A Sugarcane Height Estimation Model Based on Multi-Source Satellite Data Fusion Using Machine Learning

Thabed Tholib Baladraf
Department of Agroindustrial Technology, IPB University, Indonesia
thabedtholib@apps.ipb.ac.id

Marimin Marimin
Department of Agroindustrial Technology, IPB University, Indonesia
marimin@apps.ipb.ac.id (corresponding author)

Irman Hermadi
School of Data Science, Mathematics, and Informatics, IPB University, Indonesia
irmanhermadi@apps.ipb.ac.id

Andes Ismayana
Department of Agroindustrial Technology, IPB University, Indonesia
andesismayana@apps.ipb.ac.id

Received: 20 November 2025 | Revised: 31 December 2025, 19 January 2026, and 7 February 2026 | Accepted: 8 February 2026

 DOI: https://doi.org/10.48084/etasr.16345

ABSTRACT

The combination of satellite and Machine Learning (ML) approaches is emerging as a new paradigm in crop height estimation, facilitating informed decision-making. This study aims to identify key factors influencing sugarcane height prediction and compare the performance of single satellite data with data fusion. The proposed height estimation model was designed with the help of satellite data from Sentinel-2 and Landsat 9. A data fusion technique was used to extract vegetation, morphological, and meteorological indicators from three land plots in 2023 and 2024. The collected data were preprocessed using the Interquartile Range (IQR). Forward chaining was utilized to split data into training and testing subsets (40:20, 50:20, 60:20, 70:20, and 80:20). Feature importance analysis was performed to identify features contributing to sugarcane height prediction. The selected features were used as inputs for three ML models: Random Forest (RF), Extreme Gradient Boosting (XGB), and Support Vector Regression (SVR). The results of sugarcane height estimation indicate that the use of multi-source data fusion (vegetation indicators, morphology, and meteorology) with XGB achieved the best and most robust performance, with a coefficient of determination (R^2) of 0.997, a Root Mean Square Error (RMSE) of 6.254 cm, and a Mean Absolute Percentage Error (MAPE) of 3.097%. The findings demonstrate a strong potential of real-time remote sensing-based height estimation for sugarcane, supporting precise monitoring and data-driven decision-making.

Keywords-fusion data; height estimation; machine learning; remote sensing; sugarcane

I. INTRODUCTION

Sugarcane is a subtropical and tropical commodity that contributes to over 70% of the global sugar production [1] and is used in various industries [2]. Crop management is crucial for maintaining the quantity and quality of sugarcane, particularly by monitoring plant height, which is an indicator of growth and yield. Plant growth is closely correlated with physiological, genetic, and microbiome factors [3]. However, monitoring of sugarcane height is not conducted regularly, even though this is important for cultivation decisions. Various predictive approaches have been used to monitor sugarcane production. Authors in [4] integrated LiDAR-derived Normalized Digital Surface Model (NDSM) data with RGB orthoimages to map sugarcane land by identifying objects based on height and texture characteristics. Using a Support Vector Machine (SVM), the model achieved an overall accuracy of 98.74% [4]. Authors in [5] also used LiDAR elevation values derived from RGB data to classify sugarcane growth stages, with prediction accuracies ranging from 87.5%

to 96.3%. However, these approaches remain classification-based. Authors in [6] applied hyperspectral imagery for wheat height estimation using vegetation indices, with a coefficient of determination (R^2) ranging from 0.78 to 0.89.

Although UAVs can produce high-resolution images, their high cost makes them less efficient for large-scale sugarcane plantations. Satellite imagery from PlanetScope, Sentinel-2, and Landsat-8/9 offers a more economical alternative with wide-area coverage [7]. Sentinel-2 excels in spatial resolution and acquisition frequency, whereas Landsat-9 has superior radiometric stability [8]. Authors in [9] designed an early detection model for sugarcane growth using Landsat 8 satellite data accessed through Google Earth Engine. This model can be used to determine the critical phenological period of sugarcane through seasonal trend analysis using linear and harmonic regression models. Similarly, authors in [10] used temporal data from Landsat 8 satellite imagery and local meteorological data to assess sugarcane water stress levels using canopy top reflectance (LST) with Random Forest (RF), achieving an R^2 of 91.08% [10]. Research on estimating sugarcane plant height using remote sensing is limited. Specifically, no study has integrated multi-source satellite imagery, namely Sentinel-2 and Landsat-9, with vegetation, morphological, and meteorological indicators to produce sugarcane height estimates that enable periodic monitoring. The current study bridges this gap by producing an accurate height estimation model that can assist in decision-making for sugarcane cultivation monitoring.

In this study, three satellite image sources (Sentinel-2, Landsat-9, and data fusion) were used as input for Machine Learning (ML) models, including RF, Support Vector Regression (SVR), and Extreme Gradient Boosting (XGB) on three sugarcane plots for 2023 and 2024. The main objective of this research is to identify the factors influencing sugarcane height prediction and compare the performance of single satellite data with data fusion. The findings of this study support more efficient and data-driven sugarcane growth monitoring for decision-making.

