Journal of Agricultural Engineering

https://www.agroengineering.org/

Research on an automated mapping method for rice aboveground biomass based on low-altitude remote sensing

Honggang Xu,[1] Xuehan Li,[2] Jia Shen,[3] Ziyi Li,[1] Zhe Li,[4] Yiming Li,[4] Pengcheng Nie[1]

[1]College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou
[2]College of Environmental and Resource Sciences, Zhejiang University, Hangzhou
[3]Zhejiang EV-Tech Co., Ltd., Hangzhou
[4]College of Information Engineering, Tarim University, Alar, China

Corresponding author: Pengcheng Nie, College of Biosystems Engineering and Food Science, Zhejiang University, 866 Yuhangtang Road, Hangzhou, China. E-mail: pcn@zju.edu.cn

Please cite this article as doi: 10.4081/jae.2026.2061

Submitted: 16 December 2025
Accepted: 27 August 2026

Research on an automated mapping method for rice aboveground biomass based on low-altitude remote sensing

Honggang Xu,[1] Xuehan Li,[2] Jia Shen,[3] Ziyi Li,[1] Zhe Li,[4] Yiming Li,[4] Pengcheng Nie[1]

[1]College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou
[2]College of Environmental and Resource Sciences, Zhejiang University, Hangzhou
[3]Zhejiang EV-Tech Co., Ltd., Hangzhou
[4]College of Information Engineering, Tarim University, Alar, China

Corresponding author: Pengcheng Nie, College of Biosystems Engineering and Food Science, Zhejiang University, 866 Yuhangtang Road, Hangzhou, China. E-mail: pcn@zju.edu.cn

Contributions: **Honggang Xu,** conceptualization, investigation, methodology, data curation, formal analysis, software, validation, visualization, writing – review & editing. **Xuehan Li,** investigation, methodology, data curation. **Jia Shen,** resources, writing – review & editing.
PN, conceptualization; **Ziyi Li,** investigation, methodology. **Zhe Li,** methodology. **Yiming Li,** investigation, methodology. **Pengcheng Nie,** supervision, funding acquisition, resources, writing – review & editing.
All authors read and approved the final version of the manuscript and agreed to be accountable for all aspects of the work.

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

Availability of data and materials: The data used to support the findings of this study are available from the corresponding author upon request.

Acknowledgments: The authors gratefully acknowledge the financial support provided by the National Key Research and Development Program of China (Grant No 2022YFD2001801).

Abstract

Accurate monitoring of rice aboveground biomass (AGB) is crucial for guiding agricultural production management. This study focuses on high-precision estimation and automated mapping of rice AGB. Field experiments were conducted in Nanxun District, Huzhou City, Zhejiang Province. We collected UAV RGB and multispectral images, rice AGB, and plant height data. By integrating vegetation indices, texture features, and plant height information, the AGB estimation model was established using algorithms such as Stacking. A framework combining "SAM + MobileNetV3-Small classification" was proposed to achieve automated paddy field extraction and phenology recognition. The results demonstrate that the rice AGB prediction model based on the Stacking ensemble algorithm performed excellently. The introduction of plant height significantly improved model accuracy. For example, during the heading stage, R^2 increased from 0.421 to 0.739, and RPIQ rose from 1.995 to 3.015. The automated paddy field extraction and phenology recognition framework developed in this study achieved a segmentation accuracy of 0.968 and a classification accuracy of 0.993 on the dataset used in this study, without requiring manual annotation. This research provides a technical reference for automated and high-precision mapping of rice AGB.

Key words: rice; AGB; UAV; farmland segmentation; phenology recognition.

Introduction

Driven by agricultural modernization, precision agriculture technologies are being deployed at an unprecedented scale (Cano *et al.*, 2025; Kroupova *et al.*, 2025). UAV remote sensing enables efficient acquisition of high-resolution multi-source data over extensive agricultural areas, significantly reducing labor costs. It also demonstrates robust adaptability to complex terrains and environments, having evolved into an indispensable tool for precision agriculture (Huang *et al.*, 2013). Particularly in crop growth monitoring (Agrawal and Arafat, 2024; Raouhi *et al.*, 2023), integrating the acquired image data with machine learning algorithms facilitates the accurate estimation of key crop growth parameters such as plant height, aboveground biomass (AGB), and canopy coverage (Dreier *et al.*, 2025; Niu *et al.*, 2025; Zhang *et al.*, 2025a). This approach achieves high-precision, dynamic monitoring, thereby providing a scientific basis for optimizing management practices such as fertilization, irrigation, and pest and disease control.

As one of the most crucial food crops globally, the accurate estimation of rice AGB holds great significance for food security, agricultural resource management, and climate change research (Liu *et al.*, 2025). Traditional AGB estimation methods mainly rely on vegetation indices (VIs), such as multispectral vegetation indices and visible vegetation indices (Shi *et al.*, 2025). These indices can reflect the growth status of vegetation;

however, during the middle and late growth stages of rice, the relationship between characteristic variables and AGB becomes complex due to high vegetation coverage, resulting in a decline in estimation accuracy (Adeluyi *et al.*, 2022). Therefore, some studies have attempted to introduce new characteristic variables, such as plant height and Leaf Area Index (LAI), to improve the accuracy of AGB estimation. As an important indicator of crop growth, plant height has a close relationship with AGB. Existing studies have confirmed that there is a significant positive correlation between plant height and AGB in crops such as maize, soybeans, potatoes, and wheat (Li *et al.*, 2020; Maimaitijiang *et al.*, 2019; Zhu *et al.*, 2024). In the prediction of rice AGB, several studies have also proven that plant height makes a significant contribution to improving the accuracy of rice AGB estimation (Adeluyi *et al.*, 2022; Shi *et al.*, 2025; Yang *et al.*, 2023). However, most current studies are confined to the scale of experimental plots, which have relatively small areas. At this scale, UAV can acquire high-resolution image data by flying at low altitudes. When the monitoring area is expanded to the regional scale, the increased flight altitude of UAV leads to a decrease in image resolution. Under such circumstances, whether the estimation accuracy of rice plant height and the contribution of plant height to AGB prediction will change requires verification through relevant experiments.

After constructing AGB inversion model, achieving automated paddy field AGB mapping at the field scale requires high-precision paddy field extraction and phenological identification. Firstly, high-precision paddy field extraction aims to accurately separate paddy and non-paddy areas, providing a reliable base map for subsequent AGB mapping. Traditional classification algorithms such as the maximum likelihood method and K-Means clustering often fail to yield satisfactory segmentation results in complex field environments. In recent years, with the rise of deep learning technology, deep learning-based segmentation algorithms have gradually become the mainstream of research. Deep learning models can automatically learn features from images, thereby identifying paddy field areas more accurately (Liu *et al.*, 2024; Xu *et al.*, 2024). However, most deep learning models require large amounts of labeled data for training, which increases the difficulty of data collection. Meanwhile, paddy field images vary across different regions and time periods. Models usually need retraining and adjustment for each region. This increases model application cost and time. Secondly, since different growth stages of rice correspond to distinct threshold ranges for AGB, clarifying the current phenological stage is essential for the quantitative assessment of inversion results and diagnosis of growth status. Therefore, it is necessary to develop a paddy field extraction and phenology recognition framework that can adapt to varying regions and conditions while reducing reliance on annotated data.

This study integrates ensemble learning regression algorithms to estimate rice AGB and explores the impact of plant height characteristics on model accuracy. Furthermore, a deep learning-based framework for paddy field extraction and phenology recognition is proposed. This framework enables the automatic identification of rice field boundaries and the recognition of growth stages for each field from UAV imagery. Through this research, we aim to provide an efficient and accurate technical pathway for rice AGB estimation and to supply critical data support for monitoring rice growth status.

Materials and Methods

The study area is located in Nanxun District, Huzhou City, Zhejiang Province (30.52°-31.11°N, 120.08°-120.43°E). The average elevation of the area is less than 5 meters, and the soil types are dominated by alluvial soil and paddy soil. It has a subtropical monsoon climate, with an annual average temperature of 15.5-16°C, an annual precipitation of 1050-1850 mm, and a frost-free period of 224-246 days, all of which are suitable for rice cultivation.

Data acquisition and processing

The data collection was conducted from July to October 2024 under clear and windless weather conditions, during which UAV images and ground data of 4 study areas were acquired. The locations of the four study areas are shown in Figure 1 a-d. In this paper, the four study areas are referred to as A, B, C, and D. The paddy field areas of these regions are as follows: Region A is 37 hectares (accounting for approximately 43.5% of the region), Region B is 32 hectares (accounting for approximately 84% of the region), Region C is 47 hectares (accounting for approximately 60%% of the region), and Region D is 32 hectares (accounting for approximately 60% of the region).

Field data

In the study areas A, B and D, a random sampling method was employed to collect above-ground rice samples daily between 10:00 and 15:00 (Figure S1). Each sample point was established using a standardized quadrat measuring 15×15 cm. Within each quadrat, plant height was measured three times with a tape measure and averaged. Geographic coordinates were recorded using OviMap (Android OS). Then samples were immediately sealed in ziplock bags. They were then transferred to cooler box (GINT AS2200, 22L nominal capacity) during transport. All samples were transported to the lab on the same day for Dry weight measurement (0.01 g precision). Due to variations in planting times across fields, the sampling dates were categorized based on the growth stage of the majority of the fields for consistency. The four sampling events correspond to: July 23 (seedling stage), July 30 (tillering stage), August 23 (jointing stage), and September 7 (heading stage). The number of sampling points in these four campaigns was 62, 76, 55, and 47, respectively, resulting in a total of 240 data points.

