	0.457	38.50
	TIR	0.446	43.49	0.38	41.20	0.451	38.53
	MS	0.446	39.94	0.458	38.41	0.461	38.72
	TIR + SPAD	0.527	36.28	0.460	39.86	0.556	34.75
	MS + TIR	0.47	39.42	0.556	34.82	0.604	33.75
	MS + SPAD	0.492	37.19	0.581	34.34	0.600	32.87
	MS + TIR + SPAD	0.567	35.12	0.712	29.68	0.726	30.41
Trumpet stage	SPAD	0.452	49.12	0.751	34.81	0.778	33.13
	TIR	0.681	39.84	0.81	30.86	0.781	33.00
	MS	0.732	35.68	0.833	29.58	0.838	28.69
	TIR + SPAD	0.788	30.83	0.829	30.42	0.841	26.65
	MS + TIR	0.742	35.19	0.847	27.60	0.851	27.16
	MS + SPAD	0.817	30.24	0.845	27.48	0.860	27.52
	MS + TIR + SPAD	0.824	28.78	0.851	26.77	0.866	26.77
Big trumpet stage	SPAD	0.721	34.97	0.733	30.23	0.784	27.65
	TIR	0.768	30.36	0.777	27.33	0.777	27.27
	MS	0.786	32.60	0.798	26.50	0.823	24.37
	TIR + SPAD	0.820	25.39	0.811	25.12	0.839	23.07
	MS + TIR	0.815	25.76	0.866	21.27	0.859	21.32
	MS + SPAD	0.815	25.65	0.872	20.94	0.871	20.44
	MS + TIR + SPAD	0.822	25.18	0.881	19.76	0.901	18.72

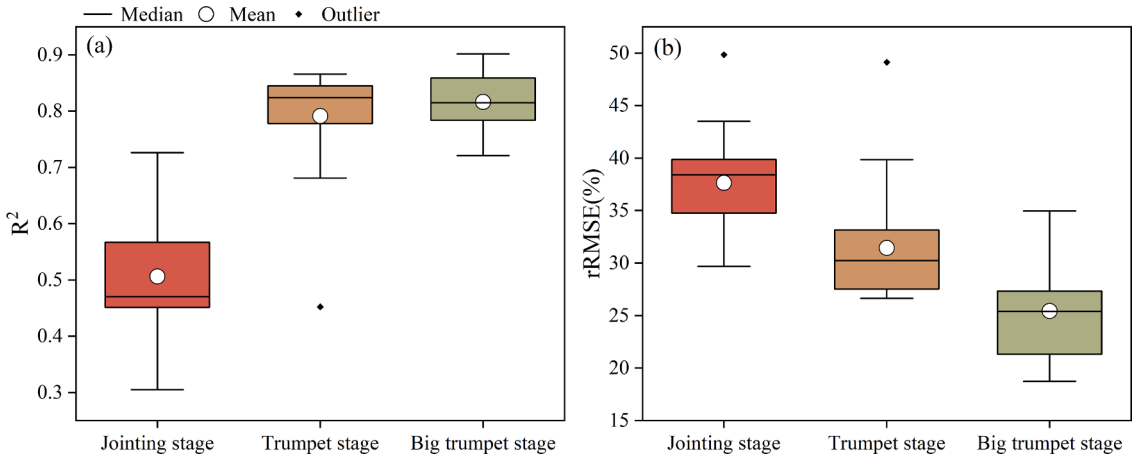

Fig. 8. Statistics on the accuracy of AGB estimation for different growth stages. (a) R^2, (b) rRMSE.

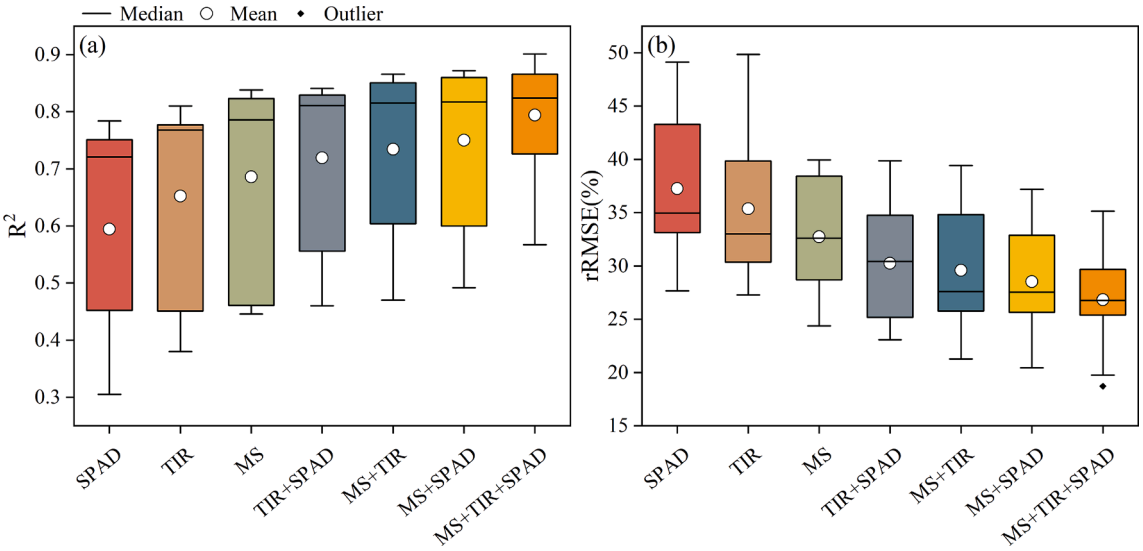

Fig. 9. Statistics on the accuracy of AGB estimation for different combinations of features at different growth stages. (a) R^2, (b) rRMSE.

