Journal of Zhejiang University-SCIENCE B (Biomedicine & Biotechnology) in press
www.jzus.zju.edu.cn; www.springer.com/journal/11585
E-mail: jzus_b@zju.edu.cn

Research Article

https://doi.org/10.1631/jzus.B2500830

Enhancing rapeseed biomass and yield estimation with ensemble learning and synergistic multidimensional features

Yanni ZHANG[1], Xiaoyu CHAI[1,2], Jinpeng HU[1], Yaxiao NIU[1]✉, Lizhang XU[1,2]✉

[1]*College of Agricultural Engineering, Jiangsu University, Zhenjiang 212000, China*
[2]*Key Laboratory for Theory and Technology of Intelligent Agricultural Machinery and Equipment, Jiangsu University, Zhenjiang 212000, China*

Abstract: Accurate rapeseed yield and biomass estimation at the meter scale prior to harvest is crucial for precision harvesting. However, there is a scarcity of structured research on the estimation of rapeseed biomass yield. This study aims to address this gap by focusing on rapeseed in Jiangsu Province. Multispectral and RGB images captured by unmanned aerial vehicles (UAVs) were taken during key growth stages (budding, flowering, and podding stages). Using the extracted multidimensional features, we developed biomass-yield estimation models using four machine learning techniques. Subsequently, we employed ensemble learning with multidimensional, multi-stage data and used Shapley additive explanations (SHAP) for feature contribution analysis, thereby constructing a framework for predicting rapeseed harvest characteristics with high estimation accuracy and interpretability. Our analysis indicates that spectral – texture is the most effective feature combination for biomass estimation, whereas the optimal combination for yield estimation includes three-dimensional spectral – texture – structural features. The synergy of these features, coupled with an ensemble learning model, significantly enhanced the accuracy of rapeseed biomass-yield estimation (biomass: coefficient of determination (R^2) = 0.72, relative root mean square error (rRMSE) = 14.35%; yield: R^2 = 0.68, rRMSE = 13.67%). The proposed model also achieved stable prediction results across the variety – density interaction. Overall, this study presents an accurate and generalizable approach for estimating rapeseed biomass yield across various planting patterns, offering new insights for precision harvesting.

Unedited

Key words: Ensemble learning; Decision-making; Feature synergy; Temporal fit; Planting pattern

1 Introduction

Rapeseed, known for its high adaptability, nutritional value, and economic benefits, is widely cultivated worldwide, particularly in Asia and Europe (Sulik and Long, 2015). Increasing rapeseed yield is essential for securing the global oilseed supply (Velde et al. 2008). To facilitate precision harvesting, accurate meter-scale biomass predictions are essential, providing data support for the real-time optimization of harvesting speed and header height (Hermann, 2018). Furthermore, the crop-to-biomass ratio (CBR), derived from yield and biomass, is a critical parameter for regulating sieve aperture settings in the threshing and cleaning systems. Therefore, accurate pre-harvest mapping of biomass and yield spatial distributions is critical for enhancing operational efficiency and informed harvesting decision-making.

Traditional field measurements are resource-intensive and labor-demanding, limiting their feasibility for large-scale data collection and efficient farm management. The advent of unmanned aerial vehicles (UAVs) has

✉ Lizhang XU, justxlz@ujs.edu.cn
Yaxiao NIU, yaxiao.niu@ujs.edu.cn
Lizhang XU, https://orcid.org/0000-0001-9996-9919
Yaxiao NIU, https://orcid.org/0009-0003-7045-1472
Yanni ZHANG, https://orcid.org/0009-0002-8112-8278

Received Dec. 17, 2025; Revision accepted Mar. 18, 2026;
Crosschecked xxx. xx, 20xx; Published online xxx. xx, 20xx

© Zhejiang University Press 2026

significantly advanced the prediction of crop yield and biomass (Adeluyi et al., 2022; Shao et al., 2022). UAV-derived vegetation indices (VIs), such as the normalized difference vegetation index (NDVI), the red-edge chlorophyll index (CIre), and the near-infrared reflectance of vegetation (NIRv), have been demonstrated as reliable predictive indicators for monitoring crop growth (Rouse et al., 1974; Dong et al., 2019; Peng et al., 2019; Fan et al., 2021). Notably, the normalized difference yellowness index (NDYI) mitigates the limitations of NDVI during the blooming phase, providing the most reliable prediction of rapeseed yield fluctuations (Sulik and Long, 2016). Multi-temporal VIs can be used to estimate crop parameters across different varieties and varying environmental conditions (Duan et al., 2021). Despite these advancements, studies have shown that due to species differences and canopy physiological changes (Feng et al., 2025), spectral indices often encounter saturation issues, reducing their effectiveness in predicting crop parameters (Mateo-Sanchis et al., 2019). Furthermore, the considerable spatial-temporal variability in climatic conditions directly influences spectral response, further complicating estimates of crop productivity (Sabut et al., 2025). Therefore, synergizing spectral indices with canopy structural and textural features is crucial to overcome spectral limitations and enhance the robustness of estimation.

Canopy structural features (SFs) are commonly used to complement spectral information, enhancing the robustness of crop parameter estimation. For example, structural indicators, when combined with linear mixed-effects modeling, have proven effective in estimating rice biomass after heading with high precision (Li et al., 2020). Similarly, structural metrics showed a stronger correlation with potato biomass than spectral features (Ye et al., 2024). Additionally, estimates of sweet corn biomass based on plant height (PH) and VIs showed good agreement with measured data (Teshome et al., 2023). Textural features (TFs) extracted from RGB images quantify pixel heterogeneity within a moving window (Nogueira et al., 2023), demonstrating effectiveness in estimating rapeseed physiological and biochemical indicators (Du et al., 2024). Texture information demonstrates increased sensitivity to variations in canopy structure induced by various experimental treatments (Zhang et al., 2021). Previous studies have demonstrated that combining VIs, TFs, and SFs can effectively improve the accuracy of crop parameter estimation (Wan et al., 2020; Liu et al., 2023; Meng et al., 2025). Therefore, it can be concluded that the fusion of multi-source indicators represents a pivotal strategy for achieving stable, precise pre-harvest yield and biomass estimation.

Once the indicators are defined, the methodology emerges as a pivotal element in accurately estimating crop parameters. Machine learning models leverage complex data relationships to enable crop growth monitoring, thereby supporting agricultural decision-making and resource optimization (Han et al., 2019; Gu et al., 2024). However, the accuracy of vegetation parameter estimation depends on several factors, including crop genotype, developmental stage, spatial resolution, and environmental conditions (Yan et al., 2022; Tao et al., 2024). A fundamental challenge remains: machine learning methods rely on patterns and relationships inherent in specific datasets, which limits their applicability to particular conditions and reduces their generality across different regions or scenarios (Meiyan et al., 2022; Ye et al., 2024). To bolster predictive robustness, ensemble learning, which aggregates the strengths of multiple base models, has been successfully applied. For example, stacking an ensemble learning model significantly improved the accuracy of apple yield estimation (Zhang et al., 2025) and physiological parameter inversion (Du et al., 2025). Nevertheless, the increased complexity of ensemble architectures often comes at the expense of model interpretability. Shapley additive explanations (SHAP) use game theory to assess the contribution of each feature, revealing directional relationships and providing actionable, context-specific insights, thereby enhancing model interpretability (Wei et al., 2026). Therefore, SHAP is further employed to evaluate feature contributions for rapeseed biomass and yield estimation, thereby improving the reliability and scalability of ensemble learning methods and making them more adaptable and applicable across a wide range of agricultural scenarios.

This study proposed a comprehensive framework combining multidimensional feature synergy with ensemble learning to enhance the applicability of biomass-yield estimation across various planting patterns. The goals of this research are to: (1) quantify the relationships between rapeseed biomass yield and multidimensional features; (2) develop a novel rapeseed biomass-yield prediction model by integrating

Unedited

multidimensional feature synergy with ensemble learning; (3) quantitatively analyze multidimensional feature combinations using SHAP under variety–density interactions to identify key drivers of biomass formation and yield accumulation; and (4) validate the model's generality across different planting scenarios over multiple years and generate meter-scale spatial distribution maps for biomass, yield, and CBR.

2 Materials and methods

2.1 Study area

The experiment was conducted in Zhenjiang, situated in the lower reaches of the Yangtze River in China (32° 04′ 23″ N, 119° 27′ 31″ E), which enjoys a warm, humid climate (Fig. 1a). The predominant soil types are paddy soils and moist soils, both of which are relatively fertile. The local climate and soil conditions make Zhenjiang highly suitable for widespread rapeseed cultivation. The soil on Shiyezhou Island in the Yangtze River is classified as permeable paddy soil, comprising silty soil and yellow sand soil.

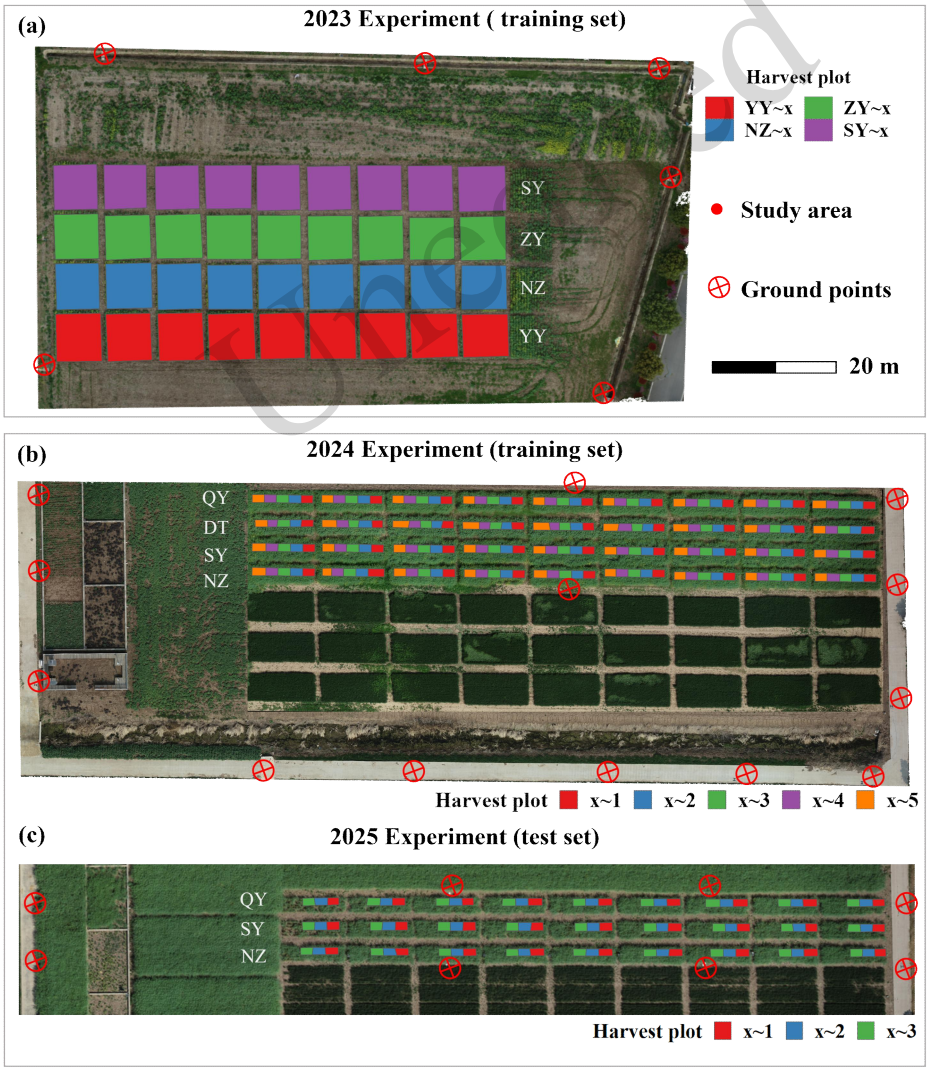


Fig. 1 Test location and experimental design. (a) Rapeseed experiment in 2023, (b) Rapeseed experiment in 2024, and (c) Rapeseed experiment in 2025, used for independent model validation. Notes: Red-circled 'x' indicates the positions of ground control points (GCPs). In 2023, six GCPs were established, followed by 13 in 2024 and eight GCPs in 2025. The

harvest plots are encoded according to the variety-x-y format. x represents different treatment plots, and y represents the subplot (1 − 5 for 2024, 1 − 3 for 2025). YY: Yangyou; ZY: Zhongyou; NZ: Ningza; SY: Suyou; QY: Qinyou; DT: Detian

Fig. 2 presents daily precipitation and maximum temperature from the budding stage to harvest for 2024 and 2025.Total precipitation from heading to maturity (February to June) in 2024 and 2025 was 336.8 mm and 2.8 mm, respectively. In 2024, rainfall was concentrated during the bud, flowering, pod, and maturity stages, providing sufficient water for rapeseed growth and dry matter accumulation. In contrast, rainfall was almost nonexistent during the same period in 2025, resulting in severe drought stress on the rapeseed crop, likely reducing both seed yield and quality. Additionally, there were no significant differences in temperature during the rapeseed growing seasons in both years. Regarding extreme temperatures, 2024 had no days exceeding 35°C, whereas two days in late May 2025 did. Furthermore, in 2024, a cold air event occurred, with the minimum temperature dropping below 5°C from late February to early March. The detailed methods for the experimental design and data collection in this paper are provided in the electronic supplementary materials.

