

Ablation-Based Climate-Aware Feature Engineering for Rice Productivity Prediction Using Multi-Source Agricultural Data

R. Hadapiningradja Kusumodestoni^{1*}, Fatchul Arifin², Mutiara Nugraheni³, Desti Setiyowati⁴

Doctoral Program in Engineering Science, Faculty of Engineering, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia^{1, 2, 3}
Faculty of Science and Technology, Universitas Islam Nahdlatul Ulama Jepara, Jepara, Indonesia^{1, 4}

Abstract—Climate-aware feature engineering is important for rice productivity prediction because rainfall effects may be delayed, accumulated, standardized, or expressed as extreme anomalies. This study evaluates the incremental contribution of rainfall-derived features for rice productivity prediction using multi-source agricultural data from districts/cities in Central Java, Indonesia, during 2010-2025. The analysis used 560 district/city-year observations and compared Random Forest and XGBoost under six sequential ablation scenarios: baseline features without engineered climate variables (S0), rainfall lags (S1), rainfall lags plus three-period moving average (S2), addition of a standardized rainfall anomaly index (S3), addition of rainfall anomaly (S4), and full climate-aware features including extreme dry and wet flags (S5). Rice productivity was defined as average harvested paddy yield per hectare at the district/city-year level, expressed in quintals per hectare, while rice production was excluded from the predictor set to prevent direct target leakage. Rainfall-anomaly parameters were estimated from the training period only, and categorical predictors were treated as nominal variables. The results show that climate-aware features improved XGBoost more strongly than Random Forest. For XGBoost, RMSE decreased from 4.8019 in S0 to 4.3068 in S5, representing a 10.31% reduction. The largest observed improvement occurred after adding rainfall lag features in S1, where RMSE decreased to 4.2121, and R2 increased to 0.3053. However, the differences among several top scenarios were small; therefore, scenario rankings are interpreted descriptively rather than as statistically significant superiority. The study demonstrates that ablation-based feature evaluation is necessary before adopting full climate-aware feature sets in agricultural prediction models.

Keywords—Ablation study; climate-aware feature engineering; multi-source agricultural data; random forest; rainfall lag; rice productivity prediction; standardized rainfall anomaly index; XGBoost

I. INTRODUCTION

Rice productivity prediction is a key analytical task for food-security planning, regional agricultural monitoring, and climate-risk mitigation. Rice remains a strategic staple crop in Indonesia, and changes in productivity can influence food availability, farmer income, market supply, and policy decisions. At the district/city level, productivity does not depend on a single factor. It is shaped by interactions among planted area, harvested area, rainfall, temperature, humidity, sunshine duration, wind speed, soil type, soil pH, drainage, irrigation systems, pest conditions, and temporal climate

patterns. Recent climate and agricultural reports also emphasize that food-production systems are increasingly exposed to rainfall variability, drought, floods, and other climate-related disruptions [1], [2]. Therefore, rice productivity prediction requires models that can integrate agricultural, climate, soil-land, agronomic, and temporal information in a valid and interpretable manner.

Machine learning has been widely applied in crop-yield and productivity prediction because it can model non-linear relationships among heterogeneous predictors. Previous studies and reviews have shown that climate indicators, soil properties, management variables, remote-sensing information, and spatial characteristics can improve agricultural prediction when they are properly represented in the feature space [3], [4], [5]. However, strong prediction is not determined only by the algorithm. Feature construction remains essential because agricultural variables often have delayed effects, accumulated effects, and context-dependent interactions. In rice systems, rainfall is especially important because it affects soil moisture, irrigation demand, field preparation, planting timing, nutrient dynamics, pest pressure, and water-management conditions.

Many prediction studies use raw climate variables directly as model inputs. This approach is useful, but it may not fully represent the agronomic meaning of rainfall. Rainfall observed in the current period may not be the only relevant signal because previous rainfall can affect soil-water carry-over, delayed field operations, residual water availability, and crop establishment. Moving averages may represent short-term rainfall tendency, standardized rainfall-anomaly measures can describe dry or wet departures relative to historical conditions, and anomaly variables can capture deviations from expected rainfall patterns. These transformations are consistent with the broader literature on agricultural prediction and climate-risk modelling, which emphasizes that the quality of climate representation can influence model reliability [6], [7], [8].

Tree-based models are suitable for evaluating the value of engineered climate features in tabular agricultural datasets. Random Forest reduces variance through bootstrap aggregation and random feature selection, making it robust for mixed numerical and categorical predictors [9]. XGBoost uses regularized gradient boosting and sequential residual correction, allowing it to exploit informative predictors in structured data [10]. Gradient boosting theory also indicates that staged additive modelling can improve predictive accuracy

*Corresponding author.

when successive learners capture residual patterns left by previous learners [11]. Because these two models are strong tabular learners, they provide a practical basis for testing whether additional climate-aware features produce measurable improvement.

Climate-aware feature engineering should be evaluated empirically rather than assumed to be beneficial. Adding more features can improve model performance when the variables contain complementary information, but it can also introduce redundancy, noise, instability, or overlapping information when extreme-event observations are sparse. This issue is particularly important for rainfall-derived features because rainfall lags, moving averages, anomaly measures, and standardized rainfall indices may partially represent similar rainfall behaviour. Previous studies on crop prediction, smart farming, and climate-aware modelling suggest that data quality, temporal resolution, and feature relevance are often as important as algorithmic complexity [12], [13], [14].

The current research gap is the limited use of ablation-based assessment for climate-aware feature engineering in rice productivity prediction. Many studies report the final set of input variables without showing how each engineered feature group changes model performance. As a result, it remains unclear whether rainfall lags, rainfall moving averages, standardized rainfall anomalies, rainfall anomaly values, and extreme flags contribute equally or whether only some of them provide substantial predictive value. An ablation design can address this problem by adding feature groups step by step and measuring their effect on RMSE, MAE, MAPE, and R^2 [15], [16], [17].

This article is intentionally positioned as a feature-engineering study rather than a model-fusion or explainability article. It is derived from the same broader doctoral research program as companion manuscripts on optimized fusion and explainable AI, but it addresses a different methodological question. The central aim is to evaluate whether climate-aware engineered features improve rice productivity prediction and to identify which feature group contributes the largest observed marginal gain under a fixed temporal evaluation design. The specific objectives are: 1) to construct a leakage-controlled multi-source rice productivity dataset at the district/city-year level; 2) to compare Random Forest and XGBoost across six feature-engineering scenarios; 3) to quantify the incremental contribution of rainfall lags, rainfall moving average, standardized rainfall anomaly, rainfall anomaly, and extreme flags; and 4) to interpret why the full feature set may not always outperform simpler climate-aware scenarios. The expected contribution is practical evidence that rainfall-memory features are valuable, while full climate-aware feature sets should be adopted only after ablation-based validation.