II. METHODOLOGY

A. Field Data Collection

The dataset spans two consecutive growing seasons (2023–2024) and comprises bi-weekly Sentinel-2 (10 m), Landsat-9 (30 m), fused satellite products, and elevation data, temporally aggregated into consistent bi-weekly composites, yielding 315 plot-level observations. The fusion process was performed in three main stages. The first stage involved the atmospheric correction of Sentinel 2 (Sen2Cor) and Landsat 9 (Lasrc). The second stage involved radiometric harmonization between the sensors through cross-sensor reflectance harmonization based on linear regression per band. The third stage involved spatial alignment using co-registration based on polynomial transformation with 30 tie points and a Root Mean Square Error (RMSE) of < 0.5 pixels. The entire image was then compiled into bi-weekly composites to ensure the temporal consistency of the data [11]. The processed plot-level dataset used for model development and evaluation is publicly available in the data repository [12].

The third set of satellite image data collected was then preprocessed by removing outliers using 1.5× Interquartile Range (IQR). Essential information was extracted, including vegetation indicators (type of sugarcane ratoon and plant age) and meteorological data (temperature, light intensity, relative humidity, and rainfall), which were used as the estimator variables. This study was conducted in Malang, East Java Province, Indonesia, on three sugarcane plots, as shown in Figure 1. The land was located in Kendalpayak District (-8.060813, 112.619979), Bakalan District (-8.113895, 112.653055), and Jambearjo District (-8.075276, 112.650779). Each plot represented a commercial sugarcane field with comparable land management practices. Variations among plots were mainly associated with ratoon stage and planting age, whereas the soil texture and irrigation regimes remained similar. The ground truth sugarcane height (cm) was measured bi-weekly by using calibrated telescopic measuring poles. Measurements were taken from the soil surface to the highest photosynthetically active leaf, excluding senescent leaves, following the established field-based sugarcane measurement procedures for ground reference data collection [13]. In each plot, five stratified random sampling points were measured bi-weekly, and individual plant heights were averaged to generate plot-level sugarcane height values, resulting in 315 plot-level observations. Accordingly, both the ground-measured height and satellite-derived indicators represent field-averaged conditions rather than independent spatial samples. Each observation reflected the temporal state of a sugarcane field, and the modeling framework captured field-scale temporal growth dynamics rather than within-field spatial variability.

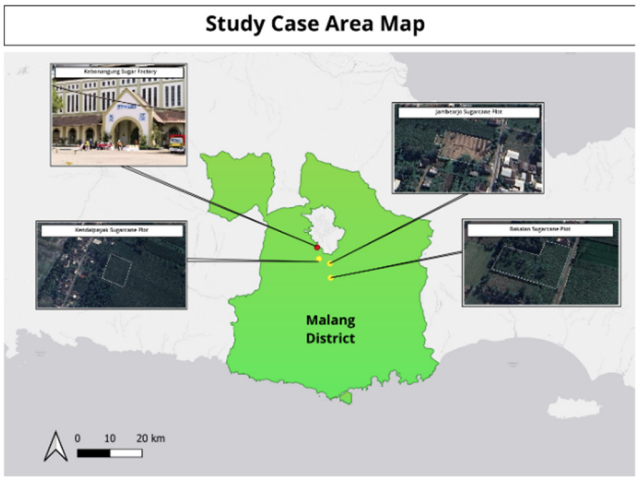


Fig. 1. Study area map with observation plots.

B. Flowchart Study

Figure 2 illustrates the workflow of sugarcane plant height prediction. Data collection was conducted by integrating various sources, including satellite imagery, company information, and supporting documents related to sugarcane cultivation. All data underwent a preprocessing stage, which included correlation analysis, multicollinearity, and data splitting (forward chain). Next, the estimation model was analyzed, and the results were interpreted.