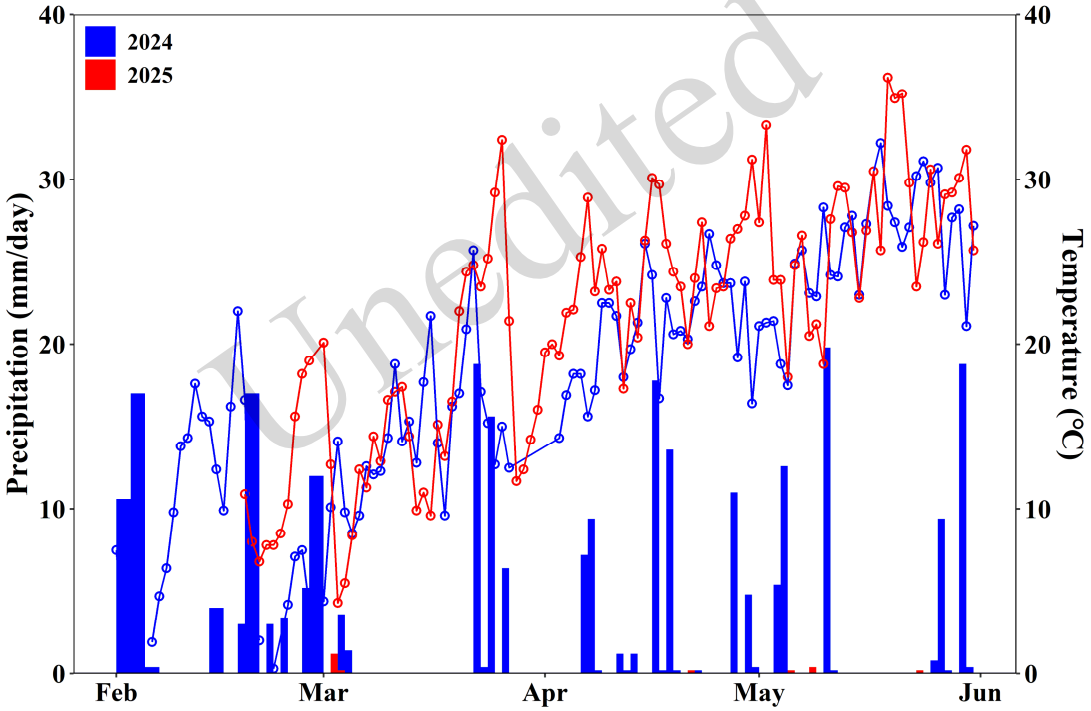


Fig. 2 Meteorological conditions across the rapeseed budding to harvesting stages in 2024 and 2025. The bar represents daily precipitation, while the line chart represents daily maximum temperature

2.2 Datasets construction

To explore the suitability of different feature combinations for estimating rapeseed biomass and yield, this study employed seven feature combinations: VIs, TFs, SFs, VIs-TFs, VIs-SFs, TFs-SFs, and VIs-TFs-SFs. Since the biomass and yield of mature rapeseed result from interconnected biophysical processes across multiple growth stages, the three feature types were subdivided by distinct growth stages—bud, early flowering, full flowering, late flowering, pod, and mature. Seven generated datasets, along with the measured yield and biomass values, were input into a random forest (RF) model, and the top 30 most important features were selected based on the model's built-in variable importance score (VIP). Next, Pearson correlation coefficients (R) were calculated to assess the relationships between the features and the biomass and yield. Features with correlation coefficients below 0.5 were discarded to enhance model stability, as low R values can lead to instability in model performance. However, if only one feature remains at this stage and its R value is below 0.5,

Unedited

it should still be retained. Auto-correlation was conducted on the multidimensional feature combinations for both yield and biomass models (Fig. S1). The final training datasets for the rapeseed biomass and yield models were generated by incorporating a small number of samples from various varieties in 2025, as detailed in Zhang et al. (2025).

2.3 Ensemble learning model construction and validation

To comprehensively explore the feature space, we employed RF, support vector machine (SVM), the instance-based K-Nearest neighbors (KNN), and partial least squares regression (PLSR). The RF model was implemented using the 'randomForest' package. Despite efforts to optimize the hyperparameters through grid search, the improvement in model accuracy was minimal. Consequently, the default parameters of 'ntree = 500' and 'mtry = 25' were retained. The SVM model was implemented using the 'e1071' package, which provides a variety of kernel function options, making it adaptable to diverse data distributions. This study utilized the 'eps-regression' kernel. The PLSR model was implemented using the 'pls' package, effectively addressing multicollinearity and being well-suited for high-dimensional data analysis, with 10 components. Additionally, the KNN model was implemented using the 'caret' package, performing regression by calculating distances between data points, making it ideal for analyzing non-linear relationships in the data.

Different base models operate within distinct hypothesis spaces and use unique structures to capture data patterns from various perspectives. By employing a stacking ensemble strategy, the approach combines the strengths of multiple base learners, reducing the biases of a single model and improving predictive accuracy and stability. Four base learners are trained on the calibration datasets, and their predictions are then used as features to create a new dataset, which is fed to the meta-learner. RF, effective at handling nonlinear data and complex feature interactions, was chosen as the meta-learner to enhance the model's ability to process complex data.

To evaluate the effectiveness of this ensemble learning approach, the biomass model was first trained on data from all plots in 2024 (180 samples), while the yield model was developed using data from 2023 and 2024 (213 samples, including 3 missing yield samples in 2024). The model's generality across different temporal scenarios was then validated using independent experimental plot data from 2025. Given that the interaction between variety and planting density influences the distribution of rapeseed biomass and yield, the 2025 validation datasets were further subdivided based on these factors to assess the model's applicability across various planting scenarios.

To comprehensively evaluate the model's performance, several assessment metrics were employed. R^2 was used to evaluate the model's goodness-of-fit, with values approaching 1 indicating a better match between predictions and the actual data. Root mean square error (RMSE) quantifies the deviation between predicted and observed values and serves as an indicator of prediction accuracy. Relative root mean square error (rRMSE), as a normalized error metric, removes the influence of datasets size and range, making it especially useful for comparing errors across different crop types or variables, such as rapeseed biomass and yield.

2.4 Model interpretation framework

Ensemble learning models can improve predictive accuracy and stability; however, their inherent complexity often compromises model transparency and interpretability. Consequently, while these models offer high performance, their utility in delivering actionable insights for agricultural decision-making remains constrained. SHAP, based on game theory, quantifies the importance of each feature while accounting for dependencies and revealing directional relationships. By leveraging a rigorous game-theoretic framework, SHAP integrates both local and global interpretability, offering mathematically sound feature attribution. SHAP analysis was carried out using the 'iml' package in R. To assess feature interpretability for ensemble learning models, weighted SHAP was employed, using the SHAP values of different base models as weights to compute the SHAP values for various features. Directional impact plots were generated to differentiate positive and negative influences on predictions under different agricultural scenarios, including varying planting densities and crop varieties.

Unedited

Overall, the comprehensive framework comprises four primary components (Fig. 3). First, the correlation between multi-source and multi-stage features and the biomass and yield of mature rapeseed is analyzed, and feature selection is performed based on VIP ranks. Second, the performance of 11 ensemble models, formed by random combinations of the four base learners, is compared, and the optimal ensemble strategy is selected based on RMSE. SHAP analysis is then applied to assess feature contributions across crop varieties and planting density scenarios, enhancing model interpretability. Third, the optimal ensemble models are tested using independent data from different varieties and planting density scenarios. Finally, meter-scale spatial distribution maps of biomass, yield, and CBR are generated, with CBR calculated as the difference between biomass and yield divided by yield.

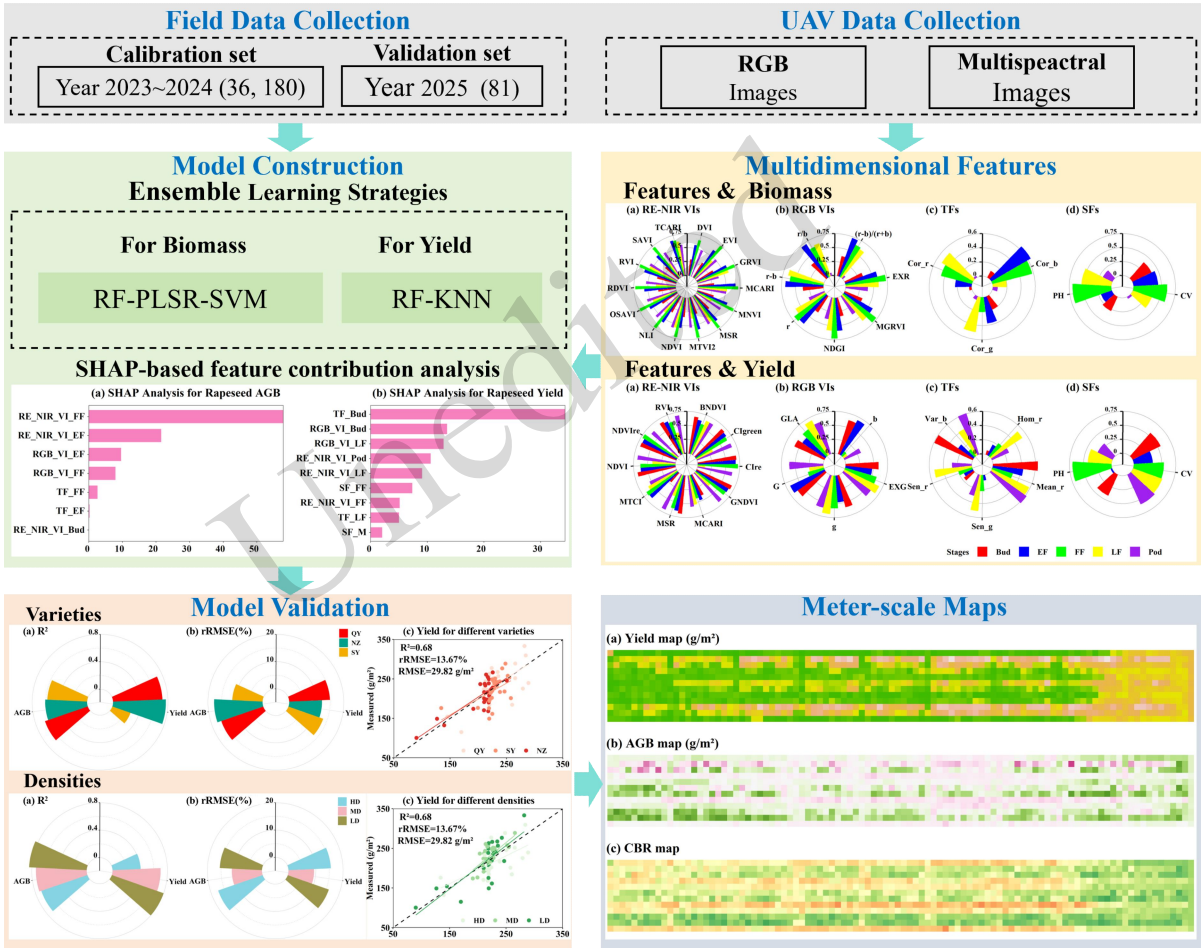


Fig. 3 Workflow of this study. Notes: RE: Red-edge; NIR: Near-Infrared; EF: early flowering stage; FF: full flowering stage; LF: late flowering stage; VIs: vegetation indices; TFs: texture features; SFs: structural features; SHAP: shapley additive explanations; RF: randomforest; KNN: K-Nearest neighbors; PLSR: partial least squares regression; SVM: support vector machine; HD: high-density; MD: medium-density; LD: low-density; AGB: above-ground biomass; CBR: crop-to-biomass ratio

3 Results

3.1 Relationship between spectral, textural and structural indices and biomass-yield

Correlation analysis visualization was performed on features from the budding to podding stages. The spectral indices were classified into The spectral indices were subdivided into Red-edge and Near-Infrared

(RE-NIR) based indices, and RGB-based indices. Fig. 4 presents the correlation coefficients between rapeseed biomass and the selected multidimensional features across these stages. The correlation distributions between RE-NIR and RGB VIs differ with biomass. Specifically, the RE-NIR VIs shows the strongest correlation with AGB during full flowering, while the RGB VIs exhibits a higher correlation with AGB at the early flowering stages. For TFs, Cor_b shows a particularly strong correlation during the early reproductive growth phase, while among SFs, canopy volume (CV) and PH are more strongly correlated with AGB in the later stages of reproductive growth.

