

Predicting Indonesia's rice production using XGBoost, GAN-XGBoost, and BiLSTM-BiGRU with ENSO-IOD integration

Reviana Siti Mardiah¹, Adang Suhendra², Fitrianiingsih³, Dian Kemala Putri⁴, Saifudin Nasir²,
Muhammad Farhan Fathurrohman²

¹Doctoral Program in Information Technology, Postgraduate Program, Gunadarma University, Depok, Indonesia

²Department of Information Engineering, Faculty of Industrial Technology, Gunadarma University, Depok, Indonesia

³Informatics Management, Information Technology Directorate of Diploma Program, Gunadarma University, Depok, Indonesia

⁴Department of Industrial Engineering, Faculty of Industrial Technology, Gunadarma University, Depok, Indonesia

Article Info

Article history:

Received Sep 29, 2025

Revised Jun 3, 2026

Accepted Sep 5, 2026

Keywords:

Agricultural yield forecasting

Deep learning

Feature selection

Generative artificial
intelligence

Machine learning

Missing data imputation

Spatio-temporal modeling

ABSTRACT

Food security programs require accurate rice production predictions to meet demand and plan reserves. However, incomplete historical data, limited predictor integration, and insufficient representation of large-scale climate variability hamper these predictions. Missing data require appropriate imputation, while feature selection retains relevant predictors. Indonesia spans 6° N to 11°08' S and 95° to 141° E, with diverse provincial weather conditions requiring integration with global El Niño-Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD) signals. Previous studies have not combined regional weather variables, ENSO-IOD signals, data imputation, and feature selection into a rice production prediction framework. To address this gap, this research developed prediction models using extreme gradient boosting (XGBoost), generative adversarial network (GAN)-augmented XGBoost, and a hybrid bidirectional long short-term memory-bidirectional gated recurrent unit (BiLSTM-BiGRU) with explicit ENSO-IOD integration. Several imputation and feature selection strategies were evaluated. XGBoost with K-Nearest Neighbors imputation and without feature selection performed best, achieving a root mean square error (RMSE) of 0.0082, a mean absolute percentage error (MAPE) of 5.86%, and an R^2 of 0.9991. Neither GAN-augmented XGBoost nor BiLSTM-BiGRU outperformed XGBoost. Ablation tests showed that ENSO-IOD integration reduced RMSE from 0.0182 to 0.0082. These findings can support rice availability estimation, early production shortfall identification, and rice reserve requirement decisions.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Fitrianiingsih

Informatics Management, Information Technology Directorate of Diploma Program

Gunadarma University

100 Margonda Raya Road, Depok, Indonesia

Email: fitrianiingsih@staff.gunadarma.ac.id

1. INTRODUCTION

In Indonesia, rice is the main staple food for over 90% of the population, making it a strategic commodity for national food security and economic stability [1]. Accurate rice production prediction is essential because production estimates inform rice reserve planning, import requirement determination, and policy interventions before supply deficits occur. Inaccurate predictions may misrepresent production dynamics and lead to incorrect supply estimates. These errors can create supply-demand imbalances, increase price volatility, and intensify socioeconomic pressure at both household and national levels [1].

Agronomic and environmental factors interact to influence rice production. Local climate variables, particularly precipitation and temperature, affect crop growth. Spanning approximately 6° N to 11°08' S and 95° to 141° E, Indonesia exhibits substantial variation in weather conditions across provinces. This climate variability is also influenced by global climate phenomena, including the El Niño-Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD), which can alter precipitation patterns, temperatures, and the onset of planting seasons [2], [3]. ENSO-IOD effects vary across regions because of differences in geographical characteristics and climate sensitivity [4]. The combined variation in provincial weather conditions and regional responses to ENSO-IOD highlights the need to integrate provincial climate variables with ENSO-IOD signals in rice production prediction. This integration enables a more comprehensive representation of the relationships among global climate conditions, local climate responses, and provincial rice production variability. Harvested area, which has tended to decline over time, is another important agronomic factor affecting rice production [5].

Data quality presents another challenge for rice production prediction in Indonesia. Indonesia's agricultural datasets frequently contain missing values due to limited reporting. Missing values can compromise data quality, reduce variable representativeness, and directly affect predictive performance [6]. Therefore, handling missing values requires more than simple mean substitution. Multivariate imputation methods should be evaluated to preserve structural relationships among agricultural and climate variables.

Several previous studies have developed rice production prediction models using various data sources and machine-learning and deep-learning techniques. Noorunnahar *et al.* [7] used extreme gradient boosting (XGBoost) to predict rice production in Bangladesh, but did not consider external climate variables. Satpathi *et al.* [8] used an artificial neural network (ANN) that outperformed several regression models, although the model showed signs of overfitting. Erlin *et al.* [9] used random forest (RF), gradient boosting (GB), decision tree (DT), support vector machine (SVM), and ANN, but the dataset was relatively small and limited. In the Indonesian context, previous studies generally relied only on local climate predictors and did not systematically evaluate imputation methods or treat feature selection strategies as part of the experimental design [9], [10]. Furthermore, bidirectional long short-term memory (BiLSTM) and bidirectional gated recurrent unit (BiGRU) architectures have shown consistent performance in sequential prediction tasks due to their ability to capture bidirectional temporal dependencies [11], [12]. However, their application to rice production prediction in Indonesia remains very limited. A review of prior studies shows that none has combined provincial climate variables, ENSO-IOD signals, and systematic evaluations of imputation and feature selection strategies within a single rice production prediction framework.

The literature indicates three research gaps; i) the integration of provincial climate variables with ENSO-IOD signals has not been widely examined within a unified prediction framework; ii) the comparative effects of imputation methods on predictive performance have not been systematically evaluated in a multivariate context combining agricultural and climate data; and iii) feature selection strategies are rarely treated as an explicit part of the experimental design, although feature selection is important for reducing variable redundancy and multicollinearity risk.

To address these gaps, this research proposes a unified rice production prediction framework for Indonesia that combines provincial climate variables, ENSO-IOD signals, imputation method evaluation, feature selection strategies, and multi-model comparison. The evaluated models include XGBoost, GAN-augmented XGBoost, and a hybrid BiLSTM-BiGRU model. Four imputation methods, namely k-nearest neighbors (KNN), RF, multivariate imputation by chained equations (MICE), and generative adversarial imputation networks (GAIN), were also evaluated to handle missing values [13], [14]. Generative adversarial networks (GAN) were also evaluated as an exploratory data augmentation approach [15]. Three feature selection strategies were evaluated, namely hybrid RF-XGBoost with variance inflation factor (VIF), hybrid RF-XGBoost with Spearman correlation, and a two-stage Spearman-based approach. This research makes four main contributions; i) integrating provincial climate variables with ENSO-IOD signals within a unified rice production prediction framework for Indonesia, a combination absent from prior studies, ii) evaluating the effects of various imputation methods on predictive performance, iii) treating feature selection strategies as an explicit part of the experimental design to reduce variable redundancy and improve model interpretability, and iv) comparing the predictive performance of XGBoost, GAN-augmented XGBoost, and the hybrid BiLSTM-BiGRU model under a uniform evaluation protocol.

2. METHOD

The rice production prediction model was developed through six stages, as illustrated in Figure 1. The first stage involved collecting secondary data from national and international agencies and harmonizing provincial-level data. Exploratory data analysis was then conducted to examine the data structure and characteristics. Data preprocessing included imputation, outlier handling, feature selection, feature

construction, encoding, transformation, and normalization. The data were then divided chronologically into training and testing datasets to preserve the temporal order of the time-series data. The next stage included model development and hyperparameter tuning using XGBoost, GAN-augmented XGBoost, and a hybrid BiLSTM-BiGRU model. The final stage included model evaluation.

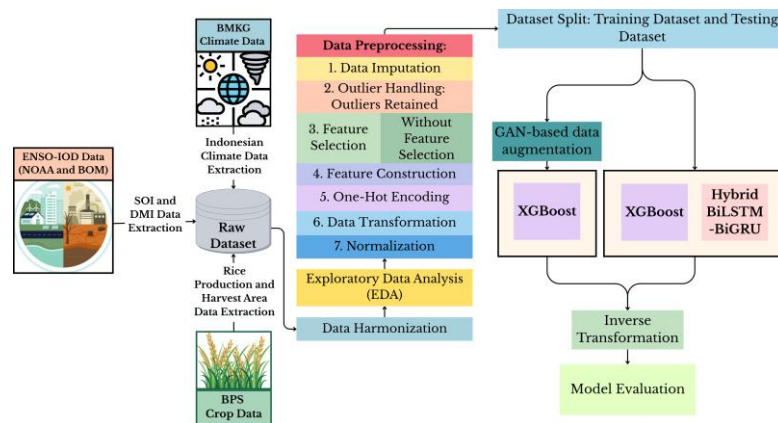


Figure 1. Proposed framework for rice production prediction in Indonesia

2.1. Data collection and processing

This research used secondary data obtained from various sources. The Indonesian Central Bureau of Statistics (BPS) provided information on rice production and harvested area, while the Indonesian Meteorology, Climatology, and Geophysics Agency (BMKG) provided local climate data. Additionally, the National Oceanic and Atmospheric Administration (NOAA) provided data on the Dipole Mode Index (DMI), and the Bureau of Meteorology (BoM) provided data on Southern Oscillation Index (SOI) anomalies. The observation period covered 1997-2024.

The dataset initially included 38 Indonesian provinces. However, during the observation period, several administrative changes occurred, including the establishment of North Kalimantan in 2012, West Sulawesi in 2004, and Bangka Belitung Islands in 2000, as well as several provinces resulting from administrative subdivisions in Papua. To maintain territorial consistency and time-series continuity, data from the newly created provinces were merged back into their parent provinces. The harmonization included reassigning Banten to West Java, Bangka Belitung Islands to South Sumatra, Riau Islands to Riau, Gorontalo to North Sulawesi, West Papua and several Papua expansion provinces to Papua, North Kalimantan to East Kalimantan, North Maluku to Maluku, and West Sulawesi to South Sulawesi. This process yields a balanced panel dataset comprising 26 provinces and spanning 28 years.

Monthly climate data were aggregated into annual averages for variables such as temperature, humidity, wind speed, and the DMI. Annual totals were calculated for precipitation and the number of rainy days. The SOI was averaged annually and classified as El Niño, La Niña, or Neutral. Measurement units were standardized. Table 1 summarizes all variable uses.