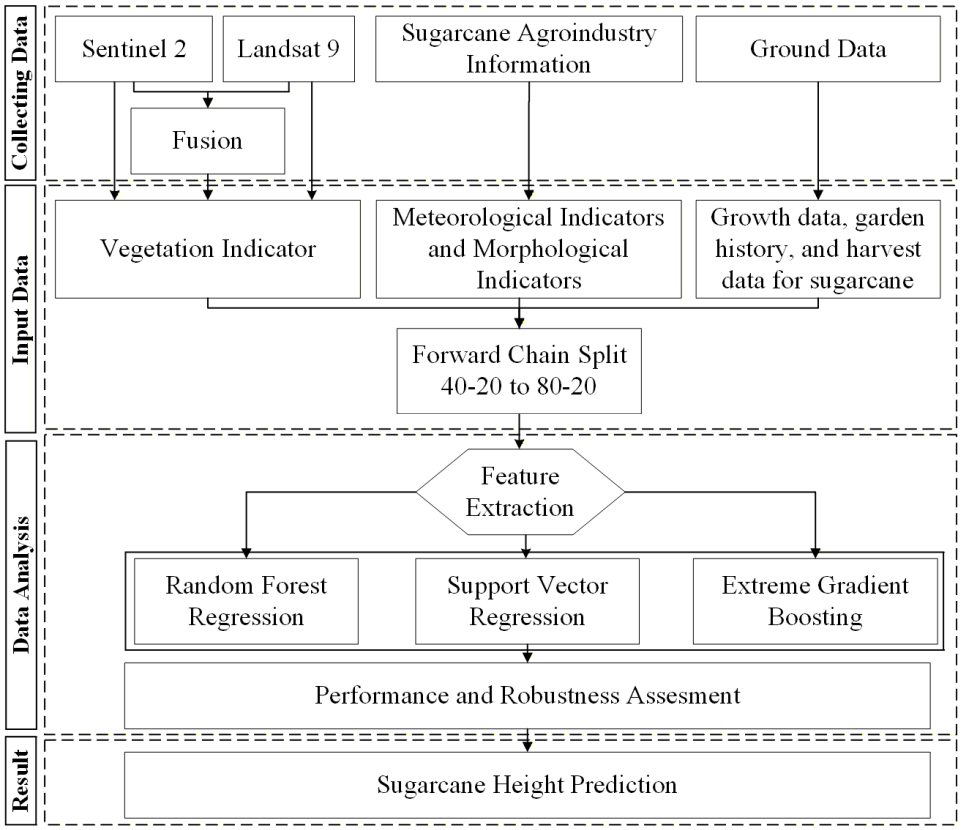


Fig. 2. Workflow for the sugarcane plant height estimation.

C. Estimator Indicator

In this study, the predictor variables consisting of vegetation, morphological, meteorological, and derivative indicators were used to estimate sugarcane height. The vegetation indicators used in the model are:

$$\text{NDVI} = \frac{\text{NIR}-\text{R}}{\text{NIR}+\text{R}} \quad (1)$$

$$\text{GNDVI} = \frac{\text{NIR}-\text{G}}{\text{NIR}+\text{G}} \quad (2)$$

$$\text{NDWI} = \frac{\text{G}-\text{NIR}}{\text{G}+\text{NIR}} \quad (3)$$

$$\text{RVI} = \frac{\text{NIR}}{\text{R}} \quad (4)$$

$$\text{DVI} = \text{NIR} - \text{R} \quad (5)$$

$$\text{EVI}=2.5\,\frac{\text{NIR}-\text{R}}{(\text{NIR}+6\text{R}+7.5\text{B})+1} \quad (6)$$

$$\text{SAVI} = (1+\text{L})\,\frac{(\text{NIR}-\text{R})}{(\text{NIR}+\text{R}+\text{L})} \quad (7)$$

$$\text{WDRVI} = \frac{(0.1\text{NIR})-\text{R}}{(0.1\text{NIR})+\text{R}} \quad (8)$$

where NDVI is the normalized difference vegetation index, GNDVI is the green normalized difference vegetation index, NDWI is the normalized difference water index, RVI is the ratio vegetation index, DVI is the difference vegetation index, EVI is the enhanced vegetation index, SAVI is the soil-adjusted vegetation index, WDRVI is the wide dynamic range vegetation index, G is the green band, L is the soil correction factor, NIR is the near-infrared, R is the red band, and RE is the red edge.

The estimation also incorporates derived variables, such as Water Stress Index (WSI), Temperature and Humidity Interaction (THI), and Vegetation Health (VH), defined in:

$$\text{WSI} = 1 - \left(\frac{\text{Light intensity}}{\text{Relative humidity}}\right) \quad (9)$$

$$\text{THI} = \frac{\text{Temperature}+\text{Relative humidity}}{100} \quad (10)$$

$$\text{VH} = \frac{\text{NDVI}+\text{EVI}+\text{SAVI}}{3} \quad (11)$$

D. Feature Importance

Feature importance was used to quantify the contribution of the estimator variables to sugarcane height and simultaneously eliminate non-contributing features. Feature importance is defined as:

$$I(j) = \frac{1}{M}\sum_{m=1}^{M} I_m(j) \quad (12)$$

$$\text{Gini Importance}(j) = \frac{I(j)}{\sum_k I(k)} \quad (13)$$

where $I_m(j)$denotes the importance of the feature $j$in the m^{th} tree, quantified by the total impurity reduction produced. Based on these formulations, the overall importance of the feature $I(j)$ was computed by averaging its importance scores across all M trees. Features yielding larger impurity reductions were

considered more influential. The averaged importance values were subsequently normalized to obtain their relative contributions to the model performance [13].