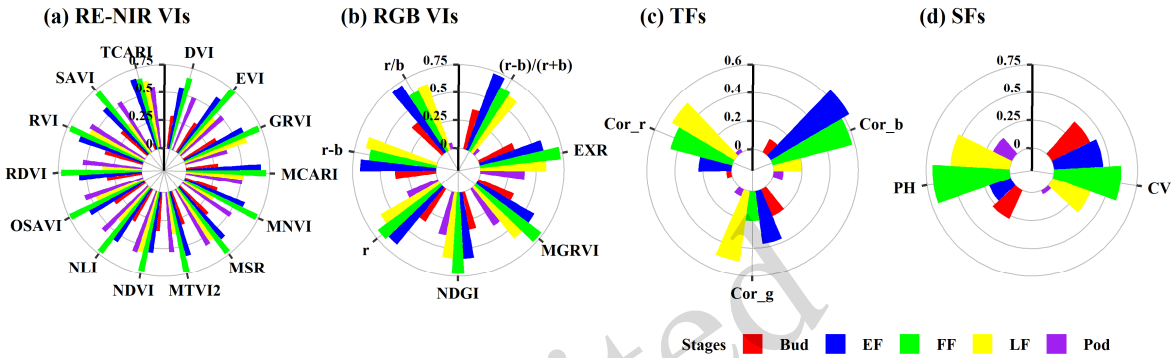


Fig. 4 Pearson correlation coefficients (R) between rapeseed biomass and Red-edge and Near-Infrared (RE-NIR) based VIs (a), RGB VIs (b), TFs (c), and SFs (d) from budding to podding stages

Similarly, this study assessed the relationship between selected features and rapeseed yield (Fig. 5). The NIR-RE VIs demonstrated a strong association with yield throughout the entire reproductive growth stage, as yield reflects the cumulative photosynthesis during critical growth phases. However, green and blue band reflectance showed a higher correlation with yield at the budding stage. The green band reflectance indicates the chlorophyll content, while the blue band is linked to nitrogen utilization. Regarding TFs, the Var_b and Mean_r show the strongest correlation with yield during the budding and podding stages. During the late flowering stage, the Sen_r, Sen_g, and Hom_r exhibit the highest correlation with yield. PH was most strongly correlated with yield at full flowering, while the relationship between CV and biomass remained stable during both the vegetative and reproductive growth stages.

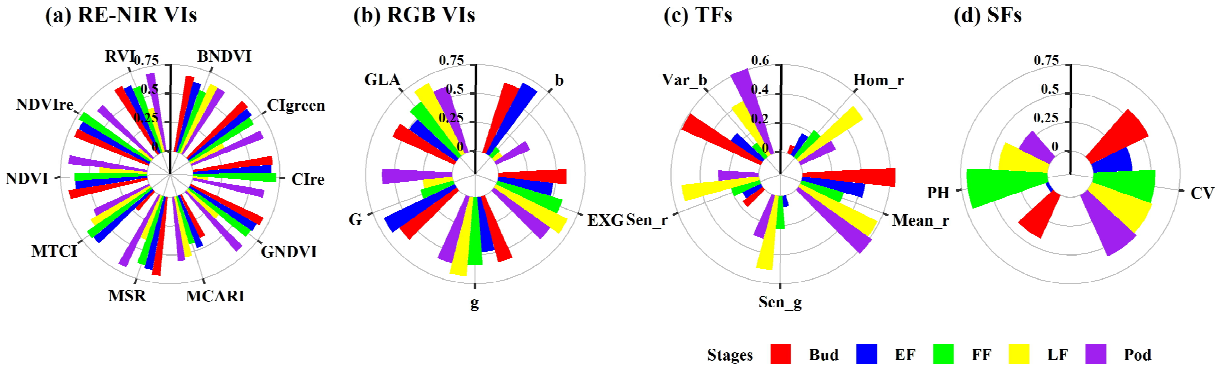


Fig. 5 R between rapeseed yield and RE-NIR VIs (a), RGB VIs (b), TFs (c), and SFs (d) from budding to podding stages

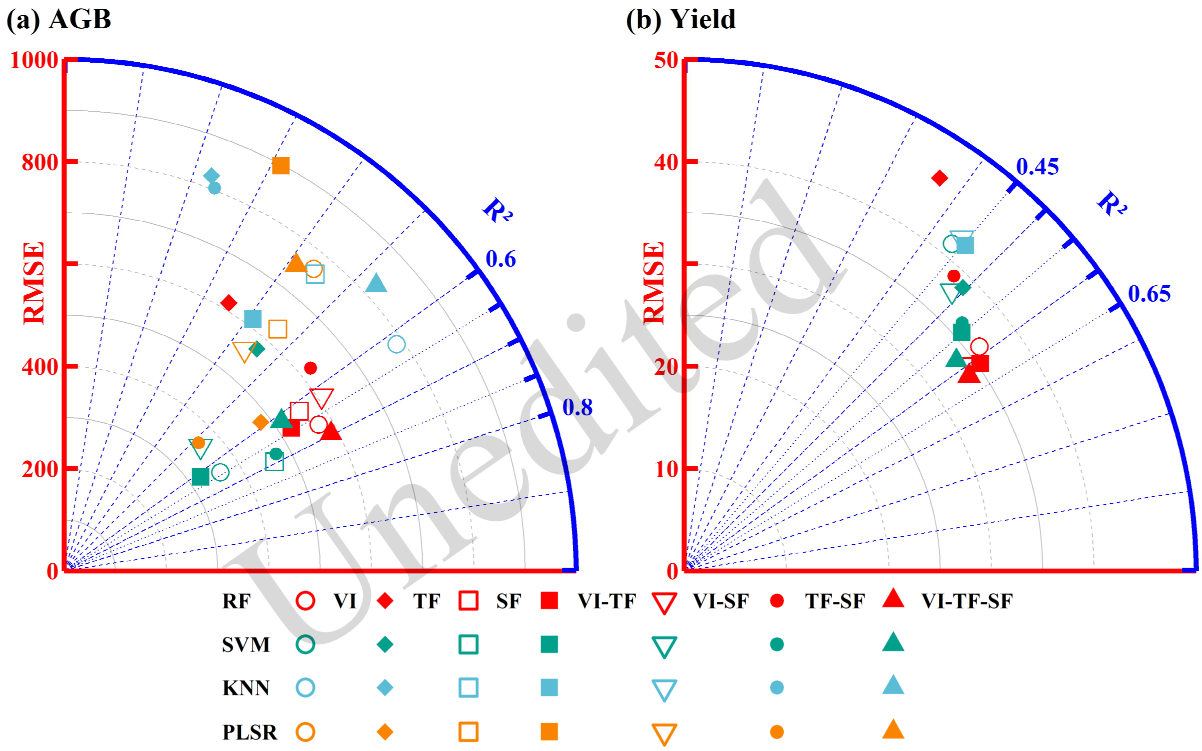
3.2 Validation results of biomass and yield estimation

3.2.1 Performance of the machine learning model

Seven feature combinations were used as inputs for four single machine learning models, validated on cross-year rapeseed yield and biomass. As shown in Fig. 6, the coupling of algorithms and features has a

Unedited

significant impact on the performance of rapeseed biomass yield estimation. The non-linear model (SVM) is more suitable for predicting AGB, while the ensemble model (RF) is better suited for predicting yield. For AGB estimation, the SVM model based on SFs achieved the highest goodness-of-fit, with a coefficient of determination (R^2) of 0.69. However, the VIs – TFs combination yielded a lower RMSE of 35.75 g/m^2. The reduction in error is likely due to the incorporation of texture information, which helps the model better handle spatial differences in spectral data across planting patterns, thereby enhancing robustness. In contrast, for yield estimation, integrating VIs, TFs, and SFs enabled the RF model to achieve optimal performance (R^2 = 0.62, RMSE = 33.72 g/m^2).



Unedited

Fig. 6 Comparison of estimation results for rapeseed AGB estimation (a) and for yield estimation (b) across four machine learning models (RF, KNN, PLSR, SVM). Notes: To facilitate a clearer comparison of high-accuracy models, models with a AGB root mean square error (RMSE) greater than 1000 g/m² and a yield RMSE greater than 50 g/m² have been excluded

3.2.2 Accuracy analysis using an ensemble learning model

KNN, RF, PLSR, and SVM were employed as base learners, with RF used to ensemble-average their predictions. Fig. 7 presents an accuracy analysis of the performance of different ensemble learning strategies, using the optimal multidimensional features selected in Section 3.2.1. The results demonstrate the effectiveness of stacking in reducing biases inherent in individual model predictions for biomass and yield estimation. Specifically, for biomass estimation, the synergy of VIs–TFs features within the RF-PLSR-SVM stacking framework yielded the best performance, with R^2 improving from 0.69 to 0.72 and RMSE decreasing from 408.62 g/m^2 to 279.95 g/m^2, compared to the best individual machine learning model. Similarly, for yield estimation, the synergy of VIs-TFs-SFs features within the KNN-RF stacking framework outperformed the best individual machine learning models, with R^2 improving from 0.62 to 0.68 and RMSE decreasing from 33.72 g/m^2 to 29.82 g/m^2. Overall, the results demonstrate that synergy among multidimensional features, combined with an ensemble learning strategy, produces accurate and reliable estimates of rapeseed biomass and yield. Furthermore, by optimizing the selection of base learners, the ensemble learning model can not only provide

more accurate estimates of rapeseed biomass and yield, but also further reduce training time.