UAV imagery

UAV RGB and multispectral images were acquired using the DJI Mavic 3M (SZ DJI Technology Co., Ltd., China) with a flight altitude of 80 meters, a forward overlap of 80%, a side overlap of 70%, and a flight speed of 7.1 m/s. The ground resolution of the RGB and multispectral images was approximately 2.26 cm and 3.7 cm, respectively. This resolution is considered low compared to the majority of existing studies (Li *et al.*, 2025a; Liu *et al.*, 2023; Xu *et al.*, 2022). The UAV data collection was conducted between 9:30 and 15:30 Beijing Time. After the flight, image mosaicking and radiometric calibration were completed using Pix4D Mapper.

Feature extraction and selection

VIs and texture features

This study calculated 50 multispectral VIs and 10 RGB VIs (names and formulas of RGB VIs are shown in Table S1). As the number of features increases, the sample size must grow exponentially to ensure statistical power. This necessitates the prior screening of excessive features. Due to the large number of multispectral VIs, we statistically analyzed their correlation with rice AGB (Table S2), preliminarily screening out 11 VIs significantly associated with rice AGB (names and formulas of the screened vegetation indices are listed in Table S3). Subsequently, we analyzed the variable autocorrelation of multispectral and RGB VIs, as shown in Figure 2. Figure S2 and S3 detail the correlation matrix diagrams for the four growth stages, which can be found in the supplementary materials. By removing highly autocorrelated VIs, we finally retained 4 multispectral VIs (NDVI, GNDVI, GSAVI, VIopt) and 4 RGB VIs (ExG, VARI, VEG, KI). Combined with 5 RGB texture features (mean, standard deviation, homogeneity, energy, entropy), a total of 13 feature variables were used to construct the rice AGB prediction model.

Plant height

In UAV remote sensing, the upper boundary of plants can be determined using the Digital Surface Model (DSM), and ground elevation variations are obtained via the Digital Terrain Model (DTM). In this study, rice plant height is derived by calculating the difference between the DSM and DTM generated from UAV RGB images. To avoid misjudgment of the maximum value caused by outliers such as weeds and occlusion, the 95th percentile is adopted as the rice plant height at the sampling points. This method can effectively reflect the actual height of most plants while eliminating a very small number of abnormally high values.

Rice AGB inversion model

To establish a robust prediction model for the rice AGB based on spectral features and to systematically evaluate the applicability of different machine learning algorithms, this study selected six categories encompassing a total of 19 regression algorithms, including linear models, tree-based models, support vector machines, kernel methods, robust regression, and neural networks. All algorithms were implemented using Python's scikit-learn (v1.3.0) and xgboost (v2.0.0) libraries under a unified training workflow and evaluation framework. The specific algorithms and their hyperparameter configurations are described below.

Linear and regularized linear models were employed to explore linear and regularized relationships between spectral features and AGB. This study selected five fundamental models: Linear Regression (LR), Ridge Regression (Ridge), LASSO, ElasticNet, and Bayesian Ridge Regression (BayesianRidge). LR used ordinary least-squares estimation; Ridge and LASSO applied L2 and L1 regularization (with α set to 1.0 and 0.1, respectively) to prevent overfitting; ElasticNet combined both penalties ($\alpha = 0.1$, l1_ratio = 0.5); and BayesianRidge automatically inferred regularization strength via a Bayesian framework. All linear models were trained after Z-score standardization of spectral features, with LASSO and ElasticNet optimized using stochastic coordinate descent.

Tree-based models can capture potential nonlinear relationships between spectral features and AGB. This study selected five ensemble algorithms based on decision trees. Random Forest (RF), ExtraTrees, and Gradient Boosting Regression (GBR) all used 100 base learners, with maximum depths of 10 for RF and ExtraTrees. GBR employed a learning rate of 0.1, a maximum depth of 5, and a subsample ratio of 0.8. AdaBoost regression used 50 weak learners with a learning rate of 1.0. XGBoost was configured with a learning rate of 0.1, a maximum depth of 6, row and column subsampling rates of 0.8, and L1 and L2 regularization terms (coefficients 0.1 and 1.0, respectively) to enhance generalization.

Support Vector Regression (SVR) are advantageous for nonlinear fitting with small sample sizes. This study tested linear kernel functions for SVR, and third-degree polynomial), with the penalty parameter C set to 1.0 and the epsilon-insensitive loss parameter set to 0.1.

In the categories of robust regression and kernel-based methods, this study introduced four algorithms: Huber Regression (Huber), Theil-Sen Regression (Theil-Sen), Kernel Ridge Regression (KRR), and Gaussian Process Regression (GPR). Huber and Theil-Sen were used to mitigate the influence of outliers; KRR (RBF kernel, $\alpha = 1.0$) and GPR (using a combination kernel of a constant kernel × RBF kernel plus white noise kernel) performed regression in high-dimensional space via the kernel trick.

Other regression methods: To broaden the scope of the method comparison, the study also included K-nearest Neighbors Regression (KNN) (K = 5, uniform weighting), Stochastic Gradient Descent Regression (SGDR) (L2 penalty, $\alpha = 0.0001$), partial least squares regression (number of components set to min (10, number of features)), and Multilayer Perceptron (MLP). The MLP had two hidden layers (100 and 50 neurons), employed L2 regularization ($\alpha = 0.01$) and early stopping to prevent overfitting, and adjusted the batch size adaptively based on the sample size.
To fully integrate the predictive advantages of the diverse base models described above, a two-level Stacking ensemble model was further constructed. The first level (base model) consisted of a subset of the aforementioned best-performing and most diverse models. According to the theory of Stacking, a simpler meta-model generally yields better performance of the ensemble (Cheng *et al.*, 2022). Therefore, a simple linear model (BayesianRidge) was adopted as the meta-model to learn the optimal linear combination of the base models' predictions. To avoid overfitting, the predictions of the base models were generated through 5-fold cross-validation. Finally, the meta-model was trained on the cross-validation outputs (out-of-fold predictions) of all base models to obtain robust ensemble predictions.
Through the systematic algorithmic configuration and comparison described above, this study aims to comprehensively evaluate the performance, stability, and applicable scenarios of various models in predicting rice AGB using spectral data, thereby providing a basis for selecting the optimal model.

Model accuracy evaluation

In this study, the coefficient of R^2, root mean square error (RMSE), residual prediction deviation (RPD) and the ratio of performance to interquartile distance (RPIQ) were used as the accuracy evaluation indexes. RPD is the ratio of sample standard deviation (SD) to RMSE:

$$RPD = \frac{SD}{RMSE} \qquad \text{(Eq. 1)}$$

When 1.5 < RPD <2.0, it is considered that the model can only roughly estimate the rice AGB; when 2.0 ≤ RPD <3.0, the model has good prediction ability and is relatively reliable; when RPD ≥3.0, the model has excellent prediction ability and is reliable.
RPIQ considers both the prediction error and the change in observation data. It is a more objective and easier index to compare in model verification. The larger the RPIQ, the stronger the prediction ability of the model. Different from the residual prediction deviation, RPIQ has no assumption on the distribution of observed values. Its formula is as shown in Equation 2:

$$\mathrm{RPIQ} = \frac{\mathrm{IQ}}{\mathrm{RMSE}} \qquad \text{(Eq. 2)}$$

where IQ is the difference between the third and first quartiles.

Automated paddy field extraction and phenology recognition framework

To accomplish the mapping of rice AGB, another task that needs to be completed is the extraction of paddy fields. Traditional classification algorithms such as maximum likelihood and K-Means do not yield satisfactory results for paddy field segmentation (as shown in Figure S4). Therefore, this study proposes a deep learning-based framework for farmland extraction (Figure 3). The framework consists of mask generation, mask classification, and paddy field segmentation. Specifically, the SAM model is first utilized to automatically generate segmentation masks. Unlike traditional deep learning models that require training and fine-tuning specific to each dataset, SAM operates in a zero-shot manner, meaning it can adapt to new images without additional training. By leveraging its pre-trained weights, SAM enables reliable segmentation across a variety of field conditions, eliminating the need to collect training samples or manually annotate large-scale datasets. All experiments were conducted using consistent parameter settings, with image resizing as the only variable. The SAM parameters were defined as follows: points per side = 8, stability score threshold = 0.92, prediction IoU threshold = 0.86, crop-n-layers = 1, and min-mask-region-area = 500. RGB images from 4 study areas across different growth stages were processed in blocks, resulting in a total of 200 images. To save computational resources, the RGB images were subjected to downsampling at various scales before being input into the SAM model for processing, and the impact on segmentation quality was examined. Tests showed that after 10× downsampling, the segmentation performance of SAM did not exhibit a significant decline. Therefore, all 200 images were downsampled by 10× and then input into the SAM model for segmentation. After the initial mask generation, the output often includes extraneous masks with areas much smaller than those of rice fields. During the mask filtering stage, these masks can be filtered out. This step allows us to roughly identify the initial plot candidates.

Then, a multi-class dataset was constructed, which includes non-farmland and paddy fields at different growth stages (seedling stage, tillering stage, jointing stage, heading stage, and maturity stage). To systematically evaluate the feature extraction capabilities of different deep learning architectures, this study selected seven representative models as backbone networks, including VGG16, ResNet18, DenseNet121, MobileNetV3-Small, EfficientNet-B0, ViT-B/16, and DeiT-Small. All models were loaded with their pre-trained weights from the ImageNet-1K dataset to serve as feature extractors, with only the final classification head replaced by a fully connected layer whose output dimension matched the number of categories in this study. The experimental dataset comprised RGB images from multiple categories, generated by randomly cropping UAV images acquired across four distinct periods. Each category contained over 500 samples, with all images resized to 256×256 pixels. A random sampling strategy was employed to split the data of each category into training and validation sets at an approximate ratio of 8:2, ensuring reliable model evaluation. The model was trained for 100 epochs using the AdamW optimizer and CosineAnnealingLR learning rate scheduler.