Table 1. Description of variables

Variable	Description	Data type
Province	Province where rice production was recorded	Categorical
Year	Observation year	Temporal/numerical
Rice production	Total annual rice production in each province	Numerical
Harvested area	Total annual harvested area in each province	Numerical
Minimum temperature	Annual average minimum air temperature in each province	Numerical
Average temperature	Annual average air temperature in each province	Numerical
Maximum temperature	Annual average maximum air temperature in each province	Numerical
Precipitation	Total annual precipitation in each province	Numerical
Rainy days	Annual number of rainy days in each province	Numerical
Humidity	Annual average air humidity in each province	Numerical
Wind speed	Annual average wind speed in each province	Numerical
Atmospheric pressure	Annual average atmospheric pressure in each province	Numerical
Sunshine duration	Annual average daily sunshine duration in each province	Numerical
SOI	Annual average SOI value representing ENSO-related atmospheric variability	Numerical
ENSO phase	Climate phase based on SOI classification (El Niño, La Niña, or neutral conditions)	Categorical
DMI	Annual average DMI value represents IOD variability	Numerical

Predicting Indonesia's rice production using XGBoost, GAN-XGBoost, and ... (Reviana Siti Mardiah)

2.2. Exploratory data analysis

Exploratory data analysis examined the dataset structure and characteristics to inform method selection. EDA also assessed dataset quality before the preprocessing stage [16]. In this research, EDA involved data visualization.

2.3. Data preprocessing

The data preprocessing stage consisted of several crucial steps to prepare the data for machine learning and deep learning model development. In this research, data preprocessing included data imputation, outlier handling, feature selection, feature construction, encoding, data transformation, and normalization.

During the data imputation stage, four methods were used to address missing values in local climate variables, namely KNN, RF, MICE, and GAIN. KNN was used because the province-year agro-meteorological panel contained observations with similar regional and temporal patterns, allowing missing values to be estimated from observations with comparable characteristics [13]. RF was used because its tree-based ensemble structure can capture nonlinear relationships and interactions between agricultural production and meteorological predictors [13], [14]. MICE was used because the dataset contained multiple interrelated meteorological and agricultural variables. Its iterative multivariate imputation process preserves conditional relationships among variables and provides a stable imputation method for tabular data [14]. GAIN was used because its generative deep learning architecture learns the joint distribution of multivariate agro-meteorological data through adversarial training and can capture complex dependencies that parametric methods may not fully represent [14]. KNN and RF were evaluated using both default and tuned settings.

In the outlier-handling stage, outliers were retained because extreme values in the dataset did not represent recording errors. Instead, these values reflected actual climatic and production conditions that should remain available for model learning [17].

This research evaluated models with and without feature selection. Feature selection was necessary because the agro-meteorological dataset contained several predictors that may convey similar information, such as minimum, average, and maximum temperature, precipitation, and rainy days. Using all overlapping features can increase model complexity and obscure the interpretation of variable contributions [18], [19].

The hybrid RF-XGBoost approach was used for feature ranking to reduce bias from single-algorithm importance scores. RF importance, measured through mean decrease in Gini impurity, may assign excessive importance to continuous variables with many unique values, whereas XGBoost gain-based importance may vary according to the order of boosting splits. Combining rankings from both algorithms produced a more stable feature order [20]. However, feature ranking alone does not eliminate redundancy. To address this, two filters were applied. Spearman's rank correlation was used because the agro-meteorological dataset exhibits markedly right-skewed production distributions and heteroscedastic variance patterns across climate predictors [21]. Additionally, VIF was used to identify multicollinearity, capturing simultaneous multicollinear dependencies among multiple predictors that pairwise correlation analysis cannot detect [19]. Evaluating both filters independently allowed the dominant sources of feature overlap to be assessed in the multivariate agro-meteorological data. Figure 2 presents the two hybrid RF-XGBoost feature selection procedures used in this research. Figure 2(a) applies Spearman's correlation to reduce feature redundancy, while Figure 2(b) uses VIF to address multicollinearity.

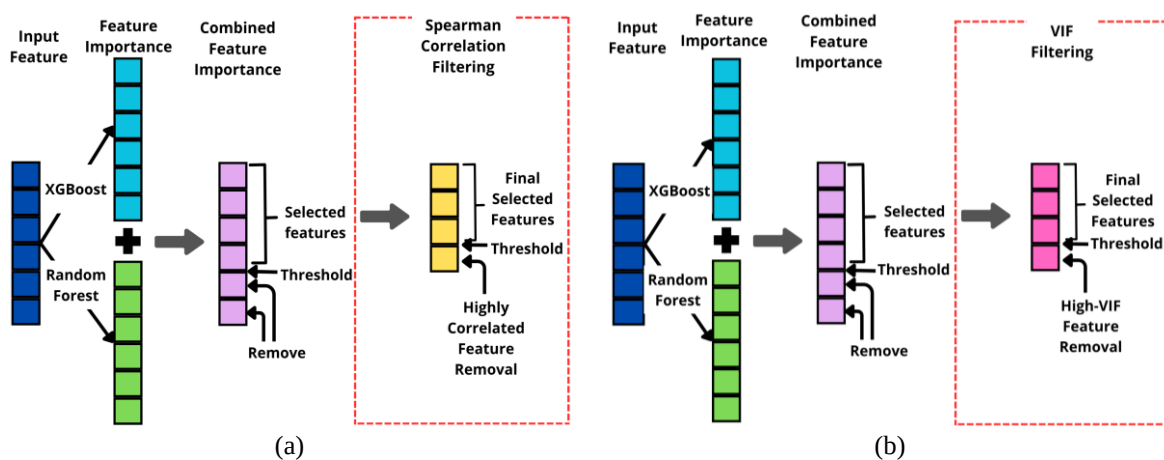


Figure 2. Hybrid RF-XGBoost feature selection method adopted from Cui *et al.* [20], with the proposed addition of; (a) a spearman correlation check and (b) a VIF check

a. Hybrid RF-XGBoost with Spearman redundancy filtering

First, the features were ranked by combining RF and XGBoost feature importance scores, as adopted from Cui *et al.* [20]. The ranking results were then filtered using Spearman's correlation to remove features that were highly correlated with one another. This process produced a less redundant subset [21]. Figure 2(a) illustrates this method.

b. Hybrid RF-XGBoost with VIF multicollinearity filtering

The initial feature-ranking stage followed the same RF-XGBoost procedure. The main difference was using the VIF to identify multicollinearity among the features. Features with high VIF values were removed to ensure model stability and reduce bias in parameter estimation [19]. Figure 2(b) illustrates this method.

c. Spearman relevance and Spearman redundancy filtering

This approach used only statistical selection methods. First, Spearman's rank correlation was used to evaluate the relationship between each feature and the target variable and identify relevant features. Spearman's correlation was then used to evaluate relationships among features and remove redundant features [21]. Figure 3 illustrates this method.

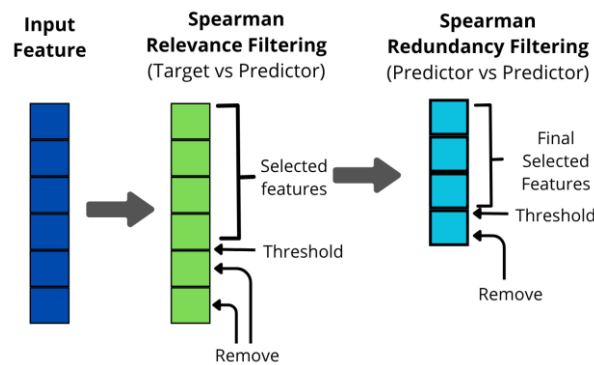


Figure 3. Feature selection using spearman relevance and redundancy filtering

The feature construction stage created new features from the existing variables to obtain a more representative variables set [22]. This stage was necessary because the original variables may not fully represent relationships among the factors. In this research, productivity variables were constructed from production and harvested area variables.

The encoding stage of this research employed the one-hot encoding method, which converts categorical variables into binary features that are appropriate for machine learning and deep learning models [23]. One-hot encoding was applied to the variables “Province” and “ENSO Phase”.

The Yeo-Johnson method was used for data transformation. This transformation reduced skewness and improved distributional symmetry, particularly for variables with skewed distributions. The method was selected because it can handle both positive and negative values [24].

The normalization used the Min-Max method, ensuring that all numerical features were scaled to the same range, specifically between 0 and 1. This process prevented features with larger scales from dominating features with smaller scales [25]. Min-Max normalization was performed using in (1), where x represents the original data, $x_{\min}$ is the minimum value, $x_{\max}$ is the maximum value in the dataset, and x' denotes the normalized result ranging from 0 to 1.

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

2.4. Data splitting

The data were divided using a year-based (temporal) split approach, which is more appropriate for time-series data than random splitting. Random splitting can leak information from future periods into the training dataset, resulting in overly optimistic performance evaluations. The temporal split approach prevented this leakage and provided a more realistic evaluation of model performance over time [26]. The training dataset covered 1997-2019, while the testing dataset covered 2020-2024, resulting in an approximate 80:20 ratio. The 2020-2024 testing period covered production dynamics from the pandemic through the post-pandemic recovery phase, as well as recent climate anomalies not present in the training dataset.

Evaluation over this period therefore assessed the model's ability to support import decisions and rice reserve planning during actual disruptions. This approach ensured that all provinces were represented in the training dataset while the models were evaluated under conditions reflecting operational use.

As a supplementary robustness check, expanding-window temporal validation was conducted using the optimal configurations of the three models. The initial training period covered 1997-2019, and 2020 was used as the first testing year. The training period was then expanded by one year in each subsequent fold, producing five folds with testing years from 2020 to 2024.

2.5. Modeling

This research developed three rice production prediction models: XGBoost, GAN-augmented XGBoost, and a hybrid BiLSTM-BiGRU model. The hybrid BiLSTM-BiGRU and GAN-augmented XGBoost models were included to examine whether temporal modeling and GAN-generated training samples could improve predictive performance relative to XGBoost. Given the limited number of annual observations, the comparison assessed whether the additional model complexity provided measurable benefits. All three models were evaluated under the same data conditions and performance measures.

2.5.1. Generative adversarial network-augmented XGBoost

This research developed a GAN-augmented XGBoost framework in which a GAN generated additional training samples and XGBoost performed rice production prediction. The inclusion of a GAN was intended as an exploratory attempt to examine whether synthetic data augmentation could improve prediction performance. Figure 4 presents the proposed GAN-augmented XGBoost framework developed in this research.

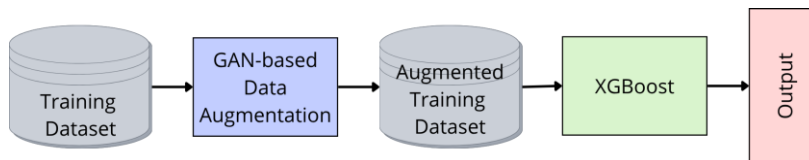


Figure 4. Proposed GAN-augmented XGBoost framework

a. GAN-based data augmentation

The data augmentation stage used a GAN. A GAN is a generative model capable of learning the distribution of samples from a specific dataset. In this research, a GAN was used to generate additional training samples intended to approximate the distribution of the original data. These generated samples were then used to examine whether synthetic augmentation could improve predictive performance. This research used a vanilla GAN and a conditional GAN. Figure 5 presents the architecture of a vanilla GAN [27].