II. RELATED WORK

A. Machine Learning for Crop Productivity Prediction

Machine learning is widely used in crop yield and productivity prediction because it can learn complex relationships from climate, soil, management, and spatial variables. Reviews on crop prediction emphasize that no single algorithm is universally superior; performance depends on data

scale, temporal resolution, feature quality, and the match between the model and the data structure [10], [14], [18]. In structured agricultural datasets, tree-based models are often competitive because they can handle non-linear relationships and interactions among heterogeneous variables.

Random Forest and XGBoost represent two important families of tree-based learning. Random Forest uses bagging and random feature selection to reduce variance [9], [19], [20]. XGBoost uses regularized gradient boosting to improve prediction by fitting residual errors sequentially [10], [14]. In agricultural modelling, these algorithms are useful for analysing district/city-level tabular data where the number of observations may be limited, and predictors come from different institutional sources.

B. Climate-Aware Feature Engineering

Climate-aware feature engineering transforms raw climate observations into variables that better represent agronomic processes. Rainfall lags can represent temporal water-memory effects, moving averages can capture accumulated rainfall tendencies, rainfall anomalies can express departures from historical means, and standardized rainfall anomalies can describe dry or wet conditions relative to a local reference distribution. These engineered variables are particularly relevant for rice because water availability and drainage conditions directly affect field preparation, vegetative growth, and yield formation.

Previous studies often use climate variables as predictors, but they do not always isolate the marginal contribution of each rainfall-derived feature group. This makes it difficult to determine whether improvement is driven by raw rainfall, delayed rainfall memory, standardized rainfall condition, or extreme-event flags. The present study addresses this limitation by using a sequential ablation design that adds climate-aware feature groups one at a time.

C. Ablation Study, Leakage Control, and Evaluation Stability

Ablation analysis is useful for identifying whether a feature group contributes measurable predictive value. However, a sequential ablation design measures order-dependent marginal contribution rather than independent causal importance. Therefore, ablation results should be interpreted as performance changes under a predefined feature-addition sequence, not as definitive evidence that one feature group is universally more important in all modelling contexts [21], [22], [23].

Data leakage is another important issue in predictive modelling. Leakage can occur when target-related information or future-period information is unintentionally made available during model training [24], [25], [26]. In rice productivity modelling, production is problematic because productivity is mathematically derived from production and harvested area. This study therefore excludes production from the predictor set. The role of harvested area is treated more carefully in this revision because it is part of the productivity denominator and may not always be available at the intended prediction time [27], [28], [29], [30].

Model rankings based on a single temporal split should also be interpreted cautiously. Small differences in RMSE may

reflect sampling variability, scenario order, or the specific train-test period. Repeated seeds, rolling-origin validation, or spatial-temporal validation would provide stronger evidence of stability. Because the present dataset version reports one fixed temporal test period, this manuscript avoids overstating small margins and frames scenario rankings as observed results under the specified split [31], [32].

Table I positions the proposed study within related work. It contrasts common research streams in crop prediction, climate-variable modelling, ablation analysis, and leakage-controlled prediction, then identifies the remaining gap addressed by this manuscript.

TABLE I. POSITION OF THE PROPOSED STUDY WITHIN RELATED WORK

Research stream	Typical contribution	Remaining gap addressed in this study
Tree-based crop prediction	Captures non-linear tabular relationships in agricultural data	Often reports final performance without isolating rainfall-derived feature contributions
Climate-variable modelling	Uses rainfall, temperature, humidity, or drought indicators as predictors	Does not always distinguish raw rainfall, rainfall memory, anomaly, and standardized anomaly effects
Ablation analysis	Evaluates incremental effects of feature groups or model components	Needs cautious interpretation when differences are small and based on one temporal split
Leakage-controlled prediction	Avoids target-related and future-period information in model training	Requires explicit distinction between mathematical leakage and prediction-time availability

Note: The table summarizes how the present study differs from prior crop-prediction and feature-engineering research. The contribution is focused on ablation-based feature assessment rather than developing a new fusion architecture or SHAP interpretation framework.

III. METHODOLOGY

A. Study Area and Dataset

The study used district/city-level agricultural and environmental data from Central Java Province, Indonesia, covering 2010-2025. The research site is in Central Java, which lies approximately between 5.67 degrees S and 8.50 degrees S and 108.50 degrees E-111.50 degrees E, with an approximate provincial centroid at 7.15 degrees S and 110.14 degrees E. The final dataset contained 560 district/city-year observations. Rice productivity was defined as average harvested paddy yield per unit harvested area for each district/city and year, expressed in quintals per hectare (q ha-1).

Agricultural statistics were compiled from BPS-Statistics Indonesia, BPS Province of Central Java, and relevant

provincial or district agricultural records. Climate indicators were compiled from meteorological records provided by the Meteorology, Climatology and Geophysical Agency (BMKG) or equivalent station/gridded climate products. The use of BPS and BMKG records is important because these institutions provide official statistical and climate information for Indonesia. Soil-land and agronomic variables were compiled from institutional land-information and agricultural-agency records. All variables were harmonized to a district/city-year unit of analysis before modelling.

Table II presents the main components of the multi-source dataset. The table clarifies how agricultural, climate, soil-land, agronomic, and temporal rainfall features contribute to the ablation experiment before model training.

TABLE II. MAIN COMPONENTS OF THE MULTI-SOURCE DATASET

Data component	Variables or examples	Role in the ablation study
Agricultural data	Planted area, harvested area, productivity	Defines land-use scale and target construction
Climate data	Rainfall, humidity, temperature, sunshine duration, wind speed, extreme rainfall	Represents environmental pressure and climate variability
Soil and land data	Soil type, soil texture, soil pH, soil moisture, drainage type	Represents land suitability and water-management capacity
Agronomic data	Irrigation system and dominant pest type	Represents management and disturbance conditions
Temporal rainfall features	Rain_Lag_1, Rain_Lag_2, Rain_Lag_3, Rain_MA_3, SRAI, Rain Anomaly and extreme flags	Represents delayed, accumulated, standardized, and abnormal rainfall effects

Note: BPS = Statistics Indonesia; BMKG = Meteorology, Climatology and Geophysical Agency; SRAI = standardized rainfall anomaly index; MA = moving average; Rain_MA_3 = three-period rainfall moving average.

B. Data Source, Preprocessing, and Leakage Control

The unit of analysis was district/city-year. Daily or monthly climate observations were aggregated into annual district/city indicators. When source data had different spatial resolutions, observations were matched using administrative identifiers, year labels, and variable definitions. Quality assurance included duplicate-record checking, unit consistency checking, missing-value inspection, outlier screening, and comparison with official annual trends. The final modelling dataset did not contain missing values.

