Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Integrating spectral, texture, soil and fertilization information for plot-level prediction of sugarcane yield, millable stalk population and Brix from Jilin-1 imagery

Yun Zhang
Anhui Normal University
Guangtao Xu
Anhui Normal University
Dan Li
lidan@gdas.ac.cn

Guangzhou Institute of Geography, Guangdong Academy of Sciences
Yinglin Lu
Institute of Nanfan and Seed Industry, Guangdong Academy of Sciences
Hao Jiang
Guangzhou Institute of Geography, Guangdong Academy of Sciences
Shasha Luo
Institute of Nanfan and Seed Industry, Guangdong Academy of Sciences
Mingfu Wen
Zhanjiang Research Center, Institute of Nanfan and Seed Industry, Guangdong Academy of Sciences
Jing Zhang
Guangzhou Institute of Geography, Guangdong Academy of Sciences
Xingda Chen
Guangzhou Institute of Geography, Guangdong Academy of Sciences
Qiong Zheng
Changsha University of Science and Technology
Chongyang Wang
Guangzhou Institute of Geography, Guangdong Academy of Sciences
Zihan Liu
Anhui Normal University
Mengjun Ku
Guangzhou Institute of Geography, Guangdong Academy of Sciences
Junliang Chen

Guangzhou Institute of Geography, Guangdong Academy of Sciences

Research Article

Keywords: Sugarcane phenotypic traits, Jilin-1, Texture indices, Machine learning, SHAP interpretability

Posted Date: April 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9371174/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Purpose

The primary objective of this study was to evaluate the potential of high spatial resolution Jilin-1 satellite imagery for plot-level prediction of sugarcane yield, millable stalk population, and Brix, and to assess whether integrating spectral, texture, soil, and fertilization information could improve prediction performance for precision sugarcane management.

Methods

Jilin-1 satellite imagery acquired at four growth stages, from seedling to maturity, was used to derive vegetation indices (VIs) and texture indices (TIs), including the normalized difference texture index (NDTI), enhanced vegetation texture index (EVTI), and double-difference ratio texture index (DDRTI). Soil chemical properties (SCPs) and fertilization information (FI) were further incorporated with the remotely sensed variables. Machine learning models were developed for plot-level prediction of sugarcane traits across plant cane and first ratoon cane, and texture window size was optimized to improve TI extraction and model performance.

Results

For yield, the combination of VIs and TIs outperformed VIs alone at the tillering stage (R^2_{CV} = 0.65, RMSECV = 15.06 t/ha, RPDCV = 1.68). Adding SCPs and FI further improved yield prediction across plant cane and first ratoon cane (R^2_{CV} = 0.70, RMSECV = 13.84 t/ha, RPDCV = 1.83). Millable stalk population was best predicted at the maturation stage by VIs and Tis, achieving the best performance (R^2_{CV} = 0.63, RMSECV = 6602 stalks/ha, RPDCV = 1.66). The best Brix model integrated VIs, TIs, SCPs, and FI at the maturation stage (R^2_{CV} = 0.44, RMSECV = 0.53 °Bx, RPDCV = 1.33). SHAP analysis identified VIs as the dominant features for sugarcane traits prediction. And, DDRTI contributed more than NDTI and EVTI in yield and Brix prediction.

Conclusion

It is concluded that integrating spectral, texture, soil, and fertilization information from high spatial resolution Jilin-1 imagery is a promising approach for improving plot-level prediction of key sugarcane traits.

1. Introduction

Sugarcane (*Saccharum* spp.) is one of the most important crops in tropical and subtropical regions, with a global planting area of more than 27 million hectares (FAO, 2025). It contributes more than 80% of global raw sugar production and plays a major role in bioenergy development, accounting for approximately 40% of global fuel ethanol production (Canton, 2021; Rosa et al., 2026). Stable high-yield and high-quality sugarcane production is important for the sustainable development of both the sugar and bioenergy industries. Sugarcane is a semi-perennial crop with a growth cycle of approximately 12–18 months, typically including the seedling, tillering, elongation, and maturation stages (Som-ard et al., 2021). Across these stages, substantial changes occur in canopy structure, biomass accumulation, leaf area index (LAI), and physiological status, which lead to pronounced variation in remote sensing signals (Yang et al., 2025). During the seedling stage, new ratoons emerge, and biomass and LAI remain low. At the tillering stage, canopy cover expands rapidly and biomass accumulation accelerates. During the elongation stage, stalk elongation and leaf expansion dominate, leading to rapid biomass accumulation, high canopy closure and an increasingly complex vertical canopy structure. At the maturation stage, leaves progressively senesce and chlorophyll content declines, while sucrose accumulates in the stalks (Li et al., 2019). This phenological transition may weaken the direct relationship between canopy greenness and yield-related traits, which makes conventional greenness-based indicators less reliable for monitoring final production status (Som-ard et al., 2021).

Sugarcane traits, such as millable stalk population, stalk weight, plant height, biomass and sucrose content, directly determine yield and economic return (Meier et al., 2016; Li et al., 2019). Early season acquisition of these traits is important. It enables timely management interventions, supports early screening in breeding programs, and improves decision making before harvest (Akbarian et al., 2024). However, conventional sugarcane phenotyping still relies heavily on manual field surveys and destructive sampling. Estimating yield and the millable stalk population usually requires large-scale field harvesting and manual counting (Zhu et al., 2026). Determining sugar content depends on destructive crushing of mature stalks (de França e Silva et al., 2024). These approaches are laborious-intensive, costly and destructive, which limits the application of precision agriculture. Remote sensing provides a non-destructive and large-scale alternative for crop monitoring and has been widely used in sugarcane mapping, growth monitoring, stress detection, and yield estimation (Jimenez-Sierra et al., 2020; Xu et al., 2025; Kambli & Palkar, 2025).

Many studies have used vegetation indices such as NDVI, GNDVI and EVI, to characterize canopy greenness and vigor (Akbarian et al., 2022). However, some commonly used broadband vegetation indices (e.g., NDVI) may become less sensitive under medium to high vegetation cover (Gao et al., 2023). In addition, many inversion models were originally developed based on medium and low spatial resolution imagery in which each pixel generally represents the average condition of a relatively large canopy area (Luciano et al., 2021; Amorim et al., 2022). As spatial resolution increases from medium or low resolution to the meter scale, the physical meaning of a pixel changes. Canopy heterogeneity, geometric effects, and local radiometric fluctuations become much more pronounced (Fan et al., 2025). The applicability of these models to high spatial resolution data has not been thoroughly evaluated. Thus, traditional models may lead to greater retrieval uncertainty. Furthermore, canopy saturation is

prone to occur when vegetation indices are applied to medium to high vegetation coverage (Zhao et al., 2012). Relying solely on spectral information is insufficient to capture the spatial complexity of sugarcane canopies, which further constrains the prediction of canopy traits (Liu et al., 2023a).

High spatial resolution remote sensing data offer spectral information and spatial details (including texture, shape, edge, and canopy arrangement patterns). Texture features derived from high-resolution imagery have the potential to characterize canopy spatial organization and structural variability (Zhu et al., 2026). The influences of canopy gaps, shadows, row structure, and adjacency-related radiometric interactions become more evident (Fan et al., 2025). Some studies have shown that integrating spectral and texture features from UAV imagery, can improve the estimation of crop biomass (Xie & Yang, 2020; Liu et al., 2023a), leaf area index (Yuan et al., 2023) and yield (Yu et al., 2025). However, the performance of texture based methods is also sensitive to feature extraction settings, such as window size, and direction, which may substantially affect parameter inversion accuracy and model robustness (Ozkan & Demirel, 2021). Moreover, the limited spatial coverage and operational scalability of UAV imagery restrict its use in routine large-area applications (Phang et al., 2023). Meter-level satellite imagery is a critical observational scale that bridges ultra fine observations and operational agricultural monitoring (Aleman-Montes et al., 2023; Fan et al., 2025). Open-access satellite constellations such as Sentinel have played an important role in crop monitoring (Luciano et al., 2021). However, these data are often constrained by revisit intervals and cloud contamination in tropical and subtropical regions. Increasing attention has been directed toward high resolution commercial imagery for more detailed agricultural monitoring and improved yield estimation (Wang et al., 2025). Nevertheless, systematic investigations of texture features under high spatial resolution conditions, including the optimal parameters for texture feature extraction and effective combination strategies remain limited. The potential of integrating spectral and texture data has not yet been fully explored for sugarcane phenotyping.

Among these platforms, the Chinese Jilin-1 satellite constellation provides a balance among spatial resolution, revisit frequency and cost-effectiveness. Its application potential has been evaluated in leaf area index inversion for corn and rice (Du et al., 2023), environmental quality assessment in flood-prone areas (Chen et al., 2023), and wetland vegetation classification using UAV-satellite fusion technology (Fu et al., 2022). Although Jilin-1 imagery perform effectively in these fields, its potential in field-level crop phenotyping monitor has not been fully investigated. For sugarcane yield estimation, previous studies have adopted process-based crop models combined with data assimilation to improve prediction accuracy at regional scales (Xu et al., 2025). However, applying these process-based framework to meter level imagery is often data- and parameter-intensive. Acquiring such fine scale inputs is challenging to, which limits their practical applicability in fine-scale crop monitoring (de França e Silva et al., 2024; Rosa et al., 2026). Machine learning (ML) algorithms are good at modeling complex non-linear relationships. They have shown great potential in dealing with multi-source heterogeneous data (Paudel et al., 2021; Li et al., 2026). In addition, explainable artificial intelligence has also attracted increasing attention due to the transparency of model decision-making processes (Xuan et al., 2025; Wu et al., 2025). Integrating high-resolution spectral features, texture information, and multi-source agricultural auxiliary data for plot level sugarcane trait prediction by ML algorithm has not yet been fully established.

This study aims to evaluate the potential of meter-level Jilin-1 imagery for plot-level sugarcane phenotyping. Plot-level field observations were collected for key sugarcane traits, including yield, millable stalk population, and Brix. Three objectives were addressed: 1) trait estimation, assessing the ability of spectral and texture indices to estimate key sugarcane traits; 2) data integration, evaluating whether the incorporation of multi-source agricultural auxiliary data improves prediction accuracy; and 3) model interpretation, identifying the most important predictors for different traits using explainable machine learing.

2. Materials and methods

2.1. Study area and experimental design

A two-year trial was conducted at the Zhanjiang Research Center (Nanfan Seed Industry Research Institute of Guangdong Academy of Sciences) in Suixi County, Guangdong Province (Fig. 1). This location has a tropical monsoon climate, with an annual average temperature of about 23°C and precipitation exceeding 1400 mm. There is sufficient sunlight, heat and water, which are highly suitable for sugarcane cultivation. The study spanned two years, focusing on plant cane in 2024 and first ratoon cane in 2025. The experimental field consisted of 32 plots, and the sugarcane variety planted was ‘Yutang 03373’. Each 100-square-meter plot contained 9 rows, with each row approximately 10 m in length.

Basal and topdressing fertilizers were applied at the seedling and tillering stages, respectively. The experiment comprised eight fertilization regimes (T1–T8) using a compound fertilizer (N: P_2O_5: K_2O = 15: 15: 15), as detailed in Table S1.

2.2. Data acquisition and preprocessing

2.2.1. Field data acquisition

Phenotypic data were collected on November 19, 2024 (plant cane) and December 2, 2025 (first ratoon cane), at the early maturity stage. In addition, for each plot, 20 plants were randomly selected to measure plant height (cm), stem diameter (cm) and single stem weight (g). To minimize edge effects, sampling was restricted to the three central rows (rows 4–6) of each plot. Twenty stalks were randomly selected from this zone for yield component analysis. Brix was measured from the juice of middle internodes using a handheld refractometer. Millable stalk population were defined as healthy, mature stalks meeting agronomic standards for height and diameter. The total number of stalks was estimated by counting the number of stalks in the middle row and extrapolating the density to the plot level. Yield was calculated based on the formula for stem fresh weight per hectare. The calculation formulas are as follows:

$$W = \pi \times \left(\frac{D}{2}\right)^2 \times L \times \rho$$

1

$$Y = W \times MS$$

2

where *Y* is the yield (t/ha), *W* is the theoretical single stalk weight (kg), *D* and *L* denote stalk diameter (cm) and length (cm), respectively, *MS* is the number of millable stalks, and *ρ* is the stalk density (constant at 1.05 g/cm³).

Soil samples were collected from the 0–20 cm soil layer of each plot prior to planting. The soil chemical properties were analyzed in the laboratory (Table S2).

2.2.2. Remote sensing data acquisition and preprocessing

Jilin-1 satellite imagery for the seedling, tillering, elongation and maturation stages was obtained via the official data portal (https://www.jl1mall.com/). Each scene included a panchromatic band and four multispectral bands (Blue, Green, Red, and Near-Infrared). Table S3 details the sensor specifications and acquisition dates. We digitized plot boundaries using QGIS 3.34. The preprocessing workflow was conducted in ENVI 5.6, involving radiometric calibration, atmospheric correction, geometric registration, and image fusion to derive surface reflectance. The spectral response functions for atmospheric correction were obtained from the website (https://www.jl1mall.com/resrepo/).