Subsequently, each candidate mask generated by SAM was classified as follows: A 256×256 pixel sliding window was randomly selected inside the mask, and the number of windows was proportional to the mask size. For masks with an area smaller than 4 times the window area, 1 window was selected; for masks with an area larger than 4 times the window area, the number of windows increased linearly up to a maximum of 10. All sampled windows were overlaid on the UAV visible light images for sampling, and the collected sample images were named following the rule "mask n-m", where "n" represents the mask serial number and "m" denotes the sampling order under the same mask. This naming convention facilitates identifying which mask a sample belongs to via the image name after classification.

All collected samples were fed into the classifier to obtain class posterior probabilities. If a single mask contained multiple windows, a majority vote was used to determine the final label, thereby suppressing false positive results. Finally, all masks labeled "paddy field" were extracted, subjected to morphological smoothing and edge regularization, and ultimately overlaid back onto the original UAV images. Pure RGB and multispectral images of paddy fields were cropped out, completing the paddy field extraction process.

Results

Rice AGB inversion model

To systematically evaluate the predictive capability of machine learning algorithms at different crop growth stages, each dataset was randomly split 100 times according to an 8:2 ratio, and independent experiments were conducted on each split dataset. The mean and standard deviation (StD) of each evaluation metric presented in Table S4 were calculated from the results of these 100 independent experiments on the test sets, aiming to comprehensively reflect the accuracy and stability of the models. The results indicate that the model prediction accuracy shows a significant declining trend as the growth period advances, and the standard deviation of the predictions generally increases, revealing that the models face issues of weakened generalization capability and increased prediction uncertainty in the later growth stages of the crop.

During the seedling stage, all models exhibited good predictive performance. Linear models (such as ElasticNet, Ridge, and BayesianRidge) and PLS performed most prominently at this stage, with relatively high prediction accuracy (R^2 >0.78, RMSE ≈ 1.80-1.85). Moreover, the StD of their various metrics remained at very low levels (e.g., R^2 StD ≤0.071), indicating high robustness in their prediction results. In contrast, some models (such as AdaBoost and XGBoost and KRR) showed relatively greater variability in their prediction outcomes. During the tillering stage, model performance began to diverge, and stability decreased compared to the seedling stage. ExtraTrees achieved the best balance between accuracy and stability at this stage (R^2 =0.773, StD =0.092). SVR and RF also maintained relatively high prediction accuracy, but their results had relatively high standard deviations, indicating that their performance was more sensitive to changes in data subsets. Notably, Theil-Sen and MLR exhibited extremely poor stability in this stage (R^2 StD >0.43), making their prediction results less reliable. During the jointing stage, the predictive ability of the models decreased significantly, with the mean R^2 of most models dropping below 0.6. BayesianRidge and SVR became the most reliable models at this stage, maintaining relatively high prediction accuracy ($R^2 \approx 0.58$) while exhibiting the lowest variability (R^2 StD ≈ 0.12). In contrast, models such as KRR, MLR, and Theil-Sen showed severe performance degradation and result instability (e.g., KRR: R^2 =0.047, StD =0.392) and were no longer suitable for prediction at this stage. During the heading stage, the predictive capability of the models was the poorest, with predictions from most models nearly failing. BayesianRidge, ExtraTrees, and SVR performed relatively good, with BayesianRidge showing a particularly significant stability advantage (R^2 StD = 0.166, RPIQ StD = 0.494). In stark contrast, more than one-third of the models, including GPR, KRR, and KNN, had mean R^2 values dropping to zero or negative, meaning their prediction results were not only inaccurate but also completely unreliable.

Comprehensive analysis indicates that the prediction accuracy and stability of the models are significantly dependent on the growth stage. In the early stages of crop growth, various models can achieve high-accuracy, high-stability predictions. As the growth process advances, the saturation effect of the spectra intensifies, and the correlation between spectral features and rice AGB declines sharply, leading to a general decrease in model performance and increased variability in results. Against this background, this study constructed a Stacking ensemble model (using BayesianRidge, ExtraTrees, SVR, RF, and XGBoost as base learners, which showed the best overall performance) and conducted experiments under the same conditions, with the results shown in Table 1. From the table, it can be seen that Stacking algorithm demonstrated comprehensive advantages over the previously evaluated individual models across all four growth stages. Its prediction accuracy and stability were systematically improved at different growth stages.

During the seedling stage, the Stacking model achieved the best predictive performance among all models, with the lowest mean RMSE (1.748) and the highest mean R^2 (0.793). Moreover, its R^2 StD (0.066) was lower than that of the previously best-performing regularized linear models (e.g., ElasticNet R^2 StD = 0.060, PLS R^2 StD =0.071), indicating that while maintaining the highest accuracy, its prediction stability was also enhanced. The mean RPD and RPIQ reached 2.302 and 3.124, respectively, also higher than other models. During the tillering stage, the mean R^2 of the Stacking model (0.781) was higher than that of the previously best-performing individual model, ExtraTrees (0.773), and its mean RMSE (2.983) was lower than ExtraTrees (2.934). The StD of the Stacking model for R^2 (0.079) and RMSE (0.822) were both smaller than those of ExtraTrees, effectively mitigating potential significant fluctuations in results. During the jointing stage, model performance generally declined, but the StD of R^2 for the Stacking model (0.101) was lower compared to BayesianRidge (0.123) and SVR (0.111), indicating better prediction consistency. This suggests that, through the integration strategy, the Stacking model effectively combined the strengths of different algorithms, achieving more robust and accurate predictive capabilities than any single base model. During the heading stage, the mean R^2 of the Stacking model (0.421) remained higher than that of the best individual models (BayesianRidge and ExtraTrees), and it had the lowest StD across all metrics. This demonstrates that, compared to any previously evaluated individual machine learning algorithm, the Stacking ensemble algorithm not only maintained the highest prediction accuracy across different growth stages but, more importantly, significantly reduced the uncertainty of the prediction results. Figure 4 more intuitively reflects the superior performance of the Stacking model. The violin plot of the Stacking model is more compact and symmetric than those of the other five base models, indicating greater stability. This provides an effective technical pathway for constructing predictive models suitable for real agricultural environments, combining high accuracy with high robustness.

Gain of plant height in AGB prediction models

Figure 5 presents the results of a correlation analysis between rice plant height extracted via UAV technology and that measured in the field. As shown in the figure, the correlation between extracted and measured plant heights during the seedling stage is relatively low, with R^2 of only 0.577. This may be attributed to the small size of rice plants at the seedling stage; when the UAV flies at an altitude of 80 meters, the resolution of the captured images is limited, resulting in low accuracy of extracted plant height. However, as rice enters the tillering stage, the correlation between extracted and measured plant heights improves significantly, with R^2 all above 0.820, and reaching 0.878 at the heading stage. This indicates that starting from the tillering stage, rice plant height extracted by UAV can accurately reflect the actual plant height and thus serve as an effective method for plant height acquisition.

We systematically compared Stacking model performance before and after including plant height to quantify the contribution of plant height to AGB prediction across rice growth stages (Table 2). Overall, the incorporation of plant height led to a significant improvement in predictive accuracy at every stage, underscoring its critical importance. At the seedling and tillering stages, all evaluation metrics increased when plant height was added. The R^2 rose from 0.793 to 0.819 and RPD from 2.302 to 2.432 at the seedling stage. It is R^2 from 0.781 to 0.844 and RPD from 2.298 to 2.827 at tillering. During the jointing and heading stages the gain became even more pronounced. R^2 values exceeded 0.7 once plant height was included. Stacking rose from R^2 = 0.421 to 0.739 and RPIQ from 1.995 to 3.015 at heading, demonstrating that plant height is a decisive predictor in mid- and late-season AGB estimation.

Figure 6 illustrates the dynamic importance of each variable within the five base models. It can be observed that the importance of plant height is relatively low at the seedling stage. During this stage, rice growth primarily focuses on root and leaf expansion, resulting in a weak correlation between plant height and AGB. The importance of plant height increases significantly from the tillering stage onward, and it becomes the most important feature during the jointing and heading stages. This corresponds to the developmental phase of rapid stem elongation and maximum canopy height. This growth transition tightly couples the accumulation of plant height with AGB gain, thereby making plant height the dominant explanatory variable for predicting AGB during the reproductive growth stages. In summary, since the accuracy of plant height extraction is relatively low at the seedling stage and its contribution to rice AGB prediction is limited, the tillering stage can be regarded as a demarcation point. Starting from this stage, plant height should be incorporated as a feature variable into the rice AGB inversion model.