E. ML Model

Three ML models (RF, SVR, and XGB) were employed to estimate sugarcane height, using vegetation, morphological, meteorological, and derived data extracted from Sentinel 2, Landsat 9, and data fusion. The RF has a decision tree structure that is trained using different data subsets through bootstrap aggregation [13]. SVR constructs an optimal hyperplane in high-dimensional space [14, 15]. However, XGB uses boosting, which combines multiple decision trees step-by-step by assigning weights to subsequent models to correct errors [16]. The three models were trained using the parameters determined by deploying Randomized SearchCV on the training data in each fold to prevent data leakage. Optimization was performed with 50 random iterations, and the best parameters were selected based on the RMSE value of the validation subset within the training fold. The parameters utilized in the RF approach were n_estimators = 1000, max_depth = 20, min_samples_split = 5, min_samples_leaf = 2, random_state = 42, and n = -1. In the SVR approach, the regularization parameter C = 100, gamma = scale, epsilon = 0.1, and the cache size = 2000 were employed. The parameters used in the XGB approach were n_estimators = 500, max_depth = 8, learning_rate = 0.03, subsample = 0.8, colsample_bytree = 0.8, reg_alpha = 0.1, reg_lambda = 0.1, random_state = 42, and n_jobs = −1.

During the testing phase, data splitting was performed using forward chaining across five folds, expanding the training window, and employing a fixed test window spanning 20 timestamps. The schemes used were 40–20, 50–20, 60–20, 70–20, and 80–20, and each test window was placed immediately after the training window to maintain the temporal consistency and prevent data leakage. The performance of sugarcane height estimation was evaluated using three key metrics: the R^2, RMSE, and Mean Absolute Percentage Error (MAPE), defined in (14-16), respectively. The ML model and satellite-based approach were compared by identifying the highest R^2 (%), along with the lowest RMSE (cm) and MAPE (%):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i^m - y_i^p)^2}{\sum_{i=1}^{n}(y_i^m - y_i^a)^2} \quad (14)$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i^m - y_i^p)^2} \quad (15)$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i^p - y_i^m}{y_i^m}\right| x\ 100 \quad (16)$$

where y_i^m is the actual value, y_i^p is the estimate, and y_i^a is the actual average. The closer the R^2 value is to 1, the better the model is. In contrast, for the RMSE and MAPE approaches, the lower the value, the better the model.

III. RESULTS

A. Feature Importance

The results of the feature importance analysis, as shown in Table I, yielded the top five features to be used as inputs for the estimator based on three image data sources (Sentinel-2, Landsat-9, and data fusion).

TABLE I. FEATURE IMPORTANCE ANALYSIS

Sentinel 2 (68.2%)	**Landsat 9 (75.9%)**	**Data fusion (75.9%)**
Planting age (0.22)	Planting age (0.25)	Planting age (0.24)
GNDVI (0.13)	NDVI (0.15)	NDVI (0.15)
NDVI (0.13)	VH (0.13)	DVI (0.13)
VH (0.12)	GNDVI (0.12)	VH (0.12)
SAVI (0.09)	DVI (0.11)	SAVI (0.12)

The results of the feature importance on Sentinel 2 showed that the selected variables were planting age (0.222), GNDVI (0.130), NDVI (0.126), plant health (0.117), and SAVI (0.087), with a cumulative contribution of 68.2%. The variables obtained for Landsat 9 were planting age (0.249), NDVI (0.152), plant health (0.127), GNDVI (0.124), and DVI (0.106), with a cumulative contribution of 75.9%. For the fusion satellite, the variables were planting age (0.244), NDVI (0.152), DVI (0.127), plant health (0.119), and SAVI (0.117), with a cumulative contribution of 75.9%. The selected features were then processed and analyzed using an ML approach to estimate the current sugarcane height on the land.

B. Performance of ML Models

This study evaluated the performance of RF, SVR, and XGB using Sentinel-2, Landsat-9, and fused data from 2022 and 2023 harvest periods across three sugarcane plots. The actual height data from the sugar industry was used for validation, with comparative estimates presented in Table II. A high predictive performance was observed across different algorithms. Model evaluation using a forward chaining strategy ensured a strict temporal separation between the training and testing data, thereby preventing information leakage. The consistently small performance gap between the training and testing sets ($\Delta R^2 < 0.01$) indicates strong internal consistency under the forward-chaining scheme, while also indicating that the model predominantly captures smooth temporal growth dynamics. The use of plot-level averaging inherently reduces intra-plot spatial variability, thereby limiting the learning of complex spatial relationships.

The test results showed that the fusion technique had the best performance. The data fusion yielded R^2 values ranging from 0.973 to 0.997, RMSE values from 6.254 to 19.375 cm, and MAPE values from 3.097 to 9.973%. Sentinel-2 recorded R^2 values from 0.965 to 0.997, RMSE values from 6.716 to 22.567 cm, and MAPE values from 3.466 to 10.911%. Landsat-9, however, achieved R^2 values from 0.973 to 0.996, RMSE from 8.141 to 19.241 cm, and MAPE from 3.749 to 10.355%.

In contrast, XGB outperformed the other models when compared with ML approaches. XGB yielded R^2 values of 0.996 to 0.997, RMSE of 6.254 to 8.141 cm, and MAPE of 3.097 to 3.749%. Meanwhile, RF achieved R^2 values of 0.965 to 0.975, RMSE of 19.241 to 22.567 cm, and MAPE of 6.779 to 8.148%, whereas SVR achieved R^2 values of 0.973 to 0.974, RMSE of 16.971 to 17.408 cm, and MAPE of 9.973%. A comparison of the ML performance is displayed in Figure 3.

