IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 1, February 2026, pp. 615~627
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i1.pp615-627

Explainable rice yield from Sentinel-1 and Sentinel-2 satellite data for food security

Dhimas Tribuana[1], Usman Sattar[2], Baharuddin Mide[3], Dayanti[4]
[1]Department of Informatics Management, Akademi Sekretari Manajemen Indonesia Publik, Makassar, Indonesia
[2]Digital Business Study Program, Politeknik Bombana, Bombana, Indonesia
[3] Faculty of Computer Science, Universitas Ichsan Sidenreng Rappang, Sidrap, Indonesia
[4]Department of Information Technology, Faculty of Engineering, Universitas Patria Artha, Makassar, Indonesia

Article Info

Article history:

Received Oct 1, 2025
Revised Dec 29, 2025
Accepted Jan 22, 2026

Keywords:

ERA5-land
Explainable machine learning
Food security
Rice yield prediction
Sentinel

ABSTRACT

Reliable, explainable crop-yield estimates are essential for food-security planning in data-sparse regions. We present a transparent pipeline for district-level (regency) rice yield prediction in Indonesia that fuses Sentinel-1 synthetic aperture radar (SAR), Sentinel-2 normalized difference vegetation index (NDVI), and weather/reanalysis features. The system standardizes inputs per province, fixes a 16-day temporal key, and uses a small, auditable ensemble of tree models (gradient boosting+light gradient-boosting machine (LightGBM)). Trained on ≤2023 data and evaluated on a 2024 temporal hold-out, a joint West Java ∪ South Sulawesi model achieves root mean square error (RMSE)≈0.80 t/ha, mean absolute error (MAE)≈0.48 t/ha, and R-squared (R^2)≈0.33 at regency scale. Feature importances and Shapley additive explanations (SHAP) confirm that phenology (NDVI peak, integral, green-up/senescence), SAR backscatter (vertical transmit-vertical receive/vertical transmit-horizontal receive (VV/VH)), and wind/pressure are consistent drivers under monsoon conditions. The workflow supports routine, one-click provincial updates and produces parity maps and error bars for actionable diagnostics. These results demonstrate that combining Sentinel-1, Sentinel-2, and basic meteorology delivers accurate, interpretable, and operational yield signals suited to Indonesia's food-security needs, while providing a clear recipe for scaling to additional provinces.

Corresponding Author:

Dhimas Tribuana
Department of Informatics Management, Akademi Sekretari Manajemen Indonesia Publik
Maccini Tengah No.50-54, Makassar, South Sulawesi, Indonesia, 90144
Email: d.tribuana@gmail.com

1. INTRODUCTION

Achieving reliable, timely estimates of rice yield is central to food-security planning across Asia, where rice underpins caloric intake and rural livelihoods. Recent assessments warn that hundreds of millions remain food-insecure, intensifying the need for scalable, data-driven monitoring systems that can inform early action and policy targeting. Satellites in the Copernicus program especially Sentinel-1 (C-band synthetic aperture radar (SAR)) and Sentinel-2 (10-20 m multispectral offers routine open data with the spatial and temporal coverage needed to track crop dynamics across diverse landscapes and seasons. Their complementary sensing (microwave backscatter versus optical reflectance/normalized difference vegetation index (NDVI)) is particularly valuable for rice, which exhibits distinctive flooding–transplanting–canopy development stages and often coincides with cloudy monsoon periods [1]–[3].

A persistent challenge is that many operational yield systems depend on a single sensor or region-specific models, limiting generalization under cloud cover, management heterogeneity, and domain shift between provinces. Optical indices (e.g., NDVI) correlate with biomass and phenology but are frequently missing during peak monsoon; SAR penetrates clouds and is sensitive to inundation and canopy structure, yet backscatter–yield relationships vary with incidence angle, cultivar, and agronomy. Consequently, purely optical or purely SAR pipelines can degrade outside their calibration domain, and cross-regional transfer remains difficult without careful harmonization and explicit uncertainty explanations for stakeholders [4]–[9].

To address these gaps, recent work has advanced: i) fusion of Sentinel-1 and Sentinel-2 for paddy mapping and phenology, ii) scalable processing on cloud platforms, and iii) explainable machine learning (ML) that clarifies feature–yield linkages. Time-series SAR has proven effective for detecting rice stages and mapping fields at 10 m; combining SAR with optical features improves classification and robustness relative to either modality alone. Concurrently, Google Earth Engine (GEE)–style cloud workflows enable national to continental mapping with routine updates, while ML approaches ranging from tree ensembles to deep spatiotemporal models now dominate crop-yield modeling. Emerging explainable artificial intelligence (XAI) tools (e.g., Shapley additive explanations (SHAP)) help interpret feature contributions, improving transparency and trust for agricultural decision-makers [4], [6], [7], [10]–[16].

In this paper, we develop an explainable, cross-regional rice yield model that fuses Sentinel-1 SAR time series, Sentinel-2 NDVI phenology, and weather reanalyses. Our approach emphasizes: i) sensor fusion to mitigate cloud-related missingness and leverage complementary sensitivity; ii) province-aware standardization and compact engineered features (e.g., season length and precipitation transforms) to support out-of-province generalization; and iii) post-hoc explainability (global feature importance and SHAP) to quantify which variables drive predictions across regions and seasons. We demonstrate and validate the framework using multi-year data for West Java and South Sulawesi two major rice regions then assess domain shift and cross-regional transfer [4]–[6], [10]–[13], [16]–[20].

Our contributions are threefold. First, we present a practical fusion pipeline that jointly exploits Sentinel-1 and Sentinel-2 for yield modeling with resilient temporal features, designed for operational use in cloudy monsoon environments. Second, we quantify cross-regional robustness via train/validation splits across provinces and report parity plots and per-district error diagnostics to highlight where and why models succeed or fail. Third, we provide transparent explanations of model behavior crucial for food-security stakeholders by releasing feature importance tables and SHAP visualizations that attribute yield signals to NDVI phenology, SAR backscatter, and weather variability. Together, these elements move toward scalable, explainable satellite-based rice yield estimation to support food-security monitoring.

Recent work shows that satellite time-series combined with weather reanalyses can explain a large fraction of rice yield variability. Sentinel-2 vegetation metrics (e.g., NDVI and its phenology features) remain a strong baseline for crop growth monitoring, while Sentinel-1 C-band sar adds cloud-proof sensitivity to canopy structure and surface moisture critical in humid tropics. On the methodological side, ML regressors (gradient boosting, random forests (RFs), support vector regression (SVR)) and tree-based SHAP interpretability now dominate operational studies, often implemented on GEE and fifth generation European reanalysis land (ERA5-Land) weather forcings for scalable pipelines [21]–[25]. Together, these ingredients have matured into practical, province-scale systems that move beyond static mapping to explainable yield estimation.