Automated paddy field extraction and phenology recognition results

Rice AGB mapping aims to monitor areas with abnormal rice growth. Field surveys reveal inconsistent growth stages of rice across different plots; therefore, it is necessary to first clarify the growth stage of rice in each plot when assessing rice growth status. Additionally, since the decision to incorporate plant height into rice AGB prediction depends on the rice's growth stage, so the classification model was defined as a multi-classification model, with categories designated as non-paddy field, seedling stage, tillering stage, jointing stage, heading stage, and maturity stage. To evaluate the framework for paddy field extraction and phenological stage recognition proposed in this study, the acquired UAV RGB images were first cropped into patches and input into the SAM to obtain category-agnostic masks. After obtaining the segmentation results, non-paddy field masks were removed by comparing them with the original RGB images. Meanwhile, paddy field masks for each image were manually delineated via visual interpretation to serve as the ground truth masks. Finally, the segmentation performance of the SAM was evaluated using these visually interpreted ground truth masks, and the mean values of relevant evaluation metrics for the 200 images are listed in Table 3. As shown in the table, the SAM performed excellently, with the mean values of all metrics approaching 1, indicating that the model could achieve accurate segmentation in most cases. However, the presence of standard deviations suggests a certain degree of variability in the data. For instance, the standard deviations of IoU and Recall were relatively large, which may imply that the model's performance could decrease significantly under certain circumstances. Thus, the modified Z-Score method was employed to detect outliers in the IoU values, and the results are presented in Figure S5. As illustrated in the figure, the modified Z-Score values of most data points fell within the normal range, indicating no significant deviation from the central tendency of the dataset. In contrast, 20 data points exhibited extremely high modified Z-Score values exceeding the threshold of 3, which were marked as outliers, meaning they deviated significantly from the rest of the dataset. The images corresponding to these 20 data points and their SAM segmentation results were inspected to analyze the causes of anomalies, with examples plotted in Figure 7. The reasons for low segmentation accuracy were categorized into two types: one was under-detection, which mostly occurred in incomplete small paddy field patches at the edges of images and could be ignored in practical applications. The other type is over-segmentation, which typically occurs in images predominantly featuring pure paddy fields, leading to mask adhesion (where adjacent masks merge into large connected regions). In the context of this study's application scenario, images are highly unlikely to consist solely of pure paddy fields. Consequently, such over-segmentation phenomena do not manifest in practical applications.

Each classification model was trained using the custom-built dataset and evaluated on the validation set. The resulting performance metrics are reported in Table 4. All models successfully completed the classification task, except VGG16. Among them, the MobileNetV3 Small model exhibited the best comprehensive performance. Its Top-1 Accuracy score was only 0.004 lower than that of the DenseNet121 model (which achieved the highest score), while its number of parameters was significantly smaller than that of other models. Meanwhile, by examining the loss rate changes during the training process of each model (Figure S6), it can be observed that the MobileNetV3-Small model converged rapidly and maintained stability. Thus, the MobileNetV3-Small model was selected as the classification model in the paddy field extraction framework proposed in this study. To further verify the performance of the MobileNetV3-Small, a test dataset was constructed, with 100 samples per category. Table 5 shows the performance of the MobileNetV3-Small model on the test dataset, evaluated comprehensively using overall accuracy (OA), producer's accuracy (PA), user's accuracy (UA), and the Kappa coefficient.

The OA reached 99.5% and the Kappa coefficient was 0.994, indicating that the model met the requirements for mapping. For paddy fields, the PA for the seedling stage was 100%. This is crucial for rice AGB mapping, as it enables accurate determination of whether the plant height variable should be incorporated into AGB prediction for each plot.

Rice AGB mapping

Finally, the automated mapping of rice AGB in each study area was conducted using the methodology described in this study. The proposed framework can complete paddy field extraction and phenological stage recognition for a single study area within minutes (Figure 8). Visual verification confirmed accurate identification of field boundaries and zero errors in phenological staging, demonstrating the robustness of the framework. Subsequently, the AGB model was invoked to generate AGB distribution maps (Figure 9). By comparing the predicted values with the normal AGB thresholds for each growth stage, anomalous growth areas can be rapidly identified, thereby providing timely and reliable guidance for variable-rate management.

Discussion

Although substantial research has been conducted on remote sensing monitoring of rice AGB, most existing studies have focused on small geographical scales, such as experimental plots or small fields (Li *et al.*, 2025b; Xu *et al.*, 2022). When the monitoring scope expands to a field scale covering hundreds or even thousands of acres, automated monitoring of rice AGB remains relatively understudied. It is argued that achieving monitoring at this scale requires addressing three core challenges. Firstly, the trade-off between accuracy and efficiency. To meet the needs of large-area monitoring at field scales, the flight altitude of UAV must be increased, which directly reduces image resolution. Low-resolution data can easily miss key details related to crop growth (Hu *et al.*, 2021; Zhang *et al.*, 2025b). Under these conditions, how to develop a stable and high-accuracy rice AGB prediction model based on low-resolution imagery remains a challenge. Secondly, the cost-effective transferability of automated paddy field extraction. UAV images at field scales inevitably include non-paddy areas such as woodlands, buildings, and water bodies. Automated rapid extraction of paddy fields is essential before AGB monitoring can be performed. More critically, existing extraction methods often rely on annotated data from specific regions, making it difficult to transfer them to new monitoring areas at low cost, which hinders large-scale application of the technology (Li *et al.*, 2025b; Liu *et al.*, 2024). Thirdly, the differentiated recognition of rice phenological stages. Due to variations in planting times and rotation patterns between early and late rice crops, rice phenology often varies significantly within a region. Because growth stages and AGB accumulation patterns differ greatly across phenological phases, directly comparing AGB values without accounting for these differences is not scientifically sound (Garcia *et al.*, 2025). Therefore, accurately identifying the phenological stage of rice is necessary to support reasonable estimation and comparison of AGB. To address these issues, this study proposes a deep learning-based framework for paddy field extraction and phenological recognition, combined with ensemble learning regression algorithms and plant height features to estimate rice AGB. Ultimately, an automated rice AGB mapping approach is developed to support the advancement of UAV-based remote sensing automated monitoring technology.

In terms of rice AGB estimation, this study compared various regression algorithms. The results indicated that the Stacking model performed the best, exhibiting high prediction accuracy and stability. In this study, DSM and DTM were generated via 3D reconstruction technology from UAV RGB images, and rice plant height was derived by calculating the difference between the two models. Experimental results showed that starting from the tillering stage, the correlation between UAV-extracted plant height and field-measured plant height improved significantly, enabling accurate reflection of actual plant height. The effective contribution of plant height to crop AGB inversion has also been confirmed in other study (Wang *et al.*, 2023). However, the remote sensing data acquired in this study were limited to RGB and multispectral imagery, which restricted the exploration of additional features such as the three-dimensional structural information from LiDAR (Meng *et al.*, 2025), canopy temperature and plant water content from thermal infrared data (Liu *et al.*, 2025), as well as more detailed spectral information from hyperspectral sensors (Xu *et al.*, 2023). Future research could incorporate these sensors to further analyze features that significantly contribute to the prediction of rice AGB at field scales, thereby enabling the development of more accurate estimation models.

The paddy field extraction framework proposed in this study demonstrated good performance on UAV images from multiple study areas. As the core component of the framework, the segment anything model (SAM) can automatically and rapidly generate segmentation masks without additional training for each dataset. This greatly reduces the workload of data annotation and enhances the model's adaptability and generality. After mask generation, a classification model was used to classify the masks, accurately excluding non-paddy field masks and conducting phenological identification of paddy fields. Experimental results showed that the framework can effectively extract paddy field areas, providing a reliable foundation for subsequent AGB mapping.

Although the paddy field extraction framework and AGB estimation model proposed in this study demonstrated good performance in experiments, there are still some limitations. For instance, during the segmentation process of the SAM model, despite its overall good performance, outliers (such as under-detection and over-detection) may still occur in certain cases. This may be related to image quality, complexity, and model parameter settings. Future work can further optimize the parameters of the SAM model and explore more effective outlier detection and processing methods to improve the model's robustness. Additionally, although this study conducted experimental validation in multiple study areas, the sample size and regional scope remain limited. A recent study utilized UAV imagery with a ground resolution comparable to ours (Zhang *et al.*, 2026). This research, covering multiple growing seasons and geographical regions, also achieved satisfactory inversion performance. However, it did not incorporate plant height, and each study plot was relatively small (several tens of square meters). Therefore, future work should involve the accumulation of multi-year and multi-region data to further validate the reliability of the framework proposed in this study. Meanwhile, integrating more feature variables (e.g., soil moisture and meteorological data) may further improve the accuracy of rice AGB estimation (Li *et al.*, 2024). Finally, with the ongoing advancement of deep learning technologies, exploring more advanced model architectures and algorithms (e.g., novel zero-shot learning approaches) will represent an important and promising direction for future research.

Conclusions

i) The rice AGB prediction model built based on the Stacking ensemble algorithm exhibits good performance. The model's prediction accuracy is significantly improved after incorporating the plant height feature, and this improvement is particularly notable during the middle and late stages of rice growth. For instance, at the heading stage, the model's R^2 increases from 0.421 to 0.739 and RPIQ from 1.995 to 3.015.

ii) The automated paddy field extraction and phenology recognition framework ("SAM + MobileNetV3-Small") constructed in this study achieves good performance in both automatic paddy field extraction and phenology recognition. The SAM model attains a segmentation accuracy of 0.968, the classifier reaches an IoU of 0.968, and the total number of parameters is merely 1.524M. Moreover, this framework eliminates the need for manual annotation, significantly outperforming traditional methods and heavyweight networks.

References

Adeluyi O, Harris A, Foster T, Clay GD, 2022. Exploiting centimetre resolution of drone-mounted sensors for estimating mid-late season above ground biomass in rice. Eur J Agron 132:126411.

Agrawal J, Arafat MY, 2024. Transforming farming: a review of AI-powered UAV technologies in precision agriculture. Drones 8:664.

Cano PB, Carcedo AJP, Hernandez CM, Gomez FM, Gimenez VD, Kyveryga PM, et al., 2025. Trends in agricultural technology: a review of US patents. Precis Agric 26:59.