The predictive behavior of the best-performing model (XGB using fused data) was further evaluated using scatter and residual analysis, as illustrated in Figure 4. The scatter plot demonstrated a strong alignment between the predicted and observed sugarcane height values, with most data points closely following the 1:1 reference line across the entire height range. In addition, the residual distribution is centered around zero without evident systematic patterns. The high R^2 value of 0.997 demonstrates strong predictive performance and internal consistency under the evaluated conditions, characterized by homogeneous soil properties and similar management practices.

TABLE II. COMPARISON OF ML AND SATELLITE MODEL PERFORMANCE IN ESTIMATING SUGARCANE CROP HEIGHT

Source	ML	R^2 (train)	R^2 (test)	RMSE (train)	RMSE (test)	MAPE (train)	MAPE (test)
L9	RF	0.978	0.975	19.229	19.241	6.824	6.827
	XGB	0.997	0.996	3.264	8.141	1.162	3.749
	SVR	0.976	0.973	13.251	17.060	7.645	10.355
S2	RF	0.965	0.965	21.872	22.567	8.067	8.148
	XGB	0.998	0.997	3.628	6.716	2.731	3.466
	SVR	0.975	0.974	13.219	17.408	9.802	10.911
Fusion	RF	0.975	0.974	18.924	19.375	6.334	6.779
	XGB	0.998	0.997	2.889	6.254	2.102	3.097
	SVR	0.977	0.973	11.822	16.971	6.980	9.973

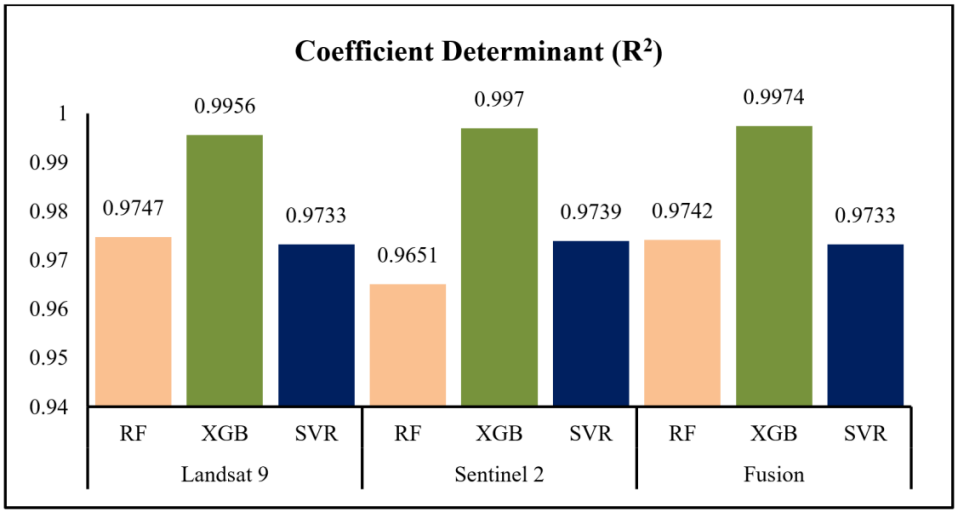
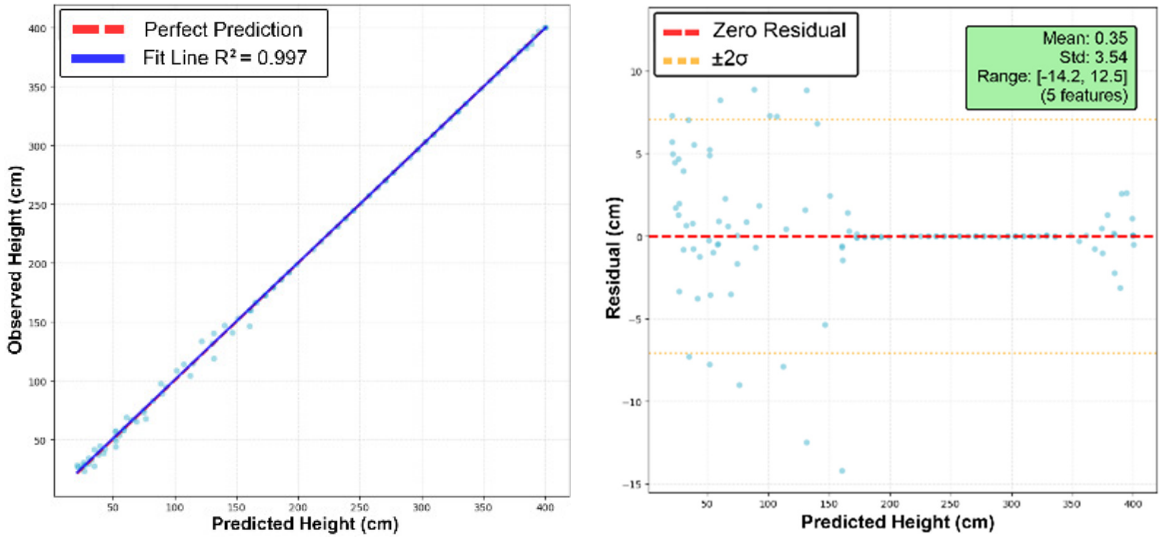


Fig. 3. Performance comparison of RF, XGB, and SVR for sugarcane height estimation using Landsat 9, Sentinel 2, and data fusion.