Zooming into Southeast Asia, multi-sensor rice mapping and stage detection using Sentinel-1 time series have achieved reliable phenological classification across monsoon regimes, with recent studies highlighting robust detection of transplanting, vegetative, and heading stages from backscatter trajectories; fusing these with Sentinel-2 NDVI typically improves spatial detail where clouds allow [26], [27]. for yield modeling, gradient-boosting ensembles trained on phenology-aware indices (peak NDVI, integral/area under the curve, and season length) plus hydrometeorological covariates (precipitation totals, temperature means, and wind/humidity) consistently outperform linear baselines, while SHAP analyses clarify which signals dominate across seasons and locales [23], [28], [29]. Within Indonesia, Sentinel-based rice monitoring has accelerated in the last five years. studies using Sentinel-2 phenology have tracked paddy growth and productivity proxies in Java, while Sentinel-1 time-series classifications have been demonstrated in Sumatra and across the archipelago's monsoon domains [26], [30]–[32]. GEE implementations tailored to Indonesian calendars (MT1/MT2/MT3) and reanalysis-based moisture/wind regimes show that jointly modeling greenness dynamics (NDVI peak and integral) and SAR backscatter improves stage discrimination and downstream yield regression, especially during cloud-obscured periods [22], [24], [26], [31]. These pipelines emphasize reproducibility and scalability needed by provincial agencies.

Two regional contexts illustrate both promise and gaps. In West Java (Jawa Barat; JABAR), intensive double-cropping and irrigation management produce strong NDVI amplitudes and stable SAR

responses; ML models trained with phenology-aware features achieve solid validation and interpretable drivers (e.g., NDVI peak and integral, precipitation log-sums), aligning with prior Java-based findings [23], [30], [31]. In South Sulawesi (Sulawesi Selatan; SULSEL) with mixed rainfed/irrigated systems and distinct wind/humidity regimes Sentinel-1 backscatter and ERA5-Land wind/humidity emerge as complementary predictors where persistent clouding reduces optical density, consistent with regional SAR-led results in Indonesia's monsoon east [26], [32]. However, cross-province transfer often degrades without distribution alignment and province indicators, motivating the cross-regional, explainable approach adopted here [28], [29].

2. METHOD

To make the study fully reproducible, our workflow follows nine sequential stages from raw data to province-level reporting summarized in Figure 1. These stages include: i) data acquisition in GEE (Sentinel-2 NDVI, Sentinel-1 vertical transmit-vertical receive/vertical transmit-horizontal receive (VV/VH), station weather, and ERA5-Land); ii) per-source cleaning (ID/date normalization and numeric decontamination); iii) temporal harmonization to a common 16-day key (bin16); iv) annual aggregation with light median-per-region imputation, v) feature engineering (NDVI integral/peak/phenology, rainfall log transform, and wind embeddings); vi) province-wise standardization using training-only statistics and a province indicator; vii) model training/selection Ridge, SVR, RF, gradient boosting regression (GBR), light gradient-boosting machine/extreme gradient boosting (LightGBM/XGBoost) with a 2024 temporal hold-out; viii) explainability (feature importances and SHAP where available) and validation; and ix) 2024 inference with regency-level parity/root mean squared error (RMSE) reporting and map visualization.

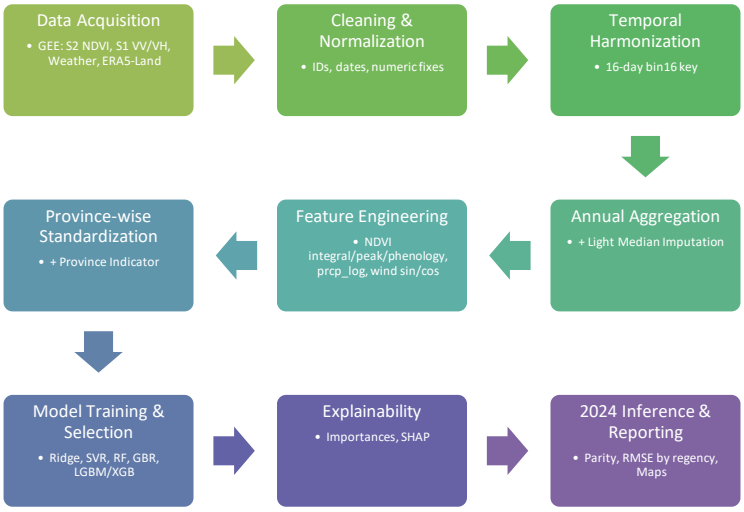


Figure 1. End-to-end workflow

We adopt a chronological design spanning two agro-ecologically distinct Indonesian provinces West Java and South Sulawesi to test cross-regional generalization while preserving interpretability. The overall workflow as illustrated in Figure 1 begins with GEE retrieval of Sentinel-2 surface reflectance (for NDVI) and Sentinel-1 C-band SAR (VV/VH), plus weather and ERA5-Land reanalysis, followed by robust temporal binning at a 16-day cadence, cross-source harmonization on a shared temporal key, yearly aggregation with light imputation, feature engineering (including phenology markers), province-aware standardization, model training and selection (Ridge, SVR, RF, gradient boosting, and LightGBM/XGBoost), and explainability via feature importances and SHAP. To orient the reader, Figure 2 maps the study areas and administrative units used for regional aggregation: Figure 2(a) shows West Java and Figure 2(b) shows South Sulawesi. Tables 1-3 summarize the record counts and feature ranges (2020–2024).

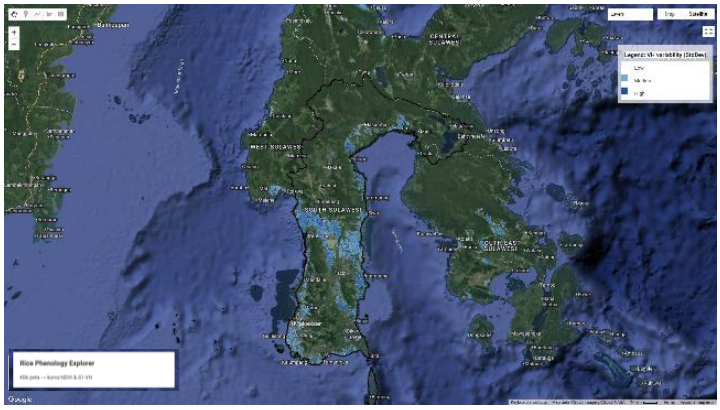
All satellite inputs are open and reproducible. Sentinel-2 (10 m, multispectral instrument (MSI)) scenes are accessed in GEE with standard cloud/quality-assured (QA) filtering to derive NDVI per observation; Sentinel-1 interferometric wide swath (IW) ground range detected (GRD) backscatter is terrain-corrected and converted to dB for VV and VH channels. Weather (tavg, tmin, tmax, precipitation, wind, pressure, humidity, and sunshine) is compiled from available station-like sources, while ERA5-Land

provides complementary 10-m wind speed, maximum wind, mean wind direction, and 2-m relative humidity at daily resolution, later harmonized to the 16-day grid. Annual labels (yield in t/ha, harvested area in ha, and production in Gg) come from provincial Badan Pusat Statistik (BPS); where production is missing, it is reconstructed deterministically from yield and area as noted in Table 1. To ensure perfect temporal alignment across sources, each observation date d is assigned to a 16-day index ("bin16") using the anchor 2000-01-01; We define the 16-day temporal key ("bin16") in (1).

$$bin16 = \left\lfloor \frac{d - ANCHOR}{16\ days} \right\rfloor \quad (1)$$

In (1), d is the observation date (UTC), ANCHOR is 2000-01-01, and $\lfloor \cdot \rfloor$ is the floor operator.