Cheng Q, Xu H, Fei S, Li Z, Chen Z, 2022. Estimation of maize LAI using ensemble learning and UAV multispectral imagery under different water and fertilizer treatments. Agriculture 12:1267.
Dreier A, Lopez G, Bajracharya R, Kuhlmann H, Klingbeil L, 2025. Structural wheat trait estimation using UAV-based laser scanning data: analysis of critical aspects and recommendations based on a case study. Precis Agric 26:18.
Garcia AD, Islam S, Prudente VHR, Sanches ID, Cheng I, 2025. Irrigated rice-field mapping in Brazil using phenological stage information and optical and microwave remote sensing. Appl Comput Geosci 25:100223.
Hu PC, Chapman SC, Zheng BY, 2021. Coupling of machine learning methods to improve estimation of ground coverage from unmanned aerial vehicle (UAV) imagery for high-throughput phenotyping of crops. Funct Plant Biol 48:766-779.
Huang Y, Thomson SJ, Hoffmann WC, Lan Y, Fritz BK, 2013. Development and prospect of unmanned aerial vehicle technologies for agricultural production management. Int J Agric Biol Eng 6:1-10.
Kroupova ZZ, Aulova R, Rumankova L, Bajan B, Cechura L, Simek P, et al., 2025. Drivers and barriers to precision agriculture technology and digitalisation adoption: meta-analysis of decision choice models. Precis Agric 26:17.
Li B, Xu X, Zhang L, Han J, Bian C, Li G, et al., 2020. 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.
Li J, Li J, Zhao D, Cao Q, Yu F, Cao Y, et al., 2025a. High-throughput method for improving rice AGB estimation based on UAV multi-source remote sensing image feature fusion and ensemble learning. Front Plant Sci 16:1576212.
Li J, Wei Y, Wei T, He W, 2025b. A comprehensive deep-learning framework for fine-grained farmland mapping from high-resolution images. IEEE Trans Geosci Remote Sens 63:5601215.
Li W, He J, Yu M, Su X, Wang X, Zheng H, et al., 2024. Multisource remote sensing data-driven estimation of rice grain starch accumulation: leveraging matter accumulation and translocation characteristics. IEEE Trans Geosci Remote Sens 62:1-18.
Liu Y, Feng H, Yue J, Jin X, Fan Y, Chen R, et al., 2023. Improved potato AGB estimates based on UAV RGB and hyperspectral images. Comput Electron Agric 214:108260.
Liu Y, Pei J, Zou Y, Tan S, He Y, Zheng X, et al., 2025. A novel Hybrid-DCNN-based framework for enhanced rice aboveground biomass estimation under limited samples. IEEE Trans Geosci Remote Sens 63:1-16.
Liu Y, Zhang T, Huang Y, Shi F, 2024. An edge-aware multitask network based on CNN and Transformer backbone for farmland instance segmentation. IEEE J Sel Top Appl Earth Obs Remote Sens 17:13765-13779.
Maimaitijiang M, Sagan V, Sidike P, Maimaitiyiming M, Hartling S, Peterson KT, et al., 2019. Vegetation index weighted canopy volume model (CVMVI) for soybean biomass estimation from unmanned aerial system-based RGB imagery. ISPRS J Photogramm Remote Sens 151:27-41.

Meng L, Ming B, Liu Y, Nie CW, Fang L, Zhou LL, et al., 2025. Maize biomass estimation by integrating spectral, structural, and textural features from unmanned aerial vehicle data. Eur J Agron 168:127647.

Niu YX, Song XY, Zhang LY, Xu LZ, Wang AC, Zhu QZ, 2025. Enhancing model accuracy of UAV-based biomass estimation by evaluating effects of image resolution and texture feature extraction strategy. IEEE J Sel Top Appl Earth Obs Remote Sens 18:878-891.

Raouhi EM, Lachgar M, Hrimech H, Kartit A, 2023. Unmanned aerial vehicle-based applications in smart farming: a systematic review. Int J Adv Comput Sci Appl 14:1150-1165.

Shi J, Yang K, Yuan N, Li Y, Ma L, Liu Y, et al., 2025. UAV-based rice aboveground biomass estimation using a random forest model with multi-organ feature selection. Eur J Agron 164:127529.

Wang D, Li R, Zhu B, Liu T, Sun C, Guo W, 2023. Estimation of wheat plant height and biomass by combining UAV imagery and elevation data. Agriculture 13:9.

Xu F, Yao X, Zhang K, Yang H, Feng Q, Li Y, et al., 2024. Deep learning in cropland field identification: a review. Comput Electron Agric 222:109042.

Xu L, Zhou L, Meng R, Zhao F, Lv Z, Xu B, et al., 2022. An improved approach to estimate ratoon rice aboveground biomass by integrating UAV-based spectral, textural and structural features. Precis Agric 23:1276-1301.

Xu TY, Wang FM, Shi Z, Xie LL, Yao XP, 2023. Dynamic estimation of rice aboveground biomass based on spectral and spatial information extracted from hyperspectral remote sensing images at different combinations of growth stages. ISPRS J Photogramm Remote Sens 202:169-183.

Yang K, Mo J, Luo S, Peng Y, Fang S, Wu X, et al., 2023. Estimation of rice aboveground biomass by UAV imagery with photosynthetic accumulation models. Plant Phenomics 5:0056.

Zhang J, Yuan N, Liu Y, Wu X, Fang S, Peng Y, et al., 2026. Multi-temporal UAV rice AGB estimation with dynamic biomass allocation and interpretable deep learning. Eur J Agron 177:128096.

Zhang PP, Lu B, Ge JY, Shang JL, Tan CW, Wang XY, et al., 2025. A sustainable learning framework: UAV-based oat chlorophyll monitoring using radiative transfer models and deep learning techniques. IEEE Trans Geosci Remote Sens 63:4416613.

Zhang RN, Jin SC, Wang Y, Zang JR, Wang Y, Zhao RF, et al., 2025. PhenoSR: enhancing organ-level phenotyping with super-resolution RGB UAV imagery for large-scale field experiments. ISPRS J Photogramm Remote Sens 228:582-602.

Zhu S, Zhang W, Yang T, Wu F, Jiang Y, Yang G, et al., 2024. Combining 2D image and point cloud deep learning to predict wheat above ground biomass. Precis Agric 25:3139-3166.

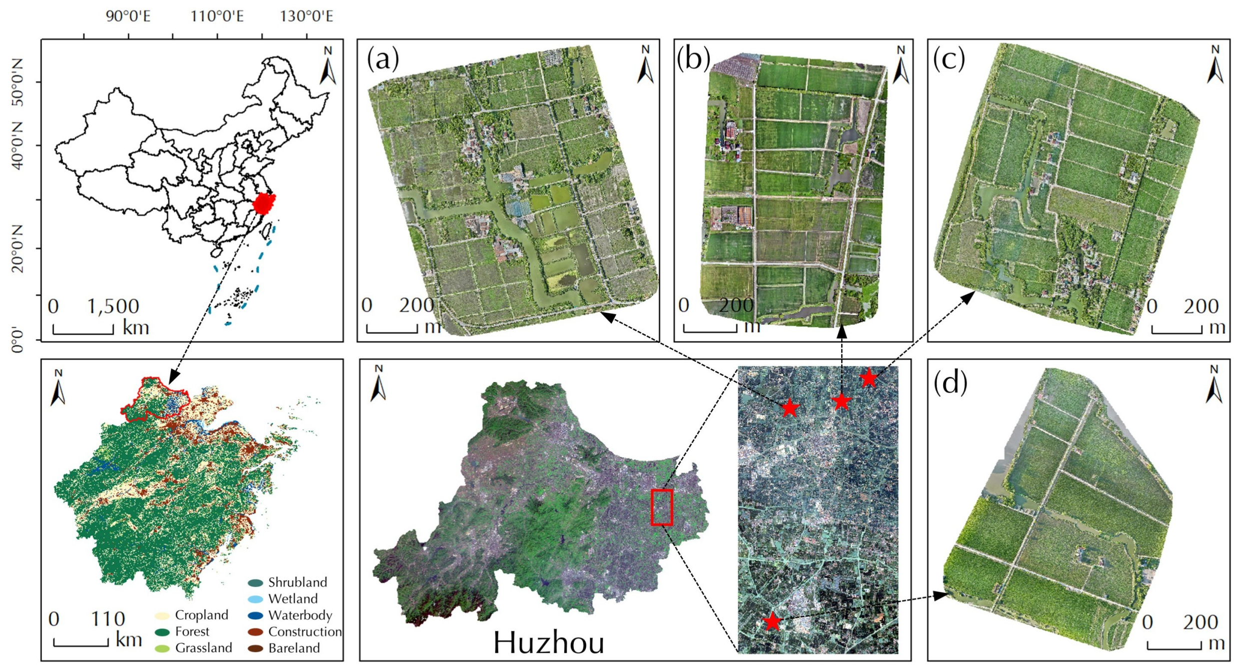

Figure 1. Location and UAV remote-sensing images of the study area.