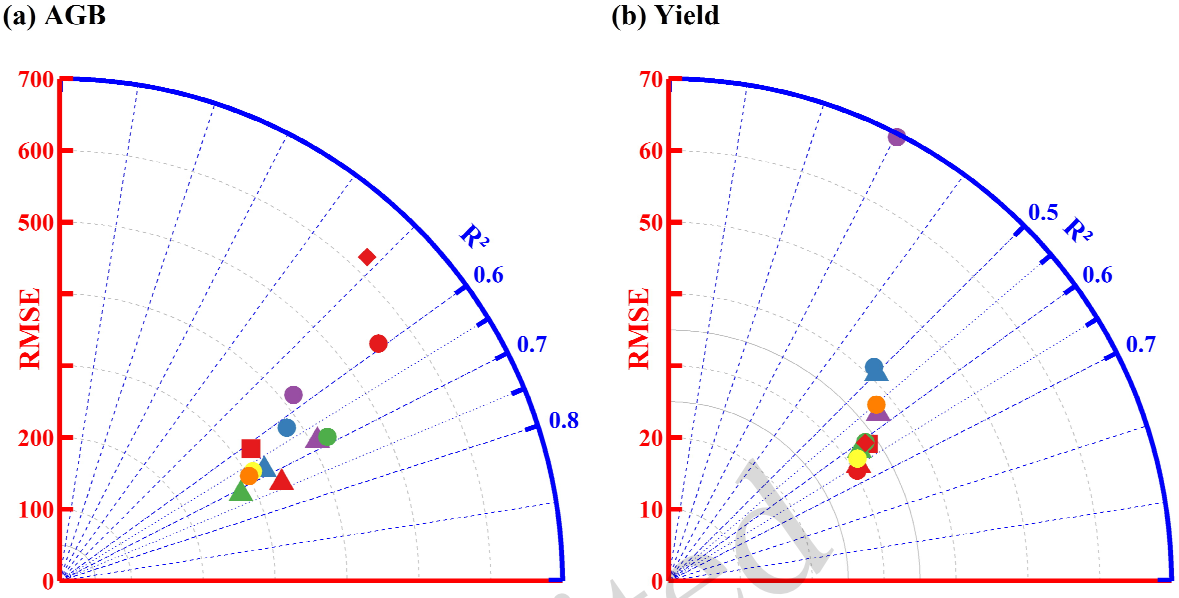


Fig. 7 Performance comparison of ensemble learning strategies for AGB estimation (a) and yield (b) estimation. Notes: The red square denotes the best machine learning models and feature combinations selected in Section 3.2.1. ML: a machine learning model

Unedited

3.3 Contribution of multidimensional features under variety − density interactions

Based on the results presented in Section 3.2, ensemble learning models are highly accurate in predicting rapeseed biomass and yield. However, their inherent complexity limits the model's interpretability. To enhance the transparency of ensemble learning models, this study employs SHAP analysis to evaluate the feature contributions for rapeseed biomass and yield across different planting scenarios. As shown in Fig. 8, the five most influential features for biomass formation are NDVI, MCARI at the early flowering stage, and MTVI2 (positive directional effects), along with TCARI, MCARI at the full flowering stage (negative directional effects). Spectral VIs at the full flowering stage contribute over 60% to the accurate estimation of rapeseed biomass. However, the impact of VIs varies across planting scenarios, demonstrating their context-dependence. For example, the RVI of the Suyou (SY) variety exerts the strongest positive influence on biomass formation, whereas the RVI of the Ningza (NZ) variety has a negative effect. The interaction of variety and planting density may influence rapeseed growth, thereby complicating the spectral mapping of crop parameters. Using a single spectral index has limitations for predicting rapeseed biomass, whereas employing multiple indices may effectively address this issue.

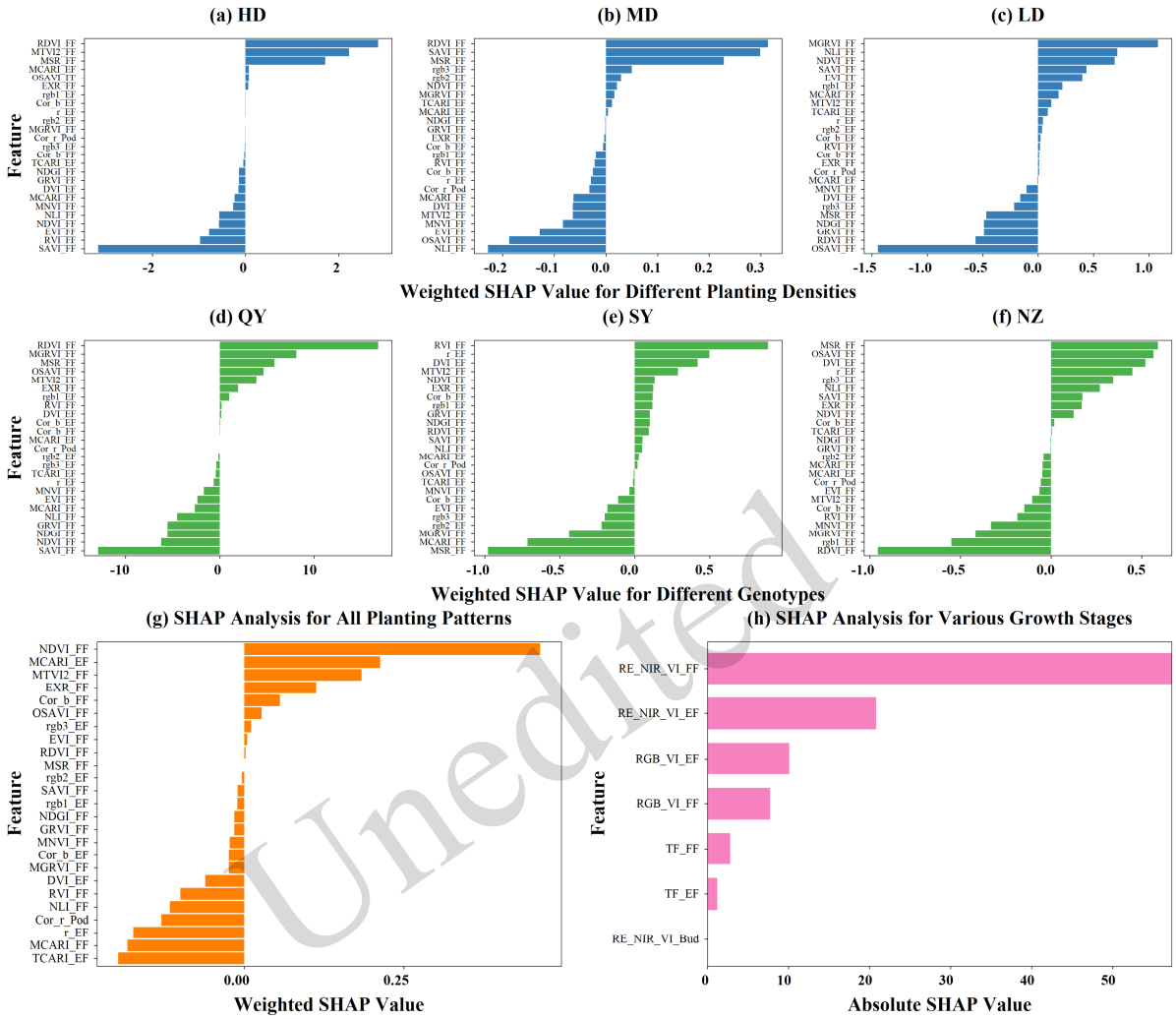


Fig. 8 Feature importance from the ensemble learning model in AGB estimation: (a-c) weighted SHAP for different planting densities; (e-f) weighted SHAP for different genotypes; (g) weighted SHAP values for all planting patterns; (h) absolute SHAP at different growth stages

Fig. 9 highlights the feature importance ranking for the rapeseed yield estimation model, which accounts for the interaction between variety and planting density. SHAP can distinguish the magnitude and directionality of individual features, revealing key agricultural driving factors. As shown in Fig. 9c, SHAP values are ranked under different planting patterns; the top variables are Var_b and Mean_r at the bud stage, blue and green band reflectance during the bud stage, MTCI during late flowering, and PH during full flowering. Features including RVI and MSR during the podding stage exhibited near-zero values. Specifically, texture information during the bud stage is highlighted as a key factor across different planting density scenarios, accounting for over 40%. CV is identified as a critical predictor of yield in both Qinyou (QY) and SY varieties, likely due to the significant effect of lodging at maturity. During the late flowering stage, photosynthesis in siliques primarily contributes to dry matter accumulation, making chlorophyll content at this stage crucial for yield accumulation. Additionally, from a feature-directionality perspective, we observe that the relationship between spectral features and yield exhibits directional changes across different planting patterns, whereas the directionality of texture and structural features remains consistent across planting scenarios.