Production was removed from the input features because rice productivity is mathematically derived from production and harvested area. This step prevents direct target leakage. Harvested area was retained as a land-use scale and prediction-time administrative variable, but its interpretation was revised carefully. Because harvested area is also part of the productivity denominator, the model should not be interpreted as a pure pre-season yield forecast unless harvested area is available before prediction. For stricter pre-season deployment, a sensitivity experiment excluding harvested area should be reported using the same ablation protocol.

TABLE III. PREPROCESSING, LEAKAGE-CONTROL, AND IMPLEMENTATION SAFEGUARDS

Reviewer concern	Revision in methodology	Interpretation safeguard
Target leakage	Production excluded from predictors	Models learn from explanatory variables rather than direct target information
Harvested area availability	Harvested area treated as prediction-time administrative information	For pre-season deployment, a no-harvested-area sensitivity run should be added
Nominal categorical variables	One-hot encoding specified for region, soil, drainage, irrigation, and pest variables	Avoids artificial ordinal relationships from numeric category codes
Rainfall anomaly parameters	Mean and standard deviation estimated from training-period rainfall only	Avoids use of future test-period information

Note: The table summarizes methodological clarifications added in response to concerns about target leakage, categorical encoding, prediction-time availability, and rainfall reference-period estimation.

Categorical variables such as region, soil type, soil texture, soil moisture, drainage type, irrigation system, and dominant pest type were treated as nominal variables. The revised preprocessing protocol specifies one-hot encoding for nominal variables so that tree-based models do not receive arbitrary numerical codes as if they represented ordered categories. Numerical coding, where present in raw files, was used only as an internal label before conversion to indicator variables.

Table III summarizes the methodological safeguards applied during preprocessing and implementation. The table is included to clarify how the study addresses target leakage, harvested-area availability, nominal categorical variables, and training-period rainfall reference parameters.

C. Climate-Aware Feature Engineering

Feature engineering was conducted to represent delayed and accumulated rainfall effects. `Rain_Lag_1`, `Rain_Lag_2`, and `Rain_Lag_3` represented rainfall from previous periods. `Rain_MA_3` represented a three-period moving average of rainfall. `Rain_Anomaly` represented the deviation of rainfall from the training-period mean. The previously labelled SPI-style feature is reported in this revision as a standardized rainfall anomaly index (SRAI), because it is calculated as a rainfall z-score and does not implement the full conventional SPI procedure involving a specified accumulation scale, distribution fitting, cumulative probability transformation, and standard-normal transformation.

Rainfall anomaly and SRAI were calculated using reference parameters estimated only from the training period. Specifically, the district-specific training-period mean rainfall and standard deviation were used to transform both training and test observations. This prevents the 2023-2025 test-period

rainfall distribution from influencing feature construction. The extreme dry and wet flags were then derived from the same training-period reference logic, so their thresholds remained fixed when applied to future test observations.

D. Ablation Scenarios for Climate-Aware Feature Engineering

Six feature-engineering scenarios were constructed to isolate the incremental contribution of climate-aware variables. Scenario S0 used base features without engineered climate-aware variables. Scenario S1 added rainfall lag variables. Scenario S2 added rainfall moving average. Scenario S3 added SRAI. Scenario S4 added rainfall anomaly. Scenario S5 added extreme dry and wet flags, representing the full climate-aware feature set.

The ablation design is sequential. Therefore, the reported contribution of each feature group is conditional on its position in the sequence. The design should not be interpreted as independent feature importance or causal attribution. Instead, it measures how model performance changes when each feature group is added after the preceding groups under an identical temporal split and fixed model configuration.

Table IV defines the six ablation scenarios used in the experiment. The table shows the sequential addition of rainfall lags, rainfall moving average, standardized rainfall anomaly index, rainfall anomaly, and extreme-condition flags.

Fig. 1 illustrates the overall ablation workflow for climate-aware rainfall features. The diagram shows how the experiment starts from the base feature set and gradually adds engineered rainfall variables before evaluating Random Forest and XGBoost under the same independent temporal test period.

TABLE IV. ABLATION SCENARIOS FOR CLIMATE-AWARE FEATURE ENGINEERING

Scenario	Feature set	Purpose
S0	Base features without climate-aware engineered variables	Baseline comparison
S1	S0 + <code>Rain_Lag_1</code> , <code>Rain_Lag_2</code> and <code>Rain_Lag_3</code>	Evaluate rainfall-memory effects
S2	S1 + <code>Rain_MA_3</code>	Evaluate smoothed rainfall tendency
S3	S2 + SRAI	Evaluate standardized rainfall departure using a training-period reference
S4	S3 + <code>Rain_Anomaly</code>	Evaluate deviation from training-period rainfall mean
S5	S4 + <code>Extreme_Dry_Flag</code> and <code>Extreme_Wet_Flag</code>	Evaluate binary extreme-condition signals

Note: S0-S5 = sequential ablation scenarios; SRAI = standardized rainfall anomaly index; `Rain_MA_3` = three-period rainfall moving average.

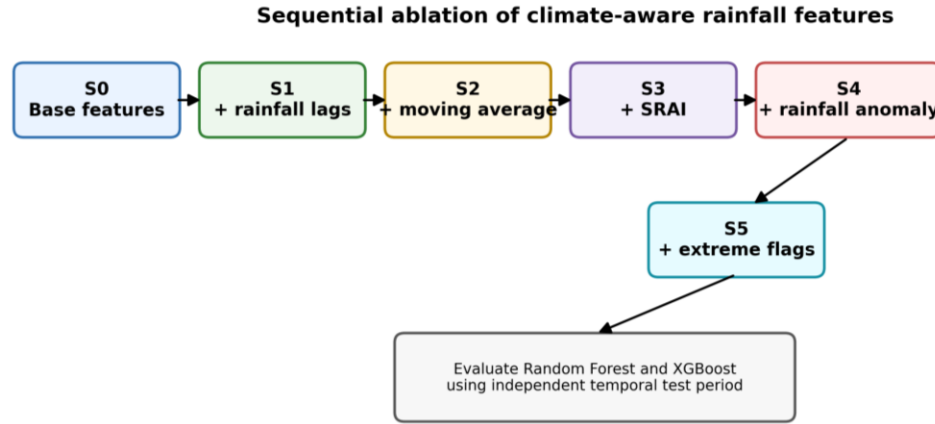


Fig. 1. Ablation design for climate-aware feature engineering.