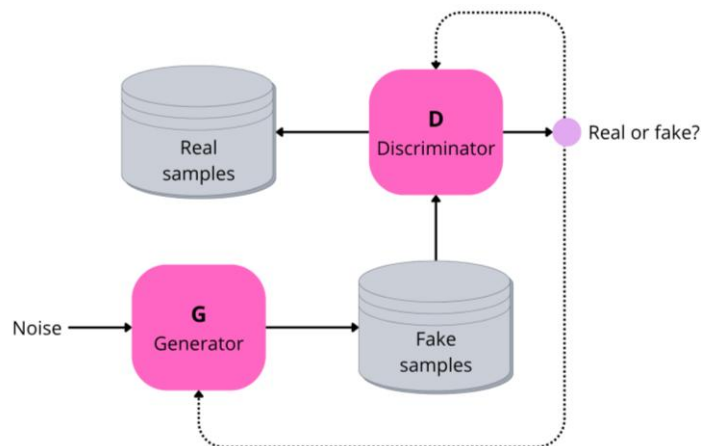


Figure 5. Vanilla GAN architecture [27]

The vanilla GAN consists of two main components, namely a generator (G) and a discriminator (D), as illustrated in Figure 5. The generator receives a random vector z drawn from a noise distribution, then generates a synthetic sample $G(z)$ with the same dimensions as the original sample x . The discriminator is trained using both real samples $\{(x_i, 1)\}_{i=1}^m$ and fake samples $\{(G(z_i), 0)\}_{i=1}^m$, then outputs a binary prediction $D(x)$ indicating whether the sample is real ($D(x) = 1$) or fake ($D(x) = 0$) as formulated in (2). On the other hand, the generator aims to reduce the likelihood that $G(z)$ will be classified as fake by D , will be classified as fake by the discriminator, as shown in (3). By combining these two functions, a MinMax game between G and D is obtained, which is formulated in (4).

$$\max_D V(D) = \mathbb{E}_{x \sim P_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim noise} [\log (1 - D(G(z)))] \quad (2)$$

$$\min_G V(G) = \mathbb{E}_{z \sim noise} [\log (1 - D(G(z)))] \quad (3)$$

$$\min_G \max_D V(G, D) = \mathbb{E}_{x \sim P_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim noise} [\log (1 - D(G(z)))] \quad (4)$$

In this implementation, the generator received a 100-dimensional latent vector as input and used a fully connected architecture with hidden layers of sizes 128, 256, 512, 256, and 128. It also incorporated LeakyReLU activation, batch normalization, dropout, and a skip connection. The discriminator was developed as a fully connected network with 256, 128, and 64 hidden units, using LeakyReLU activation and dropout. Gaussian noise was added to the input for regularization. Both components were optimized using Adam with a learning rate of 0.0001. The model was trained with a batch size of 16 for up to 300 epochs, and synthetic samples equivalent to 25% of the original training data were generated for each province.

The cGAN in this research was conditioned on the ENSO phase, which was encoded as a one-hot vector. In the generator, the ENSO condition vector was combined with the latent vector, while in the discriminator (critic), the condition vector was combined with the input data. The generator architecture in the cGAN used a 32-dimensional latent vector combined with the ENSO condition vector. The combined vector passed through fully connected layers with 64, 128, and 64 hidden units, using LeakyReLU activation, batch normalization, and dropout. The discriminator received either real or synthetic data together with the ENSO condition vector. The combined input was then processed using gaussian noise and two fully connected layers with 64 and 32 units, each using LeakyReLU activation and dropout.

The cGAN was trained using the WGAN-GP strategy with the Adam optimizer, a generator learning rate of 0.0002, a discriminator learning rate of 0.00005, two critic updates per generator update, up to 300 epochs with adaptive adjustments based on the number of samples, and a batch size of 16. The generated synthetic data targeted an augmentation ratio of 25% of the 1997-2019 training data, while the 2020-2024 test data remained unchanged. Dropout and gaussian noise were applied to reduce overfitting, while batch normalization was used to stabilize training given the limited sample size for each province.

b. XGBoost

XGBoost is a machine learning algorithm based on decision-tree ensembles, designed to optimize the loss function while effectively handling class imbalance and preventing overfitting. Hyperparameter tuning was conducted to identify the best-performing configuration [28]. The hyperparameters included `n_estimators` (ranging from 100 to 250), `max_depth` (between 1 and 14), and `learning_rate` (from 0.1 to 1), which were adopted from the research by Ørebæk and Geitle [29]. GridSearchCV with 10-fold cross-validation was used to evaluate the hyperparameter combinations. Ten-fold cross-validation was selected because the limited dataset allowed each training iteration to use a relatively large proportion of the available data, thereby reducing estimation bias associated with small sample sizes [30].

2.5.2. Hybrid bidirectional long short-term memory-bidirectional gated recurrent unit

This research developed a hybrid model based on a recurrent neural network (RNN) that combining BiLSTM and BiGRU to analyze multivariate time-series data from multiple provinces. The architecture of the proposed hybrid BiLSTM-BiGRU model is illustrated in Figure 6.

The model takes multivariate time-series data arranged using a 12-step sliding window as input, where each time step represents one full year of aggregated agricultural data. This configuration corresponds to 12 years of historical observations, which is appropriate for capturing inter-annual variability in rice productivity driven by climatic phenomena such as ENSO [2]. The selection of this window size was confirmed through comparative experiments, with the 12-step configuration yielding the best predictive performance.

The proposed architecture consists of two parallel recurrent branches. The BiLSTM branch consists of a 256-unit BiLSTM layer followed by a dropout layer with a rate of 0.2 and a second 128-unit BiLSTM

layer. This branch ends with a dense layer containing 256 units with ReLU activation. In parallel, the BiGRU branch includes a BiGRU layer with 128 units, a dropout layer with a rate of 0.1, and a dense layer with 256 units and ReLU activation.

The outputs of the two branches were concatenated to produce a 512-dimensional feature representation. This combined representation is then passed to a dense layer with 64 units and ReLU activation, followed by a final output layer with linear activation to generate continuous predictions of the target variable. To improve model performance, hyperparameter tuning was conducted by evaluating combinations of learning rates ranging from 0.0001 to 0.001 and batch sizes of 32, 64, and 128 [31]. The final model was compiled with the Adam optimizer, a learning rate of 0.0003, and a batch size of 32, and was trained for up to 200 epochs. Dropout and early stopping were applied to reduce overfitting because of the limited number of annual observations.

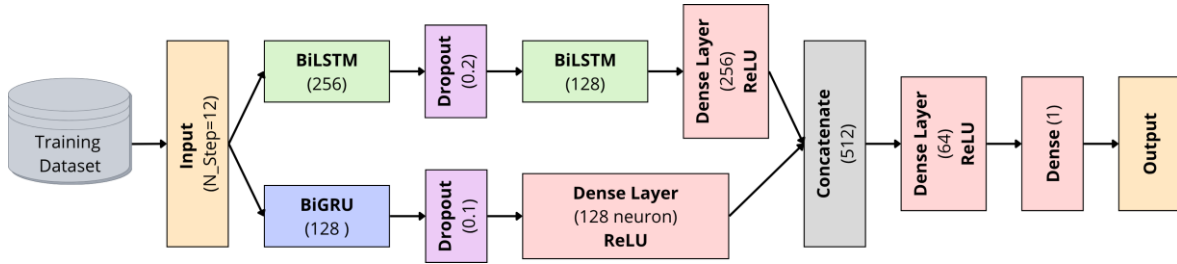


Figure 6. Proposed hybrid BiLSTM-BiGRU rice production prediction architecture

2.6. Model evaluation

Before model evaluation, the prediction results were inverse-transformed to their original scale to allow direct comparison with the actual values. For this research, three evaluation error metrics were used, namely root mean squared error (RMSE), mean absolute percentage error (MAPE), and the coefficient of determination (R^2).

R^2 measures the proportion of variance in the dependent variable, with a range from 0 to 1. A value closer to 1 indicates a better fit of the model to the data. RMSE quantifies the magnitude of the prediction error, where a lower value (closer to zero) indicates more accurate predictions. Conversely, MAPE expresses the average prediction error as a percentage [32]. High R^2 values combined with low RMSE and MAPE values indicate better predictive performance for rice reserve planning and import decision-making. For the expanding-window temporal validation, RMSE, MAPE, and R^2 were calculated for each of the five testing years and summarized using the mean and standard deviation.

The formulas for RMSE, MAPE, and R^2 are presented in (5) to (7). In these equations, y_i represents the actual value of the i -th, $\hat{y}_i$ denotes the prediction for the i -th sample, $\bar{y}$ is the average value of all actual data, and n is the total number of samples in the dataset.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (6)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (7)$$

Statistical tests were also applied to assess the significance and magnitude of the observed performance differences among models across all experimental configurations. The Friedman test was used to evaluate whether the performance differences among models were statistically significant. If the results were significant, the analysis proceeded with Nemenyi's post hoc test to identify pairs of models that differed significantly at $\alpha=0.05$. Cliff's delta was then calculated to measure the magnitude of these differences, which were classified as negligible ($|\delta| < 0.147$), small (< 0.330), medium (< 0.474), or large (≥ 0.474) [33].

Additionally, this research conducted diagnostic residual analysis, including residual plots and autocorrelation function (ACF) and partial autocorrelation function (PACF) values, to ensure that the model's predictive performance stems from generalization ability rather than overfitting. Furthermore, learning curves were analyzed to diagnose overfitting or underfitting during model training.

3. RESULTS AND DISCUSSION

3.1. Dataset

The dataset consisted of 902 rows and 16 variables, with each row representing rice production data and related variables for one province in a given year. This formed a panel time series structure. The dataset includes variables such as rice production, harvested area, and various climatic conditions, including minimum, average, and maximum air temperatures; precipitation; number of rainy days; humidity; wind speed; atmospheric pressure; and sunshine duration. For the global climate variables, this research used SOI anomalies to assess ENSO. SOI values below -7 were categorized as the El Niño phase, while values above 7 were categorized as the La Niña phase [34]. The DMI was used to represent the IOD. The target variable was rice production.

Figure 7 shows that provincial data availability was inconsistent throughout the observation period due to administrative expansions. The blank areas in Figure 7 do not indicate data collection errors, but reflect years in which newly established provinces did not yet have separate historical records. After the harmonization process, the final dataset comprised a balanced panel of 26 harmonized provincial units over 28 years. This adjustment allowed the analysis to be based on a uniform regional structure, enabling the model to recognize historical patterns accurately [35], [36]. However, harmonizing the newly established provinces with their parent provinces may introduce variations in agronomic characteristics.