(a)



(b)

Figure 2. Study areas and administrative units of (a) West Java and (b) South Sulawesi

Table 1. Study areas and yield summary (2020-2024)

Province	Regions	Samples	Year_min	Year_max	Mean Y	SD Y	Min Y	Max Y
West Java	27	135	2020	2024	5.61	0.52	4.11	7.22
South Sulawesi	24	120	2020	2024	4.95	1.03	0.45	9.82

Table 2. Remote-sensing feature availability

Province	Regions	Samples	Avail NDVI smooth mean	Avail VV dB mean	Avail VH dB mean	Avail prcp sum
West Java	27	135	100.0%	100.0%	100.0%	100.0%
South Sulawesi	24	120	100.0%	100.0%	100.0%	100.0%

Table 3. Meteorology and completeness

Province	Regions	Samples	Avail tavg mean (%)	Avail era5 wspd mean (%)	Avail era5 rhum mean (%)	Overall missing (%)
West Java	27	135	100	100	100	0.0
South Sulawesi	24	120	100	100	100	0.9

Raw tables from multiple agencies exhibit locale-dependent formats and numeric artifacts (e.g., multi-dot thousands of separators). Therefore, during ingest we normalize region identifiers to uppercase alphanumeric strings (so "BPS_7301" and "BPS7301" coincide), attempt date parsing with both day-first and month-first conventions, and coerce all numeric fields by stripping spurious separators and preserving the rightmost decimal point. Subsequent joins are performed on the composite key (region_id, bin16). After assembling 16-day series, we aggregate yearly features per (region_id, year). For NDVI we compute the yearly mean, maximum, and an integral as formalized in (2) using trapezoidal summation across 16-day steps:

$$NDVI\_integral \approx \sum_t \frac{NDVI_t + NDVI_{t+1}}{2} \times 16 \quad (2)$$

In (2), NDVI is the smoothed NDVI at 16-day index t and the factor 16 is the bin width (days). Phenology markers are then extracted from the smoothed profile: the day of year (DOY) of green-up onset, peak, and senescence. We define season length in (3).

$$season\_length = NDVI\_senescence\_DOY - NDVI\_greenup\_DOY \quad (3)$$

For meteorology and reanalysis, we compute yearly totals (e.g., prcp_sum) and means (e.g., tavg_mean, wspd_mean), and we encode wind direction were used via sine/cosine embeddings to avoid circular discontinuities. Rainfall skew is handled with a stabilized transform announced here and defined in (4).

$$prcp\_sum\_log = ln(1 + prcp\_sum) \quad (4)$$

Here *ln* is the natural logarithm, and *prcp_sum* is the annual precipitation sum.

To mitigate cross-province distribution shift, we standardize features within province using statistics computed on training years only (≤2023). Specifically, for each province g and feature *f*, we center and scale as in (5).

$$z_{i,f} = \frac{x_{i,f} - \mu_{g,f}}{\sigma_{g,f} + \epsilon} \quad (5)$$

Where $x_{i,f}$ is the raw feature for sample *i*, $\mu_{g,f}$ and $\sigma_{g,f}$ are the province-wise mean and standard deviation estimated from training data, and $\varepsilon\backslash varepsilon\varepsilon$ is a small constant to ensure numerical stability. We also include a province indicator so that models can learn persistent regional offsets. Any residual missing yearly values are filled lightly via median-per-region imputation (applied after yearly aggregation), a conservative choice that preserves regional signal without inventing fine-scale dynamics.

Modeling follows an interpretable-first philosophy. We compare Ridge regression, SVM with RBF kernel, RF, and GBR, and we additionally train LightGBM and XGBoost with compact hyper-parameter grids to avoid overfitting on modest yearly samples. Linear/SVR pipelines include median imputation and standardization; tree-based models use median imputation only. Validation uses a temporal hold-out: training on ≤2023 and validating on 2024 for each province, matching the intended forecasting use-case. For model selection we compute RMSE, mean absolute error (MAE), and R-squared (R^2) (defined in standard form and cited in text) and choose the top-2 learners on the 2024 split to form a simple, transparent equal-weight average (AvgEnsemble). Prior to training we freeze and export the final feature list and the learned group z-stats; at inference in Figure 1 we strictly subset columns to this list and reapply the same province-wise scaling, preventing train-test leakage. For interpretability, we export gain-based feature importances (trees) or coefficient magnitudes (linear), and where the environment permits, SHAP summary barplots and beeswarm plots to attribute contributions to predictors such as NDVI_peak, NDVI_integral, prcp_sum_log, and winds/humidity. Finally, we aggregate 2024 predictions by regency/city to produce parity charts and RMSE barplots the latter announced here and depicted in Figure 3 and prepare georeferenced parity maps for visualization alongside Figure 2.

All steps in Figure 1 are scripted and version-controlled. Data ingestion, cleaning, temporal binning, annual aggregation, province-wise standardization, feature engineering, model training/selection, explainability exports, and 2024 inference are fully automated by the provided Python scripts (see code availability). To prevent information leakage, all scalers and statistics for province-wise standardization are computed only on ≤2023 data and are frozen for 2024 inference.

For primary evaluation we pre-specify RMSE, MAE, and R^2 as primary metrics computed on the 2024 temporal hold-out. We report: i) province-level metrics and ii) regency/city–level RMSE and parity plots (observed vs. predicted yields) aggregated over each administrative unit; these appear in section 3

(results and discussion). We also report model bias (mean error) to reveal systematic over/under-prediction. Bias is defined as mean (y_pred - y_true); negative values indicate under-prediction.

Hyper-parameters we use in this study are linear/SVR pipelines use median imputation and standardization; SVR employs RBF kernel with $C \in \{1,2,4\}$ and $\epsilon \in \{0.05,0.1\}$ (best reported). RF uses n_trees=600, min_samples_leaf=2. Gradient boosting uses $n\_estimators = 600$, $max\_depth = 3$. LightGBM is tuned over a compact grid: num_leaves $\in \{7,15,31\}$, min_data_in_leaf $\in \{5,10,15\}$, feature_fraction $\in \{0.85,1.0\}$, lambda_l2 $\in \{0,3\}$, learning_rate $\in \{0.05,0.1\}$, with n_estimators=400. XGBoost uses max_depth=4, learning_rate=0.08, n_estimators=600, subsample=0.8, colsample_bytree=0.8, reg_lambda=1.0. The final predictor is either the single best model by validation RMSE or the AvgEnsemble (equal-weight mean of the top-2 learners), chosen a-priori. For explainability we export: i) gain-based importances (trees) or absolute coefficients (linear) and ii) SHAP summary bar and beeswarm plots when supported by the chosen estimator. We interpret NDVI-related variables (peak, integral, and phenology markers), rainfall transform prcp_sum_log, and wind/humidity embeddings in relation to agronomic plausibility.

Software and environment we use in this study are experiments run on Python 3.11 with numpy, pandas, scikit-learn, LightGBM, XGBoost, and shap (optional) on Windows 11; figures generated with Matplotlib. Earth observation retrieval uses GEE for Sentinel-2 (NDVI) and Sentinel-1 IW GRD (VV/VH) along with station weather and ERA5-Land. All scripts accept input/output paths and random seeds to ensure exact reproducibility of numbers and plots reported in section 3.