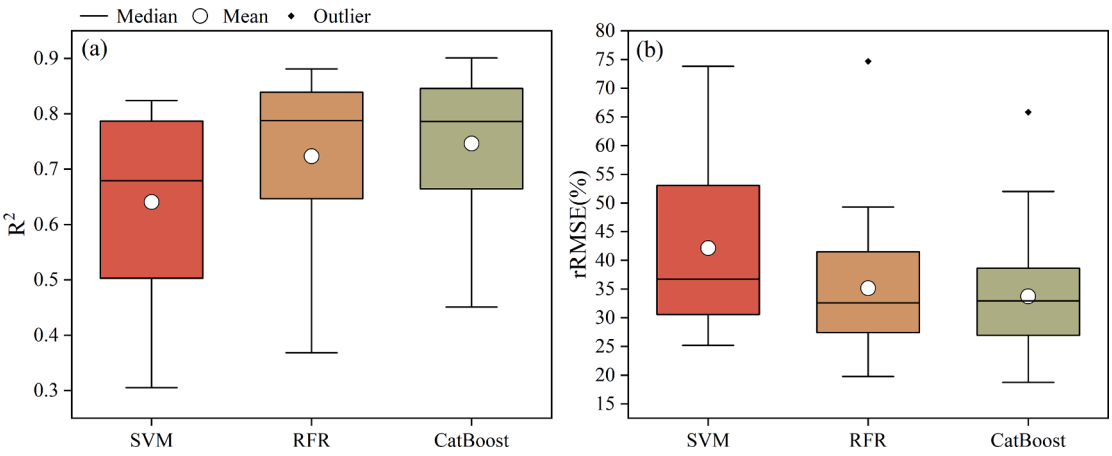

Fig. 10. Statistics on the accuracy of AGB estimation for different algorithms. (a) R^2, (b) rRMSE.

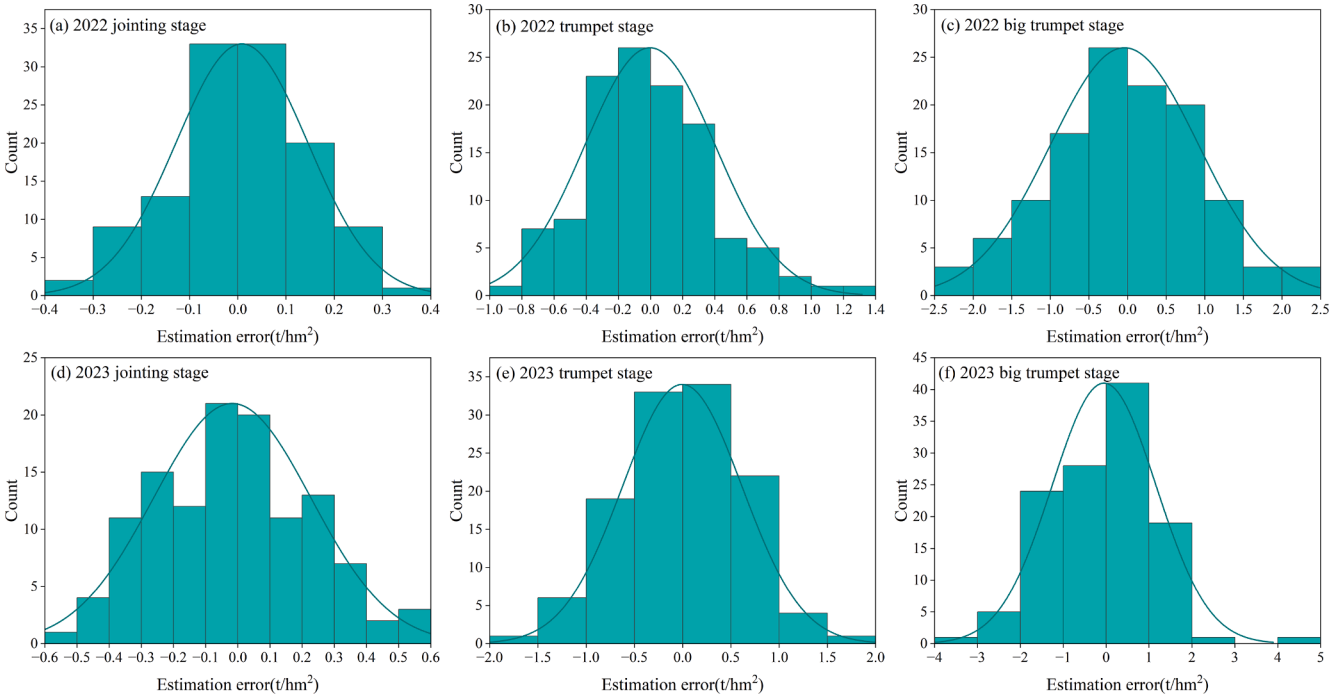

Fig. 11. AGB estimation errors of the fusion of multi-source sensor data with SPAD values based on CatBoost. (a) 2022 jointing stage, (b) 2022 trumpet stage, (c) 2022 big trumpet stage, (d) 2023 jointing stage,(e) 2023 trumpet stage, (f) 2023 big trumpet stage.