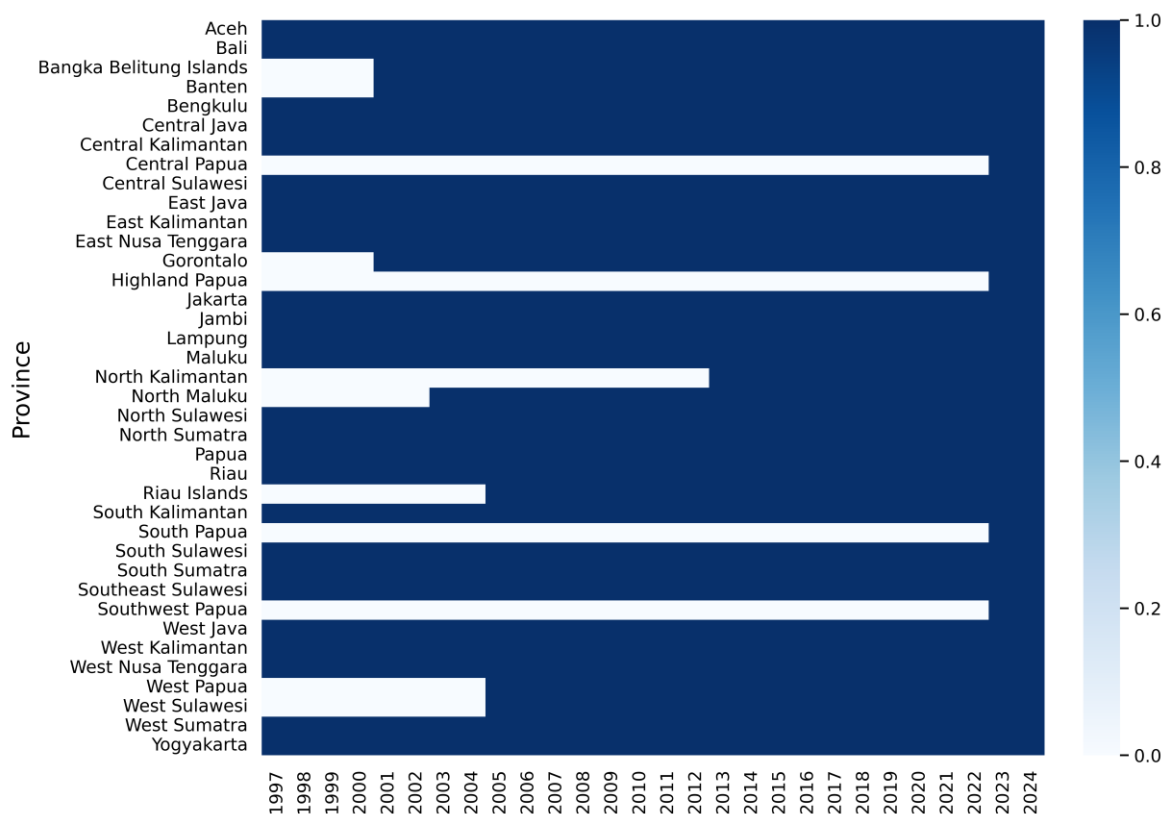


Figure 7. Provincial data availability in Indonesia, 1997-2024

To observe long-term production trends, a 3-year moving average was applied, as illustrated in Figure 8. This method reduced annual fluctuations and clarified spatial and temporal patterns [37]. The figure indicates that East Java, Central Java, and West Java consistently recorded the highest production levels. However, production in these provinces declined after 2017, indicating lower production in the main national rice-producing regions.

Conversely, several provinces outside Java, such as South Sulawesi and North Sumatra, experienced stable production growth until around 2016. Production subsequently stagnated or declined in these provinces. These differences reflect varying provincial rice production dynamics. The observed provincial variations support the use of spatiotemporal analysis for rice production prediction.

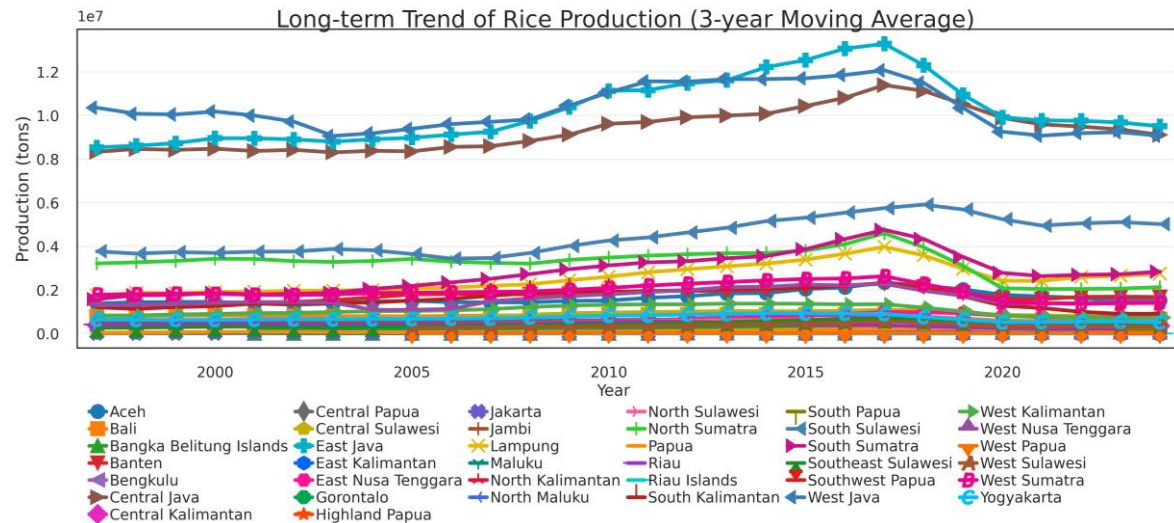


Figure 8. Long-term rice production trends across provinces

3.2. Exploratory data analysis

After harmonizing the data, the dataset consisted of 728 rows and 16 variables. The duplicate data detection results indicated that no identical rows were found across all variables, confirming zero duplicates. This result confirmed that duplicate records did not affect the statistical analyses or prediction model development.

Missing-value detection showed that the variables province, year, harvested area, and production had no missing values. However, the minimum temperature variable had 37 missing values (5.1%), the average temperature had 41 (5.6%), and the maximum temperature had 40 (5.5%). Humidity had 64 missing values (8.8%), while precipitation had 72 (9.9%). The number of rainy days had a significant percentage of missing values, totaling 264 (36.3%). Wind speed also had 86 missing values (11.8%), atmospheric pressure had 197 (27.1%), and sunshine had 241 (33.1%). These results show that missing values were concentrated in the climate variables, making imputation necessary before model development. High levels of missing values in Indonesian climate panel data are common due to limited observational coverage. However, imputed values cannot fully replace actual observations. Figure 9 illustrates the relationship between production and factors such as harvested area and climate.

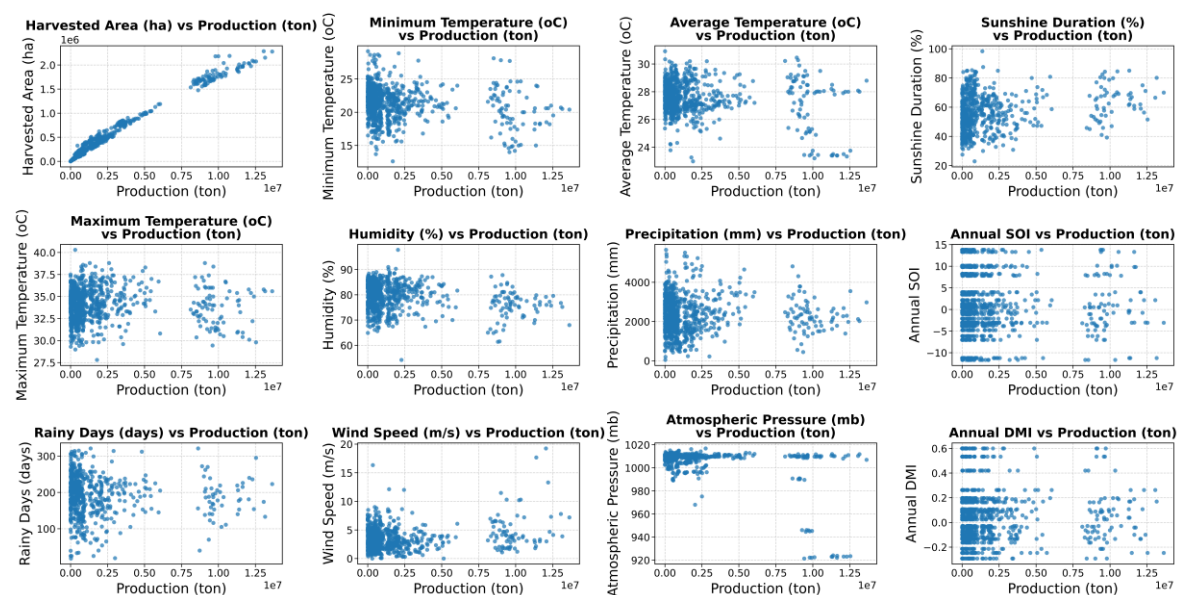


Figure 9. Relationship between rice production and other variables

The data show a very strong linear relationship between harvested area and production, as shown by the clear straight line through the data points. This indicates that a larger harvested area correlates with greater production. This result is expected because the harvested land area directly influences production volume. Therefore, the harvested area was included as an important baseline predictor.

In contrast, climate variables do not show a strong linear relationship with production. The scatterplot reveals that most data points are scattered randomly without a discernible pattern. This suggests that the influence of climate factors on production is more complex and cannot be explained through a simple linear relationship. The intricate relationship between climate variables and production highlights the need for analytical and modeling approaches that can effectively capture the dynamics of non-linear and multivariate data [37]. Therefore, methods such as XGBoost, GRU, and LSTM were evaluated to capture patterns that may not be adequately represented by traditional linear models such as ARIMA [38].

3.3. Data imputation

This research evaluated MICE, KNN imputer, RF imputer, and GAIN. These methods were selected because the dataset contains multivariate agricultural and climate variables, so the missing values should be estimated by considering relationships among variables rather than using simple mean or median replacement [13], [14].

Figure 10 shows that the precipitation distribution after imputation generally followed the original distribution across all methods. The resulting curves have similar shapes and peak locations, indicating that imputation does not alter the precipitation distribution. Significant distribution changes can affect the prediction model's reliability. However, distribution similarity alone does not identify the optimal imputation method. Therefore, the final imputation method was selected based on its contribution to model performance.