Unedited

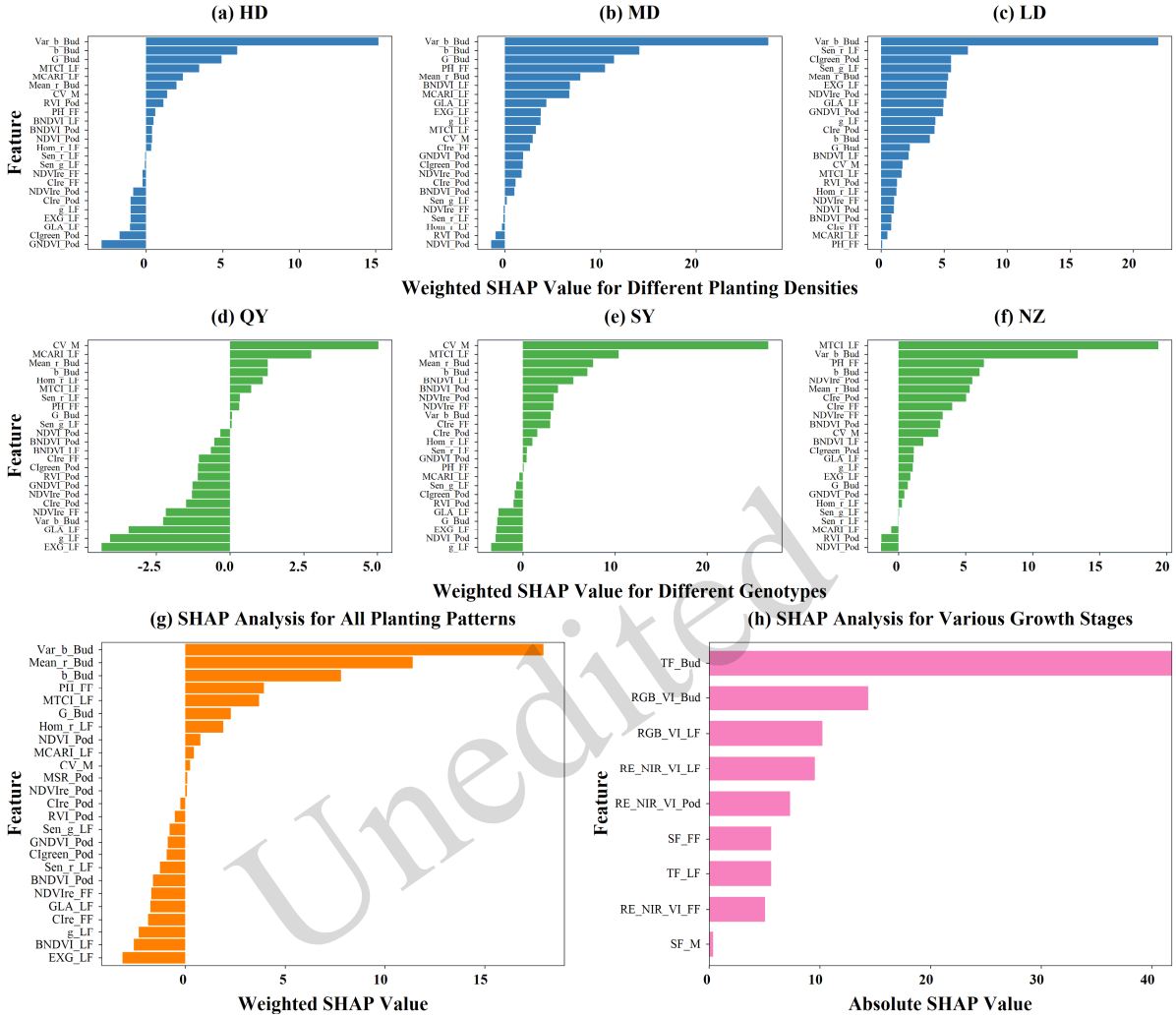


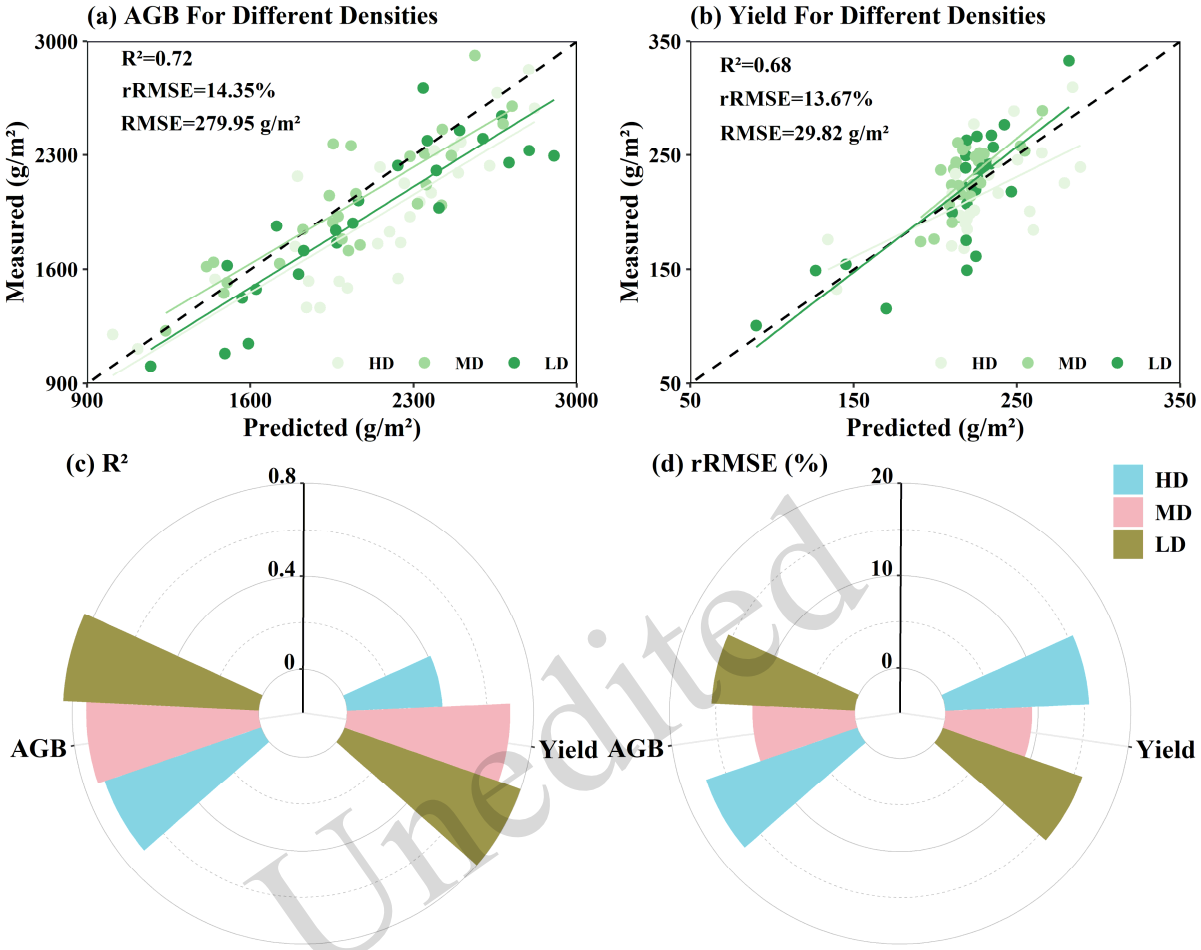
Fig. 9 Feature importance from the ensemble learning model in yield estimation: (a-c) weighted SHAP for different planting densities; (e-f) weighted SHAP for different genotypes; (g) weighted SHAP values for all planting patterns; (h) absolute SHAP at different growth stages

3.4 Model generality in different agricultural scenarios

3.4.1 Performance in the temporal-fit scenario

Fig. 10a and 10b illustrate the temporal fitting accuracy for rapeseed biomass and yield estimation using multidimensional feature synergy and ensemble learning models. These models were trained on two years of data (AGB: 2024; yield: 2023 to 2024) and tested on 2025 data. The AGB prediction model achieved an R^2 of 0.72 and an RMSE of 279.95 g/m², while the yield prediction model achieved an R^2 of 0.68 and an RMSE of 29.82 g/m². It is important to note that training the model on 2024 data alone yielded a validation R^2 of 0.53; incorporating the 2023 datasets increased R^2 by 15% and decreased the RMSE by 10.5 g/m². In this study, where farm management practices remained consistent across two years of rapeseed biomass estimation, accurate temporal-fit estimation can be achieved by training just one year of biomass monitoring data. However, accurate rapeseed yield prediction requires a longer time frame, with yield-monitoring data from both 2023 and 2024 necessary for model training. A plausible explanation is that training the model on multi-year data enables it to capture inter-annual variations (Ma et al., 2024), such as temperature, rainfall, and soil conditions, which are crucial for accurate yield prediction.

Unedited



Unedited

Fig. 10 AGB and yield performance across different planting density scenarios. Scatterplots of measured vs. estimated AGB (a) and yield (b); Coefficient of determination (R^2) (c) and relative root mean square error (rRMSE) (d) across densities

3.4.2 Performance in variety–density-fit scenarios

We further evaluated the stability and accuracy of the ensemble learning model across different planting scenarios. Figs. 10c and 10d illustrate the accuracy of the feature synergy and ensemble learning in estimating biomass and yield under varying planting densities. The estimation errors (rRMSE) for rapeseed biomass and yield were comparable in both high-density (HD) and low-density (LD) areas. In the LD area, rRMSE was 15.54% for biomass and 17.27% for yield, while in the HD area, rRMSE was 15.49% for biomass and 15.95% for yield. In contrast, the prediction errors were relatively lower in the medium-density (MD) range, with rRMSE values of 11.04% for biomass and 9.37% for yield. This can be attributed to the uniform canopy distribution in the MD area, which allows remote sensing signals to more effectively capture crop growth status. Although the estimation errors (rRMSE) for rapeseed biomass and yield varied with planting density, the estimation biases consistently fell within the 10%–20% range for all density levels, indicating that the ensemble learning model exhibits relatively stable predictive performance across varying planting densities.

Fig. 11 presents an accuracy analysis of biomass and yield prediction among different rapeseed genotypes. However, the yield prediction for the SY genotype exhibited a lower goodness-of-fit ($R^2 = 0.28$), likely due to the lower coefficient of variation in the observed SY yield data (Table 1). Nonetheless, the model's estimation error for the SY genotype was not greater than that for the other two varieties. The rRMSE for rapeseed biomass and yield across different varieties ranges from 11.12% to 17.75% (Fig. 11d). This indicates that the synergy of

VIs, TFs, and SFs in the ensemble learning model demonstrated stable and reliable prediction performance for biomass and yield estimation across rapeseed varieties.

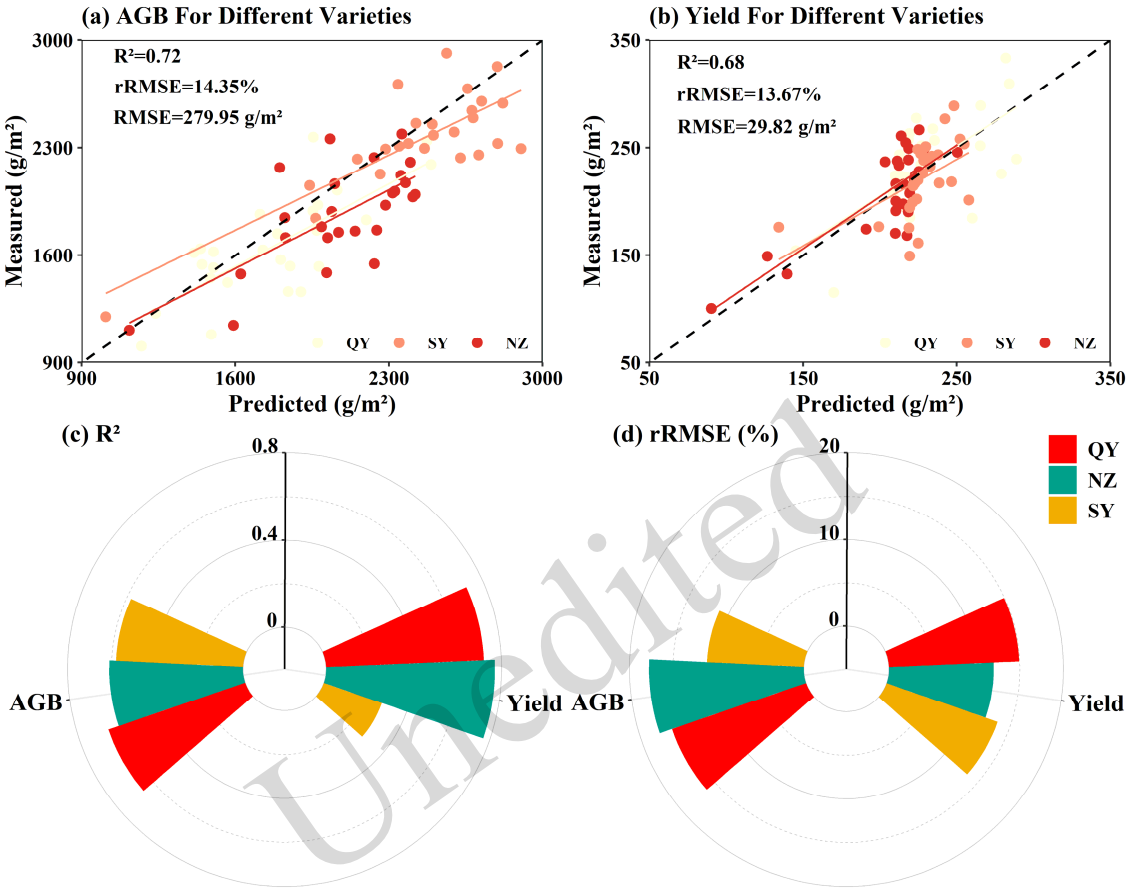


Fig. 11 AGB and yield performance across different varieties. Scatterplots of measured vs. estimated AGB (a) and yield (b); R^2 (c) and rRMSE (d) across varieties

Table 1 Statistics of measured yield and above-ground biomass (AGB) data across varieties and densities

Class	Variable	Maximum (g/m^2)	Minimum (g/m^2)	Range (g/m^2)	Mean (g/m^2)	Medium (g/m^2)	Standard deviation (g/m^2)	Variation coefficient (%)
Yield	QY	311.8	67.1	244.7	214.4	220.2	47.9	22.4
	SY	237.0	70.9	166.2	222.0	226.7	31.2	14.0
	NZ	226.6	67.4	159.3	189.9	205.6	48.0	25.3
	HD	277.3	67.4	210.0	212.9	221.2	46.2	21.7
	MD	311.8	171.9	139.9	215.0	217.1	27.6	12.9
	LD	269.3	67.1	202.2	198.5	221.2	55.6	28.0
AGB	QY	2492.0	905.8	1586.3	1735.5	1792.8	369.7	21.3
	SY	2902.3	1009.6	1892.7	2464.5	2562.3	383.1	15.5
	NZ	2419.3	1099.6	1319.6	2045.0	2070.5	357.7	17.5
	HD	2818.5	1009.6	1808.9	2138.4	2233.5	450.4	21.1
	MD	2722.9	1237.5	1485.5	2069.7	2031.4	416.4	20.1
	LD	2902.3	905.8	1996.6	2036.9	2039.1	554.6	27.1

QY: Qinyou; SY: Suyou; NZ: Ningza; HD: high-density; MD: medium-density; LD: low-density

3.5 Spatial distribution maps of rapeseed biomass and yield at the meter scale

The spatial variation of yield and AGB at the meter scale (Fig. 12) clearly reflects the differences between the treatment areas, which align with the actual planting patterns. Additionally, the CBR spatial distribution

Unedited

map (Fig. 12) was calculated based on the biomass and yield distribution. CBR is a key indicator for reducing harvest losses and balancing material flow. From a spatial perspective, the yield and biomass distribution in the area is slightly higher at the eastern field edges. This trend can be attributed to the proximity of the eastern region to water sources, which results in better moisture availability for the plots, thereby leading to higher yield and biomass production. Through coordinate matching, meter-scale maps can provide real-time, high-throughput data to support decision-making systems in smart harvesters, thereby minimizing harvesting losses and enhancing rapeseed yield.