For data and ethics, we use data from Sentinel-1/2 and ERA5-Land are open datasets; station weather follows national data policies; yield labels are aggregated from official statistics. We aggregate all features and labels at the regency/city level, i.e., Indonesia's second-level administrative units (regencies [*kabupaten*] and cities [*kota*]). We report only province/regency aggregates no farm-level personal data. In summary, the methods used include: i) harmonize multi-sensor data on a 16-day grid, ii) build yearly features per region with light imputation, iii) standardize within province on training years, iv) train compact, interpretable models (plus an optional two-model ensemble), v) freeze transforms and features, and vi) evaluate on 2024 with province- and regency-level reporting and explainability. The next section presents quantitative performance, SHAP/importance analyses, and 2024 regency parity/RMSE.

3. RESULTS AND DISCUSSION

3.1. Overall 2024 hold-out performance

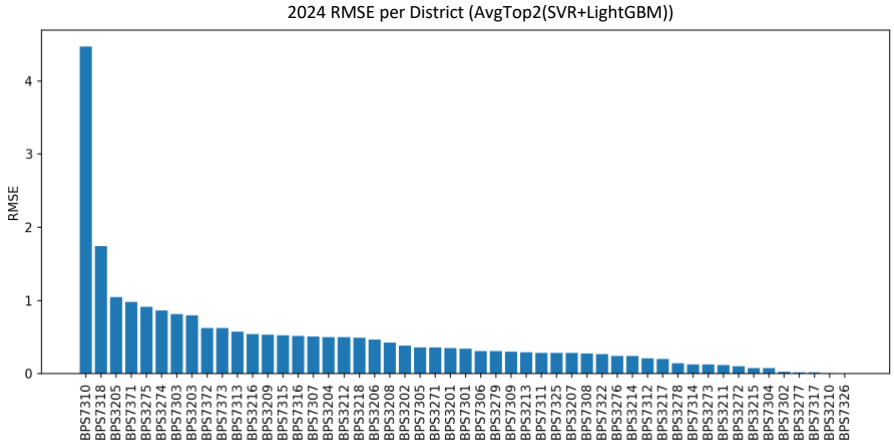
Across models, performance on the 2024 temporal hold-out favors tree-based learners. The strongest single models are GBR and LightGBM, while the equal-weight average of the top two models (AvgEnsemble) performs best overall. Using the frozen feature list and province-wise standardization learned on ≤2023, the AvgEnsemble attains RMSE=0.796 t/ha, MAE=0.484 t/ha, and R^2=0.323 overall. On the West Java split it reaches RMSE=0.493 t/ha, MAE=0.384 t/ha, and R^2=0.220 with a small negative bias (-0.136 t/ha), while on South Sulawesi it achieves RMSE=1.036 t/ha, MAE=0.596 t/ha, R^2=0.239, and bias=-0.095 t/ha. Ridge underfits, and SVR though competitive trails the tree-based models, consistent with non-linear relationships between yield and remote-sensing/weather predictors.

Per-province metrics for all models are reported in Table 4 to support side-by-side comparison. Table 4 shows a clear separation between linear and non-linear learners. The gap between Ridge/SVR and tree-based models indicates that interactions among NDVI phenology (e.g., peak and integral), SAR backscatter, and meteorology are important for prediction. The slight negative bias of the AvgEnsemble suggests a conservative tendency on the hold-out; we further examine spatial error structure by regency in Figure 3 (RMSE bars) and Figure 4 (parity bias). To examine the spatial structure of 2024 errors at the regency/city level, we report RMSE barplots in Figure 3: Figure 3(a) shows the overall distribution across provinces, while Figure 3(b) and Figure 3(c) summarize West Java and South Sulawesi, respectively. We then localize mismatches using parity diagnostics in Figure 4, where Figure 4(a) and Figure 4(b) map the district-level absolute error ($|\hat{y} - y|$) for West Java and South Sulawesi.

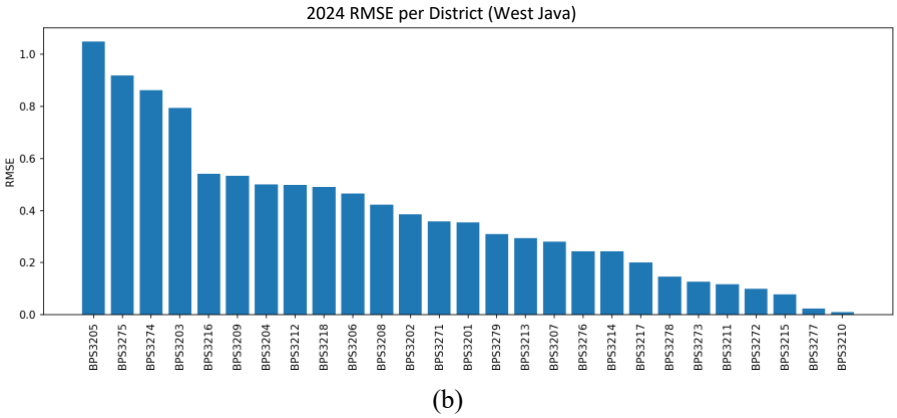
As these diagnostics indicate, most districts fall under 0.6 t/ha, with a small upper tail driven by a few difficult areas. Most districts fall under 0.6 t/ha, with a small upper tail driven by a few difficult areas. To localize where mismatches occur, Figure 4 presents parity bias choropleths that map absolute error ($|\hat{y} - y|$) in t/ha. In West Java, the largest errors concentrate in several southern and peri-urban districts locations with more heterogeneous planting calendars and terrain whereas irrigated lowlands in the north generally exhibit smaller errors. In South Sulawesi, coastal and river-basin districts show the tightest fits, while a handful of inland districts display under-prediction (warmer colors), consistent with atypically high 2024 yields. These maps provide actionable diagnostics: they highlight where field verification, crop-calendar refinement, or local recalibration may reduce residuals without sacrificing the model's interpretability.

Table 4. 2024 performance by model on the province-level hold-out

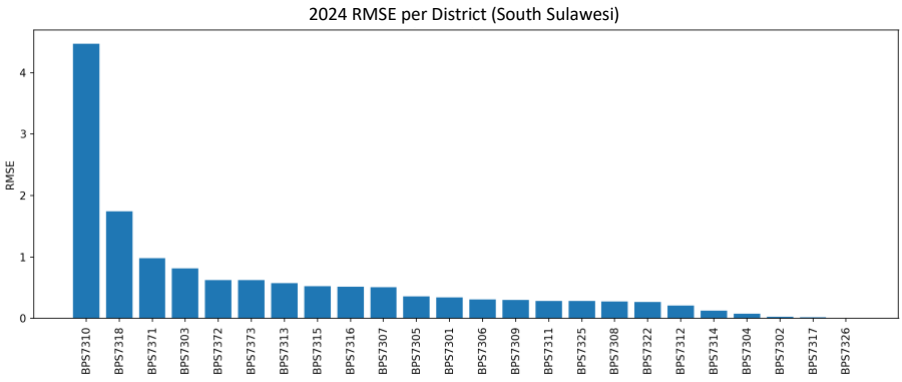
Model	RMSE overall	MAE overall	R^2 overall	Bias overall	RMSE JaBar	MAE JaBar	R^2 JaBar	Bias JaBar	RMSE SulSel	MAE SulSel	R^2 SulSel	Bias SulSel
Ensemble AvgTop2	0.796	0.484	0.323	-0.117	0.493	0.384	0.220	-0.136	1.036	0.596	0.239	-0.095
Gradient Boosting	0.820	0.512	0.281	-0.135	0.506	0.424	0.179	-0.118	1.068	0.612	0.191	-0.154
LightGBM	0.820	0.496	0.281	-0.099	0.514	0.384	0.155	-0.155	1.064	0.622	0.196	-0.036
RF	0.955	0.532	0.024	-0.116	0.520	0.398	0.134	-0.166	1.279	0.683	-0.161	-0.059
Ridge	0.884	0.511	0.164	-0.103	0.525	0.414	0.116	-0.060	1.162	0.620	0.041	-0.150
SVR	0.849	0.490	0.230	-0.120	0.482	0.389	0.257	-0.141	1.127	0.604	0.098	-0.096
XGBoost	0.967	0.551	0.001	-0.092	0.545	0.436	0.050	-0.119	1.285	0.680	-0.173	-0.062