2.2.3. Calculation of vegetation and texture indices

The preprocessed Jilin-1 imagery and sugarcane plot vector boundaries were imported into the GEE platform. Twenty widely used VIs were computed, as listed in Table S4. Texture features (TFs) were extracted using the GLCM algorithm (Haralick et al., 1973). This process employed three moving window sizes (3×3, 5×5, and 7×7 pixels), and the texture features were averaged across four directional angles (0°, 45°, 90°, and 135°) for each window size. Consequently, eight standard TFs (Table S5) were generated for each of the four spectral bands Furthermore, three advanced indices, namely the normalized difference texture index (NDTI), enhanced vegetation texture index (EVTI) and double difference ratio texture index (DDRTI), were constructed by combining these TFs. Table 1 lists the calculation formulas for the TIs.

Table 1
Summary of texture indices used in this study.

Parameters	Formulas	Reference
Normalized Difference Texture Index (NDTI)	$NDTI = \frac{T_1 - T_2}{T_1 + T_2}$	(Zheng et al., 2019)
Enhanced Vegetation Texture Index (EVTI)	$EVTI = \frac{2.5 \times (T_1 - T_2)}{(T_1 + 6 \times T_2 - 7.5 \times T_3 + 1)}$	(Yuan et al., 2023)
Double Difference Ratio Texture Index (DDRTI)	$DDRTI = \frac{T_1 - T_2}{T_3 - T_4}$	This study
Note: T_1–T_4 denote the same texture feature derived from arbitrary spectral bands.		

2.3. Feature selection

Pearson correlation analysis was performed to investigate the relationships between input variables, revealing the presence of high multicollinearity and redundancy. To address this issue, SVR-RFE (Guyon et al., 2002) was used to identify the optimal feature subset. This method iteratively removed the features with the lowest model contribution, retaining only those with the highest predictive power for each growth stage.

2.4. Machine learning models

This study evaluated the ability of Jilin-1 satellite imagery to predict phenotypic traits of four growth stages of sugarcane (seedling, tillering, elongation, and maturation stages). Four regression models were compared, namely ridge regression (RR) (Hoerl & Kennard, 1970), Gaussian process regression (GPR) (Rasmussen and Williams, 2006), categorical boosting (CatBoost) (Friedman, 2002; Zhang et al., 2020) and support vector regression (SVR) (Drucker et al., 1997; Durbha et al., 2007). All analyses and modeling were conducted using Python 3.12.

2.5. Model optimization and evaluation

Grid search with five-fold cross-validation was used to optimize the model hyperparameters. The specific parameters and search ranges are listed in Table S6.

Leave-one-out cross-validation (LOOCV) (Brovelli et al., 2008) was used to evaluate the model. Subsequently, the bootstrap method (Efron, 1979) was applied to further verify the robustness and reliability of the model. Model performance was evaluated using the coefficient of determination (R^2), root mean square error (RMSE) and ratio of performance to deviation (RPD). The RPD values were interpreted based on the classification criteria proposed by Gaston et al. (2010): 1.0–1.4 indicates poor prediction; 1.4–1.8, fair; 1.8–2.0, good; and 2.0–2.5, very good. The calculation formulas are as follows:

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

3

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

4

$$RPD = \frac{SD}{RMSE}$$

5

$$SD = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \bar{y}_i\right)^2}{n}}$$

6

where y_i is the actual value; $\hat{y}_i$ is the predicted value; $\bar{y}_i$ is the mean of the actual values; *n* is the number of samples; and *SD* is the standard deviation of the target variable.

2.6. Feature importance assessment

SHAP (Lundberg and Lee, 2017) was used to interpret model predictions. It assigned each input feature an importance value and ranked the features at a global level. All calculations were performed using the SHAP library in Python 3.12.

2.7. Research framework

Figure 2 illustrates the workflow of this study, which consists of the following key steps:

1. (a) Data Acquisition and Preprocessing. Obtain high-quality surface reflectance data from Jilin-1 satellite imagery through radiometric calibration, atmospheric correction, geometric correction, and band fusion; collect field survey data including soil chemical properties (SCPs), fertilization information (FI), and sugarcane phenotypic data, and complete data preprocessing and accuracy verification to form a multi-source dataset.
2. (b) Optimization of Optimal Texture Window Size. Calculate multiple vegetation indices (VIs) and extract texture features using the Gray Level Co-occurrence Matrix (GLCM). Construct texture indices (TIs) under different window sizes (3×3, 5×5 and 7×7), then build machine learning models (SVR, RR, GPR and CatBoost) and evaluate performance via Leave-One-Out Cross Validation

(LOOCV) with RPD_{CV}, to determine the optimal texture window size for sugarcane phenotype prediction.

3. (c) Determination of the Optimal Prediction Model. Construct multi-dimensional input feature schemes (VIs, VIs + TIs and VIs + TIs + SCPs + FI) based on the optimal texture window size, train multiple machine learning models, and assess model performance using LOOCV and Bootstrap resampling with evaluation metrics (R^2_{CV}, $RMSE_{CV}$, and RPD_{CV}) to select the optimal prediction model.
4. (d) Model Interpretability Analysis. Apply the SHAP interpretable method to the optimal prediction model, quantify feature importance, analyze the response patterns of key features, and compare the contribution of multi-source features to reveal the driving mechanisms of model predictions for sugarcane phenotypes.

3. Results and analysis

3.1. Variations of sugarcane traits, NDVI and $DDRTI_MEA_{(B,R,G,NIR)}$

The statistical results and scatter plots for yield, millable stalk population, and brix data for 2024 and 2025 are presented in Fig. S1 and Table 2. The mean values of yield, millable stalk population and brix in 2024 were higher than those in 2025. Variation analysis results indicated that there were no significant differences between the two years of yield data and brix ($p > 0.05$), whereas significant differences were observed for millable stalk population ($p < 0.05$). The first ratoon maintained yield with fewer stalks by increased individual stalk weight. The stalk quality and yield were preserved even the stalk number declined.

The coefficient of variation (CV) for brix remained below 4% across two years. The CV for millable stalk population increased from 14.93% in 2024 to 20.50% in 2025, and the minimum value decreased from 36900 to 22500 stalks/ha. It seems that this variety compensated for the loss of millable stalks through increased individual stalk weight in ratoon cane with relatively stable yield and Brix.

In this study, there are a large number of VIs and TIs were used. NDVI and $DDRTI_MEA_{(B,R,G,NIR)}$ are presented as representative examples for seasonal pattern analysis. The seasonal variation of NDVI in 2024 and 2025 was presented in Fig. S2a. It is clear that NDVI increased more rapidly in 2025 than 2024 from the seedling to the tillering. During the subsequent growth period, NDVI continued to rise and reached its maximum in early September 2025, whereas the peak occurred in mid-September 2024. In addition, during maturation, NDVI declined less and remained higher in 2025. The seasonal dynamics of $DDRTI_MEA_{(B,R,G,NIR)}$ in 2024 and 2025 were given in Fig S2b. In 2024, $DDRTI_MEA_{(B,R,G,NIR)}$ increased slightly from the seedling stage to the tillering stage and then gradually declined after elongation. In 2025, this index was low during the early stages, followed by a rapid from late tillering to elongation with a pronounced peak in early September before dropping during maturation.

Table 2
Descriptive statistics of the measured sugarcane traits.

Sugarcane Traits	Year	Mean	SD	Min	Max	CV(%)
Yield (t/ha)	2024	70.53	27.45	25.67	136.01	38.92
	2025	66.54	23.77	25.47	106.17	35.73
Millable Stalk Population (stalks/ha)	2024	57825	8636	36900	74700	14.93
	2025	44803	9186	22500	58500	20.50
Brix (°Bx)	2024	20.87	0.77	19.54	22.64	3.73
	2025	20.61	0.61	19.28	21.84	2.97

3.2. Correlation analysis of input features and sugarcane traits

3.2.1 Correlation between vegetation indices and sugarcane traits

Twenty VIs were evaluated by Pearson correlation analysis across four growth stages (Fig. S3). Correlation coefficients (r) and significance levels (p) were calculated. For sugarcane yield (Fig. S3a), the strongest correlations were observed during the tillering and elongation stages. At the tillering stage, GLI exhibited the highest positive correlation ($r = 0.60$, $p < 0.001$), followed by CVI and GRVI ($r = 0.52$). At the elongation stage, NGRDI and WDRVI showed the strongest correlations ($r = 0.57$, $p < 0.001$). Correlations at the seedling and maturation stages were generally weaker. The correlation between VIs and millable stalks showed distinct temporal patterns (Fig. S3b). The seedling stage was the most critical window for monitoring stalk number. Almost all VIs showed uniformly strong positive correlations (approx. $r = 0.60$) with high statistical significance ($p < 0.001$). Interestingly, an inverse relationship was observed in later stages. During the elongation and maturation stages, most VIs were significantly negatively correlated with millable stalk population ($|r|$ values ranging from 0.50 to 0.60). In contrast to yield and millable stalk population, Brix showed no significant correlation with any of the tested VIs across all growth stages (Fig. S3c). The correlation coefficients remained low ($|r| < 0.25$) and statistically insignificant, suggesting that standard VIs may not be sensitive enough to detect variations in sugarcane Brix.

3.2.2 Correlation between texture indices and phenotypic traits

To evaluate TIs suitable for modeling, Pearson correlation analyses were conducted between TIs (calculated using different window sizes of 3×3, 5×5, and 7×7) and sugarcane phenotypic traits across four growth stages (seedling, tillering, elongation, and maturation). In this study, TIs were derived by combining the same TFs across different spectral bands. Eight TFs were selected, and the three major categories of TIs (NDTI, EVTI, and DDRTI) were divided into eight sub-categories, respectively. From each

sub-category, the texture index exhibiting the highest absolute correlation coefficient (|*r*|) was selected. Figs. S4, S5, and S6 illustrate the Pearson correlation coefficients between TIs and yield, millable stalk population, and Brix. TIs calculated with different window sizes performed differently for the three phenotypic traits. For sugarcane yield, the correlation of texture indices first increased and then decreased during growth. It rose from 0.40 ($p < 0.001$) at the seedling stage to 0.67 ($p < 0.001$) at the tillering stage, then dropped to 0.53 ($p < 0.001$) at the maturation stage. For millable stalk population, the correlation of texture indices increased gradually. It rose from 0.47 ($p < 0.001$) at the seedling stage to 0.68 ($p < 0.001$) at the maturation stage. For Brix, the correlation showed a downward trend. The highest correlation coefficient was 0.33 ($p < 0.05$) at the seedling stage.

3.2.3 Correlation between agronomic factors and phenotypic traits

Pearson correlation analysis was conducted to evaluate the relationship between SCPs, FI, and sugarcane phenotypic traits (Fig. S7). Significant correlations were observed between yield and various SCPs as well as FI. In terms of SCPs, AP showed the strongest negative correlation ($r = -0.43$, $p < 0.001$), while soil pH exhibited a significant positive correlation ($r = 0.40$, $p < 0.001$). Regarding FI, there was a significant positive correlation between yield and variables such as BF_CF, BF_KCl, TD_CF, and TD_KCl ($r = 0.46$, $p < 0.001$). Millable stalk population, a key component of yield, showed the highest correlation with FI ($r = 0.35$, $p < 0.05$), while its correlation with SCPs was low. In contrast, the correlations between Brix and these indicators were all weak and statistically insignificant.

3.3. Determination of optimal texture calculation window size

Four machine learning models were employed to determine the optimal window size, with RPD_{CV} scores illustrated in Fig. 3. Table S7 summarized the best parameter combinations for each stage. The results indicate that the optimal texture combination varied across growth stages.

3.4. Prediction of sugarcane traits based on multi-source features

3.4.1. Yield estimation model

Table S8 summarizes the yield prediction results. The performance of four models (SVR, GPR, RR, and CatBoost) with the SVR-RFE feature selection method across four growth stages were compared. Figure 4(a1–c1) presents the prediction results of the best models based on LOOCV for the three different input data combinations.

When using only VIs, the SVR model at the tillering stage achieved the highest accuracy. The R^2_{CV} was 0.60, the $RMSE_{CV}$ was 16.02 t/ha, and the RPD_{CV} was 1.58 (Fig. 4a1). After adding TIs, the RR model

performed the best at the tillering stage. The R^2_{CV} increased to 0.65, with an $RMSE_{CV}$ of 15.06 t/ha and an RPD_{CV} of 1.68 (Fig. 4b1). Finally, with the addition of SCPs and FI, the SVR model at the tillering stage reached the highest accuracy. The R^2_{CV} rose to 0.70, the $RMSE_{CV}$ decreased to 13.84 t/ha, and the RPD_{CV} reached 1.83 (Fig. 4c1). The RPD value shifted from the “fair” category (1.4–1.8) to the “good” category (1.8–2.0).