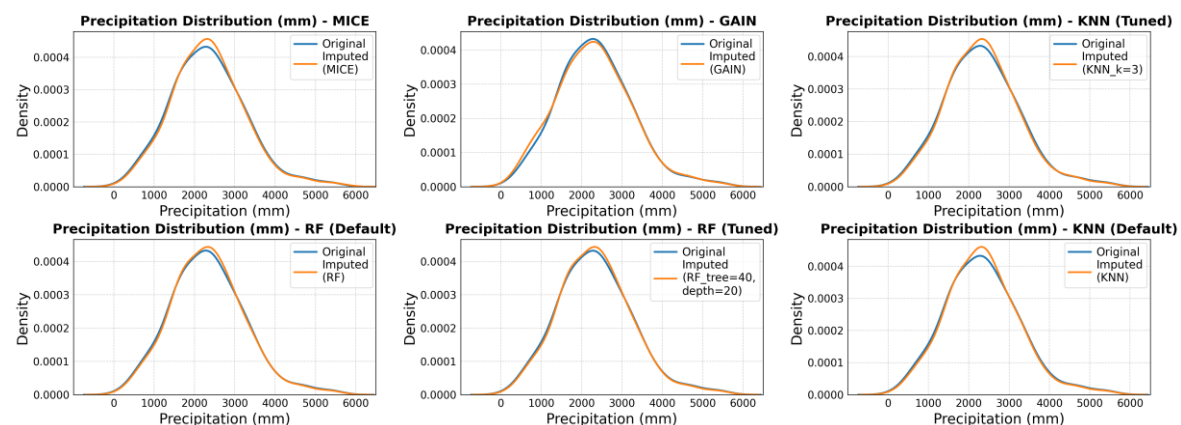


Figure 10. Precipitation distribution before and after imputation

3.4. Model performance

The initial experiments compared six imputation configurations and four feature selection approaches using XGBoost and a hybrid BiLSTM-BiGRU model. The results for these combinations are summarized in Table 2.

Table 2 shows that the XGBoost model consistently outperformed the BiLSTM-BiGRU model across nearly all combinations. The best XGBoost performance was achieved with the KNN imputation method and without feature selection, with an RMSE of 0.0082 and an R^2 of 0.9991. This result indicates that feature removal reduced predictive performance in this experimental setting. Local climate variables, harvested area, and ENSO-IOD indicators may show limited individual effects but provide useful predictive information when evaluated together. For example, two-stage spearman feature selection increased RMSE by 41%. This condition fits the nature of rice production prediction. Many factors affect rice production simultaneously, rather than a single dominant factor.

In contrast, BiLSTM-BiGRU performance depended more strongly on feature selection. Its best configuration combined GAIN imputation with the two-stage Spearman feature selection, resulting in an RMSE of 0.0209 and an R^2 of 0.9941. This configuration reduced RMSE by approximately 45% relative to the baseline configuration without feature selection.

The difference in results can be attributed to the basic architecture of the two models. XGBoost, a tree-based ensemble model, internally selects the most informative features at each splitting stage, effectively

overcoming inter-feature correlations [39]. Conversely, sequential architectures such as BiLSTM-BiGRU are more sensitive to noise and high correlations among input features. For the BiLSTM-BiGRU model, feature selection plays a crucial role in stabilizing training by reducing gradient variability, thereby improving convergence. Furthermore, GAIN imputation is more suitable for this model, as it reconstructs data by considering the interrelationships between variables, an approach that aligns well with the functionality of neural networks [14].

Table 2. The impact of combining imputation with feature selection on model performance

Imputation/feature selection	Without selection (XGBoost)	Without selection (BiLSTM-BiGRU)	RF-XGB-spearman (XGBoost)	RF-XGB-spearman (BiLSTM-BiGRU)	RF-XG-VIF (XGBoost)	RF-XG-VIF (BiLSTM-BiGRU)	Two-stage spearman (XGBoost)	Two-stage spearman (BiLSTM-BiGRU)
KNN (default)	RMSE 0.0082; R² 0.9991	RMSE 0.0382; R ² 0.9801	RMSE 0.0089; R ² 0.9989	RMSE 0.0328; R ² 0.9853	RMSE 0.0116; R ² 0.9982	RMSE 0.0308; R ² 0.9871	RMSE 0.0116; R ² 0.9982	RMSE 0.0243; R ² 0.9920
KNN (tuned)	RMSE 0.0106; R ² 0.9985	RMSE 0.0337; R ² 0.9846	RMSE 0.0119; R ² 0.9981	RMSE 0.0445; R ² 0.9731	RMSE 0.0110; R ² 0.9983	RMSE 0.0260; R ² 0.9908	RMSE 0.0119; R ² 0.9981	RMSE 0.0227; R ² 0.9930
RF (default)	RMSE 0.0118; R ² 0.9981	RMSE 0.0321; R ² 0.9860	RMSE 0.0095; R ² 0.9988	RMSE 0.0300; R ² 0.9878	RMSE 0.0121; R ² 0.9980	RMSE 0.0245; R ² 0.9918	RMSE 0.0118; R ² 0.9981	RMSE 0.0218; R ² 0.9935
RF (tuned)	RMSE 0.0117; R ² 0.9981	RMSE 0.0319; R ² 0.9861	RMSE 0.0093; R ² 0.9988	RMSE 0.02360; R ² 0.9924	RMSE 0.0106; R ² 0.9985	RMSE 0.0237; R ² 0.9923	RMSE 0.0094; R ² 0.9988	RMSE 0.0249; R ² 0.9915
MICE	RMSE 0.0108; R ² 0.9986	RMSE 0.0327; R ² 0.9854	RMSE 0.0099; R ² 0.9987	RMSE 0.0228; R ² 0.9929	RMSE 0.0089; R ² 0.9989	RMSE 0.0241; R ² 0.9921	RMSE 0.0100; R ² 0.9986	RMSE 0.02244; R ² 0.9919
GAIN	RMSE 0.0089; R ² 0.9989	RMSE 0.0269; R ² 0.9901	RMSE 0.0116; R ² 0.9982	RMSE 0.0264; R ² 0.9905	RMSE 0.0112; R ² 0.9983	RMSE 0.0247; R ² 0.9917	RMSE 0.0099; R ² 0.9987	RMSE 0.0209; R² 0.9941

Table 3 compares the optimal configurations of XGBoost and hybrid BiLSTM-BiGRU. The results show that XGBoost with the optimal configuration achieved an RMSE of 0.0082 (corresponding to 92,135 tons on the original scale), a MAPE of 5.86%, and an R² of 0.9991. These metrics were substantially better than those of the optimal BiLSTM-BiGRU configuration, which achieves an RMSE of 0.0209 (equivalent to 234,185 tons), a MAPE of 10.74%, and an R² of 0.9941. XGBoost effectively reduced the RMSE by approximately 60.8% and the MAPE by about 45.4%.

Table 3. Comparison of XGBoost and BiLSTM-BiGRU models using the optimal proposed method

Optimal configuration	RMSE	MAPE (%)	R ²
XGBoost (KNN default+without feature selection)	0.0082	5.86	0.9991
BiLSTM-BiGRU (GAIN+two-stage spearman)	0.0209	10.74	0.9941

The Friedman test on 24 XGBoost model configurations confirmed a significant difference in overall performance, with $\chi^2(23)=43.0640$ and $p=0.006806$. However, the Nemenyi test found no significant differences between model pairs. This occurred due to the extensive correction burden from 276 paired comparisons, where the performance difference must be very large to overcome the statistical significance penalty [40]. Additionally, this lack of significance is also affected by high cross-fold variation, with coefficients of variation ranging from 0.53 to 0.63. The highest-performing model has an average cross-fold RMSE of 0.0237, a standard deviation of 0.0136, and a coefficient of variation of 0.57. This relatively large cross-fold variation can make the performance differences between models appear insufficiently robust in pairwise comparisons, particularly in annual time-series data that may exhibit pattern shifts across periods [41]. Cliff's Delta results prove that most of the inter-model distances are negligible. Therefore, the configuration with KNN imputation and without feature selection is designated as the highest-performing model based on the lowest error metric, rather than due to a significant advantage over other configurations [42].

A sensitivity analysis was conducted to evaluate the impact of data harmonization and mitigate the risk of ecological fallacy. This analysis used three dataset scenarios, namely newly established provinces merged into their parent provinces (harmonized), entities separated according to actual boundaries (actual boundary), and a dataset containing only provinces with complete historical data (complete-period). A comparison of performance across the three scenarios is presented in Table 4.

Table 4. Sensitivity analysis of provincial boundary harmonization

Dataset scenario	RMSE	MAPE (%)	R ²
Harmonized provincial dataset	0.0082	5.86	0.9991
Actual administrative boundary dataset	0.0184	23.23	0.9950
Complete-period administrative dataset	0.0197	7.23	0.9977

According to Table 4, the actual administrative boundary scenario yields the highest error, with a MAPE of 23.23%. This increase is driven by newly established provinces, which have an average MAPE of 67.09%, far exceeding that of parent provinces (8.41%) and provinces without administrative changes (6.42%). This decrease in performance occurs because the historical data range for the newly established provinces is too short, which ultimately limits the model's learning process. These results indicate that while data pooling carries the risk of ecological fallacy, regional separation actually reduces model accuracy. The impact of this risk proved to be minimal due to the high climatic similarity between the parent and newly established provinces. For example, the Papua and West Papua regions recorded the same minimum temperature (21 °C). Other weather variables also showed only minor differences. The average temperature is 27.44 °C compared to 27.12 °C, the humidity percentage is 86.98% compared to 85.92%, annual precipitation is 2,471.6 mm compared to 2,385 mm, and the number of rainy days is 306 compared to 299 days.

The very high R² value achieved by XGBoost (R²=0.9991) should be interpreted in light of the dataset structure. As shown in Figure 9, harvested area has a strong linear relationship with rice production because total production is closely associated with the extent of harvested land. This dominant relationship helps explain why the model can capture a large proportion of the overall variation in rice production. However, the model's strong performance should not be attributed solely to the harvested area. Integrating local climate variables and global climate signals, particularly ENSO and IOD, remains important for capturing interannual variability and climate-related deviations not fully explained by the main agronomic pattern. The ablation results support this interpretation, showing that model performance declines when additional predictors or ENSO-IOD variables are removed. Therefore, the very high R² value is plausible, but it reflects both the dominant agronomic relationship and the added contribution of climate-related information. In addition, this result should be viewed in light of the temporal split strategy used in this research, in which earlier years were used for training and later years for testing, thereby reducing the risk of data leakage [26].

Table 5 examines the contribution of additional features and ENSO-IOD variables through ablation experiments. The ablation experiments with XGBoost show that optimal performance is achieved only when additional features are used alongside the ENSO-IOD index. When the model used only the ENSO-IOD index, performance declined, yielding an RMSE of 0.0165, a MAPE of 10.51%, and an R² of 0.9963. This indicates an increase in RMSE of approximately 101% and a rise in MAPE of 79%. When neither the additional features nor the ENSO-IOD index was applied, model performance deteriorated, with RMSE increasing by 122% compared with the best-performing model. These results highlight a synergistic effect, where the global climate index elucidates the primary climate patterns (El Niño/La Niña/Neutral), while the other features are essential for modeling seasonal change thresholds and inter-provincial variations. Boosting effectively leverages the nonlinear interactions among agronomic, local-climate, and global-climate predictors.