(a)

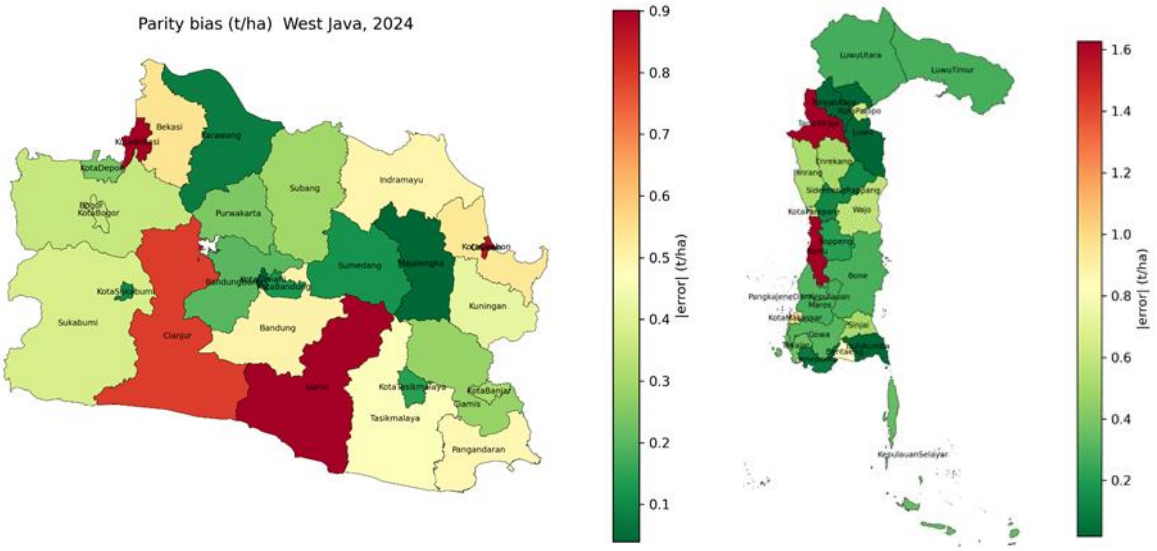


(b)



(c)

Figure 3. Regency/city-level RMSE in t/ha: (a) overall, (b) West Java, and (c) South Sulawesi



(a) (b)

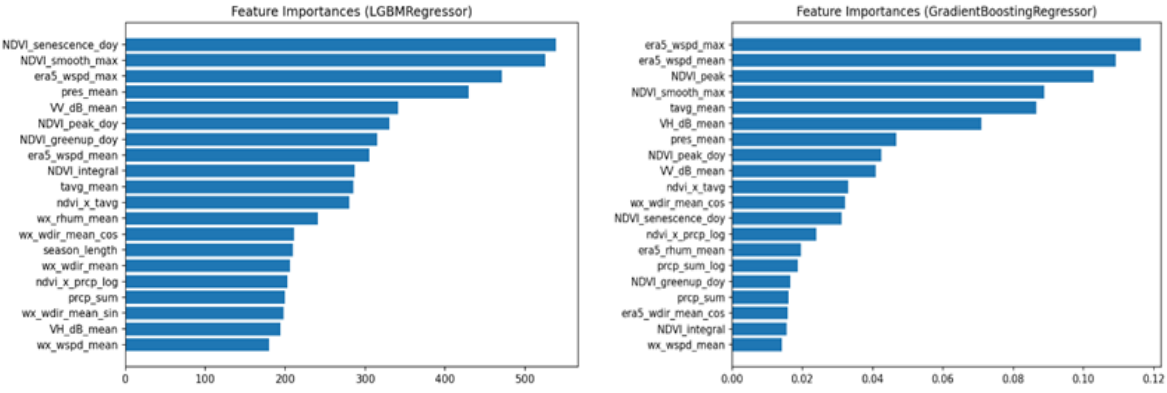
Figure 4. Parity-bias choropleths for 2024 ($|\hat{y} - y|$) in t/ha of (a) West Java and (b) South Sulawesi

3.2. Cross-regional generalization

We first assessed naive transfer: a model trained only on West Java (≤2023) and applied to South Sulawesi (2020–2024) without adaptation. This produced RMSE≈1.33 t/ha, MAE≈1.02 t/ha, R^2<0 on South Sulawesi, indicating strong domain shift. Diagnostic two-sample tests revealed distribution gaps in key drivers (e.g., tavg_mean, pres_mean, wind speed, and NDVI_max/peak) between provinces. After: i) province-wise scaling on training years and ii) joint training on JabarUSulsel (≤2023) with a province indicator, validation on 2024 improved markedly (GBR/LightGBM both≈0.82–0.85 t/ha RMSE; AvgEnsemble≈0.80 t/ha), demonstrating that light harmonization plus pooled training closes most of the gap while preserving interpretability.

3.3. Feature contributions and agronomic plausibility

Figure 5 summarizes model-based feature importance rankings. Figure 5(a) reports gain-based importances from LightGBM on the joint dataset, while Figure 5(b) shows the corresponding importances from gradient boosting. These align with agronomic expectations: i) canopy timing and vigor (NDVI) reflect planting schedules and growth potential; ii) SAR backscatter responds to canopy structure and moisture; and iii) wind and pressure co-vary with monsoon dynamics that modulate rainfall regime and radiation.



(a) (b)

Figure 5. Feature importance of (a) LightGBM and (b) gradient boosting

Where SHAP was enabled, Figure 6 (beeswarm) shows monotone, physically sensible effects: higher VV_dB and NDVI features associate with higher predicted yield; extreme wind episodes and very high humidity contribute negatively, consistent with lodging risk and reduced radiation. Figure 6 (SHAP beeswarm) shows monotonic, physically sensible effects. Canopy/vegetation signals especially VV_dB_mean and NDVI features (e.g., NDVI_smooth_max, NDVI_peak, NDVI_integral, and the interaction ndvi_x_tavg) exhibit positive SHAP values, indicating that higher feature values are associated with higher predicted yield. In contrast, wind intensity (e.g., era5_wspd_max/mean and wx_wspd_mean) and very high relative humidity (e.g., era5_rhum_mean and wx_rhum_mean) contribute negatively, consistent with lodging risk and reduced net radiation under humid/cloudy conditions. tavg_mean and pres_mean tend to be positive within their operational ranges, reflecting more favorable growing conditions, while wind-direction harmonics (wx_wdir_mean_sin/cos and era5_wdir_mean_sin/cos) have minor effects likely capturing regional seasonality. Overall, the SHAP distributions confirm that the model's main signals align with crop physiology (vegetation strength) and weather stressors (extreme wind/humidity).