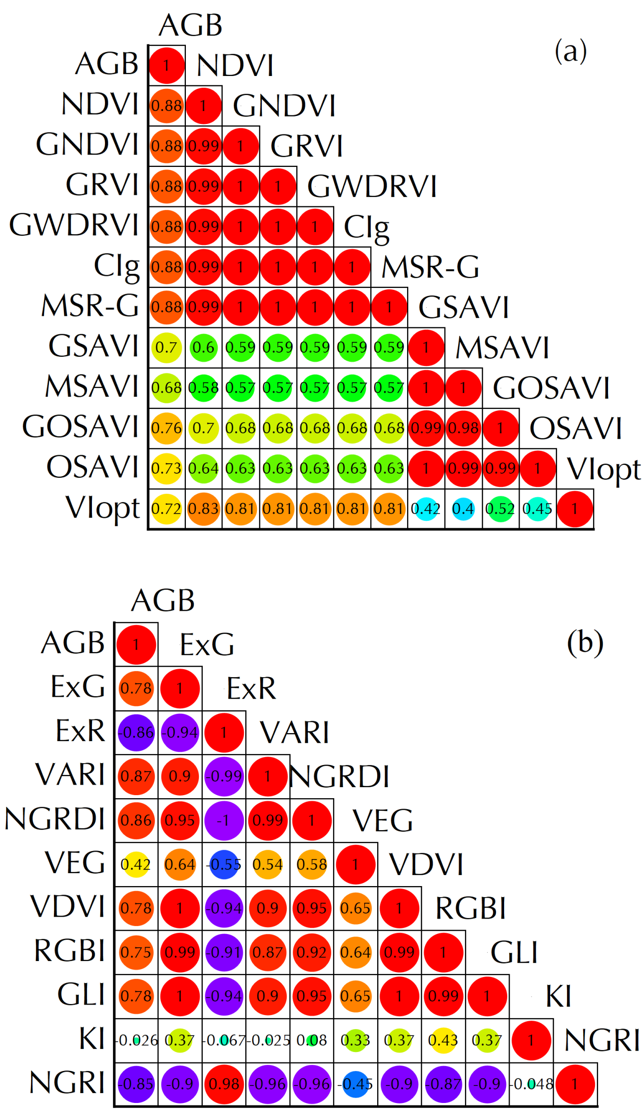

Figure 2. VIs autocorrelation matrix diagram. a) Multispectral Vis. b) RGB VIs.

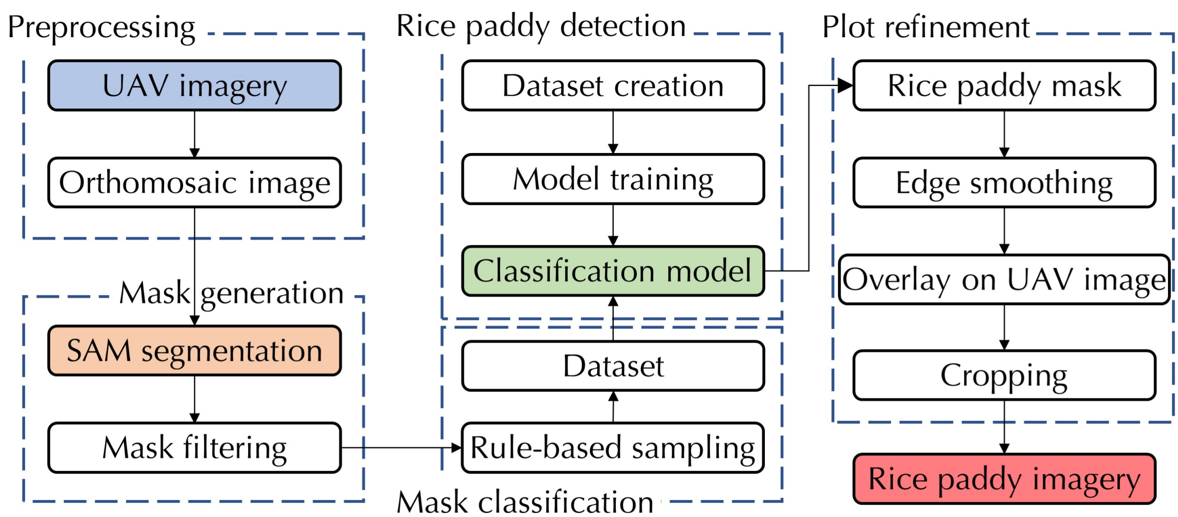

Figure 3. Flowchart of the proposed plot extraction framework.

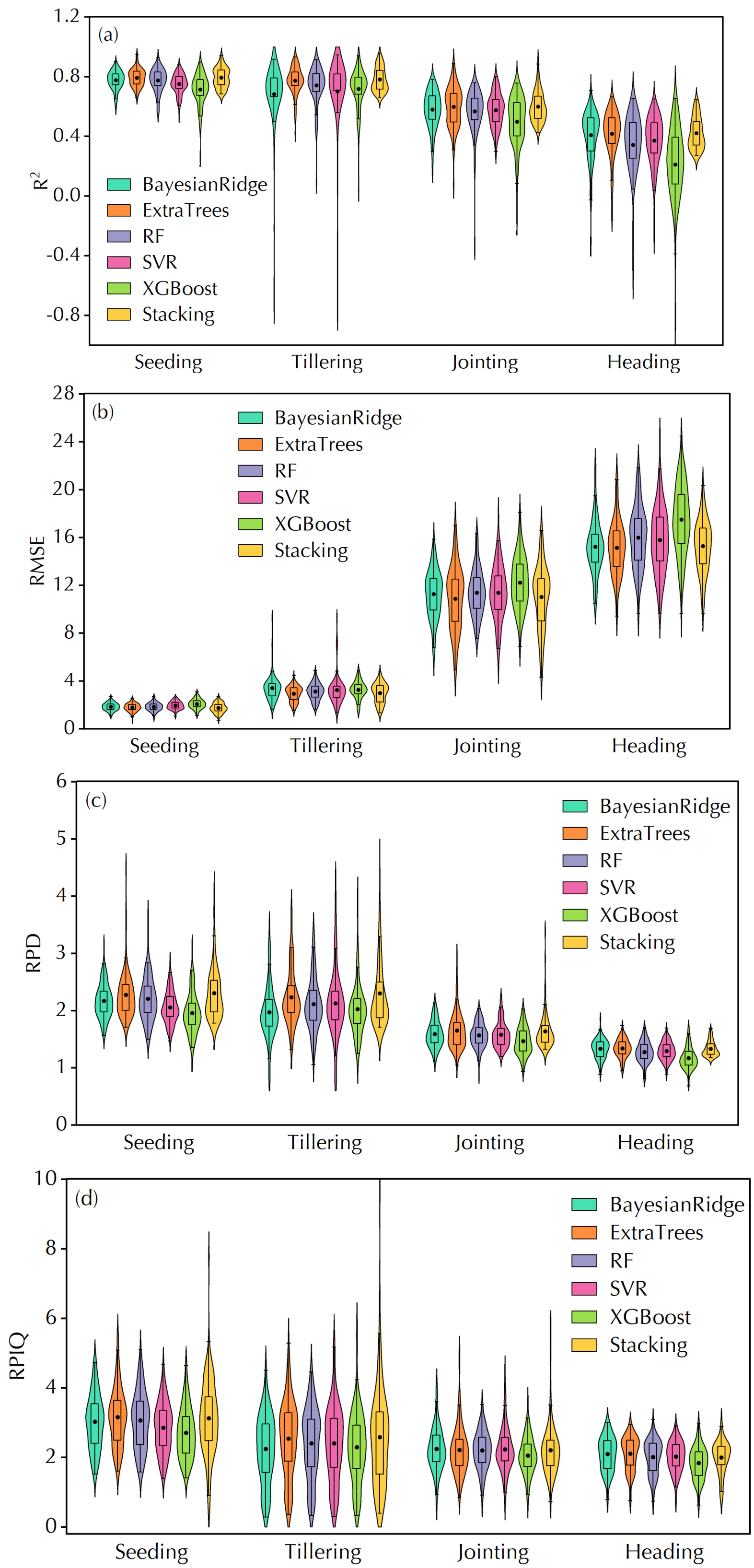

Figure 4. AGB inversion model estimation accuracy on the test set.