E. Models and Hyperparameter Settings

Random Forest and XGBoost were selected because both models are strong tabular learners and are suitable for assessing the contribution of additional feature groups. Random Forest was used as a variance-reduction learner for non-linear tabular patterns [9], [33]. XGBoost was used as a regularized boosting learner for structured-data prediction [11], [34]. Each model was trained and evaluated under all six scenarios using the same temporal split so that differences in performance could be attributed primarily to feature-set changes rather than changes in validation design.

Data from 2010-2022 were used for model training, and data from 2023-2025 were used for independent temporal testing. Hyperparameters were kept stable across ablation scenarios to focus the comparison on feature contribution rather than repeated model re-tuning. This design supports a fair ablation comparison, but it does not replace repeated temporal validation or statistical significance testing across multiple splits [19], [20], [35], [36].

Table V reports the model configuration used for all ablation scenarios. The table confirms that Random Forest, XGBoost, temporal validation, and categorical encoding settings were kept consistent so that performance changes mainly reflect feature-set changes.

TABLE V. MODEL CONFIGURATION FOR THE ABLATION EXPERIMENT

Model or process	Configuration used for reporting
Random Forest	n_estimators = 500; max_depth = None; min_samples_split = 2; min_samples_leaf = 1; max_features = sqrt; bootstrap = True; random_state = 42. Purpose: robust tabular baseline and comparison learner.
XGBoost	objective = reg:squarederror; n_estimators = 300; max_depth = 3; learning_rate = 0.05; subsample = 0.80; colsample_bytree = 0.80; reg_lambda = 1.0; random_state = 42. Purpose: regularized boosting learner for structured agricultural data.
Validation design	Training period = 2010-2022; test period = 2023-2025; identical split applied across S0-S5. Purpose: temporal evaluation without future-information leakage.
Categorical encoding	Nominal predictors converted into one-hot indicators before model fitting. Purpose: avoid artificial ordinal relations from numeric category codes.

Note: XGBoost = Extreme Gradient Boosting; sqrt = square root; reg:squarederror = regression with squared-error objective.

F. Mathematical Formulation and Evaluation Metrics

To strengthen reproducibility, this subsection presents the equations used for target construction, climate-aware feature engineering, and performance evaluation. Let d denote district/city, t denote year, i denote a district/city-year observation, y_i denote observed productivity, and $\hat{y}_i$ denote predicted productivity. Eq. (1)-(10) are cited sequentially according to their function.

Productivity is calculated as harvested paddy yield per hectare at the district/city-year level as shown in Eq. (1). Production was excluded from the input features because it is a direct component of the target construction.

$$\text{Productivity}_{\{d,t\}} = (\text{Paddy production}_{\{d,t\}} / \text{Harvested area}_{\{d,t\}}) \times 10 \quad (1)$$

$$\text{Rain_Lag}_{\{k,d,t\}} = \text{Rainfall}_{\{d,t-k\}}, \text{ where } k = 1, 2, 3 \quad (2)$$

$$\text{Rain_MA_3}_{\{d,t\}} = (\text{Rainfall}_{\{d,t\}} + \text{Rainfall}_{\{d,t-1\}} + \text{Rainfall}_{\{d,t-2\}}) / 3 \quad (3)$$

$$\text{Rain_Anomaly}_{\{d,t\}} = \text{Rainfall}_{\{d,t\}} - \mu_{\{d,\text{train}\}} \quad (4)$$

$$\text{SRAI}_{\{d,t\}} = (\text{Rainfall}_{\{d,t\}} - \mu_{\{d,\text{train}\}}) / \sigma_{\{d,\text{train}\}} \quad (5)$$

$$\text{RMSE} = \sqrt{(1/n) \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (6)$$

$$\text{MAE} = (1/n) \sum_{i=1}^n |y_i - \hat{y}_i| \quad (7)$$

$$\text{MAPE} = (100/n) \sum_{i=1}^n |(y_i - \hat{y}_i) / y_i| \quad (8)$$

$$R^2 = 1 - [\sum_{i=1}^n (y_i - \hat{y}_i)^2 / \sum_{i=1}^n (y_i - \bar{y})^2] \quad (9)$$

$$\text{RMSE reduction (\%)} = [(\text{RMSE}_{S0} - \text{RMSE}_{Sj}) / \text{RMSE}_{S0}] \times 100 \quad (10)$$

Eq. (2)-(5) define the engineered rainfall features. Eq. (6)-(9) define RMSE, MAE, MAPE, and R^2 . Eq. (10) calculates the percentage reduction in RMSE relative to the baseline scenario S0. The terms $\mu_{\{d,train\}}$ and $\sigma_{\{d,train\}}$ represent district-specific rainfall mean and standard deviation estimated exclusively from the training period.

IV. RESULTS AND DISCUSSION

A. Overall Ablation Results

The ablation results show that climate-aware feature engineering has different effects on Random Forest and XGBoost. For Random Forest, RMSE decreased from 4.4466 in S0 to 4.3732 in S5, equivalent to a 1.65% reduction. The lowest observed Random Forest RMSE occurred in S4 after adding rainfall anomaly, with RMSE of 4.3578. This suggests that Random Forest benefited from climate-aware variables, but the gain was modest and the performance curve remained relatively flat across scenarios.

For XGBoost, the effect was stronger. The base scenario produced an RMSE of 4.8019 and a MAPE of 6.0520%. After adding rainfall lag features in S1, RMSE decreased to

4.2121, MAPE decreased to 5.3318%, and R^2 increased to 0.3053. This was the largest observed marginal gain across all scenarios. The lowest observed XGBoost RMSE occurred in S1, whereas the lowest observed XGBoost MAPE occurred in S3 after adding SRAI. These results indicate that rainfall-memory features are the most important engineered climate feature group in the present temporal split.

Table VI reports the complete ablation results for Random Forest and XGBoost. The table allows direct comparison of RMSE, MAE, MAPE, and R^2 across S0-S5, while also showing that the full climate-aware feature set is not automatically the best observed scenario.

Fig. 2 visualizes RMSE across the six climate-aware feature-engineering scenarios. The figure highlights the sharp XGBoost improvement after adding rainfall lags and the relatively flat Random Forest curve across subsequent feature additions.

Fig. 3 presents MAPE across the same ablation scenarios. The figure shows that XGBoost obtains its lowest observed relative error after SRAI is added, while Random Forest remains comparatively stable across climate-aware feature sets.