influenced by the different measured AGB data. However, overall, these errors remained within relatively small ranges.

Fig. 12 displays the spatial distribution of estimated maize AGB from the jointing stage to the big trumpet stage in both 2022 and 2023. This study's findings indicate a gradual increase in AGB as the maize growth stages progress. This observation aligns with the measured AGB shown in Fig. 3. Regarding the distribution of AGB in 2022, among the four fertilization treatments, N1 and N2 treatments exhibit higher AGB levels, indicating better growth trends. However, AGB in the N3 treatment is notably lower. This discrepancy could stem from the fact that the N3 plot was previously used for apple tree cultivation before maize planting. This prior land use might have led to an excessive depletion of soil nutrients. Despite applying the highest nitrogen fertilizer level in the N3 treatment, the AGB remains lower than that in the N1 and N2 treatments. In practical applications, timely additional fertilization or irrigation should be considered for this plot to improve crop growth conditions. For the distribution of AGB in 2023, an increasing trend in AGB is evident with rising fertilizer levels, notably reaching the highest AGB in the N3 treatment. In summary, the methodology employed in this study facilitates rapid and non-destructive estimation of maize AGB. This enables decision-makers to promptly adjust fertilizer application strategies to accommodate the changing nutrient demands of crops during different growth stages.

4. Discussion

4.1. Effect of AGB estimation throughout fusion of multi-source sensor data with SPAD values

A comparison of multi-source sensor data fusion (MS + TIR) with single sensor data (MS or TIR) shows that multi-source sensor data fusion always results in better estimation in the entire growth stage and different growth stages (Table 5, Fig. 5, and Fig. 9). In addition, this approach can address the estimation error caused by spectral saturation in areas of high vegetation and limited sample size (Fig. 7). This is mainly because different sensors can represent different aspects in