To further verify the stability and robustness of the prediction models, the distribution of RPD values were also obtained from the bootstrap method (Fig. 4(a2–c2)). The mean RPD in Fig. 4c2 was 1.56, slightly lower than the 1.58 in Fig. 4b2. However, the 75th percentile in Fig. 4c2 reached 1.75, surpassing the 1.73 in Fig. 4b2.

3.4.2. Millable stalk population estimation model

For millable stalk population, the performance of tour models (SVR, GPR, RR, and CatBoost) is presented in Table S9. Three input combinations were compared. Figure 5(a1–c1) presents the prediction results of the best models based on LOOCV for the three different input data combinations.

When using only VIs, the SVR model at the maturation stage achieved the highest accuracy. The R^2_{CV} was 0.57, the $RMSE_{CV}$ was 7181 stalks/ha, and the RPD_{CV} was 1.52 (Fig. 5a1). After adding TIs, the SVR model still performed the best at the maturation stage. The R^2_{CV} increased to 0.63, with an $RMSE_{CV}$ of 6602 stalks/ha and an RPD_{CV} of 1.66 (Fig. 5b1). Finally, with the addition of SCPs and FI, the optimal prediction window shifted to the elongation stage. However, the overall accuracy decreased compared to the previous step. The R^2_{CV} dropped to 0.62, the $RMSE_{CV}$ increased to 6733 stalks/ha, and the RPD_{CV} fell to 1.62 (Fig. 5c1). The model performance remained within the “fair” category (1.4–1.8) and did not reach the “good” standard (1.8–2.0).

To further verify the stability and robustness of the prediction models, the distribution of RPD values was obtained from the bootstrap method. As shown in Fig. 5(a2–c2), the model in Fig. 5b2 performed best. Its 25% percentile (1.37), mean (1.51), and 75% percentile (1.63) were higher than those of the other two models. This indicates that the model with VIs + TIs (Fig. 5b2) achieved the highest accuracy and stability.

3.4.3. Brix estimation model

Table S10 shows the Brix prediction results. Preliminary results indicated that the SVR model consistently outperformed the other three models across all growth stages. Figure 6(a1–c1) presents the prediction results of the SVR model based on LOOCV for three different input data combinations.

Using VIs alone, the SVR model at the seedling stage achieved the highest accuracy. The R^2_{CV} was 0.21, the $RMSE_{CV}$ was 0.63 °Bx, and the RPD_{CV} was 1.12 (Fig. 6a1). After adding TIs, the SVR model still performed the best at the seedling stage. The R^2_{CV} increased to 0.40, with an $RMSE_{CV}$ of 0.55 °Bx and an

RPD_{CV} of 1.29 (Fig. 6b1). Finally, with the addition of SCPs and FI, the optimal prediction window shifted to the maturation stage. Notably, the overall accuracy improved compared to the previous step. The R^2_{CV} rose to 0.44, the $RMSE_{CV}$ decreased to 0.53 °Bx, and the RPD_{CV} reached 1.33 (Fig. 6c1). The model performance remained within the “poor” category (1.0–1.4) and did not reach the “fair” standard (1.4–1.8).

To further verify the stability and robustness of the SVR prediction model, the distribution of RPD_{CV} values was obtained from the bootstrap method (Fig. 6a2–c2). The model in Fig. 6b2 exhibited slightly higher stability metrics (a mean RPD_{CV} of 1.18). However, the model in Fig. 6c1 achieved the highest explanatory power (R^2_{CV} = 0.44) in the LOOCV assessment. Considering the priority of prediction accuracy, the SVR model developed by VIs + TIs + SCPs + FI was identified as the final model.

3.5. Model interpretation based on SHAP

SHAP analysis was used to explain feature contributions to sugarcane trait prediction. Figure 7 presents the SHAP summary and feature importance for the optimal models. For yield prediction, Fig. 7a1 ranks the features by their importance. GNDVI was the most important feature, contributing 24.7% to the model. CI and RVI were the next most important features, contributing 22.7% and 22.5%, respectively. BF_CMF ranked fourth (8.1%), followed by CVI (7.9%) and $DDRTI_CON_{(G,NIR,B,R)}$ (4.7%) occupied the fifth and sixth positions. The inset pie chart shows the contribution of different categories (Fig. 7a2). VIs were the dominant drivers, accounting for 77.8% of the total contribution, which was significantly higher than that of FI (11.7%) and TIs (10.5%). Figure 7a3 illustrates how these features affect yield; red dots represent high feature values, while blue dots represent low values. Yield prediction was positively correlated with GNDVI (ranked first) and CVI (ranked fifth). Conversely, CI and RVI (ranked second and third) showed a negative correlation, where lower values corresponded to higher yield predictions. Finally, BF_CMF (ranked fourth) had a positive impact on yield, whereas $DDRTI_CON_{(G,NIR,B,R)}$ (ranked sixth) showed a negative correlation.

For the prediction of millable stalk population, Fig. 7b1 ranks the features by importance. CVI was the most critical feature, contributing 24.7% to the model. $NDTI_HOM_{(B_G)}$, $EVTI_VAR_{(NIR_R_B)}$, and $EVTI_MEA_{(G,NIR,R)}$ followed, contributing 18.6%, 18.0%, and 16.9%, respectively. NDTI_VAR(G,NIR) (9.9%) and $EVTI_ASM_{(G,B,NIR,G)}$ (6.1%) ranked fifth and sixth. The inset pie chart (Fig. 7b2) shows that TIs were the dominant drivers, accounting for 69.5% of the total contribution, which was much higher than that of VIs (30.5%). Figure 7b3 displays the feature effects on millable stalk population. The prediction was positively correlated with CVI (ranked first), $NDTI_VAR_{(G,NIR)}$ (ranked fifth), and $EVTI_ASM_{(G,B,NIR,G)}$ (ranked sixth). In contrast, $NDTI_HOM_{(B,G)}$, $EVTI_VAR_{(NIR,R,B)}$, and $EVTI_MEA_{(G,NIR,R)}$ (ranked second, third, and fourth) showed negative correlations, as lower values corresponded to higher millable stalk population predictions.

For Brix prediction, Fig. 7c1 ranks the features by importance. GDVI was the top feature, contributing 35.9% to the model, followed by BNDVI with 27.8%. $DDRTI_VAR_{(G,R,B,NIR)}$ (8.9%) and $DDRTI_HOM_{(G,R,B,NIR)}$

(4.5%) ranked third and fourth, while WDRVI took the fifth spot (4.1%). Notably, soil pH also played a role in the model. The inset pie chart (Fig. 7c2) indicates that VIs were the dominant drivers (67.9%), far exceeding the contribution of TIs (30.8%) and SCPs (1.3%). Figure 7c3 shows the feature effects on Brix. Brix prediction was negatively correlated with GDVI (ranked first), $DDRTI_VAR_{(G,R,B,NIR)}$ (ranked third), and $DDRTI_HOM_{(G,R,B,NIR)}$ (ranked fourth). Conversely, BNDVI and WDRVI (ranked second and fifth) showed positive correlations, where higher values corresponded to higher Brix predictions. Soil pH also showed a positive correlation with Brix prediction.

3.6. Prediction mapping based on optimal models

The respective models were applied to predict the phenotypic traits of sugarcane. Figure 8 illustrates the spatial distribution of sugarcane yield, millable stalk population, and Brix across the study area for the plant cane (year 2024) and ratoon cane (year 2025) seasons.

There was a marked decline in sugarcane yield from the plant cane to first ratoon cane. In 2024 (plant cane), the majority of the field exhibited high yields, ranging between 63 and 105 t/ha (Fig. 8a1). Conversely, in 2025 (first ratoon cane), yields decreased significantly, with most areas shifting to a lower range of 21–63 t/ha (Fig. 8a2). A similar downward trend was observed for the number of millable stalks. The plant cane showed high stalk density, predominantly exceeding 45000 stalks/ha (Fig. 8b1). However, in 2025, the density dropped visibly with a large portion of the field recording values below 45000 stalks/ha (Fig. 8b2). Finally, Brix levels slightly declined from plant cane to first ratoon cane. As illustrated in Fig. 8c1, the 2024 map displays a mix of moderate to high Brix levels (mostly 20–23 °Bx). In contrast, the 2025 map (Fig. 8c2) is dominated by lower values (mostly 19–21 °Bx), suggesting a reduced sucrose accumulation potential in the ratoon year.

4. Discussion

4.1. Comparison of sugarcane traits estimation models

Based on the plot level seasonal variation analysis in 2024, NDV of plant cane was generally lower than that of the ratoon during the seedling and tillering stages, which may contribute to the advantage of ratoon cane in canopy establishment. Singla et al. (2018) indicated that the temporal values of NDVI data of Landsat 8 during the first week of April and last week of August to the end of September can be used to discriminate of ratoon sugarcane in local area. Ratoon sugarcane has a relatively high seedling density, especially in the late emergence stage (Xu et al., 2021). The canopies overlap completely, which makes it difficult to clearly distinguish the stem-leaf boundaries of individual plants (Zhu et al., 2026). From late tillering to the elongation stage, $DDRTI_MEA_{(B,R,G,NIR)}$ was markedly higher in ratoon cane than in plant cane, and this difference was more pronounced in the mid-to-late season. This is likely associated with the greater variation of first ratoon to accumulated below-ground pests and diseases, drought in seedling, as well as missing stools, and uneven emergence establishment (Xu et al., 2021; Dlamini et al., 2024). As the sugarcane enters the rapid growth phase, within-plot differences in canopy

closure, shadowing, and exposed soil patches tend to be amplified, which resulted in increased spatial heterogeneity of the canopy in high spatial resolution imagery (Li et al., 2025). $DDRTI_MEA_{(B,R,G,NIR)}$ is constructed from multi-band GLCM mean texture features and may be more sensitive to within plot spatial structural differences, which makes it better suited to capturing texture changes driven by heterogeneous growth in ratoon cane (Zhu et al., 2026). Based on 20 plants randomly selected from each plot at early maturation stage, the plant height, stem diameter, and single stem weight were compared between two years. Plant height did not differ significantly with mean values of 197.25cm in 2024 and 199.42 cm in 2025. Mean stem diameter increased from 2.66cm in 2024 to 2.80cm in 2025. Single stem weight differed significantly with the mean value of 1199 g in 2024 and 1334 g in 2025 (Table S11).

Four algorithms (SVR, RR, GPR, and CatBoost) were evaluated in this study. Across all input features, SVR achieved the highest accuracy. This aligns with our previous findings for rice (Yuan et al., 2023) and other crops (Shi et al., 2022). Indeed, SVR consistently demonstrates superior stability and fitting capabilities in agricultural applications. This is primarily attributed to the principle of structural risk minimization, which ensures that SVR maintains strong generalization ability even with high-dimensional and small-sample datasets. While GPR is also an established approach for such problems, and showed similar performance to RR, both were outperformed by SVR in this study. In contrast, CatBoost exhibited instability and yielded poorer results. Unlike SVR, CatBoost fits residuals via iterative boosting. In small sample scenarios, this approach can be susceptible to noise and prone to overfitting (Friedman, 2002). Consequently, lacking the robust regularization of SVR, CatBoost demonstrated lower stability, highlighting the importance of selecting models tailored to the specific properties of the data.

During the early growth stage, the R^2_{CV}s of sugarcane traits prediction model using VIs alone were relatively low. The R^2_{CV}s for yield varied from 0.14 to 0.29. Those for millable stalk population ranged from 0.27 to 0.33, and the Brix prediction using the SVR model reached only 0.21. Akbarian et al. (2024) reported the R^2 of 0.53 in training of SVR yield model in early growth stage of sugarcane and R^2 of 0.57 in middle stage based on UAV multispectral data. Vasconcelos et al. (2025) achieved a R^2 of 0.89 in the local area sugarcane yield prediction based on high spatial resolution data (PlantScope), genetic information and metrological data via the heteroskedastic gamma regression. And the mid-term yield prediction model based on VIs outperformed the early-stage model, which is consistent with the findings of Akbarian et al. (2022). When VIs were used alone, yield prediction accuracy showed no significant improvement at the elongation and maturity stages, which may indirectly reflect the saturation effect of vegetation indices under dense canopy conditions (Yue et al., 2019; Sofonia et al., 2019). In the middle and late growth periods, sugarcane canopy coverage approaches saturation, and conventional vegetation indices exhibit reduced sensitivity to biomass, making it difficult to detect subtle canopy variations (Shendryk et al., 2020). Sofonia et al. (2019) reported that the VI based model had the R^2 of 0.322 compared to the R^2 of 0.707 of LiDAR based model in sugarcane yield estimation. This results also indicates that relying solely on VIs is insufficient to characterize the complex process of yield formation. The combination of LiDAR and VIs could improve the yield model performance (Sofonia et al.,

2019; Shendryk et al., 2020). Integrating auxiliary data, such as meteorological variables, genetic data, phenology info, soil texture info, with vegetation indices (VIs) also had the potential to enhance the accuracy of sugarcane yield estimation (Amorim et al., 2022; Aleman-Montes et al., 2023).