TABLE VI. ABLATION RESULTS ACROSS CLIMATE-AWARE FEATURE-ENGINEERING SCENARIOS

Scenario	Model	Features	RMSE	MAE	MAPE (%)	R2
S0 Base features	Random Forest	17	4.4466	3.3885	5.9629	0.2259
S0 Base features	XGBoost	17	4.8019	3.5047	6.0520	0.0972
S1 Base + rain lags	Random Forest	20	4.3690	3.3267	5.8614	0.2526
S1 Base + rain lags	XGBoost	20	4.2121	3.0644	5.3318	0.3053
S2 S1 + Rain_MA_3	Random Forest	21	4.3699	3.3411	5.8996	0.2523
S2 S1 + Rain_MA_3	XGBoost	21	4.2942	3.1249	5.4375	0.2780
S3 S2 + SRAI	Random Forest	22	4.3692	3.3527	5.9190	0.2526
S3 S2 + SRAI	XGBoost	22	4.2310	3.0563	5.3250	0.2991
S4 S3 + anomaly	Random Forest	23	4.3578	3.3399	5.8981	0.2565
S4 S3 + anomaly	XGBoost	23	4.4175	3.1873	5.5101	0.2359
S5 Full climate-aware	Random Forest	25	4.3732	3.3530	5.9226	0.2512
S5 Full climate-aware	XGBoost	25	4.3068	3.2007	5.5767	0.2737

Note: RMSE = root mean squared error; MAE = mean absolute error; MAPE = mean absolute percentage error; R^2 = coefficient of determination. Values are reported to four decimal places to avoid overstating precision from a single temporal split.

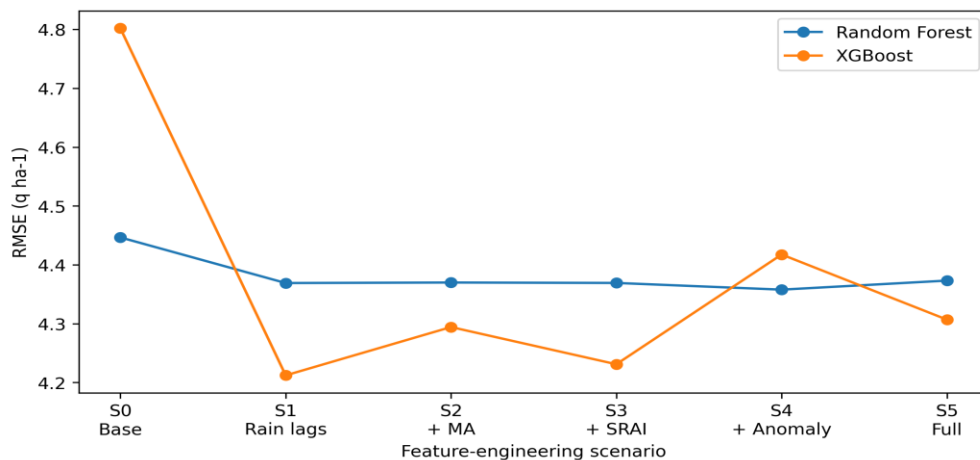


Fig. 2. RMSE across climate-aware feature-engineering scenarios.

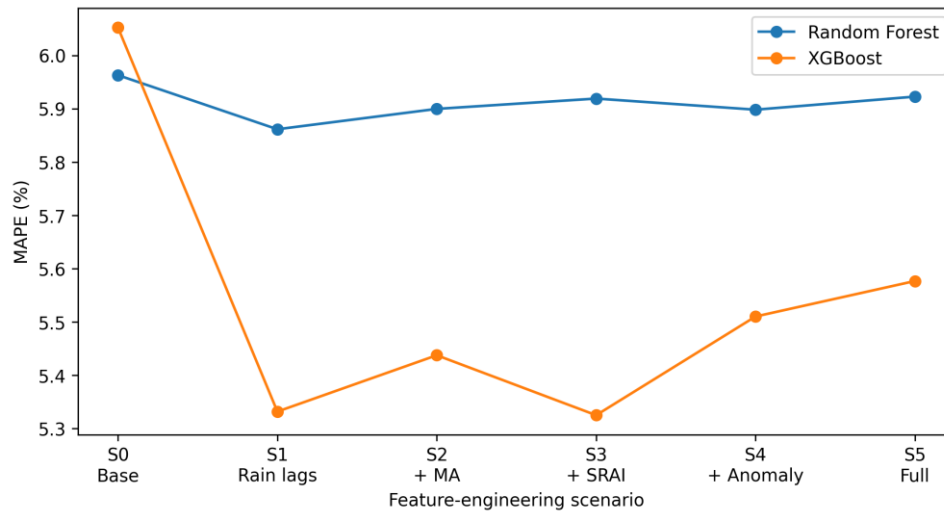


Fig. 3. MAPE across climate-aware feature-engineering scenarios.

B. Improvement from Baseline to Full Climate-Aware Features

Table VII compares the baseline scenario with the full climate-aware scenario. XGBoost achieved an RMSE reduction of 0.4951, corresponding to a 10.31% improvement. Random Forest achieved a smaller RMSE reduction of 0.0734, corresponding to a 1.65% improvement. This difference indicates that XGBoost is more responsive to engineered climate features than Random Forest in the present dataset.

The improvement should be interpreted carefully. Although S5 improved substantially over S0 for XGBoost, S5 did not outperform S1 in RMSE or S3 in MAPE. Therefore, the full feature set provides evidence that climate-aware information is

useful, but it does not prove that every engineered feature group contributes independent predictive value.

Table VII compares the baseline feature set with the full climate-aware feature set. This table quantifies the RMSE reduction from S0 to S5 and demonstrates that XGBoost is more responsive to engineered rainfall features than Random Forest in the present dataset.

Fig. 4 shows incremental RMSE reduction relative to the S0 baseline. The figure helps explain that the largest observed gain occurs immediately after adding rainfall lag variables, whereas later additions produce smaller or less stable improvements.

TABLE VII. RMSE REDUCTION FROM BASELINE TO FULL CLIMATE-AWARE FEATURE SET

Model	RMSE S0	RMSE S5	RMSE reduction	RMSE reduction (%)
Random Forest	4.4466	4.3732	0.0734	1.65
XGBoost	4.8019	4.3068	0.4951	10.31

Note: S0 = base features without climate-aware engineered variables; S5 = full climate-aware feature set; RMSE reduction was calculated using Eq. (10).