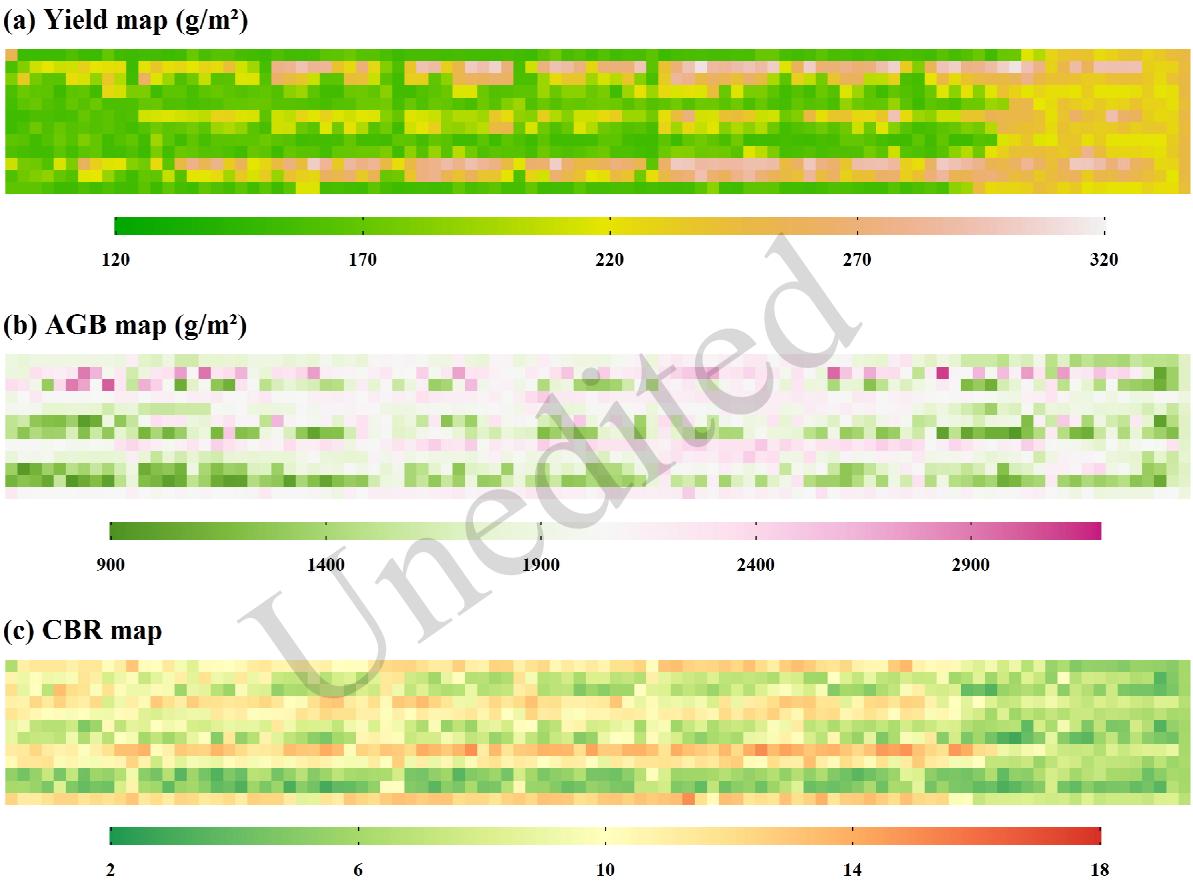


Fig. 12 Spatial distribution maps of rapeseed AGB (a), yield (b), and CBR (c) at the meter scale

Unedited

4 Discussion

4.1 Contribution of multidimensional features in yield and AGB prediction

Spectral indices exhibit saturation effects under dense canopy conditions (Wang et al., 2021; Li et al., 2022), limiting the stability of models relying solely on spectral data for biomass and yield prediction. Incorporating texture and structural features enhances model robustness by capturing canopy structural characteristics influenced by the interaction between variety and planting density (Brill et al., 2016; Ma et al., 2023), thereby improving estimation accuracy across various planting scenarios. However, traditional methods fail to provide new insights for data-driven models.

SHAP analysis reveals the complex relationships among multidimensional features within specific contexts, thereby enhancing the transparency and interpretability of black-box models. The results indicate that texture information during the bud stage, green and blue band at the bud stage, and chlorophyll content during late flowering are key drivers of yield accumulation. Consistent with these findings, Peng et al. (2019) also

identified a significant relationship between leaf structure at the bud stage and rapeseed yield. Additionally, blue band reflectance during the bud stage, which reflects nitrogen content, emerges as a critical factor influencing rapeseed yield. During late flowering, as flowers shed and siliques begin to form, silique photosynthesis plays a crucial role in yield accumulation, making chlorophyll content at this stage a key determinant of the final yield. Notably, CV at the mature stage exhibits the highest feature importance for both QY and SY varieties, likely due to their higher susceptibility to lodging. Collectively, these findings provide new insights into agronomic management and variety selection, demonstrating that the synergy of VIs, TFs, and SFs across different growth stages provides comprehensive multidimensional information that maximizes yield estimation accuracy (Fig. 6).

On the other hand, spectral data during full flowering were found to be the most influential factor for biomass prediction based on SHAP analysis. However, the importance of spectral indices varied across planting patterns, potentially leading to model instability when relying solely on spectral features. Our results indicate that the synergy of VIs and TFs significantly enhanced the accuracy of biomass prediction, with an RMSE of 323.88 g/m² (Fig. 6). Overall, SHAP analysis revealed the complex interactions that drive rapeseed yield and biomass, identifying key factors and providing valuable ecological insights for dynamic modeling and agricultural management.

4.2 Ensemble learning model performance in variety − density interactions

The modeling approach significantly influences the reliability and accuracy of agricultural monitoring (Chen et al., 2024; Gu et al., 2024; Ji et al., 2024). Ye et al. (2024) noted that different models have distinct advantages; simple linear regression models offer better interpretability, while multivariate models provide higher accuracy. Our results also showed that RF models are better at capturing broad, transferable patterns, whereas models like SVMs are more suitable for predictions within a specific domain. Thus, the two-layer stacking ensemble framework effectively combines the strengths of individual models, resulting in improved predictive performance (Zhai et al., 2023; Du et al., 2025; Zhang et al., 2025). Similarly, our investigation found that the ensemble model outperformed the single machine learning model, increasing R^2 from 0.69 to 0.72 and reducing RMSE from 408.62 g/m^2 to 279.95 g/m^2 in rapeseed biomass estimation. For yield estimation, the model also showed significant improvements, with R^2 rising from 0.62 to 0.68 and RMSE decreasing from 33.72 g/m^2 to 29.82 g/m^2. Results demonstrate that the synergy of multidimensional features with ensemble strategies effectively captures more comprehensive information, leading to improved accuracy and robustness in rapeseed yield and biomass estimation.

Planting density and variety varied across the plots. To further evaluate the generality of the ensemble learning strategies, a comparison was conducted between the observed and predicted rapeseed harvest parameters across different planting patterns. Analyzing rapeseed biomass and yield estimation results within a variety − density ecological framework can enhance the model's scalability. Fig. 10 indicated that planting density influenced model performance. In estimating rapeseed biomass and yield across different planting densities, the rRMSE was lowest in the MD area. A possible explanation is that the uniform canopy at medium planting density allows spectral and textural features to effectively reflect crop growth, whereas saturation effects in dense canopies and the uneven canopy structure at low density reduce their effectiveness. Furthermore, different varieties exhibit varying adaptability to planting density, which influences their growth and yield. Compared to other varieties, SY has more stable yields across different planting densities, with a yield variation of only 14.0% (Table 1). This can likely be attributed to its compact growth structure, in which the small leaves and dense canopy reduce interplant shading, optimizing light and nutrient utilization. Therefore, the model struggles to capture the complex relationship between features and yield (SY), resulting in lower goodness-of-fit (Fig. 11c). Nevertheless, the proposed model demonstrated stable predictive accuracy across various rapeseed varieties and planting densities (Figs. 11–12). Overall, this study demonstrated that combining multidimensional features with an optimal ensemble learning strategy yielded high accuracy and robustness in predicting rapeseed biomass and yield across multiple years and diverse planting scenarios.

Unedited

4.3 Future perspectives

This study provides an accurate estimation of rapeseed biomass and yield across various variety–density plots by integrating synergistic multidimensional features with ensemble learning. To enhance model interpretability, SHAP analysis was employed to evaluate the impact of each feature on predictions, enabling identification of key drivers and interactions influencing yield and biomass. This approach enhances transparency and effectively narrows the gap between the model's complexity and practical agricultural insights. However, it is important to note that these findings are specific to the climate and management conditions of the study site, and may therefore have limited applicability to other contexts. Water stress and nitrogen content can also affect crop growth, thereby impacting rapeseed yield and biomass (Mokhtari et al., 2025). Future studies integrating thermal infrared and hyperspectral sensors hold the potential to reveal the ecological responses of crops to water and nutrient stress. These sensors have the potential to measure physiological parameters, such as canopy temperature and water content, which could be directly linked to crop stress levels. By incorporating these data into predictive models, the interactions between the driving factors of crop growth and yield can be better captured, enhancing model interpretability and providing more valuable insights for agricultural management. Furthermore, to continuously refine and improve the model's practicality, future studies could explore additional machine learning algorithms and ensemble strategies that account for different crop types, regions, and environmental conditions.

5 Conclusions

Accurate pre-harvest estimation of rapeseed biomass and yield using remote sensing is a prerequisite for precision harvesting. However, temporal transfer and diverse planting patterns complicate the mapping between spectral data and crop parameters, limiting the model's applicability across different agricultural scenarios. To address these challenges, this study developed a robust rapeseed biomass and yield estimation model by integrating multidimensional feature synergy with an RF ensemble strategy. Furthermore, an interpretable framework is constructed using SHAP analysis, achieving a balance between predictive accuracy and model transparency. The main findings are as follows: (1) The synergy of multidimensional features and ensemble learning significantly improved the accuracy of rapeseed yield and biomass predictions in a temporal-fit scenario (AGB: R^2 = 0.72, RMSE = 279.95 g/m² ; Yield: R^2 = 0.68, RMSE = 29.82 g/m²). (2) Key features for yield and biomass models were identified via SHAP: texture features at the bud stage, chlorophyll content during late flowering, CV at the mature stage as key factors for yield prediction; canopy spectral data during full flowering is the most important for biomass prediction, although its impact varies with planting pattern. (3) Validation results showed that the model performed consistently and accurately across different varieties and planting densities, demonstrating its strong generality. Furthermore, the generated meter-scale biomass and CBR maps offer critical spatial data for precision agriculture and intelligent harvesting, providing precise guidance for machinery operation and harvest strategy optimization.

Unedited

Data availability statement

The raw data supporting the conclusions of this article will be made available upon request.

Acknowledgments

This work are supported by the National Key R&D Program of China [2023YFD2001003]; the National Natural Science Foundation of China [grant number 32401695]; the Natural Science Foundation of Jiangsu Province [grant number BK20240878]; and the Key Laboratory of Spectroscopy Sensing, Ministry of Agriculture and Rural Affairs, P.R. China [2025ZJUGP002]. We thank the local agricultural experts, Xueji LIU and Fugen SUN, for their invaluable assistance.

Author contributions

Yanni ZHANG performed original draft, investigation, methodology, software, conceptualization, review and editing. Xiaoyu CHAI performed software. Jipeng HU performed investigation. Yaxiao NIU performed review, editing, investigation, supervision, and funding acquisition. Lizhang XU performed supervision and funding acquisition. All authors have read and agreed to the published version of the manuscript.

Compliance with ethics guidelines

Yanni ZHANG, Xiaoyu CHAI, Jinpeng HU, Yaxiao NIU, and Lizhang XU declare that they have no conflicts of interest. This article does not contain any studies with human or animal subjects performed by any of the authors.

Declaration on the use of generative AI tools

During the preparation of this work, the authors used ChatGPT in order to improve language and readability, check for grammatical errors. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

Adeluyi, O., A. Harris, T. Foster and G. D. Clay (2022). "Exploiting centimetre resolution of drone-mounted sensors for estimating mid-late season above ground biomass in rice." European Journal of Agronomy 132.

Brill, R. D., M. L. Jenkins, M. J. Gardner, J. M. Lilley and B. A. Orchard (2016). "Optimising canola establishment and yield in south-eastern Australia with hybrids and large seed." Crop & Pasture Science 67(4): 409.