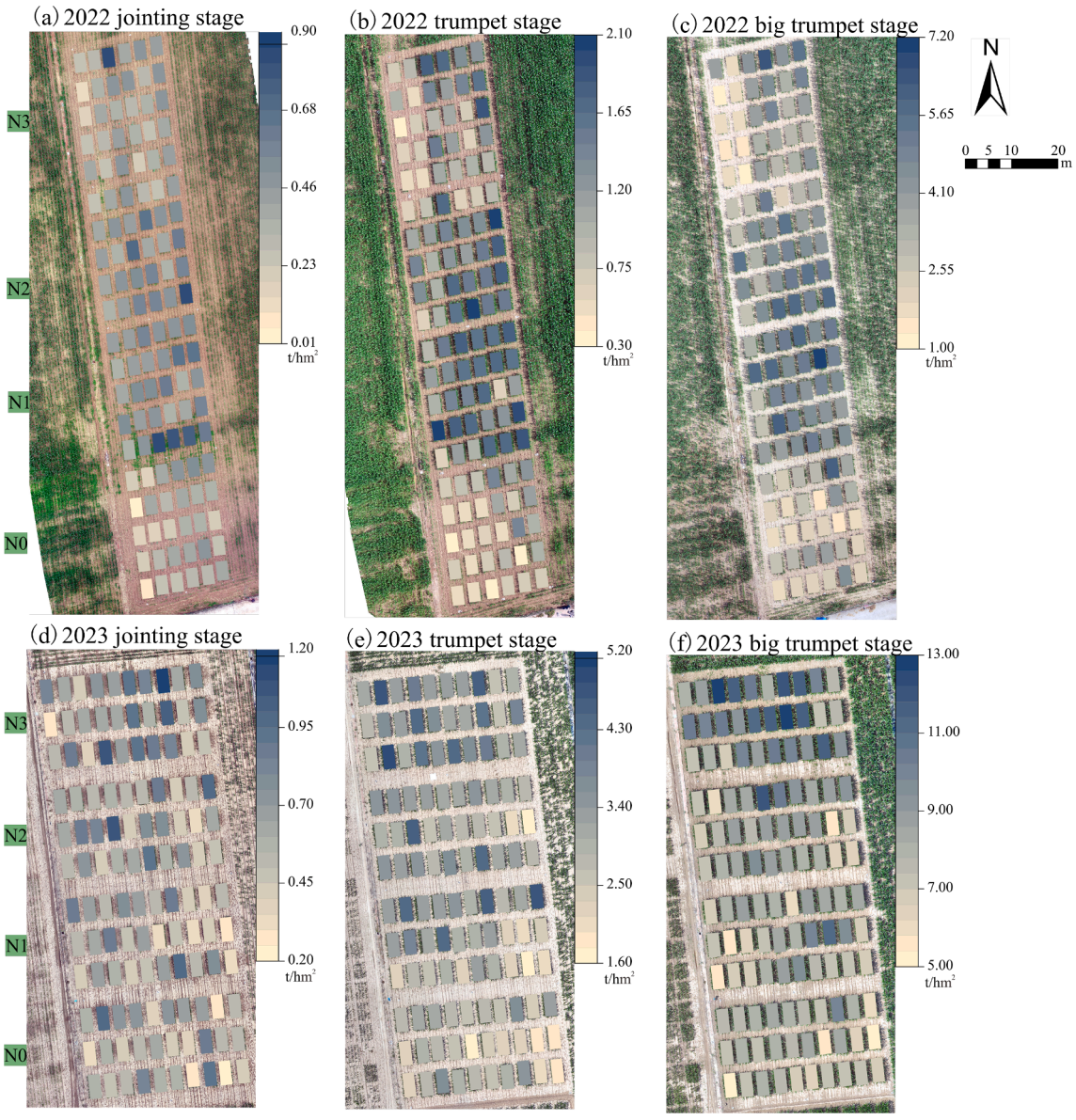

Fig. 12. Spatiotemporal distribution for estimated AGB of the fusion of multi-source sensor data with SPAD values based on CatBoost. (a) 2022 jointing stage, (b) 2022 trumpet stage, (c) 2022 big trumpet stage, (d) 2023 jointing stage, (e) 2023 trumpet stage, (f) 2023 big trumpet stage.

relation to AGB (Maimaitijiang et al., 2020). The vegetation index of MS data consists of a combination of spectral reflectance of different bands, which is an important parameter for crop growth analysis. The canopy temperature included TIR data is the results of the joint action of crop genetic characteristics and environmental conditions, which are not only closely related to the functional phase and transpiration rate of leaves but also influence the growth of dry organs and starch synthesis of crops (Fei et al., 2022). There is heterogeneity between MS data and TIR data. Combining the two can form complementary information, which is conducive to improving the AGB estimation accuracy.

In addition, the stability of sensors can be affected by many external environmental factors, such as crop growth status and soil conditions (Li et al., 2020b). The interaction of these factors can have an impact on AGB estimates, making it necessary to include remote sensing auxiliary data to address sensor bias. SPAD values, as an indicator of plant growth and development and leaf nitrogen content, provide a good reflection of the nutritional status and senescence process of the crop and help to reduce sensor bias. In this study, the combination of SPAD values and sensor data further improved the accuracy of AGB estimation compared to using sensor data only (Tables 5 and 6, Figs. 5 and 9). This is because SPAD values are closely related to the strength of crop photosynthesis, which can explain changes in crop AGB and yield (Peng, 2000; Yoshida and Horie, 2009). Compared with other remote sensing auxiliary data such as crop moisture content, crop nitrogen content, and soil factor, SPAD values are easily obtained in field experiments using the SPAD-502 plus, and have the advantages of low price, simple operation, and non-destructive sampling. It has a broad application scenario for crop parameter inversion.

4.2. *Effects of different growth stages on AGB estimation*

There were differences in the effectiveness for estimating AGB across different growth stages, with an overall upward trend from jointing satge to big trumpet stage (Table 6 and Fig. 8). This is because the spatial heterogeneity of the ground changes with the growth and development of the crop, leading to different correlations between features and AGB at different growth stages (Qiao et al., 2022b). The maize jointing stage is dominated by nutritional growth and the various types of features

from this stage do not reflect the process of organ dry matter accumulation. As maize growth enters the big trumpet stage, the dry matter mass of maize organs increases rapidly due to the growth of canopy leaves and the accumulation of nutrients, and the chlorophyll content of leaves rises (Qiao et al., 2022a). At this time maize AGB correlates well with MS data, TIR data, and SPAD values, so the accuracy of the model increases.

Another reason is that early maize plants were shorter and had less canopy cover, and soil pixels were more heavily represented in the images. In spite of the fact that the plant-soil separation process was carried out according to strict standards, some of the soil pixels were not completely removed. The information extracted was affected by the mixed pixels, making the obtained data not fully express the relationship with the AGB. As the maize grows, the canopy gradually thickens and the proportion of soil pixels in the image decreases significantly. The influence of the mixed pixels is reduced when performing the vegetation separation process and feature extraction (Liu et al., 2021). Compared with the previous growth stage, the estimation accuracy is higher.