Table 5. XGBoost model: the impact of additional features and ENSO-IOD

Optimal configuration	RMSE	MAPE (%)	R ²
XGBoost (with additional feature+ENSO-IOD)	0.0082	5.86	0.9991
XGBoost (with ENSO-IOD)	0.0165	10.51	0.9963
XGBoost (without additional feature+ENSO-IOD)	0.0182	10.49	0.9979

The results are based on differences in the spatial scale between local climate and ENSO-IOD signals. Local climate variables such as precipitation and temperature reflect atmospheric conditions at the observation-station scale. In contrast, the ENSO-IOD signal represents large-scale atmospheric-oceanic circulation anomalies that simultaneously affect precipitation distribution across the entire Indonesian region. Therefore, these two predictors are complementary. The same local climate conditions can have different impacts on production depending on the current ENSO phase [4]. These results are consistent with the findings of Wannasingha *et al.* [2] who demonstrated that the ENSO phase is an important predictor of rice production in Southeast Asia.

Figure 11 shows the province-level RMSE comparison among the full-feature XGBoost model, the model with ENSO-IOD, and the model without additional features and ENSO-IOD. Including ENSO-IOD

generally reduced RMSE across most tested provinces, although the extent of improvement varied across regions. RMSE reductions were observed in several provinces, including Central Java, South Sumatra, West Nusa Tenggara, Southeast Sulawesi, and South Kalimantan, whereas only marginal changes were observed in some other provinces. This pattern indicates that the influence of ENSO-IOD on rice production is region-specific rather than uniform across Indonesia. When ENSO-IOD was combined with additional variables in the full-feature model, the reduction in RMSE became more pronounced across most provinces. The improvement was particularly visible in provinces with higher baseline errors, such as East Nusa Tenggara, Bengkulu, and South Sumatra. This pattern supports the ablation result in Table 5 and indicates that ENSO-IOD contributed not only to aggregate performance but also to improved predictive accuracy at the provincial level.

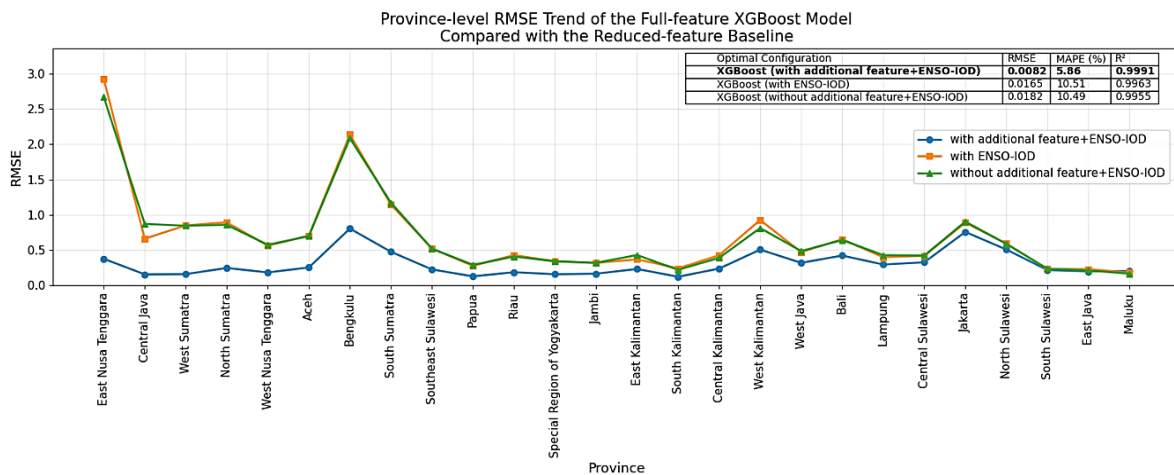


Figure 11. Province-level RMSE comparison across XGBoost models with and without ENSO-IOD integration

This research also conducted preliminary experiments on GAN-based data augmentation to evaluate the impact of adding synthetic samples on prediction accuracy. Several previous studies have shown that GANs have the potential to be used as an augmentation approach for time series data [43]. This experiment compared two approaches, namely vanilla GAN and cGAN conditioned on the ENSO phase.

Table 6 shows that cGAN slightly outperformed vanilla GAN. The cGAN model achieved an RMSE of 0.0189, which was lower than the RMSE of 0.0202 obtained by vanilla GAN. The cGAN model also achieved an R² of 0.9951, which was higher than the R² of 0.9944 obtained by vanilla GAN. These results suggest that conditioning on the ENSO phase improved the quality of the generated samples for XGBoost prediction.

Table 6. Impact of GAN-based data augmentation on XGBoost performance

Optimal configuration	RMSE	R ²
XGBoost	0.0082	0.9991
Vanilla GAN-augmented XGBoost	0.0202	0.9944
cGAN-augmented XGBoost	0.0189	0.9951

Nevertheless, neither augmentation method has yet to surpass the accuracy of the unaugmented XGBoost model. This is due to the limited sample size in the panel data characteristics used. The GAN model was trained independently for each province, so each unit of analysis had only 23 observations. With such a small dataset, the discriminator is prone to overfitting to the training samples, causing the generator to fail to learn the data distribution accurately and instead to inject noise into the model [44]. In addition, synthetic augmentation risks introducing spurious patterns that do not reflect actual variability, which can lead to overconfidence in the model [43]. This risk is even higher in multivariate panel data, where temporal and intercorrelation consistency must be strictly maintained. Thus, GAN effectiveness depends heavily on data availability and requires careful application.

The XGBoost prediction panel shown in Figure 12 shows predictions that are closer to the ideal line than those of BiLSTM-BiGRU and vanilla GAN-augmented XGBoost. This indicates smaller prediction errors across most of the production range. However, some deviations remain at higher production values, suggesting that error dispersion may increase for provinces with very large production volumes.

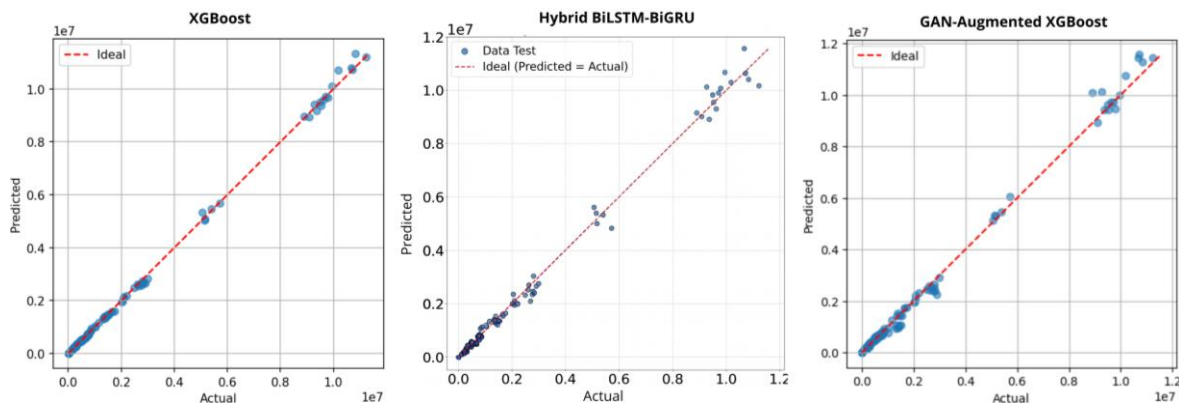


Figure 12. Parity plot: XGBoost, vanilla GAN-XGBoost, and BiLSTM-BiGRU

The high R^2 value was evaluated using the learning curve and residual plot in Figure 13 to assess potential overfitting. Figure 13(a) shows that the errors in both the training and validation data decrease consistently and converge. This indicates that the model learned the data patterns effectively. In Figure 13(b), most residual values are randomly distributed around the zero line within a range of ± 2 standard deviations. Although the error tends to increase at extreme values (heteroscedasticity), this can be explained by the high variation in production scale across provinces.

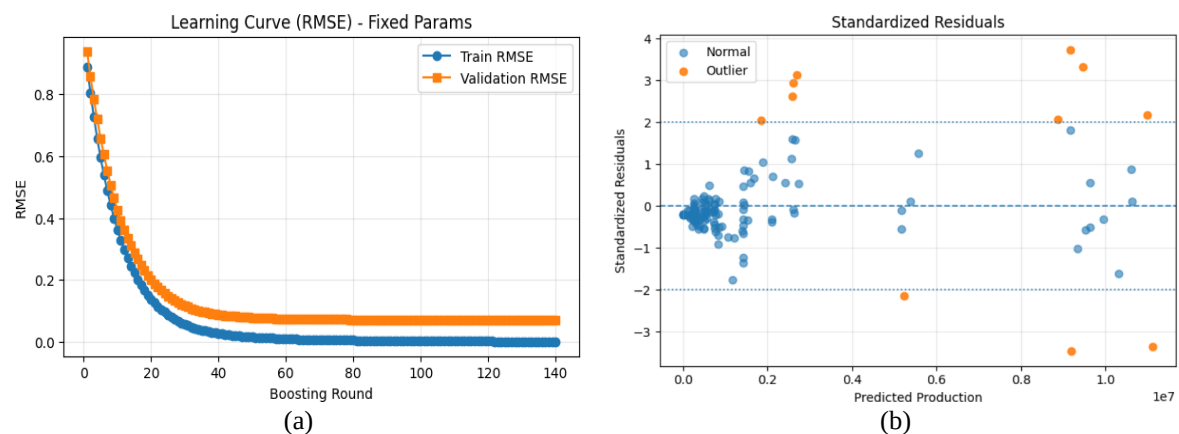


Figure 13. Diagnostic plots of the optimal XGBoost model; (a) learning curve and (b) residual plot

Additional tests resulted in ACF and PACF values at lag-1 of 0.0679. These values were within the confidence interval (± 0.8765) and were therefore not statistically significant. These results indicate that the model captured the primary temporal information in the data and eliminated autocorrelation in the residuals. These diagnostic results collectively demonstrate that XGBoost's high-performance stems from the model's predictive capability, not from overfitting.

To evaluate model generalization over unseen periods, an expanding-window temporal validation was conducted using five successive testing years from 2020 to 2024. The initial training window covered 1997-2019 and was used to predict 2020. The training period was then expanded by one year in each fold until the final model was trained on 1997–2023 to predict 2024. Table 7 presents the mean and standard deviation of the evaluation metrics across the five folds.

As shown in Table 7, the expanding-window temporal validation consistently identified XGBoost as the most stable model, with a mean RMSE of 0.0067, a MAPE of 4.99%, and an R^2 of 0.9994 across the five testing years. The hybrid BiLSTM-BiGRU model maintained a strong correlation with the actual production pattern but produced higher prediction errors, whereas cGAN-Augmented XGBoost demonstrated the weakest temporal generalization. The consistency of these results across successive temporal folds supports selecting the XGBoost model based on the primary evaluation presented in Table 6.