(a) (b)

Fig. 4. Predicted versus observed sugarcane height and residual analysis for the XGB model using fused satellite data.

IV. DISCUSSION

The performance results indicate that all ML models achieved strong predictive performance. XGB with data fusion achieved the best results with an R^2 of 0.9974, an RMSE of 6.2541 cm, and an MAPE of 3.0967%. The fusion approach and ML training are the novelties explored in this study and have not yet been applied to sugarcane height estimation. Overall, the results indicate that fusion is a better approach than using single satellite data. However, because both predictors and target values were aggregated at the plot level, the model primarily captured the temporal growth dynamics of sugarcane fields rather than the within-field spatial variability. Accordingly, the results should be interpreted as field-averaged temporal monitoring performance, which also explains the consistently high predictive accuracy observed across the models.

Authors in [17] reported that integrating two satellites in the estimation process reduced the estimation error rate of wheat yield by up to 1.8 times compared to using a single satellite. This advantage arises because data fusion provides richer information and reduces the inherent limitations of each satellite. Authors in [18] estimated soybean yield by integrating Sentinel 1 and Sentinel 2. In contrast, the fusion approach uses meteorological and morphological indicators, thereby improving the estimation performance. Authors in [19] conducted a multimodal analysis using Sentinel 2 and climatological indicators consisting of 14 estimator variables to estimate wheat yield and achieved an R^2 value of 0.771 [19].

In this study, the fusion technique yielded the most informative features, including the planting age, NDVI, DVI, crop health, and SAVI. Planting age is a significant parameter in height estimation because height increases progressively with crop maturity. NDVI and crop health also play dominant roles, as they are closely associated with canopy vigor and stem elongation, where unhealthy plants exhibit stunted growth [20]. In addition, DVI and SAVI contribute to height estimation through their sensitivity to biomass accumulation and canopy density [21]. The relative importance of these features varies across satellite sources owing to differences in spatial resolution and spectral configuration. Sentinel-2 provides higher sensitivity to canopy structure through its finer spatial resolution and red-edge bands, whereas Landsat-9 is highly affected by mixed-pixel effects. The fusion of Sentinel-2 and Landsat-9 integrates complementary spectral and spatial information, stabilizes the spectral response, and reduces sensor-specific limitations, thereby amplifying the influence of temporally driven variables such as planting age.

The selected features were then processed using the three ML models, and XGB was found to be the best. XGB can handle complex and non-linear datasets identical to those in agricultural fields. The former has L1-L2 regularization, native handling of missing values, and tree pruning to remove irrelevant branches, thereby making it more efficient and resistant to overfitting [22]. Previous studies have not used XGB to estimate crop height, but it has consistently outperformed other ML approaches in estimating similar cases. Authors in [23] compared the performance of XGB, CNN, and CNN-LSTM in estimating soybean yield using MODIS data and found that XGB provided the lowest RMSE (5.24 bu/ac). Authors in [24] estimated sugarcane yields based on land characteristics and phenology. SVR had lower performance because the relationship pattern between indicators and height was non-linear, making it more suitable for handling by XGB [15], whereas RF had lower performance because it lost stability due to highly varying feature distributions over time [13].

Research has estimated crop height. However, applications specifically targeting sugarcane and incorporating satellite data fusion remain limited. Authors in [25] estimated the height of multiple crops (wheat, potatoes, and beets), using UAV LiDAR and a 3-Dimensional Profile Index (3DPI). Their results achieved R^2 values of 0.78 for winter wheat, 0.50 for potatoes, and 0.70 for beets, with RMSE of 3.4 cm, 12 cm, and 7.4 cm, respectively. However, there was a lack of satellite approach utilization, particularly through data fusion.

The integration of satellite data and ML in this study demonstrates its strong potential for accurate sugarcane height estimation; nevertheless, several limitations should be acknowledged. Image quality can be affected by cloud cover, leading to inconsistencies in the vegetation-derived indicators. In addition, multi-sensor data fusion requires computationally intensive processing pipelines involving radiometric and atmospheric correction, resampling, temporal compositing, and spatial alignment, whereas differences in spectral configurations between Sentinel-2 and Landsat-9 necessitate cross-sensor calibration. Beyond these technical constraints, the study was conducted within a single agroecological zone using three nearby plots managed under relatively homogeneous conditions. Consequently, the observed variability primarily reflects temporal growth dynamics rather than spatial heterogeneity across soil types, crop varieties, and agroclimatic settings. Future research should address these constraints by extending the framework to pixel-level sugarcane height mapping at 10–30 m resolution to support within-field precision monitoring, exploring alternative or individualized outlier detection strategies to better preserve genuine biophysical variability, and validating the model across 10–15 fields in diverse agroecological zones to rigorously assess robustness and generalizability.