For millable stalk population, the model's predictive accuracy was highest at the maturation stage, achieving an R^2_{CV} value of 0.57. This finding aligns closely with the R^2_{CV} of approximately 0.60 reported by Khuimphukhieo et al. (2025). Sofinia et al. (2019) reported the R^2 of 0.502 for millable stalk population by UAV multispectral data-based model. With respect to Brix, the associations with VIs were generally weak or statistically non-significant. Although Some studies reported the detectable relationship between VIs and Brix (Chea et al. 2020; Aleman-Montes et al., 2023), it varied considerably across sugarcane cultivars. In this study, VIs alone model failed in Brix estimation, while in the maturation stage, the including of soil properties improved explanation of Brix model.

Previous studies have demonstrated that texture features of high spatial resolution image and UAV data could partly capture differences in crop canopy characteristics (Zhang et al., 2021; Liu et al., 2023a,b; Zhu et al., 2026). The optimum texture window sizes and types of texture parameters depended on data source, crop types, growth stage and agronomic indicators (Liu et al., 2023a,b). Furthermore, TIs were more useful than individual TFs. Yu et al. (2024) reported that the correlation between TIs and LAI in apple orchards was stronger than that of TFs. TIs could reflect more detailed features and richer information (Guo et al., 2023) and enhance the accuracy of growth simulation (Yuan et al., 2023). In this study, good predictive accuracy for sugarcane traits were obtained by the fusion of VIs and TIs. The maximum R^2_{CV} of the yield model achieved to 0.65, and those of the millable stalk population and Brix reached 0.63 and 0.40, respectively. These results demonstrate that the integration of VIs and TIs derived from Jilin-1 satellite images outperforms the use of VIs alone for predicting sugarcane traits.

At the field level, crop growth is closely related to factors such as cultivar, soil fertility, and fertilization management (Li et al., 2026). In this study, by including these variables, the R^2_{CV} for yield increased to 0.70, and the tillering stage remained the optimal prediction period. For Brix, the R^2_{CV} improved to 0.44, and the optimal prediction stage shifted to the maturity stage. Sugar content is strongly influenced by cultivar, late season crop status, temperature and management related information. In contrast, for millable stalk population, the R^2_{CV} slightly decreased with the inputs of SCPs and FI. The additional variables may introduce redundant or less informative signals for millable stalk population (Yu et al., 2025).

4.2. SHAP analysis of multi-source feature contributions

Khosravirad et al. (2019) and Rahman & Robson (2020) successfully employed GNDVI to predict sugarcane yield, achieving favorable accuracy. Akbarian et al. (2022) obtained similar results using CI. In this study, SHAP analysis revealed that GNDVI was the most influential feature, with CI ranking second in contribution for yield estimation. The higher GNDVI values tended to increase the model prediction. However, higher CI and RVI values tended to reduce the yield prediction, which may present nonlinear

response patterns with yield. TIs were less dominant than VIs, which may provide complementary information on plot level spatial structural variation. Zhu et al. (2026) indicated that the combination of VIs and texture features improved the discrimination of sugarcane ratoon. Compared with NDTI (Zheng et al., 2019; Yu et al., 2024) and EVTI (Yuan et al., 2023), DDRTIs showed a greater contribution to yield estimation. DDRTI is derived by four band GLCM texture features, which may amplify spatial variation in canopy structure across multiple spectral responses. In this study, first ratoon cane exhibited larger within plot heterogeneous (with larger CV values of stem parameters and single stem weight compared to those in 2024). In terms of fertilization, base fertilizer calcium magnesium fertilizer (BF_CMF) and base fertilizer urea (BF_Urea) were identified as important features. However, topdressing was not selected as a final predictive feature. This result indicates that early nutrient supply is critical for sugarcane yield formation. Base fertilizer served as the primary nutrient source for young plants, supporting root growth, tillering, and early canopy development. Similar results were also reported in corn growth (Li et al., 2022; Li et al., 2026).

For the prediction of millable stalk population, EVTI and NDTI contributed more than DDRTI. Although DDRTI was proved effective for yield estimation, it was less suitable for predicting the number of millable stalk population. Increased index complexity did not guarantee improved performance, and it was important to select texture indices for specific traits (Guo et al., 2023). Conversely, for Brix prediction, the contribution of DDRTI exceeded that of EVTI and NDTI. Additionally, soil pH showed a positive effect on Brix prediction. Dao et al. (2022) suggested that pH should be optimized to maximize sugarcane sugar accumulation. Soil pH influences nutrient availability and enzyme activity, such as protein kinase activity, which regulates sucrose metabolism (Huber et al., 1996). For the prediction of sugarcane yield, millable stalk population, and Brix, VIs were among the most important features. This indicates that canopy spectral reflectance characteristics can characterize photosynthetic physiological status and population structure differences during sugarcane growth and development, serving as key remote sensing indicators of crop vigor and yield formation potential. The inclusion of TIs provided complementary plot level spatial information related to within canopy heterogeneity and structural organization, which could not be fully captured by VIs alone. The formulations of NDTI, EVTI, and DDRTI in this study differed from those reported in some studies, where texture indices were derived either by combining multiple texture features from a single band (Zheng et al., 2019) or through a random texture-feature-cross of multiple band combination (Yu et al., 2024). In this study, the texture indices were constructed in a form analogous to conventional VIs. This design may enhance interpretability and reduce inconsistencies.

4.3. Limitations and future prospects

This study investigated the potential of Jilin-1 multispectral imagery for plot level sugarcane traits estimation. The promising results were reported for yield and quality related traits (millable stalk population and Brix). However, several limitations remain due to constraints in data availability, study scope and methodological design.

Firstly, this study only analyzed data from 2024 and 2025, and included only plant cane and first ratoon cane. The results showed that they did not differ significantly in yield or Brix, but significant differences

were observed in stalk weight, millable stalk population, and stalk diameter (Table S11). Second ratoon cane data was not included in this study. The conclusions remain limited in the applicability across ratoon generations. Secondly, the texture indices used in this study were able to improve the performance in sugarcane traits estimation, which may be able to capture differences in canopy structure at the plot level. Some texture indices appeared to be more sensitive to overall canopy conditions (e.g., NDRTI). Others seemed more responsive to structural variation at the individual plant level. However, the causes of these differences were not further analyzed.

Thirdly, due to the limited sample size, this study mainly relied on traditional machine learning methods. And SVR was identified as the best model. Such methods are practical under small sample size. More data driven algorithms (Akbarian et al., 2024), and hybrid framework combining crop growth models with ML (de Oliveira et al., 2020; Xu et al., 2025), were not explored. In addition, although high spatial resolution multispectral imagery (Jilin-1) was used, the lack of accurate regional-scale soil property information limited the analysis and map to the experimental plot scale. Large-area mapping and regional validation were not conducted, which constrains the application of the findings. Moreover, under high spatial resolution conditions, canopy reflectance may be affected by factors such as soil background, adjacency effects, planting density, and viewing geometry (Fan et al., 2025; Li et al., 2026?). In this study, plot-level mean values of reflectance were used for analysis, but the influence of within-plot soil background variation and related factors on reflectance was not assessed.

Finally, although the models achieved acceptable accuracy in yield and millable stalk population estimation, the performance for Brix estimation remained relatively low. This sugar-related quality trait is more difficult to monitor using the current remote sensing and modeling framework. Future research should incorporate multi-source data, such as temperature, precipitation, soil properties, varietal characteristics, management practices, thermal, and LiDAR to improve the accuracy and robustness of sugarcane trait estimation (Xie &Yang, 2020).

5. Conclusion

This study integrated VIs, TIs, SCPs, and FI to predict sugarcane phenotypic traits. SVR had the best performance in the estimation of yield, millable stalk population, and Brix. The inclusion of TIs with different window sizes in different growth stages improved the model accuracy. For sugarcane yield, combining VIs, TIs, SCPs, and FI allowed the model to reach the best prediction accuracy during the tillering period across plant cane and first ratoon cane (R^2_{CV} = 0.70, $RMSE_{CV}$ = 13.84 t/ha, RPD_{CV} = 1.83). The yield model had the potential to predict sugarcane yield in early growth stage. For millable stalk population, the inclusion of VIs and TIs achieved the highest accuracy at the maturation stage (R^2_{CV} = 0.63, $RMSE_{CV}$ = 6602 stalks/ha, RPD_{CV} = 1.66). Regarding Brix, the inclusion of SCPs and FI together with VIs and TIs achieved the highest prediction accuracy at the maturation stage (R^2_{CV} = 0.44, $RMSE_{CV}$ = 0.53 °Bx, RPD_{CV} = 1.33). These results demonstrate the contribution of VIs and texture indices based on optimal texture windows, as well as multi-source data fusion, in sugarcane traits estimation. SCPs

suitable for sugarcane growth and appropriate fertilization ratios contributed to high sugarcane yield and Brix. This study validated the applicability of Jilin-1 high-resolution satellite imagery for field-scale sugarcane traits estimation and provided an important insight for remote sensing phenotyping of other crops.

Declarations

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Declaration of Competing Interest

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

Author Contribution

Yun Zhang: Conceptualization, Data curation, Investigation, Methodology, Writing – original draft. Guangtao Xu: Data curation, Investigation, Methodology, Writing – original draft. Dan Li: Conceptualization, Methodology, Writing – original draft, Writing – review and funding acquisition. Yinglin Lu: Investigation, Data curation, Writing – review and funding acquisition. Hao Jiang: Investigation, Data curation. Shasha Luo: Investigation, Data curation. Mingfu Wen: Investigation, Data curation.Jing Zhang: Investigation, Data curation. Xingda Chen: Investigation, Data curation. Qiong Zheng: Methodology, Data curation. Chongyang Wang: Investigation, Data curation. Zihan Liu: Investigation, Data curation. Mengjun Ku: Investigation, Data curation. Junliang Chen: Investigation, Data curation.

Acknowledgements

This work was supported by Guangzhou Science and Technology Project (No. 2024B03J1321), GDAS'Project of Science and Technology Development (2022GDASZH-2022010202) and the China Association for Science and Technology (CAST) Special Talent Cultivation Project for PhD Students (2025).

Data availability

The data that support the findings of this study are available in the main text and the supplementary materials. Part of the code and test data for model training are available at https://github.com/guangtaoxu08-dev/SPT_JL. Additional data are available from the corresponding author upon reasonable request.

References

1. Akbarian, S., Xu, C., Wang, W., Ginns, S., & Lim, S. (2022). Sugarcane yields prediction at the row level using a novel cross-validation approach to multi-year multispectral images. *Computers and Electronics in Agriculture*, *198*, 107024. https://doi.org/10.1016/j.compag.2022.107024
2. Akbarian, S., Jamnani, M. R., Xu, C., Wang, W., & Lim, S. (2024). Plot level sugarcane yield estimation by machine learning on multispectral images: a case study of Bundaberg, Australia. *Information Processing in Agriculture*, *11*(4), 476–487. https://doi.org/10.xxxx/j.inpa.2024.xxxx
3. Alemán-Montes, B., Zabala, A., Henríquez, C., & Serra, P. (2023). Modelling two sugarcane agro-industrial yields using Sentinel/Landsat time-series data and their spatial validation at different scales in Costa Rica. *Remote Sensing*, *15*, 5476. https://doi.org/10.3390/rs15235476
4. Amorim, M. T. A., Silvero, N. E. Q., Bellinaso, H., Gómez, A. M. R., Greschuk, L. T., Campos, L. R., & Demattê, J. A. M. (2022). Impact of soil types on sugarcane development monitored over time by remote sensing. *Precision Agriculture*, *23*, 1532–1552. https://doi.org/10.1007/s11119-022-09893-7
5. Brovelli, A. M., Crespi, M., Fratarcangeli, F., Giannone, F., & Realini, E. (2008). Accuracy assessment of high resolution satellite imagery orientation by leave-one-out method. *ISPRS Journal of Photogrammetry and Remote Sensing*, *63*(4), 427–440. https://doi.org/10.1016/j.isprsjprs.2008.01.006
6. Canton, H. (2021). Food and Agriculture Organization of the United Nations—FAO. The Europa Directory of International Organizations, 297–305. ISBN: 978-1-00-317990-0.
7. Chea, C., Saengprachatanarug, K., Posom, J., Wongphati, M., & Taira, E. (2020). Sugar yield parameters and fiber prediction in sugarcane fields using a multispectral camera mounted on a small unmanned aerial system (UAS). *Sugar Tech*, *22*, 605–621. https://doi.org/10.1007/s12355-020-00802-5
8. Chen, Z., Zhang, Y., Zang, P., Liu, Z., & Zhao, W. (2023). Mapping the grassland quality of the flooded plain based on Jilin-01 images with a red-edge spectral band. *Frontiers in Earth Science*, *10*, 1024680. 10.3389/feart.2022.1024680
9. Dao, J., Mao, J., Li, R., Yang, S., Deng, J., Fan, X., & Zhang, Y. (2022). Effects of key climate and soil factors on sugarcane production in different ecological regions in China. *Soil and Fertilizer Sciences in China*, (12), 102–113. 10.11838/sfsc.1673-6257.21593
10. de Silva, F., Chaves, N. R., Luciano, M. E. D., Sanches, A. C. S., de Almeida, I. D., C.M. and, & Adami, M. (2024). Sugarcane yield estimation using satellite remote sensing data in empirical or mechanistic

modeling: a systematic review. *Remote Sensing*, *16*, 863. https://doi.org/10.3390/rs16050863