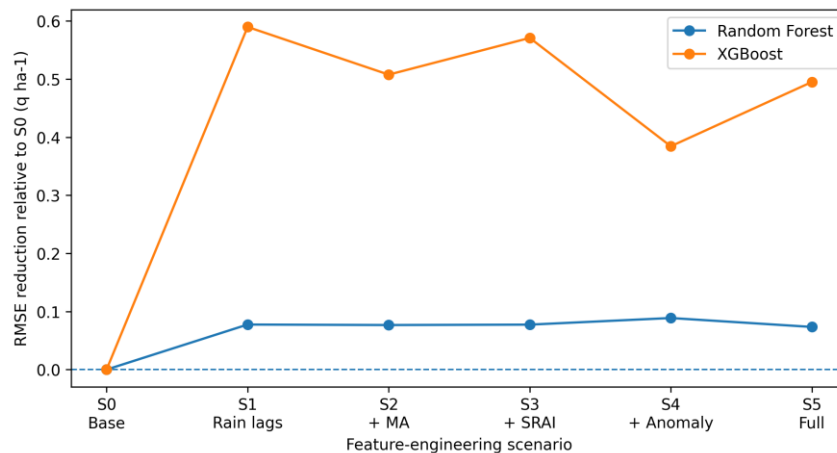


Fig. 4. Incremental RMSE reduction relative to the baseline scenario.

C. Best-Observed Scenarios and Stability Caution

The best scenario depends on both the model and the metric. For Random Forest, S4 produced the lowest observed RMSE, suggesting that rainfall anomaly provided a small additional benefit. For XGBoost, S1 produced the lowest observed RMSE, indicating that rainfall lag variables captured the strongest predictive signal. For MAPE, the best observed XGBoost result occurred in S3, suggesting that SRAI improved relative error even when the RMSE difference was small.

However, several observed differences are small. For Random Forest, S1, S2, S3, and S5 are within approximately

0.0154 RMSE of the lowest observed S4 value. For XGBoost, S3 is within approximately 0.0189 RMSE of S1. These gaps should not be interpreted as statistically significant evidence because the present experiment uses one fixed temporal split. The results therefore identify promising feature sets rather than definitive universal rankings.

Table VIII summarizes the best-observed scenarios by metric under the fixed temporal split. The table is interpreted cautiously because the scenario differences were not validated through repeated seeds, rolling-origin evaluation, or statistical significance testing.

TABLE VIII. BEST-OBSERVED SCENARIOS BY METRIC UNDER THE FIXED TEMPORAL SPLIT

Criterion	Model	Scenario	RMSE	MAE	MAPE (%)	R2	Interpretation
Lowest observed RMSE	Random Forest	S4 S3 + anomaly	4.3578	3.3399	5.8981	0.2565	Very small margin over other RF climate-aware scenarios
Lowest observed RMSE	XGBoost	S1 Base + rain lags	4.2121	3.0644	5.3318	0.3053	Strongest observed rainfall-memory gain
Lowest observed MAPE	Random Forest	S1 Base + rain lags	4.3690	3.3267	5.8614	0.2526	Observed relative-error advantage
Lowest observed MAPE	XGBoost	S3 S2 + SRAI	4.2310	3.0563	5.3250	0.2991	Small relative-error improvement over S1

Note: The term best-observed is used deliberately because repeated seeds, rolling-origin validation, and statistical significance tests were not available in the present dataset version.

D. Explanatory Power and Prediction-Time Availability

The best R^2 value across all scenarios was 0.3053 for XGBoost under S1. This means that the model explained only part of the variance in rice productivity, while most variation remained outside the available predictors. The result should not be interpreted as model failure; rather, it indicates that district/city-level rice productivity is affected by factors that were not fully represented in the dataset, such as rice variety, fertilizer dose and timing, irrigation volume, mechanization, labour availability, planting calendar, pest intensity, field-level management, and satellite-derived vegetation indices.

This explanatory ceiling also changes the interpretation of RMSE reduction. A 10.31% reduction from S0 to S5 for XGBoost indicates that engineered rainfall information improves prediction error relative to the baseline, but it does not imply that the model captures all determinants of rice productivity. Therefore, the practical contribution of this study

is a feature-engineering assessment, not a claim of complete productivity explanation.

Harvested area also requires careful interpretation. Because productivity is calculated from production divided by harvested area, harvested area is mathematically related to the target construction. In this manuscript, production is excluded to avoid direct target leakage, while harvested area is retained only when it is available at prediction time as an administrative land-use variable. For strict ex-ante forecasting before harvested area is known, the same S0-S5 ablation should be repeated without harvested area. Until that sensitivity analysis is added, conclusions should be interpreted as conditional on the included prediction-time variables.

Table IX provides interpretation safeguards for the main reviewer concerns. It summarizes how the manuscript treats the R^2 ceiling, small RMSE margins, harvested-area availability, and sparse extreme flags to avoid overclaiming from a single temporal split.

TABLE IX. INTERPRETATION SAFEGUARDS FOR REVIEWER CONCERNS

Issue	Current treatment in this manuscript	Recommended interpretation
R^2 ceiling	Best observed $R^2 = 0.3053$ for XGBoost S1	Model explains part of district/city-year productivity variation; omitted variables remain important
Small RMSE margins	Several climate-aware scenarios are separated by less than 0.10 RMSE	Use descriptive rather than statistically definitive ranking
Harvested area	Retained as prediction-time administrative land-use information	Repeat ablation without harvested area for strict pre-season forecasting
Extreme flags	Included only in S5 after anomaly features	Interpret cautiously because extreme observations are sparse

Note: These safeguards prevent overclaiming from a single temporal split and clarify the distinction between target-leakage prevention and prediction-time variable availability.

E. Interpretation of Feature-Engineering Contribution

The ablation results indicate that not all engineered climate features contribute equally. Rainfall lag features provided the most consistent and substantial improvement, particularly for XGBoost. This is agronomically meaningful because rice

productivity can be influenced by previous rainfall conditions through soil moisture, irrigation availability, field preparation, planting timing, and pest dynamics. The strong effect of S1 suggests that rainfall memory is a core climate signal for district/city-year productivity prediction.

Adding Rain_MA_3 did not improve XGBoost beyond S1, indicating that rainfall lag variables already captured much of the useful short-term rainfall information. Adding SRAI in S3 slightly improved MAPE compared with S1, suggesting that standardized rainfall status is useful for relative error control. However, adding rainfall anomaly and extreme flags did not further improve XGBoost performance and increased error relative to S1 and S3. This may be caused by redundancy among rainfall-derived variables, the small number of extreme observations, or additional feature noise.

For Random Forest, the performance curve was relatively flat across scenarios. This suggests that Random Forest was less sensitive to incremental climate-aware features. XGBoost, by contrast, showed a sharper response to rainfall lags and SRAI, which is consistent with the ability of boosting models to exploit informative residual patterns from engineered predictors. These findings support a practical recommendation: climate-aware feature engineering should be adopted selectively, and rainfall-memory variables should be prioritized before more complex climate indices are added.

F. Discussion and Relationship to the Broader Research Program

The main finding of this article is that climate-aware feature engineering improves rice productivity prediction, but the improvement is not monotonic. The strongest observed gain comes from adding rainfall lag features, while the full feature set does not necessarily produce the best performance. This finding challenges the assumption that adding more engineered climate variables always improves model accuracy. In practical modelling workflows, additional variables should be evaluated not only by theoretical relevance but also by their empirical contribution under temporal testing.