4.3. Effect of CatBoost algorithms on maize AGB estimation

In addition to the use of multi-source data fusion to improve estimation accuracy, developments in computer science have provided new methods for the estimation of crop parameters. The emergence of ensemble learning methods has effectively addressed the problem of the weak generalization ability of single learners. The two ensemble learning models, RFR and CatBoost, showed superior estimation results in this study, which outperformed the single learner SVR (Fig. 10). Previous scholars have successfully applied RFR to crop parameters estimation studies and have estimated it with higher accuracy than Ridge Regression, SVR, and ANN (Han et al., 2019; Zhai et al., 2023b). This is because each base learner in ensemble learning has a different hypothesis space. Integrating these base learners can expand the hypothesis space, thus improving the robustness of the ensemble model to unknown distributed data. RFR was less effective in estimation in this study compared to CatBoost. Most likely because RFR is not able to make predictions beyond the scope of the training set data, resulting in poorer performance when validating on certain specific noisy data. CatBoost, a new machine Learning algorithm, has proven its performance in studies of plant water demand prediction in wet areas of China, with significant improvements in accuracy and stability compared to RFR (Huang et al., 2019). CatBoost in this study showed optimal AGB estimation for different stages, the model was highly generalizable. This is because CatBoost can improve the accuracy of the model by continuously correcting and updating the sample weights and controlling the error by learning the features of the training set. It is more adaptable to the data, with strong mapping ability and resistance to interference. In addition, CatBoost optimizes the traditional gradient boosting decision tree approach to gradient estimation employing ordered boosting. This allows it to obtain an unbiased estimate of the gradient, which in turn reduces the effect of gradient estimation bias and the model has a stronger generalization capability (Prokhorenkova et al., 2018). Therefore, it is still highly adaptive and noise-resistant under the small training set of this experiment, with optimal estimation results.

4.4. Significance and constraints of the study

This study demonstrated the efficacy of fusing multi-source sensor data with SPAD values to enhance maize AGB estimation accuracy holds substantial promise for modern farming practices, and bears significant implications for both agricultural management and crop phenotypic parameters estimation. The capability to accurately monitor crop AGB across different growth stages empowers farmers and agronomists to make well-informed decisions about resource allocation, irrigation, and fertilization. This directly translates into improved resource efficiency, cost reduction, and a decreased environmental impact. Moreover, the superiority of the CatBoost algorithm over RFR and SVR highlights the crucial role of algorithm selection in precision agriculture. In an evolving artificial intelligence technology, the integration of advanced machine-learning techniques can notably elevate the accuracy of AGB estimation. As observed in this study, CatBoost has a clear superiority and therefore these advanced algorithms should be adopted more resolutely in practical agricultural management. This endorses practitioners with a reliable framework that yields dependable and high-accuracy estimation outcomes. Furthermore, the identified influence of different growth stages on estimation accuracy provides actionable insights for crop precision management. With a clearer understanding of the varying impacts of different growth stages, practitioners can adjust their data collection efforts and interventions to specific growth stages. This temporal precision in decision-making supports targeted interventions such as timely pest management, nutrient application, and harvest planning. This study also utilized UAV sensor data and SPAD values from 2022 as a training set to construct an AGB estimation model using the CatBoost algorithm. The model was then tested using UAV sensor data and SPAD values from 2023 to estimate the AGB for that year, as shown in Fig. 13. It is evident that with the incorporation of different data sources, the accuracy of estimation gradually improves. Notably, the combination of MS + TIR + SPAD maintains the highest estimation accuracy. This result strongly validates the efficacy of the proposed method in estimating maize AGB for future years. Additionally, the study using multi-source sensor data with SPAD values fusion estimate maize yield for the year 2022, as depicted in Fig. 14. It is evident that the fusion of multi-source sensor data continues to yield higher estimation accuracy compared to using a single sensor. Moreover, the addition of SPAD values on top of the multi-source sensor data further enhances estimation accuracy. This affirms the method's adaptability across various scenarios. From a practical perspective, the high-throughput crop phenotyping method proposed in this study provides valuable insights for precision agriculture management.

In practical situations with a limited number of plots, the combination of SPAD data does indeed hold significant practical value. In such scenarios, the manual collection of SPAD data can be easily managed and can provide valuable information for model calibration and validation. However, obtaining a large quantity of SPAD values extensively requires significant human and material resources, posing a challenge that needs to be addressed. To tackle this issue, this study proposes an improved approach. This study using UAV sensor data from 2022 as a training set, constructs a SPAD value estimation model (using the CatBoost algorithm). Subsequently, the 2023 UAV data was used as a testing set to estimate the 2023 SPAD values. (Fig. 15(a)). This method requires only a certain amount of SPAD value data to be modeled alongside UAV sensor data, while the rest of the SPAD data can be obtained from high-throughput UAV sensor data, reducing the labor required for manual ground-based SPAD data collection. To demonstrate the reliability of this approach, the measured SPAD values and UAV sensor data from 2022 for training to construct an AGB estimation model, and the estimated SPAD values and UAV sensor data from 2023 for testing, in order to estimate the AGB of maize for 2023. Fig. 15(b) presents a scatter plot depicting the combination of UAV sensor data with estimated SPAD values from 2023 for 2023 AGB estimation. Notably, the AGB estimated by this method achieves an R^2 value of 0.916 and an rRMSE value of 24.98%, which is significantly superior in accuracy compared to the method shown in Fig. 13 that solely utilizes UAV sensor data for AGB estimation for 2023 (MS + TIR, R^2 = 0.886, rRMSE = 29.19%). By introducing this improved approach, the challenge of expending excessive resources to acquire SPAD values in the real world is effectively addressed, providing a more reliable and efficient solution for estimating maize AGB.