11. de Oliveira, M. P. G., & Rodrigues, L. H. A. (2020). How good are the models available for estimating sugar content in sugarcane? *European Journal Of Agronomy*, *113*, 125992. https://doi.org/10.1016/j.eja.2019.125992
12. Dlamini, N. E., Franke, A. C., & Zhou, M. (2024). Indices for measuring ratooning ability of sugarcane varieties. *Crop Science*, *64*, 667–677. https://doi.org/10.1002/csc2.21191
13. Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A. J., & Vapnik, V. (1997). Support vector regression machines. *Advances in Neural Information Processing Systems 9* (pp. 155–161). MIT Press.
14. Du, Y., Zhu, R., Gong, J., Wang, D., & Zhong, X. (2023). Retrieval of crop leaf area index based on Jilin-1GP image. *Remote Sensing Technology and Application*, *38*(4), 816–826. https://doi.org/10.11873/j.issn.1004-0323.2023.4.0816
15. Durbha, S. S., King, R. L., & Younan, N. H. (2007). Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer. *Remote Sensing of Environment*, *107*(1–2), 348–361. https://doi.org/10.1016/j.rse.2006.07.032
16. Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, *7*(1), 1–26. https://doi.org/10.1214/aos/1176344552
17. Fan, W. J., Peng, N. J., Cao, B., Mu, X. H., Yang, S. Q., He, Q. C., Zhai, D. C., Ren, H. Z., Cui, Y. K., & Yan, G. J. (2025). Progress in high spatial resolution vegetation quantitative remote sensing. *National Remote Sensing Bulletin*, *29*(6), 1365–1387. https://doi.org/10.11834/jrs.20255036
18. FAO (2025). FAOSTAT: crops and livestock products. Available at: http://www.fao.org/faostat/en/#data/QCL (accessed 18 December 2025).
19. Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics and Data Analysis*, *38*, 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2
20. Fu, B., Zuo, P., Liu, M., Lan, G., He, H., Lao, Z., Zhang, Y., Fan, D., & Gao, E. (2022). Classifying vegetation communities in karst wetland by synergistic use of image fusion and object-based machine learning algorithm with Jilin-1 and UAV multispectral images. *Ecological Indicators*, *140*, 108989. https://doi.org/10.1016/j.ecolind.2022.108989
21. Gao, S., Zhong, R., Yan, K., Ma, X., Chen, X., Pu, J., Gao, S., Qi, J., Yin, G., & Myneni, R. B. (2023). Evaluating the saturation effect of vegetation indices in forests using 3D radiative transfer simulations and satellite observations. *Remote Sensing of Environment*, *295*, 113665. https://doi.org/10.1016/j.rse.2023.113665
22. Gaston, E., Frías, J. M., Cullen, P. J., O'Donnell, C. P., & Gowen, A. A. (2010). Prediction of polyphenol oxidase activity using visible near-infrared hyperspectral imaging on mushroom (Agaricus bisporus) caps. *Journal of Agricultural and Food Chemistry*, *58*(10), 6226–6233. https://doi.org/10.1021/jf100501q
23. Guo, Y., Xiao, Y., Hao, F., Zhang, X., Chen, J., de Beurs, K., He, Y., & Fu, Y. H. (2023). Comparison of different machine learning algorithms for predicting maize grain yield using UAV-based

hyperspectral images. *International Journal of Applied Earth Observation and Geoinformation*, *124*, 103528. https://doi.org/10.1016/j.jag.2023.103528

24. Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, *46*(1–3), 389–422. https://doi.org/10.1023/A:1012487302797
25. Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics, SMC-3(6), 610–621. https://doi.org/10.1109/TSMC.1973.4309314
26. Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, *12*, 55–67. https://doi.org/10.1080/00401706.1970.10488634
27. Huber, S. C., & Huber, J. L. (1996). Role and regulation of sucrose-phosphate synthase in higher plants. *Annual Review of Plant Physiology and Plant Molecular Biology*, *47*, 431–444. https://doi.org/10.1146/annurev.arplant.47.1.431
28. Jimenez-Sierra, D. A., Benítez-Restrepo, H. D., Vargas-Cardona, H. D., & Chanussot, J. (2020). Graph-based data fusion applied to change detection and biomass estimation in rice crops. *Remote Sensing*, *12*(17), 2683. https://doi.org/10.3390/rs12172683
29. Kambli, M., & Palkar, B. (2025). Remote sensing and machine learning methods to analyse the vegetation of sugarcane crop. *Multimedia Tools and Applications*, *84*, 45297–45319. https://doi.org/10.1007/s11042-025-20950-8
30. Khosravirad, M., Omid, M., Sarmadian, F., & Hosseinpour, S. (2019). Predicting sugarcane yields in Khuzestan using a large time-series of remote sensing imagery region. *Int Arch Photogramm Remote Sens Spat Inf Sci XLII*, *-4/W18*, 645–648.
31. https://doi.org/10.5194/isprs-archives-xlii-4-w18-645-2019
32. Khuimphukhieo, I., Bhandari, M., Enciso, J., & da Silva, J. A. (2025). Estimating sugarcane yield and its components using unoccupied aerial systems (UAS)-based high throughput phenotyping (HTP). Computers and Electronics in Agriculture, 237 (Part B), 110658. https://doi.org/10.1016/j.compag.2025.110658
33. Li, D., Miao, Y., Ransom, C. J., Bean, G. M., Kitchen, N. R., Fernández, F. G., Sawyer, J. E., Camberato, J. J., Carter, P. R., Ferguson, R. B., et al. (2022). Corn nitrogen nutrition index prediction improved by integrating genetic, environmental, and management factors with active canopy sensing using machine learning. *Remote Sensing*, *14*, 394. https://doi.org/10.3390/rs14020394
34. Li, D., Wang, C., Miao, Y., et al. (2026). Developing an innovative stacking ensemble machine learning and multi-source data fusion-based precision nitrogen management strategy for corn. *Precision Agriculture*, *27*, 41. https://doi.org/10.1007/s11119-026-10344-7
35. Li, J., Duan, T. T., Deng, K., Jing, Z. X., Lin, Z., Liang, Y. Q., Deng, C. N., Gao, Y., & Zheng, C. (2019). Responses of sugarcane varieties growth to nitrogen fertilizer application and related use efficiency. *Chinese Journal of Tropical Crops*, *40*(4), 649–656. https://doi.org/10.3969/j.issn.1000-2561.2019.04.001

36. Li, Q., Song, X., Meng, X., Zhang, J., Zhang, M., Chen, L. Y., Li, J., & Wang, B. (2025). Shaping future sugarcane: ideal plant architecture and breeding strategies. *Molecular Plant*, *18*, 725–728. https://doi.org/10.1016/j.molp.2025.04.004
37. Liu, Y., Feng, H., Yue, J., Fan, Y., Bian, M., Ma, Y., Jin, X., Song, X., & Yang, G. (2023a). Estimating potato above-ground biomass by using integrated unmanned aerial system-based optical, structural, and textural canopy measurements. *Computers and Electronics in Agriculture*, *213*, 108229. https://doi.org/10.1016/j.compag.2023.108229
38. Li, Y. R. (2022). Growth and development of sugarcane (Saccharum spp. hybrid) and its relationship with environmental factors. In K. K. Verma, X. P. Song, V. D. Rajput, S. Solomon, Y. R. Li, & G. P. Rao (Eds.), *Agro-industrial Perspectives on Sugarcane Production under Environmental Stress*. Springer. https://doi.org/10.1007/978-981-19-3955-6_1
39. Liu, J., Zhu, Y., Song, L., Su, X., Li, J., Zheng, J., Zhu, X., Ren, L., Wang, W., & Li, X. (2023b). Optimizing window size and directional parameters of GLCM texture features for estimating rice AGB based on UAV multispectral imagery. *Frontiers in Plant Science*, *14*, 1284235. https://doi.org/10.3389/fpls.2023.1284235
40. Luciano, A. C. S., Picoli, M. C. A., Duft, D. G., Rocha, J. V., Leal, M. R. L. V., & le Maire, G. (2021). Empirical model for forecasting sugarcane yield on a local scale in Brazil using Landsat imagery and random forest algorithm. *Computers and Electronics in Agriculture*, *184*, 106063. https://doi.org/10.1016/j.compag.2021.106063
41. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), pp. 4768–4777. https://doi.org/10.1016/j.compag.2021.106063
42. Meier, E. A., & Thorburn, P. J. (2016). Long term sugarcane crop residue retention offers limited potential to reduce nitrogen fertilizer rates in Australian wet tropical environments. *Frontiers in Plant Science*, *7*, 1017. https://doi.org/10.3389/fpls.2016.01017
43. Ozkan, U. Y., & Demirel, T. (2021). The influence of window size on remote sensing-based prediction of forest structural variables. *Ecological Processes*, *10*, 60. https://doi.org/10.1186/s13717-021-00330-4
44. Phang, S. K., Chiang, T. H. A., Happonen, A., & Chang, M. M. L. (2023). From satellite to UAV-based remote sensing: a review on precision agriculture. *Ieee Access : Practical Innovations, Open Solutions*, *11*, 127057–127076. https://doi.org/10.1109/ACCESS.2023.3330886
45. Paudel, D., Boogaard, H., de Wit, A., Janssen, S., Osinga, S., Pylianidis, C., & Athanasiadis, I. N. (2021). Machine learning for large-scale crop yield forecasting. *Agricultural Systems*, *187*, 103016. https://doi.org/10.1016/j.agsy.2020.103016
46. Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. The MIT Press. 978-0-262-18253-9.
47. Rahman, M. M., & Robson, A. (2020). Integrating Landsat-8 and Sentinel-2 time series data for yield prediction of sugarcane crops at the block level. *Remote Sensing*, *12*, 1313.

https://doi.org/10.3390/rs12081313

48. Rosa, J. M., Fattori Junior, I. M., Melo, M. L. A., & Marin, F. R. (2026). Remote sensing-driven ensemble smoother assimilation of LAI for regional sugarcane yield estimation. *Remote Sensing Applications: Society and Environment*, *41*, 101952. https://doi.org/10.1016/j.rsase.2026.101952
49. Shendryk, Y., Sofonia, J., Garrard, R., Rist, Y., Skocaj, D., & Thorburn, P. (2020). Fine-scale prediction of biomass and leaf nitrogen content in sugarcane using UAV LiDAR and multispectral imaging. *International Journal of Applied Earth Observation and Geoinformation*, *92*, 102177. https://doi.org/10.1016/j.jag.2020.102177
50. Singla, S. K., Garg, R. D., & Dubey, O. P. (2018). Sugarcane ratoon discrimination using LANDSAT NDVI temporal data. *Spatial Information Research*, *26*, 415–425. https://doi.org/10.1007/s41324-018-0184-0
51. Shi, Y., Gao, Y., Wang, Y., Luo, D., Chen, S., Ding, Z., & Fan, K. (2022). Using unmanned aerial vehicle-based multispectral image data to monitor the growth of intercropping crops in tea plantation. *Frontiers in Plant Science*, *13*, 820585. https://doi.org/10.3389/fpls.2022.820585
52. Sofonia, J., Shendryk, Y., Phinn, S., Roelfsema, C., Kendoul, F., & Skocaj, D. (2019). Monitoring sugarcane growth response to varying nitrogen application rates: a comparison of UAV SLAM LiDAR and photogrammetry. *International Journal of Applied Earth Observation and Geoinformation*, *82*, 101878. https://doi.org/10.1016/j.jag.2019.05.011
53. Som-ard, J., Atzberger, C., Izquierdo-Verdiguier, E., Vuolo, F., & Immitzer, M. (2021). Remote sensing applications in sugarcane cultivation: a review. *Remote Sensing*, *13*, 4040. https://doi.org/10.3390/rs13204040
54. Vasconcelos, J. C. S., Arantes, C. S., Speranza, E. A., Antunes, J. F. G., Barbosa, L. A. F., Cançado, G., & M.d.A. (2025). Predicting sugarcane yield through temporal analysis of satellite imagery during the growth phase. *Agronomy*, *15*, 793. https://doi.org/10.3390/agronomy15040793
55. Wang, H., Dai, Y., Yao, Q., Ma, L., Zhang, Z., & Lv, X. (2025). Multi-task learning model driven by climate and remote sensing data collaboration for mid-season cotton yield prediction. *Field Crops Research*, *333*, 110070. https://doi.org/10.1016/j.fcr.2025.110070
56. Wu, N., Zhou, Y., Yin, S., Gong, H., & Zhang, C. (2025). Revealing the nonlinear impact of environmental regulation on ecological resilience using the XGBoost-SHAP model: evidence from the Yangtze River Delta region, China. *Journal of Cleaner Production*, *514*, 145700. https://doi.org/10.1016/j.jclepro.2025.145700
57. Xie, C., & Yang, C. (2020). A review on plant high-throughput phenotyping traits using UAV-based sensors. *Computers and Electronics in Agriculture*, *178*, 105731. https://doi.org/10.1016/j.compag.2020.105731
58. Xu, F., Wang, Z., Lu, G., Zeng, R., & Que, Y. (2021). Sugarcane ratooning ability: research status, shortcomings, and prospects. *Biology*, *10*(10), 1052. https://doi.org/10.3390/biology10101052
59. Xu, J., Du, X., Dong, T., Li, Q., Zhang, Y., Wang, H., Liu, M., Zhu, J., & Yang, J. (2025). Estimation of sugarcane biomass from Sentinel-2 leaf area index using an improved SAFY model (SAFY-Sugar).