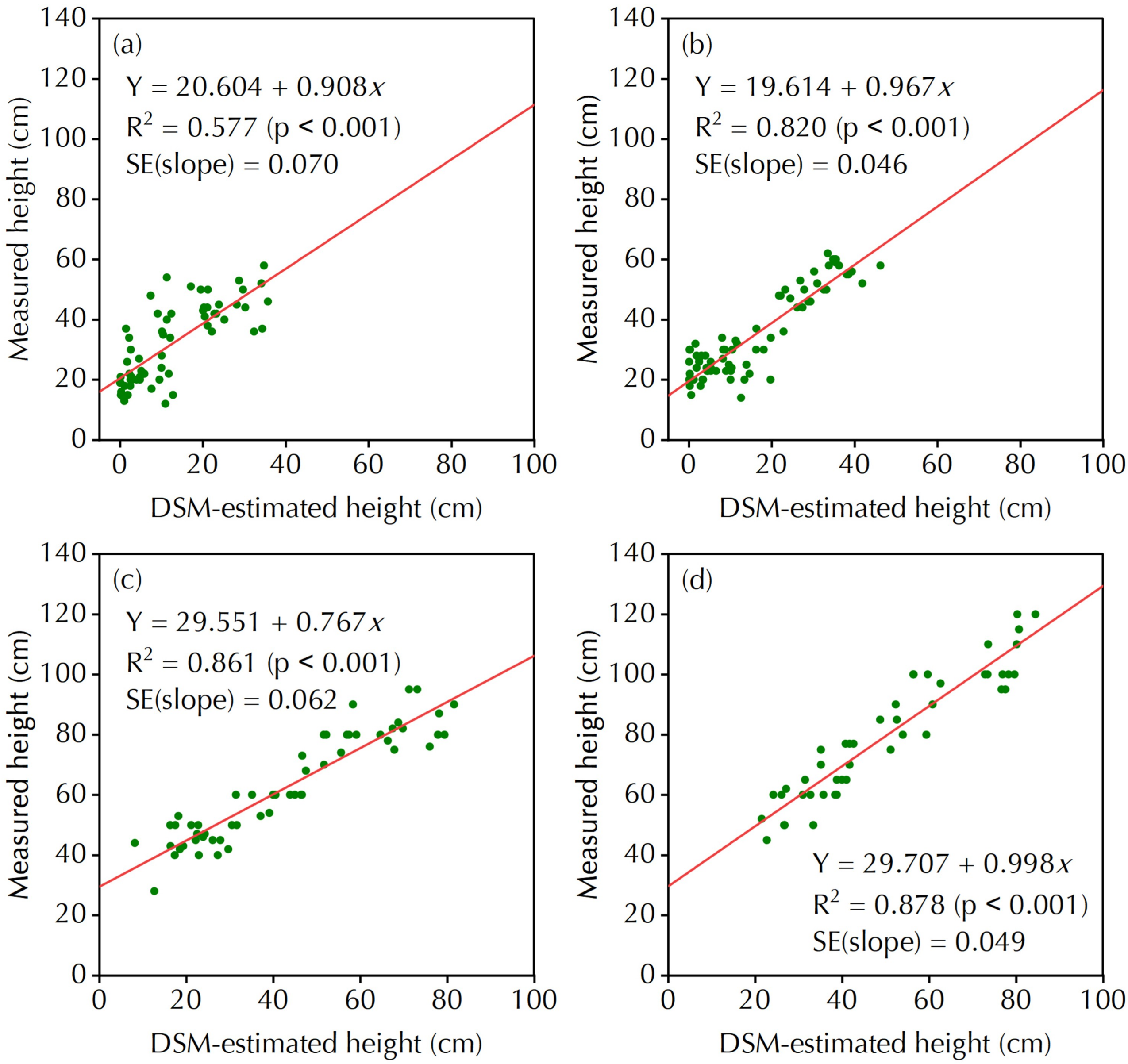

Figure 5. Accuracy statistics chart for plant height extraction from UAV imagery. a) Seedling. b) Tillering. c) Jointing. d) Heading.

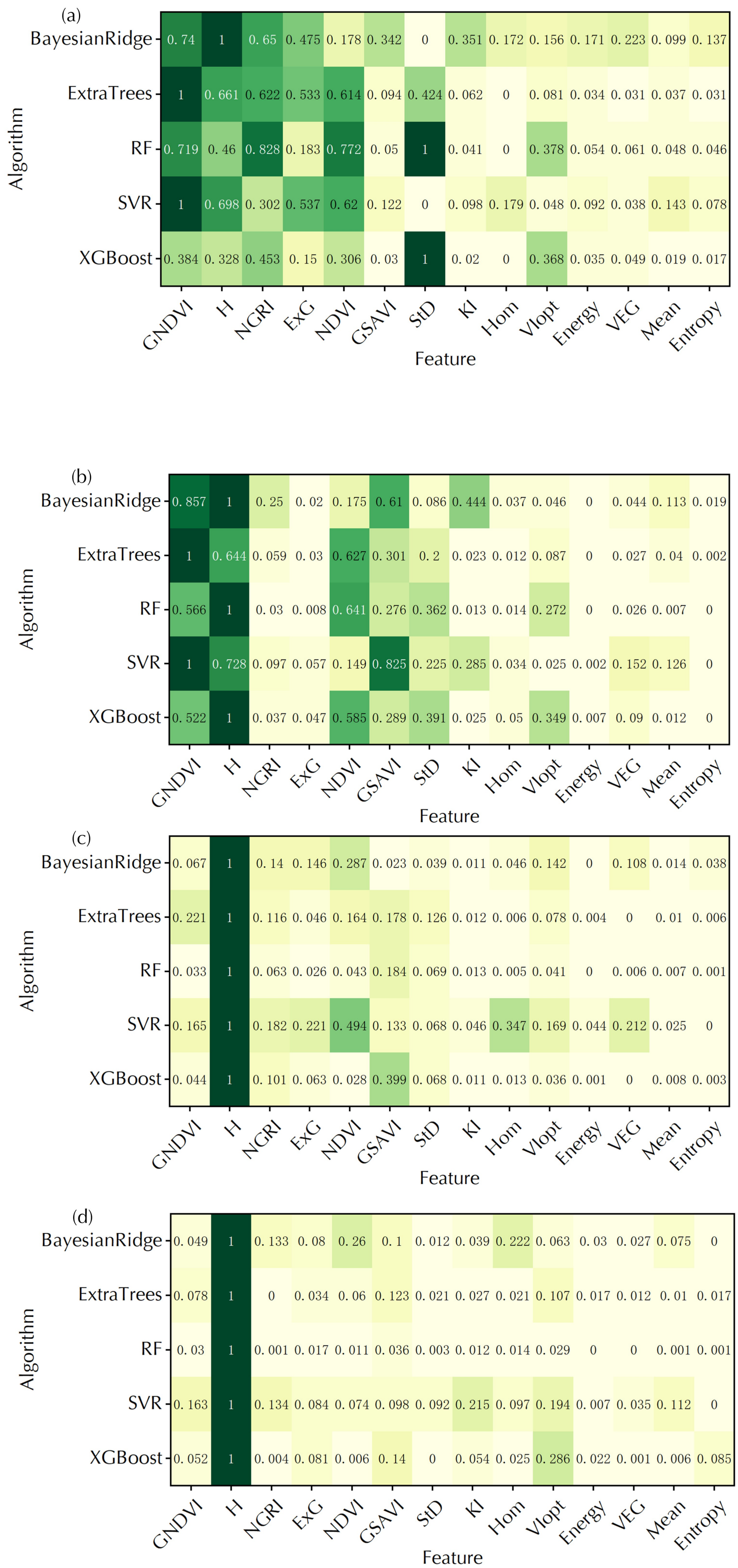

Figure 6. Heatmap of model feature importance across different growth stages. Values in the figure represent the normalized importance of each feature in different models. (a), (b), (c), and (d) denote the seedling, tillering, jointing, and heading, respectively.

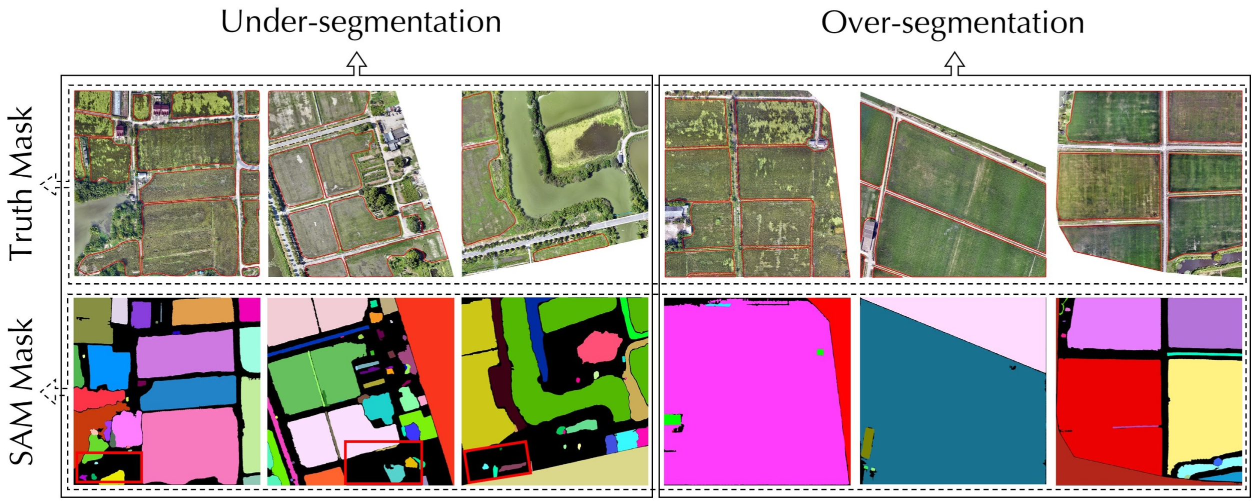

Figure 7. Example masks corresponding to SAM segmentation outliers.