V. CONCLUSION

This study employed three satellite data sources (Sentinel-2, Landsat-9, and their fusion) and three Machine Learning (ML) models: Random Forest (RF), Extreme Gradient Boosting (XGB), and Support Vector Regression (SVR) to estimate sugarcane crop height across three land plots over two planting seasons (2023–2024). The models were trained using multiple sources, including vegetation, morphological, meteorological, and derivative, with feature selection applied to identify the most influential predictors. The evaluation results indicated that the XGB model combined with multi-source data fusion achieved the best performance, with a coefficient of determination (R^2) of 0.997, a Root Mean Square Error (RMSE) of 6.254 cm, and a Mean Absolute Percentage Error (MAPE) of 3.097%. However, these results should be interpreted as preliminary findings that demonstrate strong

internal consistency under the current experimental conditions rather than as evidence of broad model reliability.

Accordingly, the proposed approach can be regarded as a proof-of-concept for satellite-based sugarcane height estimation. Further research is required to validate the model across multiple locations with diverse soil characteristics, sugarcane varieties, and agroclimatic conditions to rigorously assess its effectiveness and generalizability. Additional validation using longer time-series data and high-resolution UAV observations as benchmark references is necessary. From an operational perspective, future studies should focus on developing a dashboard to support real-time monitoring once sufficient external validation has been achieved.

ACKNOWLEDGMENT

This research was funded by the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia under Grant No. 006/C3/DT.05.00/PL/2025.

DATA AVAILABILITY STATEMENT

The processed plot-level dataset supporting the findings of this study is publicly available at the Zenodo repository at https://doi.org/10.5281/zenodo.18093974.

REFERENCES

[1] S. Yadav *et al.*, "Accelerating Genetic Gain in Sugarcane Breeding Using Genomic Selection," *Agronomy*, vol. 10, no. 4, Apr. 2020, Art. no. 585, https://doi.org/10.3390/agronomy10040585.

[2] Rohayati, Marimin, H. Hardjomidjojo, F. Fahma, and K. E. Putranto, "A Tertiary Perspective of the Sugarcane-Based Agro-Industry Sustainability Analysis," *Jurnal Teknologi Industri Pertanian*, pp. 286–301, Dec. 2024, https://doi.org/10.24961/j.tek.ind.pert.2024.34.3.286.

[3] S. D. Artikanur, Widiatmaka, Y. Setiawan, and Marimin, "An Evaluation of Possible Sugarcane Plantations Expansion Areas in Lamongan, East Java, Indonesia," *Sustainability*, vol. 15, no. 6, Mar. 2023, Art. no. 5390, https://doi.org/10.3390/su15065390.

[4] M. K. Villareal and A. F. Tongco, "Remote Sensing Techniques for Classification and Mapping of Sugarcane Growth," *Engineering, Technology & Applied Science Research*, vol. 10, no. 4, pp. 6041–6046, Aug. 2020, https://doi.org/10.48084/etasr.3694.

[5] M. K. Villareal and A. F. Tongco, "Multi-sensor Fusion Workflow for Accurate Classification and Mapping of Sugarcane Crops," *Engineering, Technology & Applied Science Research*, vol. 9, no. 3, pp. 4085–4091, Jun. 2019, https://doi.org/10.48084/etasr.2682.

[6] Z. Zhen, L. Yunsheng, O. A. Moses, L. Rui, M. Li, and L. Jun, "Hyperspectral Vegetation Indexes to Monitor Wheat Plant Height Under Different Sowing Conditions," *Spectroscopy Letters*, vol. 53, no. 3, pp. 194–206, Mar. 2020, https://doi.org/10.1080/00387010.2020.1726401.

[7] A. Matese, J. M. Prince Czarnecki, S. Samiappan, and R. Moorhead, "Are Unmanned Aerial Vehicle-Based Hyperspectral Imaging and Machine Learning Advancing Crop Science?," *Trends in Plant Science*, vol. 29, no. 2, pp. 196–209, Feb. 2024, https://doi.org/10.1016/j.tplants.2023.09.001.

[8] S. Acharki, "Planetscope Contributions Compared to Sentinel-2, and Landsat-8 for LULC Mapping," *Remote Sensing Applications: Society and Environment*, vol. 27, Aug. 2022, Art. no. 100774, https://doi.org/10.1016/j.rsase.2022.100774.

[9] S. Sudianto, Y. Herdiyeni, and L. B. Prasetyo, "Early Warning for Sugarcane Growth Using Phenology-Based Remote Sensing by Region," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 2, 2023, https://doi.org/10.14569/IJACSA.2023.0140259.

[10] A. Suharso, Y. Herdiyeni, S. D. Tarigan, and Y. Arkeman, "Water Stress Prediction Model in Rainfed Sugarcane Fields Using Temporal Landsat Data Based on Random Forest Regressor," *IAENG International Journal of Computer Science*, vol. 52, no. 5, pp. 1393–1406, May 2025.