This study is complementary to broader work on rice productivity prediction conducted within the same doctoral research program, but it has a distinct scope. The optimized-fusion manuscript focuses on combining RF, XGBoost, and LSTM outputs. The explainability manuscript focuses on interpreting model drivers using SHAP. The present manuscript focuses on upstream climate-aware feature engineering and asks which rainfall-derived feature groups improve tabular models before any final fusion or detailed explanation stage. This distinction reduces overlap and clarifies the manuscript contribution.

Several limitations should be noted. First, the dataset uses annual district/city-level observations, which may not fully represent intra-seasonal rainfall timing, rice phenology, planting-calendar variation, or field-level management. Second, the number of extreme observations is limited, so extreme flags may not provide stable incremental information. Third, the study uses a single temporal test split; repeated seeds, rolling-origin evaluation, or spatiotemporal cross-validation would provide stronger evidence for scenario stability. Fourth, the no-harvested-area sensitivity analysis should be added before claiming strict pre-season applicability. Fifth, model hyperparameters were kept stable across scenarios to focus on feature contribution; alternative tuning strategies may produce different results.

V. CONCLUSION

This study evaluated climate-aware feature engineering for rice productivity prediction using a sequential ablation design and multi-source agricultural data from Central Java, Indonesia. Six scenarios were tested, ranging from base features to a full climate-aware feature set containing rainfall lags, rainfall moving average, standardized rainfall anomaly index, rainfall anomaly, and extreme dry/wet flags. Random Forest and XGBoost were used as tabular learners to isolate the contribution of feature groups under the same temporal testing design.

The results show that XGBoost benefited substantially from climate-aware feature engineering. Full climate-aware features reduced XGBoost RMSE by 10.31% compared with the base feature set. However, the lowest observed XGBoost RMSE was achieved by S1, which added rainfall lag variables, not by the full S5 feature set. This indicates that rainfall lag features provide the largest observed marginal contribution, while adding all engineered features can introduce redundancy, sparse-signal instability, or noise. Random Forest showed smaller improvement, with only 1.65% RMSE reduction from S0 to S5.

The study also emphasizes methodological caution. The best observed R^2 was 0.3053, indicating that much productivity variation remains unexplained by the available predictors. Several differences among scenarios were small and should not be interpreted as statistically significant without repeated evaluation. Harvested area was retained only as a prediction-time administrative variable, and a sensitivity analysis without harvested area is recommended for strict pre-season forecasting. Future research should test the same feature-engineering scheme using monthly or seasonal rainfall data, alternative standardized-rainfall anomaly time scales, spatial-temporal validation, remote-sensing indicators, richer agronomic variables, and repeated-split statistical testing.

ACKNOWLEDGMENT

The authors thank Universitas Negeri Yogyakarta and the institutions providing agricultural and climate data used in this research. The authors also acknowledge the academic supervision and methodological feedback received during the dissertation development process.

DECLARATIONS

AUTHORS' CONTRIBUTIONS

Hadapiningradja Kusumodestoni Contributed to Conceptualization, Data Curation, Methodology, Analysis, Visualization, and Original Draft Preparation. Fatchul Arifin Contributed to Supervision, Methodology Validation, and Manuscript Review. Mutiara Nugraheni Contributed to Supervision, Interpretation, and Manuscript Review. Desti Setiyowati Contributed to Data Review and Manuscript Preparation Support. All Authors Read and Approved the Final Manuscript.

CONFLICT OF INTERESTS

The authors declare no conflict of interest.

DATA AVAILABILITY

The processed dataset, feature-engineering scenario table, model-output files, and Python scripts used in this study are available from the corresponding author upon reasonable request. Public release of the complete dataset may depend on data-sharing permissions from the original data providers.

Declaration of Artificial Intelligence Assistance: The authors used AI-assisted language support to improve grammar, formatting consistency, and editorial clarity during revision. No AI tool was used to create the research data, conduct the statistical analysis, or make conceptual decisions. All data analysis, interpretation, and final manuscript responsibility remain with the authors.

Ethics Approval: This study used secondary agricultural, climate, soil, and agronomic data and did not involve human participants, patient information, or animal experiments. Therefore, human or animal ethics approval was not applicable.