5. Conclusion

This study explores the fusion of UAV multi-source sensor data with

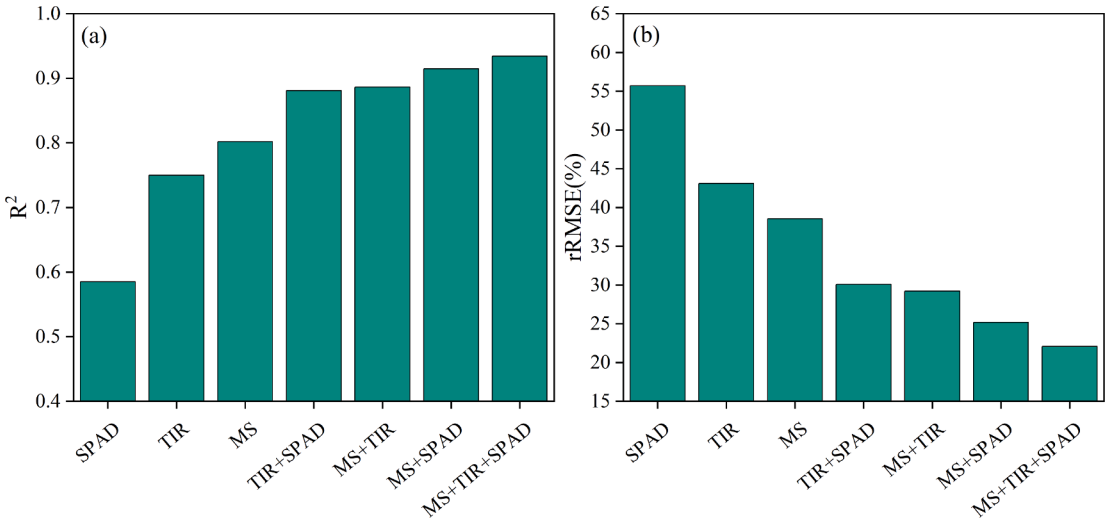

Fig. 13. Effectiveness of fusion of multi-source sensor data with SPAD values for maize AGB estimation in future years (2023). (a) R^2, (b) rRMSE.

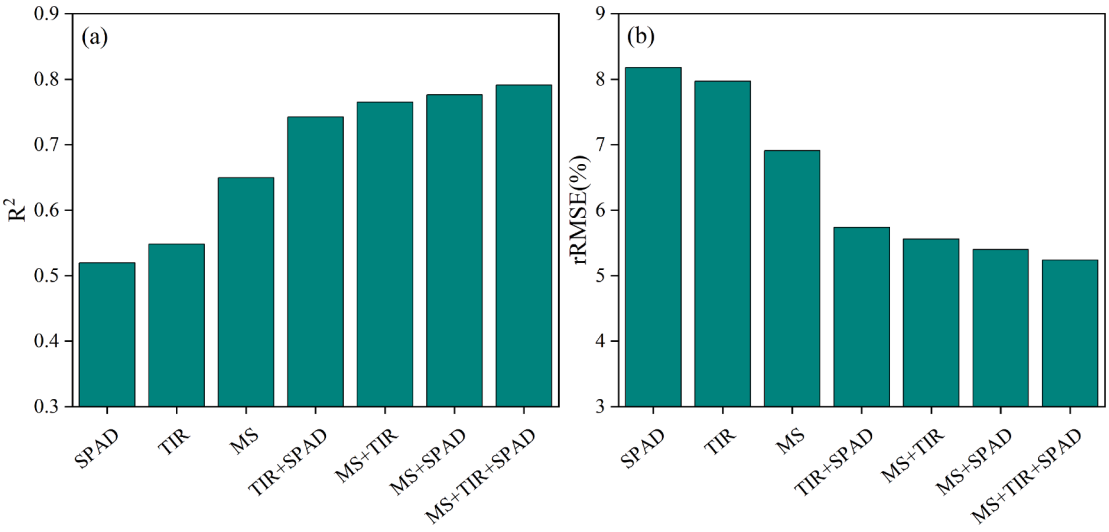

Fig. 14. Effectiveness of fusion of multi-source sensor data with SPAD values for maize yield estimation. (a) R^2, (b) rRMSE.