International Journal of Applied Earth Observation and Geoinformation, 140, 104570. https://doi.org/xxxxx.

60. Xuan, F., Su, W., Chen, Z., Huang, X., Zhai, W., Li, X., Zeng, Y., Li, Z., Li, J., & Huang, J. (2025). Performance of stacking machine learning and volume model for improving corn above ground biomass prediction. *Plant Phenomics*, *7*(3), 100068.
61. Yang, Y., Wu, Z., Wang, D., Wang, C., Yang, X., Wang, Y., Wang, J., Huang, Q., Hou, L., Wang, Z., et al. (2025). Sugarcane phenology retrieval in heterogeneous agricultural landscapes based on spatiotemporal fusion remote sensing data. *Agriculture*, *15*, 1578. https://doi.org/10.3390/agriculture15151578
62. Yu, J., Zhang, Y., Song, Z., Jiang, D., Guo, Y., Liu, Y., & Chang, Q. (2024). Estimating leaf area index in apple orchard by UAV multispectral images with spectral and texture information. *Remote Sensing*, *16*(17). 3237.https://doi.org/10.3390/rs16173237
63. Yu, Z., Chen, Q., Shan, Q., Lv, Y., Chen, Z., Zhou, Y., Zhang, P., Jiang, H., & Cao, W. (2025). Developing soybean yield prediction model based on multicolor space and texture features of UAV images. *Computers and Electronics in Agriculture*, *235*, 110345. https://doi.org/10.1016/j.compag.2025.110345
64. Yuan, W., Meng, Y., Li, Y., Ji, Z., Kong, Q., Gao, R., & Su, Z. (2023). Research on rice leaf area index estimation based on fusion of texture and spectral information. *Computers and Electronics in Agriculture*, *211*, 108016. https://doi.org/10.1016/j.compag.2023.108016
65. Yue, J., Yang, G., Tian, Q., Feng, H., Xu, K., & Zhou, C. (2019). Estimate of winter-wheat above-ground biomass based on UAV ultrahigh-ground-resolution image textures and vegetation indices. *ISPRS Journal of Photogrammetry and Remote Sensing*, *150*, 226–244. https://doi.org/10.1016/j.isprsjprs.2019.02.022
66. Zhang, J., Qiu, X., Wu, Y., Zhu, Y., Cao, Q., Liu, X., & Cao, W. (2021). Combining texture, color, and vegetation indices from fixed-wing UAS imagery to estimate wheat growth parameters using multivariate regression methods. *Computers and Electronics in Agriculture*, *185*, 106138. https://doi.org/10.1016/j.compag.2021.106138
67. Zhang, Y., Zhao, Z., & Zheng, J. (2020). CatBoost: a new approach for estimating daily reference crop evapotranspiration in arid and semi-arid regions of Northern China. *Journal of Hydrology*, *588*, 125087. https://doi.org/10.1016/j.jhydrol.2020.125087
68. Zhao, J., Li, J., Liu, Q., & Yang, L. (2012). A preliminary study on mechanism of LAI inversion saturation. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXIX-B1, 77–81. https://doi.org/10.5194/isprsarchives-XXXIX-B1-77-2012
69. Zheng, H., Cheng, T., Zhou, M., Li, D., Yao, X., Tian, Y., Cao, W., & Zhu, Y. (2019). Improved estimation of rice aboveground biomass combining textural and spectral analysis of UAV imagery. *Precision Agriculture*, *20*, 611–629. https://doi.org/10.1007/s11119-018-9600-7
70. Zhu, H., Dong, Z., Wei, L., Qin, S., Qin, X., & He, Y. (2026). From canopy segmentation to accurate prediction: an UAV-based multi-feature fusion framework for plot-scale ratoon sugarcane seedling

counting. *International Journal of Applied Earth Observation and Geoinformation*, *147*, 105183. https://doi.org/10.1016/j.jag.2026.105183

Figures

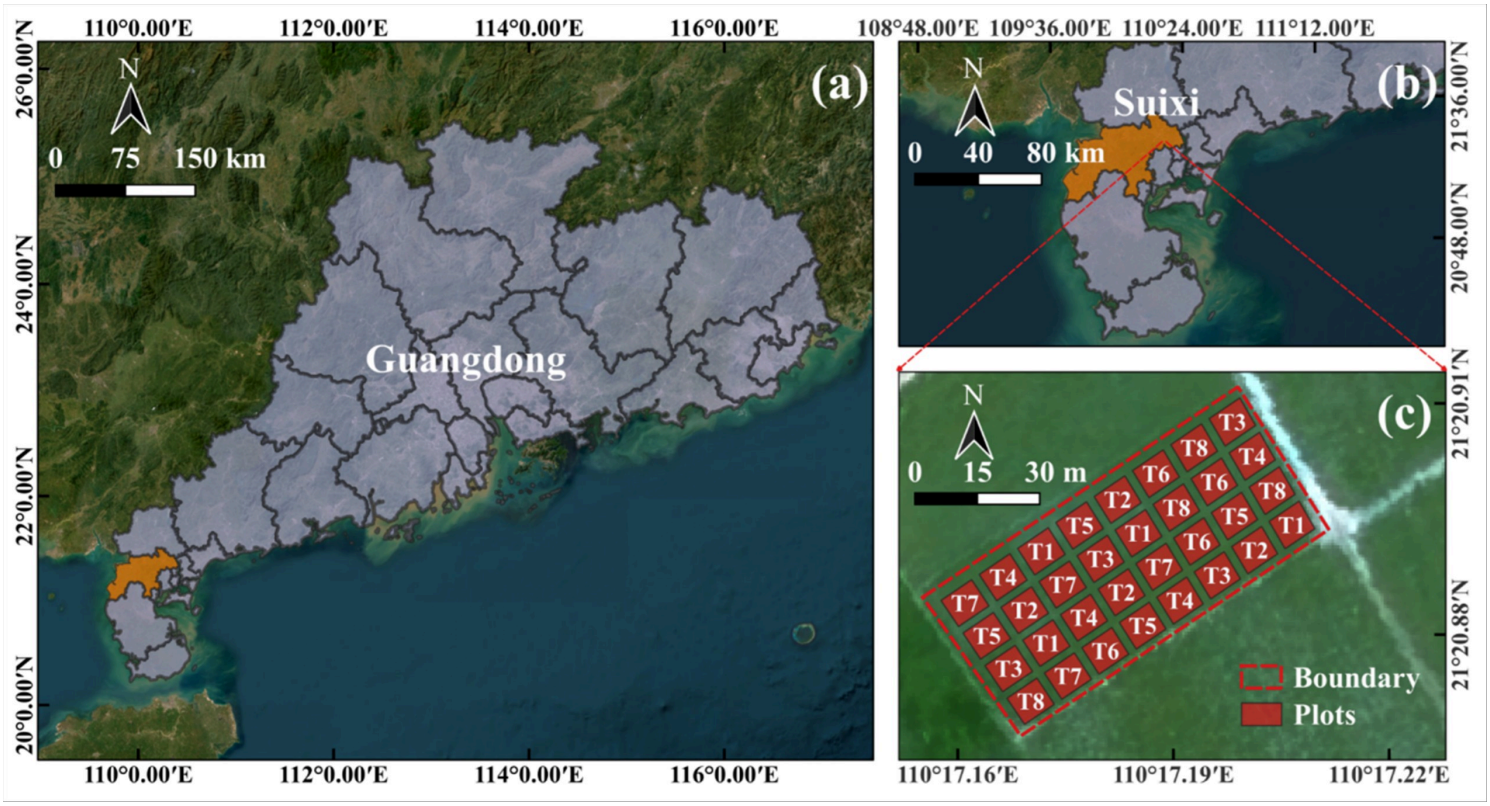


Figure 1

Geographical location of the study area. (a) Inset map showing the location of Guangdong Province; (b) Location of Suixi County; (c) Specific distribution of the experimental plots.

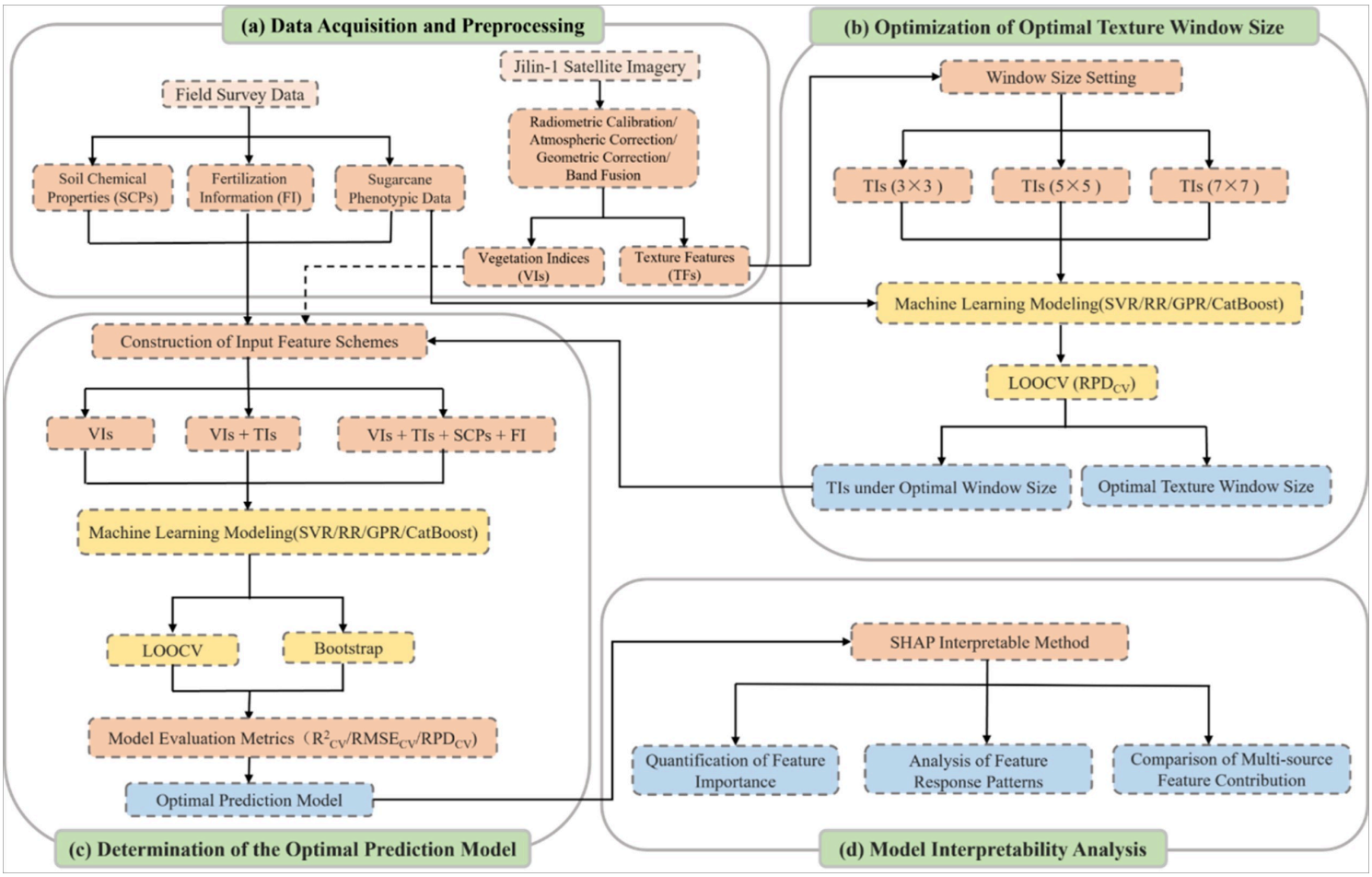


Figure 2

Workflow of this study.