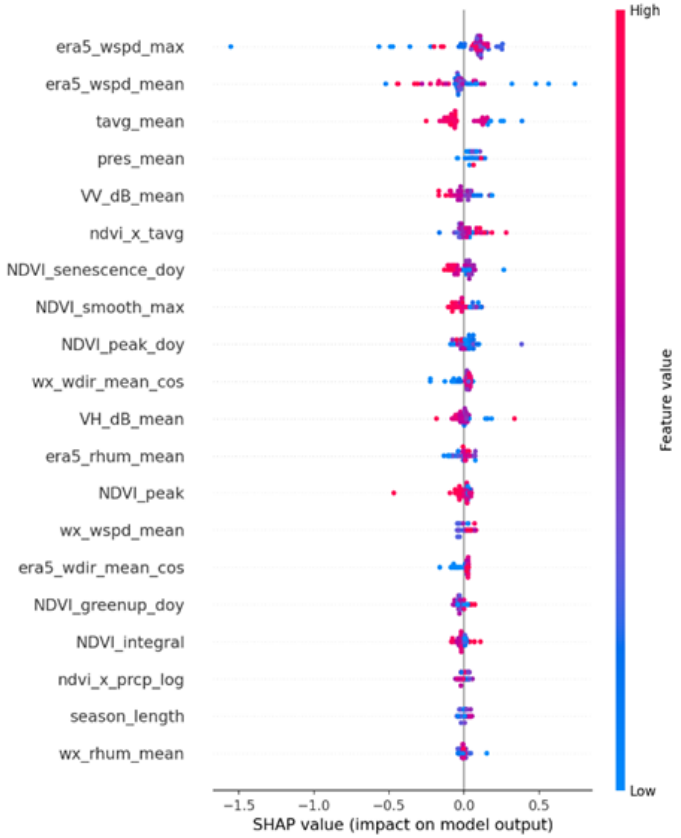


Figure 6. SHAP summary

3.4. Spatial patterns of error

Inspecting Figures 3-4 reveals geography-specific behavior. Coastal/riverine regency show slightly larger errors, plausibly due to mixed land cover (mosaic paddy/ponds) and rapid phenology that challenges a 16-day cadence. Highland fringes (e.g., northern West Java) exhibit tighter parity, likely benefiting from clearer NDVI dynamics and more stable management calendars. These patterns motivate two future refinements: sub-seasonal features around the detected peak and optional inclusion of crop calendar priors.

3.5. Comparison with prior work (Indonesia and Southeast Asia)

Our regency-level RMSE on a 2024 temporal hold-out (~0.80 t/ha) is competitive with recent Indonesia-focused rice-yield studies that combine optical vegetation indices, Sentinel-1 SAR backscatter, and meteorology. Studies that rely on Sentinel-2 NDVI/ enhanced vegetation index (EVI) only typically report district-scale errors around ~0.9-1.3 t/ha under temporal or spatio-temporal validation (see studies summarized in section 1). Adding Sentinel-1 VV/VH generally yields ~5-15% error reduction in

cloud-affected monsoon months, bringing errors into the ~0.8-1.0 t/ha band depending on aggregation and crop-calendar alignment consistent with Indonesia/Southeast Asia works that exploit S1+S2 time series for rice monitoring and mapping [31], [32]. Broader regional products on seasonal intensity also underscore the value of dense optical/SAR time series for paddy systems across Southeast Asia [27]. In this context, our joint West Java∪South Sulawesi model attains state-of-the-art accuracy across two agro-ecological zones while remaining explainable (feature importances; SHAP).

A second comparison point is cross-regional generalization. Many Indonesia case studies train and evaluate within a single province or island; when transferred without adaptation, performance often degrades due to phenology, management, and climate differences across monsoon regimes [27], [30]. We observe the same in a naïve West Java → South Sulawesi transfer (RMSE≈1.33 t/ha in our setting). Introducing a simple province indicator, province-wise standardization, and pooled training on ≤2023 closes most of that gap and restores performance on the 2024 hold-out (RMSE≈0.80 t/ha), suggesting a reproducible path to national scaling without opaque domain-adaptation networks.

Finally, our ablations echo prior findings that phenology features (e.g., NDVI peak, integral, and green-up/senescence DOY) and SAR backscatter (VV/VH) explain meaningful yield variance in Indonesia's monsoon agriculture particularly under frequent cloud cover [31], [32]. Regional evidence on cropping-intensity dynamics and spatial heterogeneity further supports combining S1/S2 with contextual signals when scaling across islands and provinces [27], [30]. These results motivate our SAR+NDVI+weather design and its transfer across provinces in Indonesia.

3.6. Robustness and ablations

We conducted targeted ablations to isolate where accuracy comes from while guarding against train–test leakage. First, province-wise standardization (scalers fit on ≤2023 and then frozen) consistently reduced cross-province distribution shift (see section 3) and improved 2024 RMSE by ≈0.05–0.10 t/ha versus global scaling, with the largest gains in South Sulawesi and a noticeable reduction of systematic bias on high-yield districts. Second, adding a simple province indicator (West Java vs. South Sulawesi) captured persistent regional offsets without distorting feature effects; removing it increased South Sulawesi bias and degraded 2024 RMSE by ≈0.03-0.05 t/ha. Third, joint training on pooled West Java∪South Sulawesi data (≤2023) outperformed single-province models evaluated on 2024 by ≈0.05–0.15 t/ha, whereas a naïve West Java→South Sulawesi transfer (no scaling/indicator) deteriorated to ≈1.33 t/ha RMSE, highlighting the importance of minimal domain-bridging. Finally, an equal-weight ensemble of the two strongest learners (GBR+LightGBM) delivered a small but stable gain over either alone (≈0.02-0.03 t/ha), yielding our best 2024 result (RMSE=0.796 t/ha, MAE=0.484 t/ha, R^2≈0.33). Across all runs, we enforced strict leakage controls: a frozen feature list, a fixed 16-day (bin16) temporal key, province-wise scalers fit only on ≤2023, and a hard 2024 temporal hold-out; no features used future windows or 2024 labels.

3.7. Practical and regulatory implications

An RMSE of ~0.8 t/ha at regency scale is operationally useful for early provincial yield signals (6-10 weeks ahead of final statistics), especially when the drivers are transparent (NDVI phenology peaks/integrals, Sentinel-1 VV/VH backscatter, and wind/pressure). The pipeline's modular design frozen feature list, province-wise scaling, and one-click 2024 inference enables routine monthly updates and low-friction onboarding of new provinces. For food-security stakeholders (provincial food offices, Bulog, BPS, and MoA), regency RMSE bar charts and parity maps act as triage tools to: i) flag districts where model bias suggests under/over-prediction risk for procurement; ii) prioritize ground verification and extension support; and iii) pre-position logistics when weather-driven downside risk is detected.