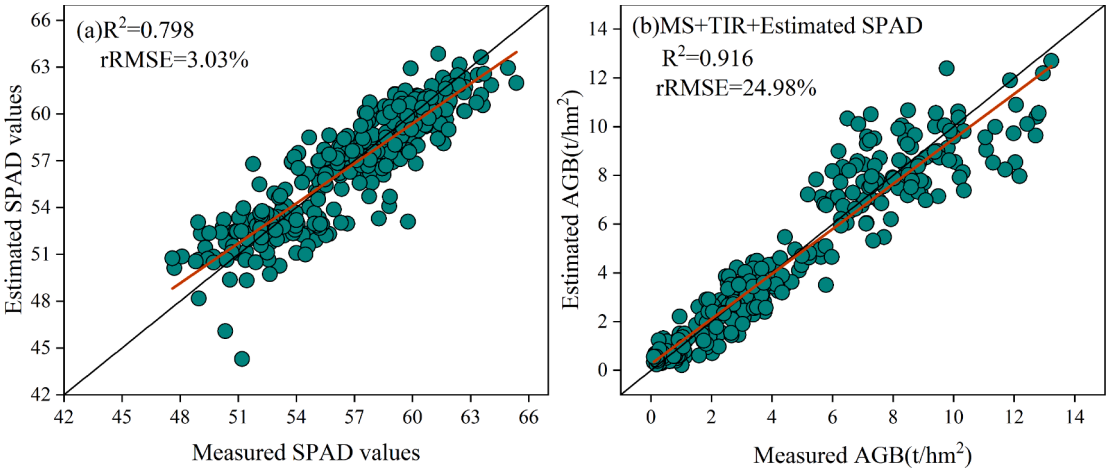

Fig. 15. Scatter plots of estimated SPAD values in 2023 (a) and the fusion of UAV multi-source sensor data with estimated SPAD values from 2023 to estimate 2023 AGB (b).

SPAD values and the application of CatBoost to maize AGB estimation. The following conclusions were obtained:

1. Whether for the entire growth stage or a single growth stage, the fusion of UAV multi-source sensor data can improve the estimation accuracy of individual sensors, and the integration of SPAD values with multi-source sensor data can further improve the estimation accuracy.
2. The estimation accuracy varies across different growth stages, with an increasing trend in accuracy from the jointing stage to the big trumpet stage.
3. CatBoost has achieved the best estimation results, which has significant potential for crop phenotyping research.

This study provides a new and effective tool for maize AGB estimation, which can meet the need for low-cost, large-scale, and repeated observations for crop growth monitoring.

Declaration of Competing 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.

Data availability

Data will be made available on request.

Acknowledgments

This research was funded by the Intelligent Irrigation Water and Fertilizer Digital Decision System and Regulation Equipment (2022YFD1900404), Central Public-interest Scientific Institution Basal Research Fund (IFI2023-29, IFI2023-01), the Key Grant Technology Project of Henan (221100110700), the Contract for Post-Disaster Reconstruction Special Tasks in Xinxiang City (21CJ001), Henan Science and Technology Project (222102110038), National Natural Science Foundation of China (41871333) and Innovation Training Program for College Students (202210460019).

References

Bian, C., Shi, H., Wu, S., Zhang, K., Wei, M., Zhao, Y., Sun, Y., Zhuang, H., Zhang, X., Chen, S., 2022. Prediction of Field-Scale Wheat Yield Using Machine Learning Method and Multi-Spectral UAV Data. Remote Sens. 14, 1474.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Broge, N.H., Leblanc, E., 2001. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. Remote Sens. Environ. 76, 156–172.

Devia, C.A., Rojas, J.P., Petro, E., Martinez, C., Mondragon, I.F., Patiño, D., Rebolledo, M. C., Colorado, J., 2019. High-throughput biomass estimation in rice crops using UAV multispectral imagery. J. Intell. Robot. Syst. 96, 573–589.

Fei, S., Hassan, M.A., Xiao, Y., Su, X., Chen, Z., Cheng, Q., Duan, F., Chen, R., Ma, Y., 2022. UAV-based multi-sensor data fusion and machine learning algorithm for yield prediction in wheat. Precis. Agric. 1–26.

Feng, L., Zhang, Z., Ma, Y., Du, Q., Williams, P., Drewry, J., Luck, B., 2020. Alfalfa yield prediction using UAV-based hyperspectral imagery and ensemble learning. Remote Sens. 12, 2028.

Gitelson, A.A., Viña, A., Arkebauer, T.J., Rundquist, D.C., Keydan, G., Leavitt, B., 2003. Remote estimation of leaf area index and green leaf biomass in maize canopies. Geophys. Res. Lett., 30.

Gitelson, A.A., Kaufman, Y.J., Merzlyak, M.N., 1996. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. Remote Sens. Environ. 58, 289–298.

Han, L., Yang, G., Dai, H., Xu, B., Yang, H., Feng, H., Li, Z., Yang, X., 2019. Modeling maize above-ground biomass based on machine learning approaches using UAV remote-sensing data. Plant Methods 15, 1–19.

Houborg, R., McCabe, M.F., 2016. High-Resolution NDVI from planet's constellation of earth observing nano-satellites: A new data source for precision agriculture. Remote Sens. 8, 768.

Huang, G., Wu, L., Ma, X., Zhang, W., Fan, J., Yu, X., Zeng, W., Zhou, H., 2019. Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. J. Hydrol. 574, 1029–1041.

Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). Remote Sens. Environ. 25, 295–309.

Jin, X., Li, Z., Feng, H., Ren, Z., Li, S., 2020a. Deep neural network algorithm for estimating maize biomass based on simulated Sentinel 2A vegetation indices and leaf area index. Crop J. 8, 87–97.