Table 7. Expanding-window temporal validation results

Optimal configuration	RMSE	MAPE	R^2
XGBoost	0.0067±0.0017	4.99±1.09	0.9994±0.0003
cGAN-augmented XGBoost	0.1988±0.0345	91.10±10.23	0.4972±0.1834
BiLSTM-BiGRU	0.0618±0.0369	16.59±6.60	0.9397±0.0581

The selected XGBoost configuration was then compared with previous studies conducted in Indonesia and South India, as presented in Table 8. The proposed model achieved an RMSE of 0.0082 and an R^2 of 0.9991, surpassing the previously reported performance. In addition, the optimal configuration produced a MAPE of 5.86%. These results suggest that for province-year predictions, the recommended approach includes KNN imputation without feature selection and a tuned XGBoost model. This configuration is not only more accurate but also simpler and more robust across the production range, making it well-suited for rice reserve planning and import decisions.

A MAPE value of 5.86% has important practical implications for food policy planning. In provinces producing around 1 million tons, a 5.86% discrepancy is about 58,600 tons. However, in major producing provinces with production exceeding 10 million tons, the same discrepancy can amount to more than 580,000 tons. This difference in scale demonstrates that improved prediction accuracy plays a crucial role in rice reserve planning, determining import requirements, and monitoring regional food security. Therefore, food policy should not rely solely on point estimates; it should also include uncertainty intervals to ensure supply estimates are more reliable as a basis for decision-making.

Table 8. Comparison with previous research

Study	Study area	Type and period of data	RMSE	R^2
Erlin <i>et al.</i> [9]	All provinces in Indonesia	Rice production, harvested area, precipitation, humidity, and temperature (2018–2023).	-	0.9113
Joshua <i>et al.</i> [45]	South India	Rice production and agro-meteorological data (June 2018–January 2019).	0.2295	0.9863
Yoshino <i>et al.</i> [10]	All provinces in Indonesia	Rice production, harvested area, and productivity (2018–2022).	0.015	0.997
This study	All provinces in Indonesia	Rice production, harvested area, local climate data, ENSO index, and IOD (1997–2024).	0.0082	0.9991

The model developed in this research still uses annual provincial-level data. Although the model has accurately mapped historical climate variability, evaluating future extreme climate anomalies that are out of distribution remains an area requiring further investigation. Experiments with synthetic data augmentation also did not improve the model's performance on this dataset, suggesting that this approach requires longer time series to affect model performance.

Future research may use agricultural and climate data with higher spatiotemporal resolution and develop spatiotemporal modeling architectures such as the temporal fusion transformer (TFT). Predictions can also be expanded to severity forecasting to map the risk of production decreases, thereby supporting faster policy responses. For implementation, real-time deployment must also be developed so that the model can serve as an operational monitoring system to support rice reserve planning, import quota determination, and the formulation of preventive policies before a supply deficit occurs. This will contribute to the development of more resilient and sustainable agricultural policies.

4. CONCLUSION

This research evaluated rice production prediction models for Indonesia using XGBoost, GAN-augmented XGBoost, and a hybrid BiLSTM-BiGRU model based on panel data from 1997 to 2024. Unlike previous studies that generally considered only local climate variables, this research integrated

ENSO-IOD signals into a unified rice production prediction framework for Indonesia and evaluated imputation methods and feature selection strategies. The results showed that the XGBoost model achieved the best performance, with an RMSE of 0.0082, an R^2 of 0.9991, and a MAPE of 5.86%, using the KNN method for data imputation and without feature selection. A MAPE of 5.86% means that the predicted rice production differed from the actual production by an average of 5.86%. This level of accuracy can support food policy formulation and implementation by helping government agencies estimate import requirements, determine rice reserve volumes, identify provinces at risk of production shortfalls, and guide reserve release timing and interregional distribution. However, policymakers should combine the model's predictions with expert assessment because even a relatively small percentage error may represent a substantial rice volume, particularly in major producing provinces. Ablation tests also showed that combining additional features with global climate indices (ENSO-IOD) improved prediction accuracy. In contrast, preliminary attempts to use GANs for data augmentation did not improve performance and tended to reduce XGBoost accuracy. These results show that prediction accuracy depends not only on model selection but also on preprocessing strategies and integrating global climate indices. This research has limitations due to the use of annual provincial-level data, which has prevented it from capturing the dynamics of the growing season and spatial variations in greater detail. Future research should examine monthly or regency-level data, develop spatiotemporal models, extend the model output to forecasts of production-shortfall severity, and implement the model in real time.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Indonesia's Ministry of Higher Education, Science, and Technology for the Doctoral Dissertation Research grant under Contract Nos. 0974/LL3/AL.04/2025 and 16.2/LP/UG/VI/2025. The authors also thank Gunadarma University for providing institutional support, administrative assistance, and research facilities that supported this study.

FUNDING INFORMATION

This research was funded by Indonesia's Ministry of Higher Education, Science, and Technology through the Doctoral Dissertation Research grant program, under Contract No. 0974/LL3/AL.04/2025 and 16.2/LP/UG/VI/2025. The funding organization did not participate in the research design, data collection and analysis, interpretation of the findings, decision to publish, or the writing of the manuscript.

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
Reviana Siti Mardiah	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓
Adang Suhendra		✓		✓	✓			✓		✓		✓		✓
Fitrianingsih	✓	✓		✓	✓		✓		✓	✓			✓	✓
Dian Kemala Putri	✓			✓			✓	✓		✓		✓		✓
Saifudin Nasir			✓			✓			✓		✓			
Muhammad Farhan			✓			✓				✓	✓			
Fathurrohman														

C : Conceptualization	I : Investigation	Vi : Visualization
M : Methodology	R : Resources	Su : Supervision
So : Software	D : Data Curation	P : Project administration
Va : Validation	O : Writing - Original Draft	Fu : Funding acquisition
Fo : Formal analysis	E : Writing - Review & Editing	

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY

The datasets supporting the findings of this research were sourced from publicly available repositories provided by the Central Bureau of Statistics of Indonesia, the Meteorology, Climatology, and Geophysics Agency of Indonesia, NOAA, and the Australian Bureau of Meteorology. The corresponding author may provide the cleaned datasets for reasonable academic purposes.

REFERENCES

- [1] G. Octania, "The Government's Role in the Indonesian Rice Supply Chain," Jakarta, Indonesia, 2021. [Online]. Available: <https://repository.cips-indonesia.org/publications/338075/the-governments-role-in-the-indonesian-rice-supply-chain>. (Accessed date: May 3, 2024).
- [2] U. H. Wannasingha, M. Waqas, S. Ahmad, A. Wangwongchai, and P. Dechpichai, "Quantification and prediction of the impact of ENSO on rainfed rice yields in Thailand," *Environmental Challenges*, vol. 19, pp. 1–15, Jun. 2025, doi: 10.1016/j.envc.2025.101123.
- [3] B. Ghose *et al.*, "Rice yield responses in Bangladesh to large-scale atmospheric oscillation using multifactorial model," *Theoretical and Applied Climatology*, vol. 146, no. 1–2, pp. 29–44, Oct. 2021, doi: 10.1007/s00704-021-03725-7.
- [4] T. Mulyaqin, R. Nurmawati, N. Kusnadi, and B. H. Trisasongko, "Spatial Heterogeneity of Rice Production Responses to ENSO Anomalies in Banten Province, Indonesia," *Journal of Applied Agricultural Science and Technology*, vol. 10, no. 1, pp. 105–122, Feb. 2026, doi: 10.55043/jaast.v10i1.503.
- [5] A. Hossain *et al.*, "Agricultural Land Degradation: Processes and Problems Undermining Future Food Security," *Environment, Climate, Plant and Vegetation Growth*, pp. 17–61, 2020, doi: 10.1007/978-3-030-49732-3_2.
- [6] T. Emmanuel, T. Maupong, D. Mpoeleng, T. Semong, B. Mphago, and O. Tabona, "A survey on missing data in machine learning," *Journal of Big Data*, vol. 8, no. 1, pp. 1–37, Oct. 2021, doi: 10.1186/s40537-021-00516-9.
- [7] M. Noorunnahar, A. H. Chowdhury, and F. A. Mila, "A tree based eXtreme Gradient Boosting (XGBoost) machine learning model to forecast the annual rice production in Bangladesh," *PLOS ONE*, vol. 18, no. 3, pp. 1–15, Mar. 2023, doi: 10.1371/journal.pone.0283452.
- [8] A. Satpathi *et al.*, "Comparative Analysis of Statistical and Machine Learning Techniques for Rice Yield Forecasting for Chhattisgarh, India," *Sustainability (Switzerland)*, vol. 15, no. 3, pp. 1–18, Feb. 2023, doi: 10.3390/su15032786.
- [9] Erlin, A. Yuniarta, L. A. Wulandhari, Y. Desnelita, N. Nasution, and Junadhi, "Enhancing Rice Production Prediction in Indonesia Using Advanced Machine Learning Models," *IEEE Access*, vol. 12, pp. 151161–151177, 2024, doi: 10.1109/ACCESS.2024.3478738.
- [10] E. Yoshino, F. I. Kurniadi, and B. Juarto, "Forecasting Rice Production in Indonesia using Regression Techniques: A Comparative Analysis of Support Vector Machine, Linear Regression, and XGBoost Regression," in *10th International Conference on ICT for Smart Society, ICISS 2023 - Proceeding*, Sep. 2023, pp. 1–5, doi: 10.1109/ICISS59129.2023.10291271.
- [11] C. Pavlatos, E. Makris, G. Fotis, V. Vita, and V. Mladenov, "Enhancing Electrical Load Prediction Using a Bidirectional LSTM Neural Network," *Electronics (Switzerland)*, vol. 12, no. 22, pp. 1–13, Nov. 2023, doi: 10.3390/electronics12224652.
- [12] I. K. Friday, S. P. Pati, and D. Mishra, "A Comparative Study of Recurrent Neural Networks for Predicting Indian Bank Stocks," in *3rd Odisha International Conference on Electrical Power Engineering, Communication and Computing Technology, ODICON 2024*, Nov. 2024, pp. 1–6, doi: 10.1109/ODICON62106.2024.10797507.
- [13] J. H. Li *et al.*, "Comparison of the effects of imputation methods for missing data in predictive modelling of cohort study datasets," *BMC Medical Research Methodology*, vol. 24, no. 1, pp. 1–9, Feb. 2024, doi: 10.1186/s12874-024-02173-x.
- [14] Y. Sun, J. Li, Y. Xu, T. Zhang, and X. Wang, "Deep learning versus conventional methods for missing data imputation: A review and comparative study," *Expert Systems with Applications*, vol. 227, p. 120201, Oct. 2023, doi: 10.1016/j.eswa.2023.120201.
- [15] L. Cui, H. Li, K. Chen, L. Shou, and G. Chen, "Tabular Data Augmentation for Machine Learning: Progress and Prospects of Embracing Generative AI," *ACM Computing Surveys*, vol. 58, no. 12, pp. 1–39, Sep. 2026, doi: 10.1145/3808692.
- [16] T. Santhoshkumar and S. Vanila, "Exploratory data analysis and energy predictions with advanced AI and ML techniques," *AI Approaches to Smart and Sustainable Power Systems*, IGI Global Scientific Publishing, pp. 336–370, 2024, doi: 10.4018/979-8-3693-1586-6.ch018.
- [17] A. Blázquez-García, A. Conde, U. Mori, and J. A. Lozano, "A Review on Outlier/Anomaly Detection in Time Series Data," *ACM Computing Surveys*, vol. 54, no. 3, pp. 1–33, Apr. 2022, doi: 10.1145/3444690.
- [18] S. Zhang, A. Khattak, C. M. Matara, A. Hussain, and A. Farooq, "Hybrid feature selection-based machine learning Classification system for the prediction of injury severity in single and multiple-vehicle accidents," *PLoS ONE*, vol. 17, no. 2, pp. 1–19, Feb. 2022, doi: 10.1371/journal.pone.0262941.
- [19] J. Y. L. Chan *et al.*, "Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review," *Mathematics*, vol. 10, no. 8, pp. 1–17, Apr. 2022, doi: 10.3390/math10081283.
- [20] J. Cui *et al.*, "Advanced Short-Term Load Forecasting with XGBoost-RF Feature Selection and CNN-GRU," *Processes*, vol. 12, no. 11, pp. 1–19, Nov. 2024, doi: 10.3390/pr12112466.
- [21] J. Jiang, X. Zhang, and Z. Yuan, "Feature selection for classification with Spearman's rank correlation coefficient-based self-information in divergence-based fuzzy rough sets," *Expert Systems with Applications*, vol. 249, p. 123633, Sep. 2024, doi: 10.1016/j.eswa.2024.123633.
- [22] S. Ozdemir, "Introduction to feature engineering," in *Feature Engineering Bookcamp*, Shelter Island, NY, USA: Manning Publications, 2022, pp. 1–12.
- [23] M. K. Dahouda and I. Joe, "A Deep-Learned Embedding Technique for Categorical Features Encoding," *IEEE Access*, vol. 9, pp. 114381–114391, 2021, doi: 10.1109/ACCESS.2021.3104357.
- [24] M. Riani, A. C. Atkinson, and A. Corbellini, "Automatic robust Box-Cox and extended Yeo-Johnson transformations in regression," *Statistical Methods and Applications*, vol. 32, no. 1, pp. 75–102, Mar. 2023, doi: 10.1007/s10260-022-00640-7.
- [25] K. M. Sujon, R. B. Hassan, Z. T. Towshi, M. A. Othman, M. A. Samad, and K. Choi, "When to Use Standardization and Normalization: Empirical Evidence from Machine Learning Models and XAI," *IEEE Access*, vol. 12, pp. 135300–135314, 2024, doi: 10.1109/ACCESS.2024.3462434.
- [26] Y. Lyu, H. Li, M. Sayagh, Z. M. Jiang, and A. E. Hassan, "An Empirical Study of the Impact of Data Splitting Decisions on the Performance of AIOps Solutions," *ACM Transactions on Software Engineering and Methodology*, vol. 30, no. 4, pp. 1–38, Oct. 2021, doi: 10.1145/3447876.