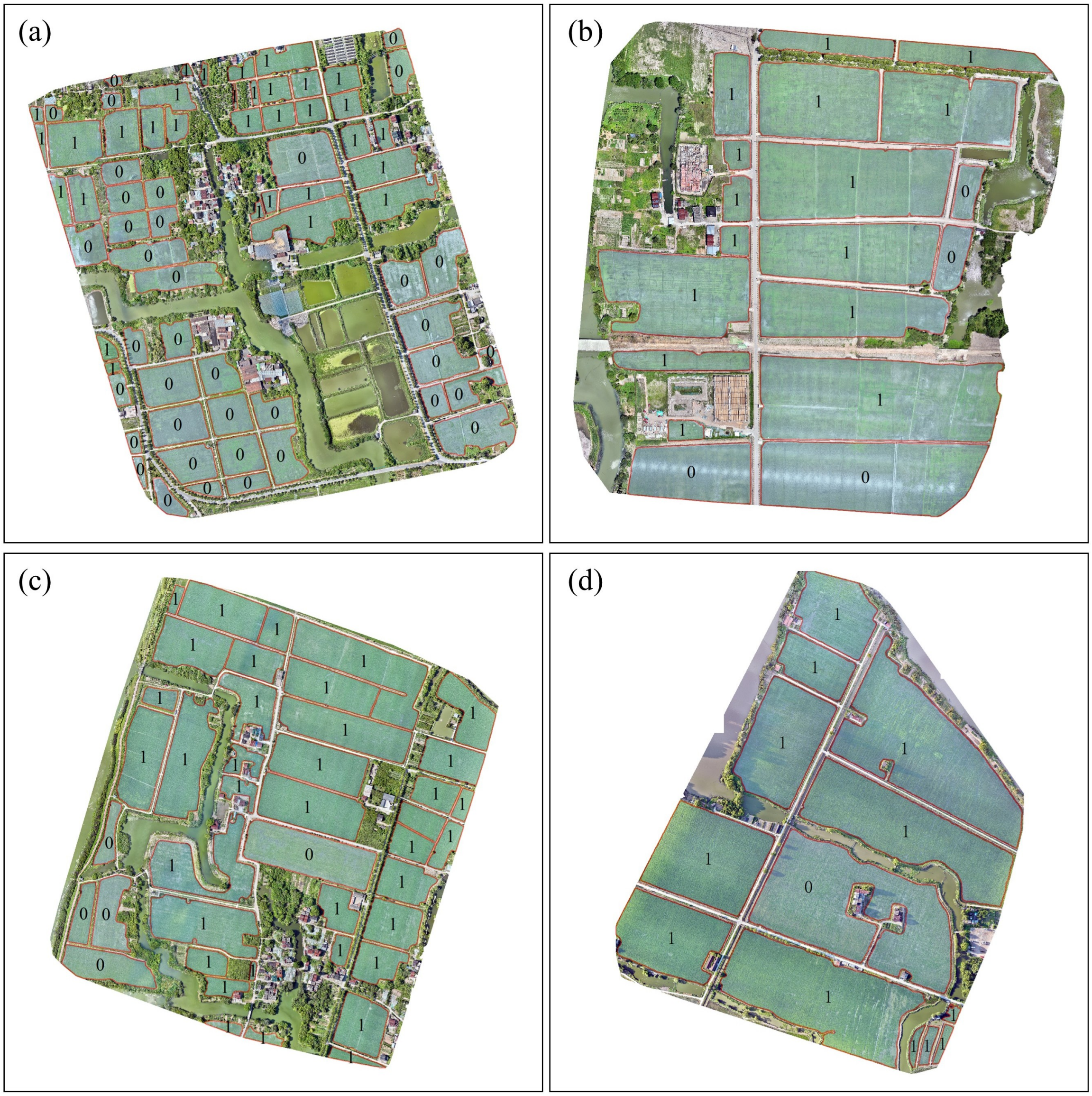

Figure 8. Results of the proposed framework applied to UAV imagery acquired on July 30th. The numbers labeled on each mask represent growth stages: 0 denotes the seedling stage, and 1 denotes the tillering stage.

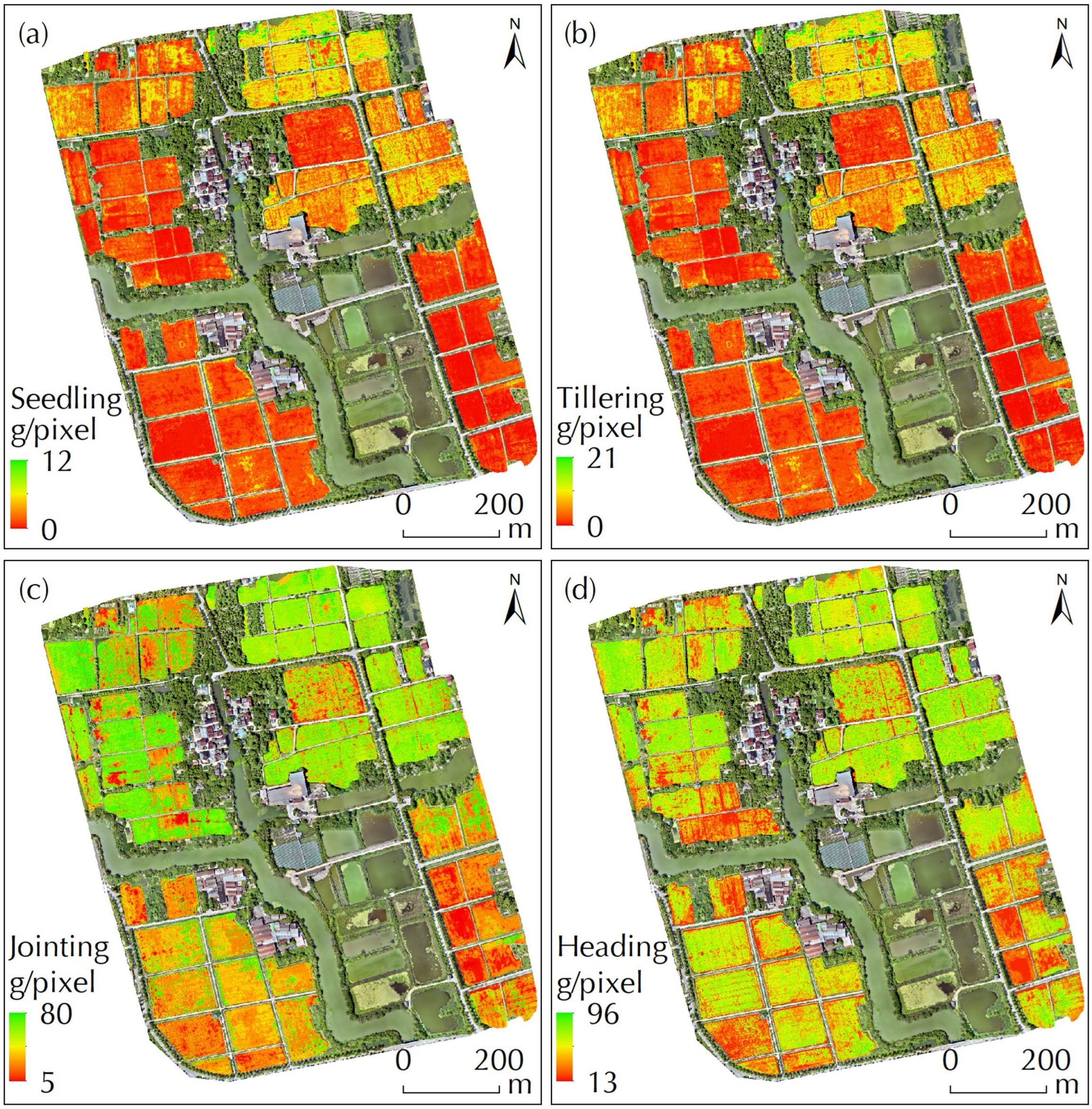

Figure 9. Schematic diagram of Rice AGB prediction results in region A. The pixel size of the biomass map was configured as 0.15 m × 0.15 m.

Table 1. Test set accuracy statistics of Stacking for AGB estimation. The table shows the mean and StD of 100 runs, with the dataset redivided in each iteration.

Period	RMSE		R^2		RPD		RPIQ	
	Mean	StD	Mean	StD	Mean	StD	Mean	StD
Seedling	1.748	0.402	0.793	0.066	2.302	0.454	3.124	1.080
Tillering	2.983	0.822	0.781	0.079	2.298	0.607	2.584	1.447
Jointing	11.019	2.649	0.598	0.101	1.635	0.310	2.208	0.720
Heading	15.277	2.173	0.421	0.103	1.332	0.131	1.995	0.480

Table 2. Accuracy statistics of models on the test set after incorporating plant height. The table shows the mean and StD of 100 runs, with the dataset redivided in each iteration.

Period	RMSE		R^2		RPD		RPIQ	
	Mean	StD	Mean	StD	Mean	StD	Mean	StD
Seedling	1.645	0.299	0.819	0.049	2.432	0.392	3.296	1.212
Tillering	2.447	0.822	0.844	0.074	2.827	0.856	3.085	1.813
Jointing	8.355	1.655	0.777	0.059	2.182	0.343	2.929	0.835
Heading	10.198	1.750	0.739	0.065	2.009	0.288	3.015	0.878

Table 3. Segmentation performance of the SAM on 200 UAV images.

	IoU	Dice	Acc	Recall
Mean	0.968	0.978	0.982	0.970
StD	0.115	0.106	0.074	0.112

Table 4 .Performance comparison of classification models after 100 epochs of training on the validation set.

	Top-1 Acc	Macro-F1	Weighted-F1	Cohen's κ	AUC	Params
VGG16	0.177	0.050	0.053	0.0000	0.500	134.285M
Resnet18	0.993	0.993	0.993	0.991	0.999	11.180M
DenseNet121	0.997	0.997	0.997	0.996	0.999	6.960M
EfficientNet-B0	0.994	0.994	0.994	0.993	0.999	4.015M
ViT-B/16	0.979	0.979	0.979	0.975	0.998	85.803M
DeiT	0.970	0.970	0.970	0.964	0.998	21.668M
MobileNetV3-small	0.993	0.993	0.992	0.991	0.999	1.524M

Table 5. Confusion matrix of classification results on validation samples.

	Non-paddy area	Seedling	Tillering	Jointing	Panicle initiation	Maturity	User' s Acc (%)
Non-paddy Area	98	2	0	0	0	0	100
Seedling	0	100	0	0	0	0	98
Tillering	0	0	99	1	0	0	100
Jointing	0	0	0	100	0	0	99
Panicle Initiation	0	0	0	0	100	0	100
Maturity	0	0	0	0	0	100	100
Producer's Acc/%	98	100	99	100	100	100	
Overall Acc/%				99.5			
Kappa Coefficient				0.994			