[11] V. Khosravi, A. Gholizadeh, and M. Saberioon, "Soil Toxic Elements Determination Using Integration of Sentinel-2 and Landsat-8 Images: Effect of Fusion Techniques on Model Performance," *Environmental Pollution*, vol. 310, Oct. 2022, Art. no. 119828, https://doi.org/10.1016/j.envpol.2022.119828.

[12] T. T. Baladraf, M. Marimin, I. Hermadi, and A. Ismayana, "Sugarcane Height Estimation." Zenodo, Dec. 30, 2025, https://doi.org/10.5281/ZENODO.18093974.

[13] R. A. Molijn, L. Iannini, J. V. Rocha, and R. F. Hanssen, "Ground Reference Data for Sugarcane Biomass Estimation in São Paulo State, Brazil," *Scientific Data*, vol. 5, no. 1, Aug. 2018, Art. no. 180150, https://doi.org/10.1038/sdata.2018.150.

[14] M. Schonlau and R. Y. Zou, "The Random Forest Algorithm for Statistical Learning," *The Stata Journal: Promoting communications on statistics and Stata*, vol. 20, no. 1, pp. 3–29, Mar. 2020, https://doi.org/10.1177/1536867X20909688.

[15] V. Hanuman, K. V. Pinnamaneni, and T. Singh, "Best Fit Radial Kernel Support Vector Machine for Intelligent Crop Yield Prediction Method," in *Machine Learning and Information Processing*, vol. 1311, D. Swain, P. K. Pattnaik, and T. Athawale, Eds. Singapore: Springer Singapore, 2021, pp. 457–467.

[16] M. Sitienei, A. Anapapa, and A. Otieno, "Application of XGBoost Regression in Maize Yield Prediction," *Asian Journal of Probability and Statistics*, vol. 24, no. 1, pp. 1–9, Aug. 2023, https://doi.org/10.9734/ajpas/2023/v24i1513.

[17] S. Skakun *et al.*, "Winter Wheat Yield Assessment Using Landsat 8 and Sentinel-2 Data," in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, València, Spain, Jul. 2018, pp. 5964–5967, https://doi.org/10.1109/IGARSS.2018.8519134.

[18] K. Amankulova, N. Farmonov, K. Omonov, M. Abdurakhimova, and L. Mucsi, "Integrating the Sentinel-1, Sentinel-2 and Topographic Data into Soybean Yield Modelling Using Machine Learning," *Advances in Space Research*, vol. 73, no. 8, pp. 4052–4066, Apr. 2024, https://doi.org/10.1016/j.asr.2024.01.040.

[19] A.-M. Li *et al.*, "Sugarcane Borers: Species, Distribution, Damage and Management Options," *Journal of Pest Science*, vol. 97, no. 3, pp. 1171–1201, Jun. 2024, https://doi.org/10.1007/s10340-024-01750-9.

[20] T. M. Susantoro, K. Wikantika, A. B. Harto, and D. Suwardi, "Monitoring Sugarcane Growth Phases Based on Satellite Image Analysis (A Case Study in Indramayu and its Surrounding, West Java, Indonesia)," *HAYATI Journal of Biosciences*, vol. 26, no. 3, Dec. 2019, Art. no. 117, https://doi.org/10.4308/hjb.26.3.117.

[21] H. Luo *et al.*, "A Framework for Montane Forest Canopy Height Estimation via Integrating Deep Learning and Multi-Source Remote Sensing Data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 138, Apr. 2025, Art. no. 104474, https://doi.org/10.1016/j.jag.2025.104474.

[22] X. Zhou, S. Chen, N. Peng, X. Zhou, and X. Wang, "Uncertainty Guided Pruning of Classification Model Tree," *Knowledge-Based Systems*, vol. 259, Jan. 2023, Art. no. 110067, https://doi.org/10.1016/j.knosys.2022.110067.

[23] F. Huber, A. Yushchenko, B. Stratmann, and V. Steinhage, "Extreme Gradient Boosting for Yield Estimation Compared with Deep Learning Approaches," *Computers and Electronics in Agriculture*, vol. 202, Nov. 2022, Art. no. 107346, https://doi.org/10.1016/j.compag.2022.107346.

[24] P. Charoen-Ung and P. Mittrapiyanuruk, "Sugarcane Yield Grade Prediction Using Random Forest and Gradient Boosting Tree Techniques," in *2018 15th International Joint Conference on Computer Science and Software Engineering*, Nakhonpathom, Thailand, Jul. 2018, pp. 1–6, https://doi.org/10.1109/JCSSE.2018.8457391.

[25] J. Ten Harkel, H. Bartholomeus, and L. Kooistra, "Biomass and Crop Height Estimation of Different Crops Using UAV-Based Lidar," *Remote Sensing*, vol. 12, no. 1, Dec. 2019, Art. no. 17, https://doi.org/10.3390/rs12010017.