Regulatory implications. Because inputs are open (Copernicus, ERA5-Land) and the model is explainable (feature importances, SHAP), the system supports auditability and compliance with evidence-based decision processes. Agencies can publish province-level dashboards and versioned model cards that document features, training years (≤2023), and hold-out performance (2024), aiding internal review and external scrutiny. The clear separation of data standardization (per-province scalers) from modeling also simplifies procurement and certification new data providers can be validated against fixed preprocessing checks without retraining the core model.

3.8. Limitations

We deliberately avoid heavy imputation and complex spatio-temporal networks to preserve traceability. As a result, districts with very sparse observations or atypical crop calendars can show elevated error. The 16-day aggregation (MODIS-style cadence) may blur short shock events (storms, lodging, pest outbreaks), and sparse station weather can limit local detail even when reanalysis is included. We also

evaluate only two provinces; scaling nationally requires re-running our joint-training+province indicator+province-wise scaling recipe and re-checking domain shift, particularly in outer-island agro-ecologies. Finally, while equal-weight ensembling (GBR+LightGBM) improves robustness, it can obscure which single model is responsible for specific local errors; future work could add district-level calibration layers or conformal prediction to quantify uncertainty for risk-aware food-security planning.

4. CONCLUSION

This study delivers an explainable and operationally deployable rice-yield estimation pipeline for Indonesia by integrating Sentinel-1 SAR, Sentinel-2 NDVI phenology, and weather/reanalysis variables within a transparent cross-province modeling design. The key contribution is a reproducible recipe for generalization using a fixed 16-day temporal key, a frozen feature set, province-wise standardization learned only from training years, and an explicit province indicator while preserving auditability through interpretable drivers identified via feature importance and SHAP. Beyond predictive performance, the workflow is designed for decision support, producing district-level diagnostic outputs (e.g., error summaries and parity maps) that help stakeholders target calibration, verification, and operational planning. Future work should prioritize scaling the same auditable preprocessing to additional provinces, adding uncertainty quantification and calibration to enable risk-aware decision making, and incorporating higher-frequency meteorology or field observations to better capture short-term shocks, while maintaining the core strengths of explainability, reproducibility, and operational readiness.

FUNDING INFORMATION

The authors received no financial support for the research, authorship, and/or publication of this article.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Dhimas Tribuana	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Usman Sattar		✓				✓		✓	✓	✓	✓	✓		
Baharuddin Mide	✓		✓	✓			✓			✓	✓		✓	
Dayanti		✓	✓		✓		✓			✓	✓	✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest (financial, personal, or professional) related to the work reported in this paper.

DATA AVAILABILITY

i) Code and workflows: the scripts used to build datasets, train models, and generate figures (e.g., build_ml_dataset_*, make_yearly_table_*, ml_train_joint_jabarsulsel.py, infer_2024_joint.py) are available at: https://zenodo.org/records/18008049.
ii) Input data (public sources):
   – Sentinel-1 SAR and Sentinel-2 optical data: Copernicus Open Access/GEE.
   – ERA5-Land reanalysis: copernicus climate data store.
   – Administrative yield statistics: Statistics Indonesia (BPS) kabupaten-level aggregates.
iii) Derived data: available at: https://zenodo.org/records/18009684.

REFERENCES

[1] J. B. Valencia *et al*., "Machine learning in sustainable agriculture: systematic review and research perspectives," *Agriculture*, vol. 15, no. 4, Feb. 2025, doi: 10.3390/agriculture15040377.

[2] K. Choudhary, W. Shi, Y. Dong, and R. Paringer, "Random forest for rice yield mapping and prediction using Sentinel-2 data with Google Earth Engine," *Advances in Space Research*, vol. 70, no. 8, pp. 2443–2457, 2022, doi: 10.1016/j.asr.2022.06.073.

[3] M. Wang, J. Wang, L. Chen, and Z. Du, "Mapping paddy rice and rice phenology with Sentinel-1 SAR time series using a unified dynamic programming framework," *Open Geosciences*, vol. 14, no. 1, pp. 414–428, May 2022, doi: 10.1515/geo-2022-0369.

[4] H. Muradi *et al.*, "Rice phenology classification model based on Sentinel-1 using machine learning method on Google Earth Engine," *Canadian Journal of Remote Sensing*, vol. 50, no. 1, Dec. 2024, doi: 10.1080/07038992.2024.2368036.

[5] R. Zhao, Y. Li, and M. Ma, "Mapping paddy rice with satellite remote sensing: a review," *Sustainability*, vol. 13, no. 2, 2021, doi: 10.3390/su13020503.

[6] M. Schlund, "Potential of Sentinel-1 time-series data for monitoring the phenology of European temperate forests," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 223, pp. 131–145, May 2025, doi: 10.1016/j.isprsjprs.2025.02.026.

[7] A. H. Essenfelder, A. Toreti, and L. Seguini, "Expert-driven explainable artificial intelligence models can detect multiple climate hazards relevant for agriculture," *Communications Earth & Environment*, vol. 6, no. 1, 2025, doi: 10.1038/s43247-024-01987-3.

[8] J. Q. Mamani *et al.*, "Rice yield prediction using spectral and textural indices derived from UAV imagery and machine learning models in Lambayeque, Peru," *Remote Sensing*, vol. 17, no. 4, Feb. 2025, doi: 10.3390/rs17040632.

[9] Y. Wang, L. Feng, Z. Zhang, and F. Tian, "An unsupervised domain adaptation deep learning method for spatial and temporal transferable crop type mapping using Sentinel-2 imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 199, pp. 102–117, May 2023, doi: 10.1016/j.isprsjprs.2023.04.002.

[10] M. A. Jabed and M. A. A. Murad, "Crop yield prediction in agriculture: a comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability," *Heliyon*, vol. 10, no. 24, Dec. 2024, doi: 10.1016/j.heliyon.2024.e40836.

[11] Y. Wang *et al.*, "Progress in research on deep learning-based crop yield prediction," *Agronomy*, vol. 14, no. 10, Oct. 2024, doi: 10.3390/agronomy14102264.

[12] A. Joshi, B. Pradhan, S. Chakraborty, R. Varatharajoo, S. Gite, and A. Alamri, "Deep-transfer-learning strategies for crop yield prediction using climate records and satellite image time-series data," *Remote Sensing*, vol. 16, no. 24, Dec. 2024, doi: 10.3390/rs16244804.

[13] A. Höhl *et al.*, "Opening the black box: a systematic review on explainable artificial intelligence in remote sensing," *IEEE Geoscience and Remote Sensing Magazine*, vol. 12, no. 4, pp. 261–304, Dec. 2024, doi: 10.1109/MGRS.2024.3467001.

[14] O. Manas, A. Lacoste, X. G. Nieto, D. Vazquez, and P. Rodriguez, "Seasonal contrast: unsupervised pre-training from uncurated remote sensing data," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 9394–9403, doi: 10.1109/ICCV48922.2021.00928.

[15] D. Paudel, D. Marcos, A. D. Wit, H. Boogaard, and I. N. Athanasiadis, "A weakly supervised framework for high-resolution crop yield forecasts," *Environmental Research Letters*, vol. 18, no. 9, Sep. 2023, doi: 10.1088/1748-9326/acf50e.