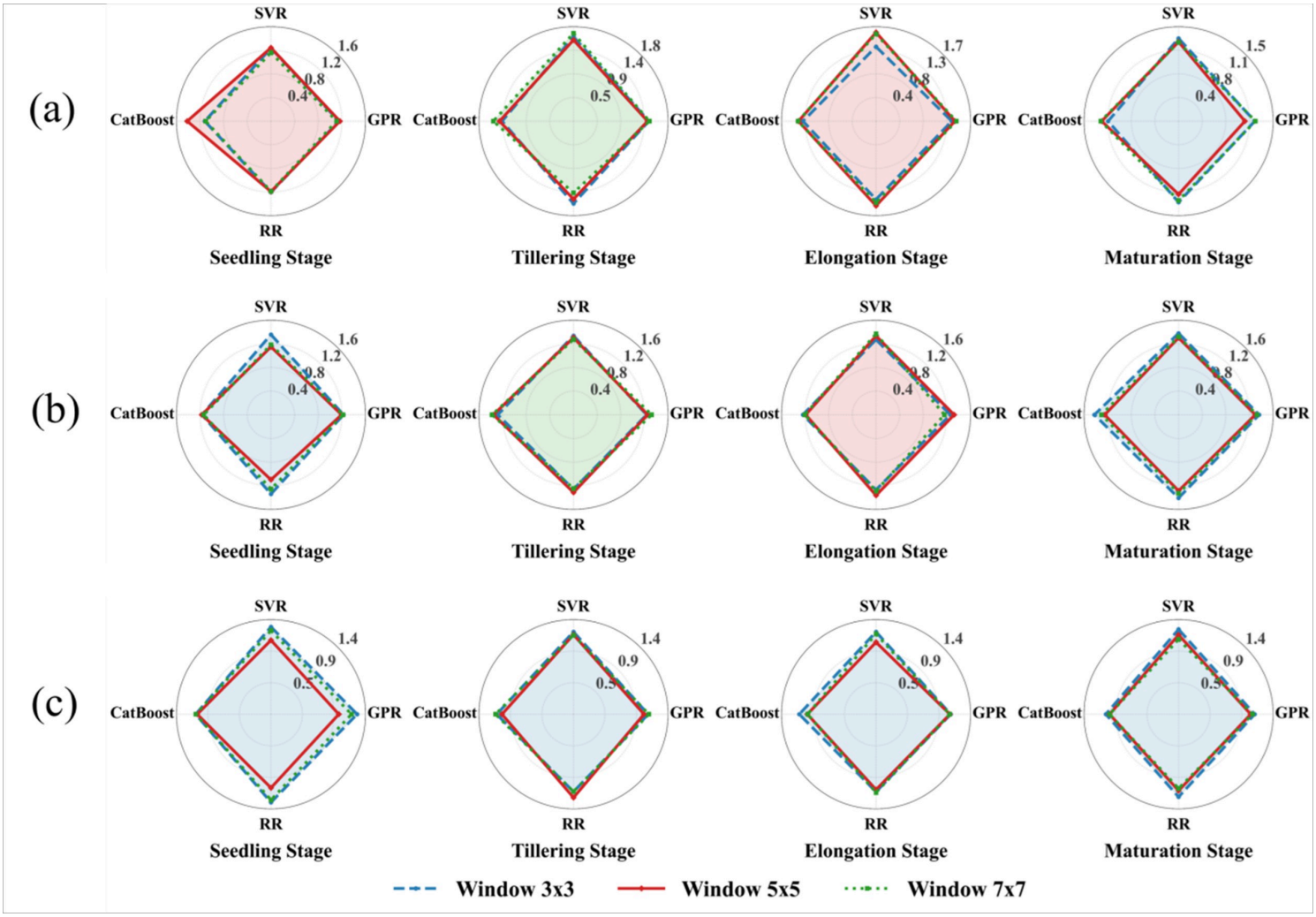


Figure 3

Model performance (RPD) under different texture window sizes across growth stages. (a) Yield; (b) Millable Stalk Population; (c) Brix.

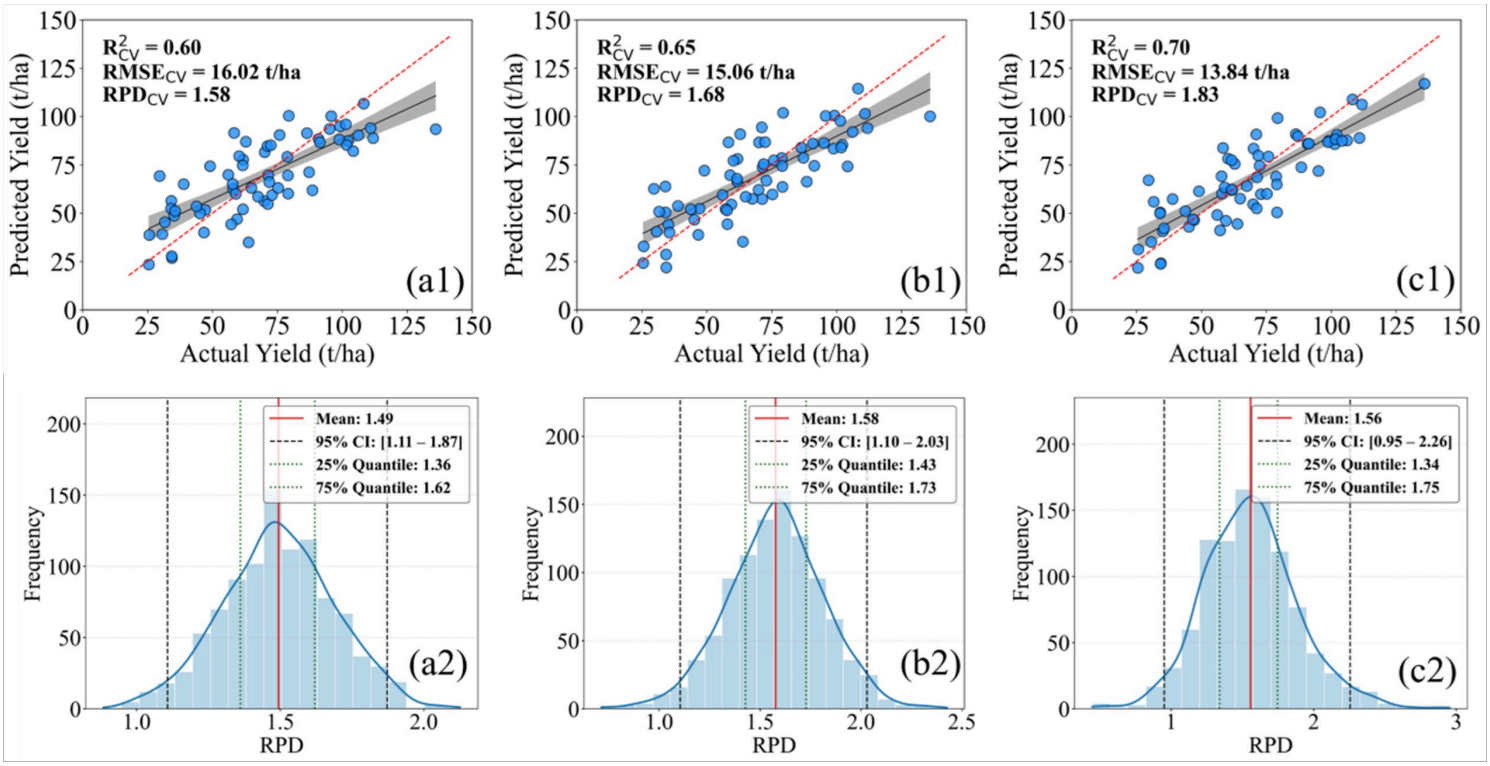


Figure 4

Prediction results of yield using the best performing models for three different input feature combinations. Top row (LOOCV results): (a1) SVR with VIs (Tillering stage); (b1) RR with VIs + TIs (Tillering stage); (c1) SVR with VIs + TIs + SCPs + FI (Tillering stage). Bottom row (Bootstrap results): (a2) SVR with VIs (Tillering stage); (b2) RR with VIs + TIs (Tillering stage); (c2) SVR with VIs + TIs + SCPs + FI (Tillering stage).

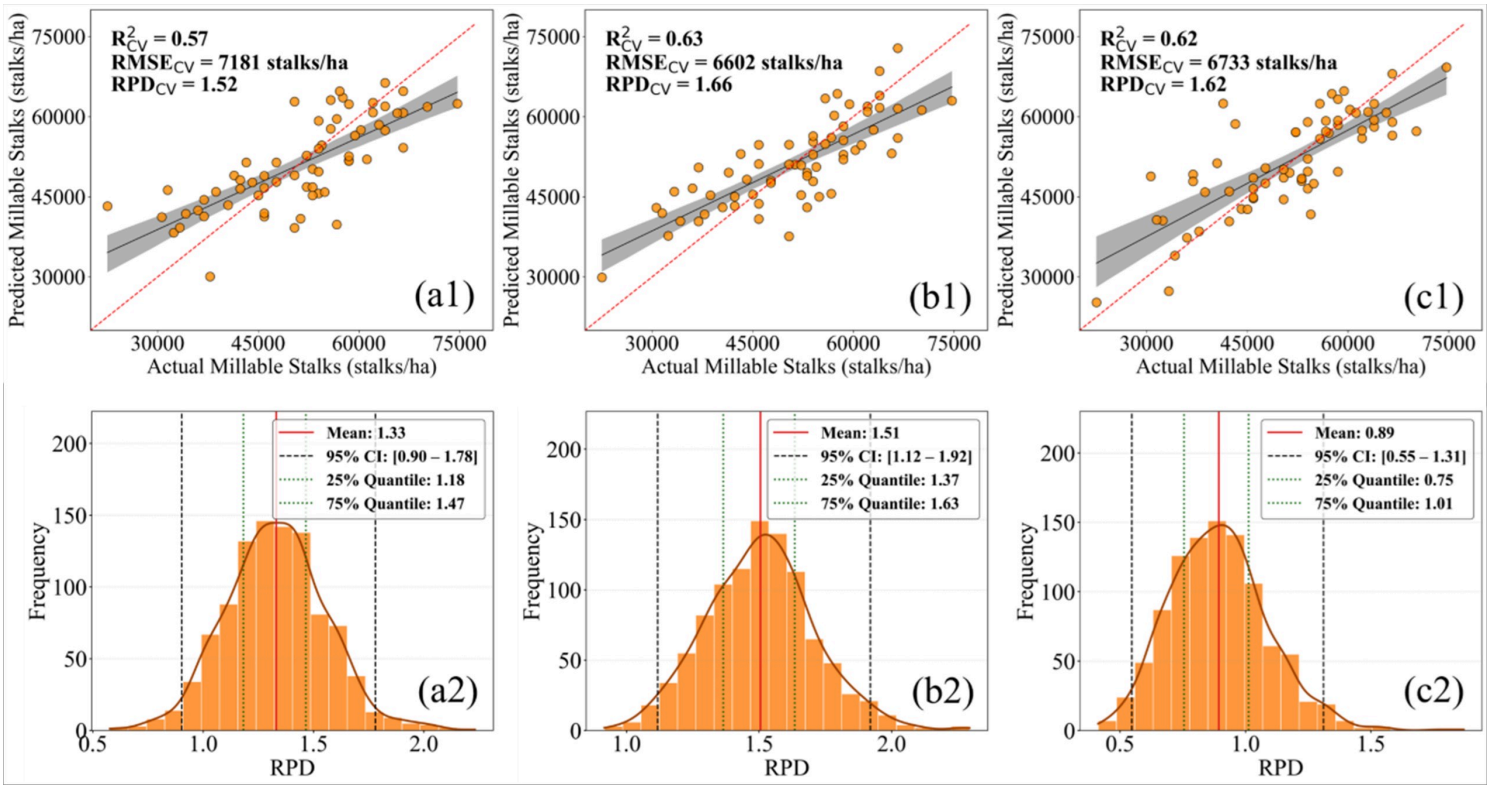


Figure 5

Prediction results of millable stalk population using the best performing models for three different input feature combinations. Top row (LOOCV results): (a1) SVR with VIs (Maturation stage); (b1) SVR with VIs + TIs (Maturation stage); (c1) SVR with VIs + TIs + SCPs + FI (Elongation stage). Bottom row (Bootstrap results): (a2) SVR with VIs (Maturation stage); (b2) SVR with VIs + TIs (Maturation stage); (c2) SVR with VIs + TIs + SCPs + FI (Elongation stage).