Jin, X., Zarco-Tejada, P.J., Schmidhalter, U., Reynolds, M.P., Hawkesford, M.J., Varshney, R.K., Yang, T., Nie, C., Li, Z., Ming, B., 2020b. High-throughput estimation of crop traits: A review of ground and aerial phenotyping platforms. IEEE Geosci. Remote Sens. Mag. 9, 200–231.

Lee, S., Vo, T.P., Thai, H.-T., Lee, J., Patel, V., 2021. Strength prediction of concrete-filled steel tubular columns using Categorical Gradient Boosting algorithm. Eng. Struct. 238, 112109.

Li, Z., Taylor, J., Yang, H., Casa, R., Jin, X., Li, Z., Song, X., Yang, G., 2020b. A hierarchical interannual wheat yield and grain protein prediction model using spectral vegetative indices and meteorological data. Field Crop. Res. 248, 107711.

Li, B., Xu, X., Zhang, L., Han, J., Bian, C., Li, G., Liu, J., Jin, L., 2020a. Above-ground biomass estimation and yield prediction in potato by using UAV-based RGB and hyperspectral imaging. ISPRS-J. Photogramm. Remote Sens. 162, 161–172.

Ling, Q., Huang, W., Jarvis, P., 2011. Use of a SPAD-502 meter to measure leaf chlorophyll concentration in Arabidopsis thaliana. Photosynth. Res. 107, 209–214.

Liu, H.Q., Huete, A., 1995. A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. IEEE Trans. Geosci. Remote Sensing 33, 457–465.

Liu, S., Jin, X., Nie, C., Wang, S., Yu, X., Cheng, M., Shao, M., Wang, Z., Tuohuti, N., Bai, Y., 2021. Estimating leaf area index using unmanned aerial vehicle data: shallow vs. deep machine learning algorithms. Plant Physiol. 187, 1551–1576.

Maes, W.H., Steppe, K., 2019. Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture.Trends. Plant Sci. 24, 152–164.

Maimaitijiang, M., Sagan, V., Sidike, P., Hartling, S., Esposito, F., Fritschi, F.B., 2020. Soybean yield prediction from UAV using multimodal data fusion and deep learning. Remote Sens. Environ. 237, 111599.

Pearson, R.L., & Miller, L.D. (1972). Remote mapping of standing crop biomass for estimation of the productivity of the shortgrass prairie. Remote Sens. Environ., VIII, 1355.

Peng, S., 2000. Single-leaf and canopy photosynthesis of rice. Studies in Plant Science, Elsevier, pp. 213-228.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2018. CatBoost: unbiased boosting with categorical features. Adv. Neural Inf. Process. Syst., 31.

Qiao, L., Gao, D., Zhao, R., Tang, W., An, L., Li, M., Sun, H., 2022a. Improving estimation of LAI dynamic by fusion of morphological and vegetation indices based on UAV imagery. Comput. Electron. Agric. 192, 106603.

Qiao, L., Tang, W., Gao, D., Zhao, R., An, L., Li, M., Sun, H., Song, D., 2022b. UAV-based chlorophyll content estimation by evaluating vegetation index responses under different crop coverages. Comput. Electron. Agric. 196, 106775.

Tang, J., Fan, B., Xiao, L., Tian, S., Zhang, F., Zhang, L., Weitz, D., 2021. A new ensemble machine-learning framework for searching sweet spots in shale reservoirs. SPE J. 26, 482–497.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8, 127–150.

Yoshida, H., Horie, T., 2009. A process model for explaining genotypic and environmental variation in growth and yield of rice based on measured plant N accumulation. Field Crop. Res. 113, 227–237.

Yu, D., Zha, Y., Sun, Z., Li, J., Jin, X., Zhu, W., Bian, J., Ma, L., Zeng, Y., Su, Z., 2023. Deep convolutional neural networks for estimating maize above-ground biomass using multi-source UAV images: A comparison with traditional machine learning algorithms. Precis. Agric. 24, 92–113.

Yue, J., Yang, G., Li, C., Li, Z., Wang, Y., Feng, H., Xu, B., 2017. Estimation of winter wheat above-ground biomass using unmanned aerial vehicle-based snapshot hyperspectral sensor and crop height improved models. Remote Sens. 9, 708.

Zha, H., Miao, Y., Wang, T., Li, Y., Zhang, J., Sun, W., Feng, Z., Kusnierek, K., 2020. Improving unmanned aerial vehicle remote sensing-based rice nitrogen nutrition index prediction with machine learning. Remote Sens. 12, 215.

Zhai, W., Li, C., Cheng, Q., Ding, F., Chen, Z., 2023a. Exploring Multisource Feature Fusion and Stacking Ensemble Learning for Accurate Estimation of Maize Chlorophyll Content Using Unmanned Aerial Vehicle Remote Sensing. Remote Sens. 15, 3454.

Zhai, W., Li, C., Cheng, Q., Mao, B., Li, Z., Li, Y., Ding, F., Qin, S., Fei, S., Chen, Z., 2023b. Enhancing Wheat Above-Ground Biomass Estimation Using UAV RGB Images and Machine Learning: Multi-Feature Combinations, Flight Height, and Algorithm Implications. Remote Sens. 15, 3653.