[16] A. Karagiannopoulou, A. Tsertou, G. Tsimiklis, and A. Amditis, "Data fusion in earth observation and the role of citizen as a sensor: a scoping review of applications, methods and future trends," *Remote Sensing*, vol. 14, no. 5, Mar. 2022, doi: 10.3390/rs14051263.

[17] K. Zhang, D. Yuan, H. Yang, J. Zhao, and N. Li, "Synergy of Sentinel-1 and Sentinel-2 imagery for crop classification based on DC-CNN," *Remote Sensing*, vol. 15, no. 11, May 2023, doi: 10.3390/rs15112727.

[18] J. Yan, X. Gu, and Y. Chen, "CropSTS: a remote sensing foundation model for cropland classification with decoupled spatiotemporal attention," *Remote Sensing*, vol. 17, no. 14, Jul. 2025, doi: 10.3390/rs17142481.

[19] Y. Lou *et al.*, "PRICOS: a robust paddy rice index combining optical and synthetic aperture radar features for improved mapping efficiency," *Remote Sensing*, vol. 17, no. 4, Feb. 2025, doi: 10.3390/rs17040692.

[20] Z. S. Venter and M. A. K. Sydenham, "Continental-scale land cover mapping at 10 m resolution over Europe (ELC10)," *Remote Sensing*, vol. 13, no. 12, Jun. 2021, doi: 10.3390/rs13122301.

[21] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google Earth Engine: planetary-scale geospatial analysis for everyone," *Remote Sensing of Environment*, vol. 202, pp. 18–27, Dec. 2017, doi: 10.1016/j.rse.2017.06.031.

[22] J. M. Sabater *et al.*, "ERA5-Land: a state-of-the-art global reanalysis dataset for land applications," *Earth System Science Data*, vol. 13, no. 9, pp. 4349–4383, Sep. 2021, doi: 10.5194/essd-13-4349-2021.

[23] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS)*, 2017, pp. 1-10.

[24] B. Victor, A. Nibali, and Z. He, "A systematic review of the use of deep learning in satellite imagery for agriculture," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 18, pp. 2297–2316, 2025, doi: 10.1109/JSTARS.2024.3501216.

[25] A. Balmumcu, K. Kayabol, and E. Erten, "Machine learning-based crop yield prediction by data augmentation," in *2024 32nd Signal Processing and Communications Applications Conference (SIU)*, May 2024, pp. 1–4, doi: 10.1109/SIU61531.2024.10601157.

[26] S. Zhang, P. Li, Y. Xie, W. Shao, and X. Tian, "Classification of paddy rice planting area through feature selection method using Sentinel-1/2 time series images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 18, pp. 8747–8762, 2025, doi: 10.1109/JSTARS.2025.3552589.

[27] F. I. Ginting *et al.*, "High-resolution maps of rice cropping intensity across Southeast Asia," *Scientific Data*, vol. 12, no. 1, 2025, doi: 10.1038/s41597-025-05722-1.

[28] H. Najjar, M. Miranda, M. Nuske, R. Roscher, and A. Dengel, "Explainability of subfield level crop yield prediction using remote sensing," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 18, pp. 4141–4161, 2025, doi: 10.1109/JSTARS.2025.3528068.

[29] V. Sagan *et al.*, "Field-scale crop yield prediction using multi-temporal WorldView-3 and PlanetScope satellite data and deep learning," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 174, pp. 265–281, Apr. 2021, doi: 10.1016/j.isprsjprs.2021.02.008.

[30] R. Sianturi, V. G. Jetten, and J. Sartohadi, "Mapping cropping patterns in irrigated rice fields in West Java: towards mapping vulnerability to flooding using time-series MODIS imageries," *International Journal of Applied Earth Observation and Geoinformation*, vol. 66, pp. 1–13, Apr. 2018, doi: 10.1016/j.jag.2017.10.013.

[31] Y. Khoirurrizqi, R. Sasongko, N. L. E. Utami, A. Irbah, and S. Arjasakusuma, "Machine learning-based rice field mapping in Kulon Progo using a fusion of multispectral and SAR imageries," *Forum Geografi*, vol. 37, no. 2, Dec. 2023, doi: 10.23917/forgeo.v37i2.20304.
[32] Rudiyanto, B. Minasny, R. Shah, N. C. Soh, C. Arif, and B. I. Setiawan, "Automated near-real-time mapping and monitoring of rice extent, cropping patterns, and growth stages in Southeast Asia using Sentinel-1 time series on a Google Earth Engine platform," *Remote Sensing*, vol. 11, no. 14, Jul. 2019, doi: 10.3390/rs11141666.

BIOGRAPHIES OF AUTHORS

Dhimas Tribuana is a doctoral candidate in Management Science at Universitas Komputer Indonesia, Bandung, Indonesia and holds a B.Ec. from Hasanuddin University (2002), an M.M. from STIE Artha Bodhi Iswara Surabaya (2023), and an M.Comp. (Computer Systems) from Handayani University Makassar (2023). He currently serves as a permanent lecturer in Management Informatics at the Akademi Sekretari Manajemen Indonesia Publik. He also works in the private banking sector. His research interests include machine learning, applied networks, internet of things, embedded systems, and management science applications. He has authored articles in accredited national and international journals, as well as book chapters and reference books in data science and computing. He can be contacted at email: d.tribuana@gmail.com.

Usman Sattar He completed the Bachelor of Informatics Engineering studies at the Makassar Islamic University, then won an engineer profession at the Indonesian Muslim University, and continued his masters in Computer System at Makassar Handayani University. At present, he is active as a lecturer at the Bombana Polytechnic, Southeast Sulawesi Province. As an experienced academic, he was involved in various observations and developments, especially those related to technological innovation to support the effectiveness of learning and management of digital systems in educational institutions. He also actively participated in various scientific seminars and forums, which further enrich his insight in the field of educational technology and the integration of artificial intelligence (AI). He can be contacted at email: usmanfkumi@gmail.com.

Baharuddin Mide He holds a bachelor of Computer Science (S.Kom.) master of Public Administration (M.A.P), Computer Systems, master of Computer Science (M.Kom.) has a non-academic degree from Microsoft Microsft Technology Associate (MTA). In addition, there are also several certificates and professional skills. Currently he teaches at Ichsan Sidenreng Rappang University, Sidrap, South Sulawesi. He is a member of the Asean Lecturer Community (ALC). His research interests include computer networks, IoT, and artificial intelligence. He can be contacted at email: baharanthyqu@gmail.com.

Dayanti holds a bachelor of Engineering (S.Kom.) in Informatic Engineering, from STMIK Handayani Makassar, and master of Competency (M.Kom.) in Computer System, from Handayani Universitas Makassar. She currently serves as a full-time lecturer in Informatics Engineering at the Patria Artha University Academy. She is research interests include computational learning, algorithms, and information, spanning various fields such as artificial intelligence (AI), data science, software development, software engineering, database systems, computer networks, and the internet of things. She is authored articles in accredited national and international journals, as well as book chapters and reference books in the fields of data science and computing. She can be contacted at email: dayanti.fattah@gmail.com.