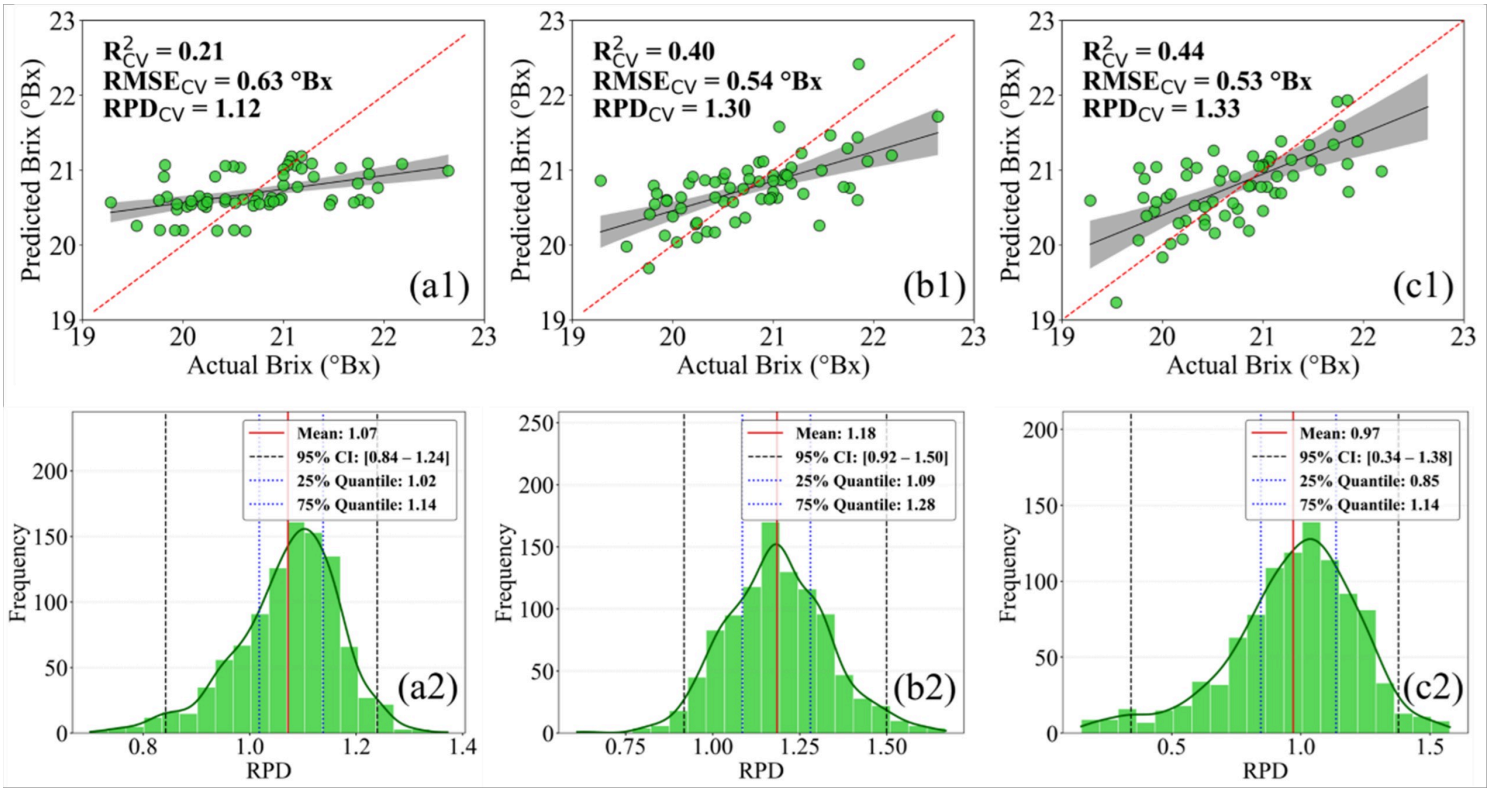


Figure 6

Prediction results of Brix by the SVR model using three different input feature combinations. Top row (LOOCV results): (a1) VIs (Seedling stage); (b1) VIs + TIs (Seedling stage); (c1) VIs + TIs + SCPs + FI (Maturation stage). Bottom row (Bootstrap results): (a2) VIs (Seedling stage); (b2) VIs + TIs (Seedling stage); (c2) VIs + TIs + SCPs + FI (Maturation stage).

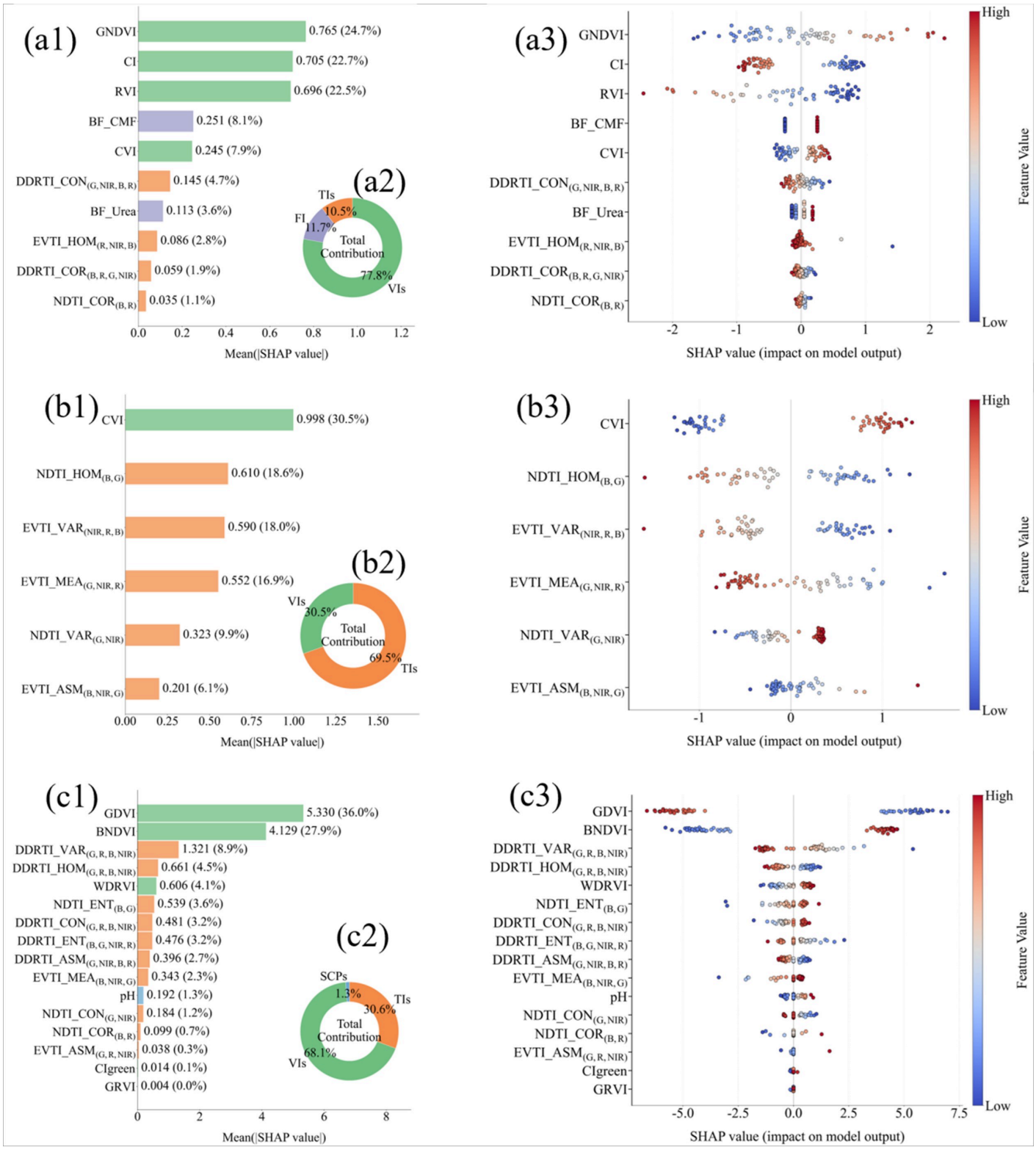


Figure 7

SHAP feature importance plots for the optimal prediction models. (a1)–(a3) Yield (SVR: VIs + TIs + SCPs + FI, Tillering stage); (b1)–(b3) Millable Stalk Population (SVR: VIs + TIs, Maturation stage); (c1)–(c3) Brix (SVR: VIs + TIs + SCPs + FI, Maturation stage).

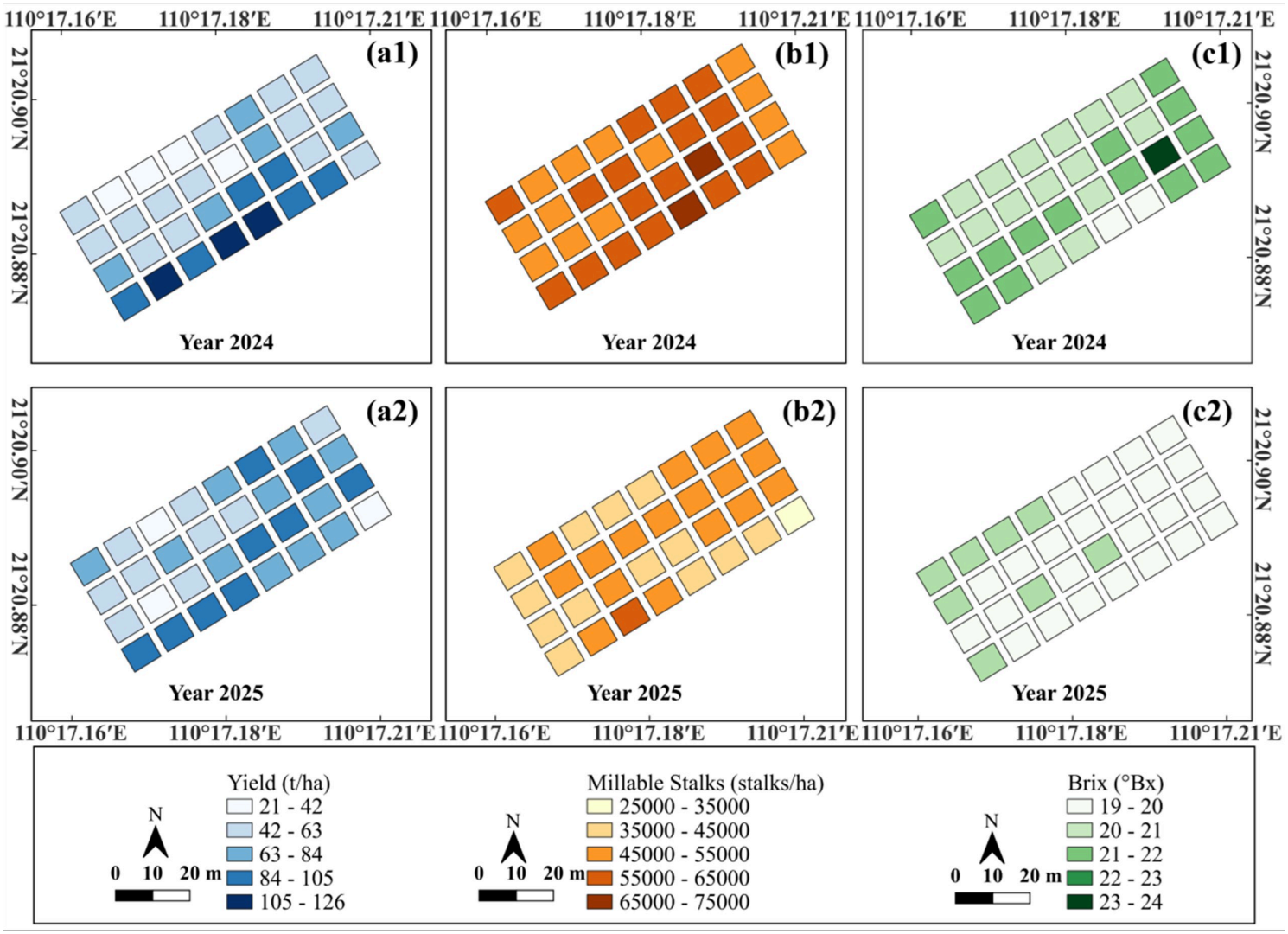


Figure 8

Prediction results of sugarcane phenotypic traits based on the optimal model. (a1) Yield (plant cane)-2024; (a2) Yield (first ratoon cane)-2025; (b1) Millable Stalk Population-2024; (b2) Millable Stalks-2025; (c1) Brix-2024; (c2) Brix-2025.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterial.docx