Chen, A., C. Xu, M. Zhang, J. Guo, X. Xing, D. Yang, B. Xu and X. Yang (2024). "Cross-scale mapping of above-ground biomass and shrub dominance by integrating UAV and satellite data in temperate grassland." Remote Sensing of Environment 304.

Dong, T., JianguiShang, JialiQian, BudongMa, BaoluoKovacs, John M.Walters, DanJiao, XianfengGeng, XiaoyuanShi, Yichao (2019). "Assessment of red-edge vegetation indices for crop leaf area index estimation." Remote Sensing of Environment: An Interdisciplinary Journal 222.

Du, R., J. Lu, Y. Xiang, F. Zhang, J. Chen, Z. Tang, H. Shi, X. Wang and W. Li (2024). "Estimation of winter canola growth parameter from UAV multi-angular spectral-texture information using stacking-based ensemble learning model." Computers and Electronics in Agriculture 222.

Du, R., X. Lu, Y. Zhang, X. Feng, Y. Xiang and F. Zhang (2025). "Improved estimation of irrigated field soil water (SWC) and salt content (SSC) from Sentinel-2 imagery by combining multi-dimensional spectra decomposition with ensemble learning." Computers and Electronics in Agriculture 239.

Duan, B., S. Fang, Y. Gong, Y. Peng and R. Zhu (2021). "Remote estimation of grain yield based on UAV data in different rice cultivars under contrasting climatic zone." Field Crops Research 267(11): 108148.

Fan, H., S. Liu, J. Li, L. Li, L. Dang, T. Ren and J. Lu (2021). "Early prediction of the seed yield in winter oilseed rape based on the near-infrared reflectance of vegetation (NIRv)." Computers and Electronics in Agriculture 186.

Feng, D., H. Yang, K. Gao, X. Jin, Z. Li, C. Nie, G. Zhang, L. Fang, L. Zhou, H. Guo, Z. Jia, B. Ming, K. Wang and S. Li (2025). "Time-series NDVI and greenness spectral indices in mid-to-late growth stages enhance maize yield estimation." Field Crops Research 333.

Gu, Q., F. Huang, W. Lou, Y. Zhu, H. Hu, Y. Zhao, H. Zhou and X. Zhang (2024). "Unmanned aerial vehicle-based assessment of rice leaf chlorophyll content dynamics across genotypes." Computers and Electronics in Agriculture 221.

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

Hermann, D. "Optimisation of Combine Harvesters using Model-based Control."

Ji, Y., Z. Liu, R. Liu, Z. Wang, X. Zong and T. Yang (2024). "High-throughput phenotypic traits estimation of faba bean based on machine learning and drone-based multimodal data." Computers and Electronics in Agriculture 227: 109584.

Li, P., X. Zhang, W. Wang, H. Zheng, X. Yao, Y. Tian, Y. Zhu, W. Cao, Q. Chen and T. Cheng (2020). "Estimating aboveground and organ biomass of plant canopies across the entire season of rice growth with terrestrial laser scanning." International Journal of Applied Earth Observation and Geoinformation 91.

Li, Z., Y. Zhao, J. Taylor, R. Gaulton, X. Jin, X. Song, Z. Li, Y. Meng, P. Chen, H. Feng, C. Wang, W. Guo, X. Xu, L. Chen and G. Yang (2022). "Comparison and transferability of thermal, temporal and phenological-based in-season predictions of above-ground biomass in wheat crops from proximal crop reflectance data." Remote Sensing of Environment 273.

Liu, Y., H. Feng, J. Yue, Y. Fan, M. Bian, Y. Ma, X. Jin, X. Song and G. Yang (2023). "Estimating potato above-ground biomass by using integrated unmanned aerial system-based optical, structural, and textural canopy measurements." Computers and Electronics in Agriculture 213.

Ma, Y., S. Chen, S. Ermon and D. B. Lobell (2024). "Transfer learning in environmental remote sensing." Remote Sensing of Environment 301(000): 22.

Ma, Z., R. Du, J. Xie, D. Sun, H. Fang, L. Jiang and H. Cen (2023). "Phenotyping of Silique Morphology in Oilseed Rape Using

Unedited

Skeletonization with Hierarchical Segmentation." Plant Phenomics 5: 0027.
Malambo, L., S. C. Popescu, S. C. Murray, E. Putman, N. A. Pugh, D. W. Horne, G. Richardson, R. Sheridan, W. L. Rooney, R. Avant, M. Vidrine, B. McCutchen, D. Baltensperger and M. Bishop (2018). "Multitemporal field-based plant height estimation using 3D point clouds generated from small unmanned aerial systems high-resolution imagery." International Journal of Applied Earth Observation and Geoinformation 64: 31-42.
Mateo-Sanchis, A., M. Piles, J. Muñoz-Marí, J. E. Adsuara, A. Pérez-Suay and G. Camps-Valls (2019). "Synergistic integration of optical and microwave satellite data for crop yield estimation." Remote Sensing of Environment 234.
Meiyan, S., S. Mengyuan, D. Qizhou, Y. Xiaohong, L. Baoguo and M. Yuntao (2022). "Estimating the maize above-ground biomass by constructing the tridimensional concept model based on UAV-based digital and multi-spectral images." Field Crops Research 282.
Meng, L., B. Ming, Y. Liu, C. Nie, L. Fang, L. Zhou, J. Xin, B. Xue, Z. Liang, H. Guo, D. Yin and X. Jin (2025). "Maize biomass estimation by integrating spectral, structural, and textural features from unmanned aerial vehicle data." European Journal of Agronomy 168.
Mokhtari, A., H. Yang, H. Croft, S. V. Luca, F. Li, M. Minceva, U. Schmidthalter and K. Yu (2025). "Satellite-based winter wheat yield estimation with a newly parameterized LUE model based on crop water status and leaf chlorophyll content." Field Crops Research 333.
Nogueira Martins, R., F. de Assis de Carvalho Pinto, D. Marçal de Queiroz, D. Sárvio Magalhães Valente, J. Tadeu Fim Rosas, M. Fagundes Portes and E. Sânzio Aguiar Cerqueira (2023). "Digital mapping of coffee ripeness using UAV-based multispectral imagery." Computers and Electronics in Agriculture 204.
Peng, Y., T. e. Zhu, Y. Li, C. Dai, S. Fang, Y. Gong, X. Wu, R. Zhu and K. Liu (2019). "Remote prediction of yield based on LAI estimation in oilseed rape under different planting methods and nitrogen fertilizer applications." Agricultural and Forest Meteorology 271: 116-125.
Rouse, J. W., R. H. Haas, J. A. Schell and D. W. Deering (1974). "Monitoring vegetation systems in the Great Plains with ERTS."
Sabut, A., K. P. Tripathy, A. Mishra, M. Anderson, M. Cosh, S. kraatz, F. Gao and R. Cirone (2025). "Assessing the impact of climate indices on corn yield in the continental USA using machine learning approach." Agricultural and Forest Meteorology 371.
Shao, G., W. Han, H. Zhang, Y. Wang, L. Zhang, Y. Niu, Y. Zhang and P. Cao (2022). "Estimation of transpiration coefficient and aboveground biomass in maize using time-series UAV multispectral imagery." The Crop Journal 10(5): 1376-1385.
Sulik, J. J. and D. S. Long (2015). "Spectral indices for yellow canola flowers." International Journal of Remote Sensing 36(10): 2751-2765.
Sulik, J. J. and D. S. Long (2016). "Spectral considerations for modeling yield of canola." Remote Sensing of Environment 184: 161-174.
Tao, Z., L. Yi, A. Bao, W. Xu, Z. Wang, S. Xiong and H. Bing (2024). "UAV or satellites? How to find the balance between efficiency and accuracy in above ground biomass estimation of artificial young coniferous forest?" International Journal of Applied Earth Observation and Geoinformation 134.
Teshome, F. T., H. K. Bayabil, G. Hoogenboom, B. Schaffer, A. Singh and Y. Ampatzidis (2023). "Unmanned aerial vehicle (UAV) imaging and machine learning applications for plant phenotyping." Computers and Electronics in Agriculture 212: 108064.
Velde, M., F. Bouraoui and A. Aloe (2008). "Pan‐European regional‐scale modelling of water and N efficiencies of rapeseed cultivation for biodiesel production." Global Change Biology 15: 24-37.
Wan, L., H. Cen, J. Zhu, J. Zhang, Y. Zhu, D. Sun, X. Du, L. Zhai, H. Weng and Y. Li (2020). "Grain yield prediction of rice using multi-temporal UAV-based RGB and multispectral images and model transfer – a case study of small farmlands in the South of China." Agricultural and Forest Meteorology 291(000).
Wang, F., Q. Yi, J. Hu, L. Xie, X. Yao, T. Xu and J. Zheng (2021). "Combining spectral and textural information in UAV hyperspectral images to estimate rice grain yield." International Journal of Applied Earth Observation and Geoinformation 102.
Wei, M. C. F., J. P. Molin and L. Longchamps (2026). "Predictive power vs interpretability: Machine learning approaches to unravel sugarcane yield drivers." Computers and Electronics in Agriculture 243(c): 111353.
Yan, X., Z. Zhang, M. Huang, X. Zhao, F. Yang and X. Wu (2022). "The impact of climate change on growth and drought-induced mortality risk of Robinia pseudoacacia plantations along a precipitation gradient on the Chinese Loess Plateau." Agricultural and Forest Meteorology 325.
Ye, Y., L. Jin, C. Bian, G. Xian, Y. Lin, J. Liu and H. Guo (2024). "Estimating potato aboveground biomass using unmanned aerial vehicle RGB imagery and analyzing its relationship with tuber biomass." Field Crops Research 319: 109657.
Zhai, W., C. Li, Q. Cheng, F. Ding and Z. Chen (2023). "Exploring Multisource Feature Fusion and Stacking Ensemble Learning for Accurate Estimation of Maize Chlorophyll Content Using Unmanned Aerial Vehicle Remote Sensing." Remote Sensing 15(13).
Zhang, J., X. Qiu, Y. Wu, Y. Zhu, Q. Cao, X. Liu and W. Cao (2021). "Combining texture, color, and vegetation indices from

Unedited

fixed-wing UAS imagery to estimate wheat growth parameters using multivariate regression methods." Computers and Electronics in Agriculture 185.

Zhang, Y., Y. Niu, Z. Cui, X. Chai and L. Xu (2025). "Cross-Year Rapeseed Yield Prediction for Harvesting Management Using UAV-Based Imagery." Remote Sensing 17(12): 2010.

Zhang, Z., J. Guo, Y. Gao, F. Zhang, Z. Hou, Q. An, A. Yan and L. Zhang (2025). "Increasing yield estimation accuracy for individual apple trees via ensemble learning and growth stage stacking." Computers and Electronics in Agriculture 237.