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

- [1] ADDIN Mendeley Bibliography CSL BIBLIOGRAPHY [1] World Meteorological Organization, State of the Global Climate 2024. Geneva, Switzerland: World Meteorological Organization, 2025, WMO No. 1368. [Online]. Available: https://wmo.int/sites/default/files/2025/03/WMO-1368-2024_en.pdf.
- [2] BPS-Statistics Indonesia, Statistical Yearbook of Indonesia 2025. Jakarta, Indonesia: BPS-Statistics Indonesia, 2025, Catalogue No. 1101001, Publication No. 03200.25004. [Online]. Available: <https://www.bps.go.id/en/publication/2025/02/28/8cfe1a589ad3693396d3db9f/statistical-yearbook-of-indonesia-2025.html>.
- [3] M. Rashid, B. S. Bari, Y. Yusup, M. A. Kamaruddin, and N. Khan, "A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches With Special Emphasis on Palm Oil Yield Prediction," vol. 9, 2021, doi: 10.1109/ACCESS.2021.3075159.
- [4] U. Shafi *et al.*, "Tackling Food Insecurity Using Remote Sensing and Machine Learning-Based Crop Yield Prediction," *IEEE Access*, vol. 11, no. October, pp. 108640–108657, 2023, doi: 10.1109/ACCESS.2023.3321020.
- [5] J. Lu *et al.*, "Deep Learning for Multi-Source Data-Driven Crop Yield Prediction in Northeast China," 2024.
- [6] C. Lin, Z. Yan, and S. Miao, "Identifying Nonlinear Thresholds and Interaction Dominance of Meteorological Drivers on Rice Yield: A SHAP-Based Approach," pp. 1–30, 2026.
- [7] K. G. Liakos, P. Busato, D. Moshou, and S. Pearson, "Machine Learning in Agriculture : A Review," no. MI, pp. 1–29, doi: 10.3390/s18082674.
- [8] T. Van Klompenburg, A. Kassahun, and C. Catal, "Crop yield prediction using machine learning: A systematic literature review," *Comput. Electron. Agric.*, vol. 177, no. January, p. 105709, 2020, doi: 10.1016/j.compag.2020.105709.
- [9] L. E. O. Breiman, "Random Forests," pp. 5–32, 2001.
- [10] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," pp. 785–794, 2016.
- [11] V. Nguyen, C. Le, M. T. Bui, T. Nguyen, and T. Le, "Ecological Informatics Multi-modal representation for UAV-based rice yield prediction: Understanding temporal redundancy and feature sufficiency," *Ecol. Inform.*, vol. 98, no. July, p. 103957, 2026, doi: 10.1016/j.ecoinf.2026.103957.
- [12] H. Demirhan, "Information Processing in Agriculture: A deep learning framework for prediction of crop yield in Australia under the impact of climate change," *Inf. Process. Agric.*, vol. 12, no. 1, pp. 125–138, 2025, doi: 10.1016/j.inpa.2024.04.004.
- [13] S. Wolfert, L. Ge, C. Verdouw, and M. Bogaardt, "Big Data in Smart Farming – A review," *Agric. Syst.*, vol. 153, pp. 69–80, 2017, doi: 10.1016/j.agry.2017.01.023.
- [14] M. Waqas, S. Ahmad, A. Wangwongchai, and P. Dechpichai, "Quantification and prediction of the impact of ENSO on rainfed rice yields in Thailand," *Environ. Challenges*, vol. 19, no. January, p. 101123, 2025, doi: 10.1016/j.envc.2025.101123.
- [15] M. A. Javed and M. A. Azmi 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, p. e40836, 2024, doi: 10.1016/j.heliyon.2024.e40836.
- [16] G. Gupta and S. Kumar, "Applications of AI in precision agriculture," *Discov. Agric.*, no. November 2024, 2025, doi: 10.1007/s44279-025-00220-9.
- [17] H. Li, J. Gao, Y. Guo, and X. G. Yuan, "Application of XGBoost model and multi-source data for winter wheat yield prediction in Henan Province of China," no. March, pp. 29–47, 2025.
- [18] R. Hadapiningradja Kusumodestoni, F. Arifin and M. Nugraheni, "Optimized Weighted Fusion for Rice Productivity Prediction Using Multi-source Agricultural Data," vol. 7, no. July, pp. 1530–1546, 2026, doi: 10.47857/irjms.2026.v07i03.012859.
- [19] R. Hadapiningradja Kusumodestoni, F. Arifin and M. Nugraheni, "Comparative analysis of random forest and XGBoost for rice yield estimation," vol. 020002, no. November 2025, 2026.
- [20] Y. Prediction, "A Temporal – Geospatial Deep Learning Framework for Crop Yield Prediction," 2024.
- [21] J. Brinkhoff, A. Clarke, B. W. Dunn, and M. Groat, "Agricultural and Forest Meteorology Analysis and forecasting of Australian rice yield using phenology-based aggregation of satellite and weather data," vol. 353, no. May, 2024.
- [22] Y. Liu, C. Ao, W. Zeng, Z. Li, Y. Zhang, and J. Ishchanov, "Assimilation of remote sensing data and crop model simulations for improved rice yield prediction under controlled irrigation and extreme precipitation in Northeast China," vol. 333, no. June, 2026.
- [23] T. Balasubramaniam, W. A. Mohotti, K. Sabir, and R. Nayak, "Ecological Informatics Feature engineering on climate data with machine learning to understand time-lagging effects in pasture yield prediction," *Ecol. Inform.*, vol. 86, no. September 2024, p. 103011, 2025, doi: 10.1016/j.ecoinf.2025.103011.
- [24] S. Kaufman, S. Rosset, C. Perlich, and O. R. I. Stitelman, "Leakage in Data Mining : Formulation , Detection , and Avoidance," vol. 6, no. 4, pp. 1–21, 2012, doi: 10.1145/2382577.2382579.
- [25] P. Probst, "Tunability : Importance of Hyperparameters of Machine Learning Algorithms," vol. 20, pp. 1–32, 2019.
- [26] Y. Li, H. Zeng, M. Zhang, B. Wu, Y. Zhao, and X. Yao, "International Journal of Applied Earth Observation and Geoinformation A county-level soybean yield prediction framework coupled with XGBoost and multidimensional feature engineering," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 118, no. August 2022, p. 103269, 2023, doi: 10.1016/j.jag.2023.103269.
- [27] L. Fan, S. Fang, J. Fan, Y. Wang, L. Zhan, and Y. He, "Rice Yield Estimation Using Machine Learning and Feature Selection in Hilly and Mountainous Chongqing , China," 2024.
- [28] M. N. Id, A. Hossain, C. Id, F. Arefeen, and M. Id, "A tree based eXtreme Gradient Boosting (XGBoost) machine learning model to forecast the annual rice production in Bangladesh," pp. 1–15, 2023, doi: 10.1371/journal.pone.0283452.
- [29] J. W. Choi *et al.*, "Recent Trends in Machine Learning , Deep Learning , Ensemble Learning , and Explainable Artificial Intelligence Techniques for Evaluating Crop Yields Under Abnormal Climate Conditions," pp. 1–26, 2025.
- [30] R. Zeng, J. Li, Z. Li, and Q. Zhang, "Crossformer-Based Model for Predicting and Interpreting Crop Yield Variations Under Diverse Climatic and Agricultural Conditions," pp. 1–26, 2025.

- [31] P. Probst, "Hyperparameters , Tuning and Meta-Learning for Random Forest and Other Machine Learning Algorithms," 2019.
- [32] M. Waqas, S. Ahmad, A. Wangwongchai, and P. Dechpichai, "Quantification and prediction of the impact of ENSO on rainfed rice yields in Thailand," *Environ. Challenges*, vol. 19, no. February, p. 101123, 2025, doi: 10.1016/j.envc.2025.101123.
- [33] A. Amarasinghe, I. Sangarasekara, N. De Silva, M. Ariyaratne, and R. Amarasinghe, "Advancing food sustainability: a case study on improving rice yield prediction in Sri Lanka using weather - based , feature - engineered machine learning models," *Discov. Appl. Sci.*, 2024, doi: 10.1007/s42452-024-06300-7.
- [34] S. Sah, D. Haldar, R. N. Singh, B. Das, and A. S. Nain, "Rice yield prediction through integration of biophysical parameters with SAR and optical remote sensing data using machine learning models," pp. 1–17, 2024.
- [35] H. Fang, S. Liang, Y. Chen, H. Ma, W. Li, and T. He, "Science of Remote Sensing A comprehensive review of rice mapping from satellite data : Algorithms , product characteristics and consistency assessment," *Sci. Remote Sens.*, vol. 10, no. July, p. 100172, 2024, doi: 10.1016/j.srs.2024.100172.
- [36] A. Clarke *et al.*, "Integrating Climate and Satellite Data for Multi-Temporal Pre-Harvest Prediction of Head Rice Yield in Australia," pp. 1–20, 2024.