Supplementary information:

Fig. S1; Materials and methods

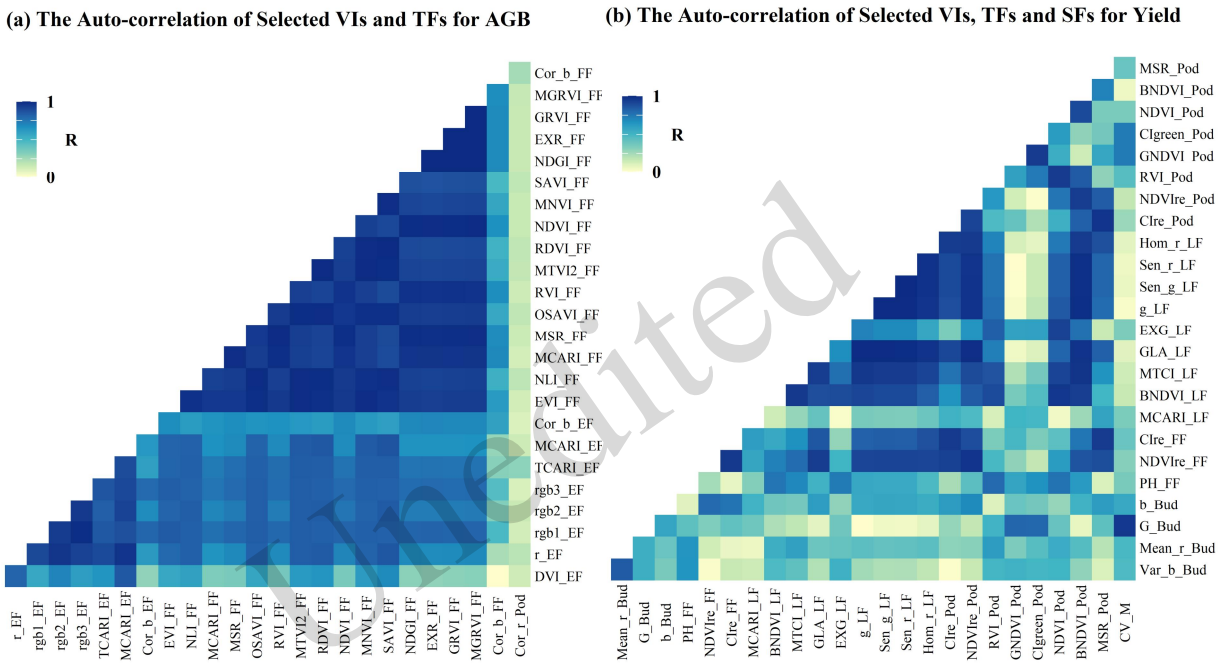


Fig. S1 The autocorrelation analysis of biomass characteristic combinations VIs-TFs and biomass characteristic combinations VIs-TFs-SFs

Unedited

Materials and methods

1.1 Experimental design

Rapeseed was transplanted on December 21, 2023 and harvested on June 2, 2024. Thirty-six plots (7 m×7 m) were established, with inter-plant spacing of 15 cm, 25 cm, and 35 cm, respectively. The four rapeseed varieties were Yangyou (YY) 9, Zhongyou (ZY) 19, Ningza (NZ) 158, and Suyou (SY) 1908. In 2024, manual sowing took place on October 17. The Detian (DT) 158 variety was manually harvested on May 18, 2025, and the remaining varieties on May 24, 2025; 108 sample plots were established (2.5 m×10 m). The rapeseed varieties were Qinyou (QY) 1718, DT 158, SY 1908, and NZ 158, with planting densities of 1.25 g/row, 1.875 g/row, and 2.5 g/row. The planting scheme for 2025 will remain consistent with that of 2024, except for the omission of the DT 158 variety. Before planting, all plots at the experimental site were applied with 375 kg/ha of green polymeric compound fertilizer. Irrigation was solely dependent on rainfall.

1.2 Data collection

For each annual experiment, above-ground biomass (AGB) and yield sampling were conducted at rapeseed maturity across all plots. In 2023, rapeseed yield was collected from the entire planting area, with a total of 36 samples collected. AGB and yield samples were collected in 2024 (36 plots×5 samples). In 2025, sampling was repeated accordingly (27 plots×3 samples). During data collection, the rapeseed plants at the plot edges were damaged. Consequently, sampling was conducted in the central three rows, with each plot further divided into five smaller subplots to ensure sample uniformity and representativeness while minimizing bias from edge

effects. The fresh AGB for each plot sample was immediately measured and recorded on-site. After air-drying in the field, the rapeseed was manually threshed, and the yield was measured using an electronic scale. Table S1 presents the analysis of variance for the interaction effects of variety and density on measured biomass and yield. It is worth noting that, under drought stress in 2025, variety was not significantly correlated with rapeseed yield, whereas the correlation between planting density and yield approached significance.

Table S1 The effects of rapeseed variety, density, and their interaction on rapeseed yield and biomass

Year	Factor	Yield	Biomass
2023	Variety	0.00073	
	Density	0.86	
	Variety&Density	0.85	
2024	Variety	3.05e-10	<2e-16
	Density	0.39	0.15
	Variety&Density	0.89	0.97
2025	Variety	0.17	2.58e-10
	Density	*0.09*	0.23
	Variety&Density	0.96	0.55

Bold p-values indicate $p < 0.05$, italicized p-values indicate $p < 0.1$

Unmanned aerial vehicles (UAVs) imagery was acquired, with sampling occurring approximately once a week on clear days. Phenological phases were systematically recorded and documented by researchers during each UAVs flight. Notable canopy variations were observed across distinct stages. Specifically, during the budding phase, rapid growth of the stalk and leaves was noted. After the budding stage, the yellow flowers gradually cover the entire rapeseed canopy. At the podding stage, its yield increases rapidly, while biomass remains relatively stable. Valid images were collected in 2023 (8 sets), 2024 (13 sets), and 2025 (16 sets), covering stages from budding to maturing (Table S2).

Table S2 Unmanned aerial vehicles (UAVs) and ground-based data collection

Growth Stages	Trails Dates (2023)	Trails Dates (2024)	Trails Dates (2025)	Data
Budding	March 19	February 17, March 8	February 24, March 1, March7, March11	MS Images, RGB Images
Early Flowering	March 26	March 16	March 18	
Full Flowering	April 7, April 12	March 22, March 29	March 24, March 30	
Late Flowering	April 16	April 9, April 15	April 6, April14	
Poddling	April 21	April 22, May 1	April 20, April28	
Maturing	May 8, May 14	May 8, May 15	May 7	
Harvesting	June 2	May 18, May 24	May 24	Fresh biomass, yield

MS: multispectral images

Using sensors mounted on DJI drones, we conducted multi-year and multi-temporal image acquisition of the rapeseed canopy, capturing its spectral characteristics and structural details. UAVs images were captured around midday under clear, sunny, and windless conditions. Real-time kinematic positioning enables drones to achieve centimeter-level positioning accuracy. The flight path, heading, lateral overlap rate (85%), forward speed, and flight altitude (30 m) are consistent across all aerial surveys. A Zenmuse L1x camera, mounted on a Matrice 300 (M300) drone, was used to capture RGB images (0.33 cm/pixel). Multispectral images were acquired using a Phantom 4 multispectral (P4M) drone. The images have a half-maximum width of 16 nm, with center wavelengths at 450 nm (Red, R), 560 nm (Green, G), 600 nm (Blue, B), 730 nm (Near-Infrared, NIR), and 840 nm (Red-edge, RE) (1.6 cm/pixel).

1.3 Multidimensional feature extraction and selection

Pix4Dmapper was employed to process the multispectral and RGB images (Pix4D, Lausanne, Switzerland). A sequence of procedures was conducted, primarily involving photo alignment, the incorporation of control points, and the generation of dense point clouds. Radiometric calibration was conducted using reflectance images obtained from a reflectance calibration panel across five spectral bands. Geometric correction was conducted using ground control points (GCPs) to enhance spatial accuracy. The spectral vegatation indices (VIs) used in this study were divided into two types: those that include near-infrared, red-edge, and visible light bands, and those that cover only the visible light band (Table S3). Structural features (SFs) were extracted from crop surface model (CSM) based on the structure-from-motion algorithm (Malambo

Unedited

et al., 2018). Plant height (PH) was derived from CSM at the maximum percentiles. Canopy volume (CV) was calculated by multiplying each pixel's area by the corresponding PH. Additionally, texture feature (TFs) were extracted from the orthophotography RGB images. We calculated eight TFs using the "glcm" package in R software. Since RGB images consist of 3 color bands (R, G, B), a total of 24 TFs were extracted for each region of interest (ROI). The naming convention appends suffixes (_r, _g, _b) to each texture to denote the specific feature corresponding to each band.

Table S3 Spectral vegetation indices (VIs) for rapeseed yield and biomass estimation

Class	VIs	Formula	References
RGB VIs	R,G,B	Raw value of each band	/
	r	R(R+G+B)	/
	g	G(R+G+B)	/
	b	B(R+G+B)	/
	r/b	r/b	/
	r-b	r–b	/
	(r-b)/(r+b)	(r–b)/(r+b)	/
	GLA	(2G–B–R)/(2G+B+R)	(Ye et al., 2024)
	NDGI	(G–R)/(G+R)	(Zhao et al., 2020)
	MGRVI	$(g^2–r^2)/(g^2+r^2)$	(Ye et al., 2024)
	EXR	1.4×r–g	(Ye et al., 2024)
	EXG	2×g–r–b	(Ye et al., 2024)
Yield RE-NIR VIs	BNDVI	(NIR–B)/(NIR+B)	(Suilk and Long 2016)
46	GNDVI	(NIR–G)/(NIR+G)	(Suilk and Long 2016)
25	NDVIre	(NIR–RE)/(NIR+RE)	(Meng et al., 2025)
24	CIre	NIR/RE–1	
	CIgreen	NIR/G–1	(Meng et al., 2025)
	MTCI	(NIR–RE)/(RE–R)	(Meng et al., 2025)
	MSR	$(NIR/R–1)/(NIR/R+1)^{1/2}$	(Lu et al., 2022)
	NDVI	(NIR–R)/(NIR+R)	(Suilk and Long 2016)
	RVI	NIR/R	(Lu et al., 2022)
	MCARI	[(RE–R)–0.2(RE–G)](RE/G)	(Meng et al., 2025)
Biomass RE-NIR VIs	TVI	60(NIR–G)–100R–G	(Meng et al., 2025)
	NLI	$(NIR^2–R)/(NIR^2+R)$	(Lu et al., 2022)
	EVI	2.5(NIR–R)/(NIR+6R–7.5B+1)	(Meng et al., 2025)
	RDVI	$(NIR–R)/(NIR+R)^{1/2}$	(Li et al., 2020)
	SAVI	1.5(NIR–R)/(NIR+R+0.5)	(Meng et al., 2025)
	OSAVI	1.16(NIR–R)/(NIR+R+0.16)	(Meng et al., 2025)
	TCARI	3[RE–R–0.2(RE–G)(RE/R)]	(Zarco-Tejada et al., 2002)
	MNVI	$1.5(NIR^2–R)/(NIR^2+R+0.5)$	(Gong et al., 2003)
	MTVI2	$1.5[1.2(NIR–G)–2.5(R–G)]/[(2NIR+1)^2–(6×NIR–5R^{1/2}–0.5)]^{1/2}$	(Meng et al., 2025)

R: Red; G: Green; B: Blue; NIR: Near-Infrared; RE: Red-edge. The bold Red-edge and Near-Infrared (RE-NIR) based VIs serve as a common index for both biomass and yield estimation

Unedited

References

Gong, P., R. Pu, G. S. Biging and M. R. Larrieu (2003). "Estimation of forest leaf area index using vegetation indices derived from Hyperion hyperspectral data." IEEE Transactions on Geoscience & Remote Sensing 41(6): 1355-1362.

Li, X. Z., C. Yang, W. Huang, J. Tang, Y. Tian and Q. Zhang (2020). "Identification of Cotton Root Rot by Multifeature Selection from Sentinel-2 Images Using Random Forest." Remote. Sens. 12: 3504.

Lu, L., J. Luo, Y. Xin, H. Duan, Z. Sun, Y. Qiu and Q. Xiao (2022). "How can UAV contribute in satellite-based Phragmites australis aboveground biomass estimating?" International Journal of Applied Earth Observation and Geoinformation 114.

Meng, L., B. Ming, Y. Liu, C. Nie, L. Fang, L. Zhou, J. Xin, B. Xue, Z. Liang, H. Guo, D. Yin and X. Jin (2025). "Maize biomass estimation by integrating spectral, structural, and textural features from unmanned aerial vehicle data." European Journal of Agronomy 168.

Sulik, J. J. and D. S. Long (2016). "Spectral considerations for modeling yield of canola." Remote Sensing of Environment 184: 161-174.

Ye, Y., L. Jin, C. Bian, G. Xian, Y. Lin, J. Liu and H. Guo (2024). "Estimating potato aboveground biomass using unmanned aerial

vehicle RGB imagery and analyzing its relationship with tuber biomass." Field Crops Research 319: 109657.

Zarco-Tejada, P., D. Haboudane, J. R. Miller, N. Tremblay and L. Dextraze (2002). "Leaf Chlorophyll a+b and canopy LAI estimation in crops using R-T models and Hyperspectral Reflectance Imagery."

Zhao, J., Y. Jin, H. Ye, W. Huang, Y. Dong, L. Fan, H. Ma and J. Jiang (2020). "Remote sensing monitoring of areca yellow leaf disease based on UAV multi-spectral images." Transactions of the Chinese Society of Agricultural Engineering 36(8): 54-61.

Unedited