- [27] H. H. N. Alrashedy, A. F. Almansour, D. M. Ibrahim, and M. A. A. Hammoudeh, "BrainGAN: Brain MRI Image Generation and Classification Framework Using GAN Architectures and CNN Models," *Sensors*, vol. 22, no. 11, pp. 1–21, Jun. 2022, doi: 10.3390/s22114297.
- [28] S. Soleymani and S. Mohammadzadeh, "Comparative Analysis of Machine Learning Algorithms for Solar Irradiance Forecasting in Smart Grids," *arXiv preprint*, 2023, doi: 10.48550/arXiv.2310.13791.
- [29] O. E. Ørebæk and M. Geitle, "Exploring the hyperparameters of XGBoost through 3D visualizations," in *AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering*, 2021.
- [30] B. G. Marcot and A. M. Hanea, "What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis?," *Computational Statistics*, vol. 36, no. 3, pp. 2009–2031, Sep. 2021, doi: 10.1007/s00180-020-00999-9.
- [31] S. Lee *et al.*, "A novel deep learning framework with artificial protozoa optimization-based adaptive environmental response for wind power prediction," *Scientific Reports*, vol. 15, no. 1, pp. 1–34, May 2025, doi: 10.1038/s41598-025-97793-8.
- [32] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Computer Science*, vol. 7, pp. 1–24, Jul. 2021, doi: 10.7717/PEERJ-CS.623.
- [33] S. Das and S. Das, "Statistical Validation of Metaheuristic-Guided Multi-Backbone Fusion: A Comprehensive Analysis for Rice Leaf Disease Detection," in *Proceedings of 2025 IEEE 22nd India Council International Conference, INDICON 2025*, Dec. 2025, pp. 1–7, doi: 10.1109/INDICON68490.2025.11392870.
- [34] Bureau of Meteorology, "Southern Oscillation Index (SOI) since 1876," *Climate, SOI*, 2016, [Online]. Available: <https://www.bom.gov.au/climate/enso/soi/>. (Accessed date: September 1, 2025).
- [35] A. Bailly *et al.*, "Effects of dataset size and interactions on the prediction performance of logistic regression and deep learning models," *Computer Methods and Programs in Biomedicine*, vol. 213, pp. 1–7, Jan. 2022, doi: 10.1016/j.cmpb.2021.106504.
- [36] F. M. Siemers, C. Feldmann, and J. Bajorath, "Minimal data requirements for accurate compound activity prediction using machine learning methods of different complexity," *Cell Reports Physical Science*, vol. 3, no. 11, pp. 1–20, Nov. 2022, doi: 10.1016/j.xcrp.2022.101113.
- [37] D. R. I. M. Setiadi, A. Susanto, K. Nugroho, A. R. Muslikh, A. A. Ojugo, and H. S. Gan, "Rice Yield Forecasting Using Hybrid Quantum Deep Learning Model," *Computers*, vol. 13, no. 8, p. 191, Aug. 2024, doi: 10.3390/computers13080191.
- [38] Z. Shi, Y. Hu, G. Mo, and J. Wu, "Attention-based CNN-LSTM and XGBoost hybrid model for stock prediction," *arXiv preprint*, 2022, doi: 10.48550/arXiv.2204.02623.
- [39] S. Bakhtiari, C. Aldrich, V. M. Calo, and M. Iannuzzi, "XGBoost model for the quantitative assessment of stress corrosion cracking," *npj Materials Degradation*, vol. 8, no. 1, pp. 1–19, 2024, doi: 10.1038/s41529-024-00538-7.
- [40] X. Bouthillier *et al.*, "Accounting for Variance in Machine Learning Benchmarks," in *Proceedings of Machine Learning and Systems*, vol. 3, no. 747–769, 2021.
- [41] A. Deng, "Time series cross validation: A theoretical result and finite sample performance," *Economics Letters*, vol. 233, p. 111369, Dec. 2023, doi: 10.1016/j.econlet.2023.111369.
- [42] E. Spiliotis, F. Petropoulos, and V. Assimakopoulos, "On the Disagreement of Forecasting Model Selection Criteria," *Forecasting*, vol. 5, no. 2, pp. 487–498, Jun. 2023, doi: 10.3390/forecast5020027.
- [43] E. Brophy, Z. Wang, Q. She, and T. Ward, "Generative Adversarial Networks in Time Series: A Systematic Literature Review," *ACM Computing Surveys*, vol. 55, no. 10, pp. 1–31, Oct. 2023, doi: 10.1145/3559540.
- [44] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila, "Training generative adversarial networks with limited data," *Advances in Neural Information Processing Systems*, vol. 33, pp. 12104–12114, 2020.
- [45] V. Joshua, S. M. Priyadharson, and R. Kannadasan, "Exploration of machine learning approaches for paddy yield prediction in eastern part of Tamilnadu," *Agronomy*, vol. 11, no. 10, pp. 1–19, Oct. 2021, doi: 10.3390/agronomy11102068.

BIOGRAPHIES OF AUTHORS



Reviana Siti Mardiah    received a Master's degree in Industrial Engineering and is currently pursuing a Ph.D. in Information Technology at Gunadarma University, Indonesia. Her earlier research focused on fatigue, alertness, and sleep. Currently, her research interests have expanded to include machine learning, deep learning, generative AI, and predictive modeling, particularly in the context of food security and supply chain systems. She can be contacted at email: revianasitimardiah@staff.gunadarma.ac.id.



Adang Suhendra    received a Ph.D. in Informatics Engineering from the Department of Electrical Engineering, University of Kassel, Germany, in 2001. He is currently a professor in the postgraduate program at Gunadarma University, Indonesia, where he teaches courses including information systems analysis and design, software engineering, cryptography, computer network management, database systems, and system testing and implementation. He also lectures in the Department of Informatics Engineering for courses such as computer graphics and object-oriented programming, and contributes to mathematics-related subjects. His research interests include mathematics and information systems. He can be contacted at email: adang@staff.gunadarma.ac.id.



Fitrianingsih    received a Ph.D. in Information Technology from Gunadarma University, Indonesia. She is currently a lecturer in Informatics Management, Information Technology Directorate of Diploma Program, Gunadarma University, Indonesia. Her research interests include system analysis, image processing, and databases. She can be contacted at email: fitrianingsih@staff.gunadarma.ac.id.



Dian Kemala Putri    received a Ph.D. in Engineering Psychology from Université de Bourgogne, France. She is currently active as a lecturer and researcher at Gunadarma University, with expertise in Engineering Psychology, Industrial Engineering, Human Factors Engineering, and Industrial-Organizational Psychology. She can be contacted at email: dian@staff.gunadarma.ac.id.



Saifudin Nasir    received his Bachelor of Engineering (B.Eng.) in Informatics Engineering from Gunadarma University, Indonesia, in 2025. His primary research interests lie in applying machine learning to environmental data analysis and developing internet of things (IoT) systems. His final project involved developing a web-based prediction system using the random forest algorithm to map mangrove carbon and biomass. He can be contacted at email: saifudinnasirs@gmail.com.



Muhammad Farhan Fathurrohman    received a Bachelor's degree in Informatics Engineering from Gunadarma University, Indonesia, in 2025. He is currently a Mobile Engineer Intern at Alfagift-Global Loyalty Indonesia in Tangerang, Indonesia, and an IT Infrastructure Support Engineer at LePKom, Gunadarma University, Jakarta, Indonesia. His research interests include mobile application development, particularly Android development using Kotlin and Jetpack Compose, API integration, and software architecture. He is also certified as a Junior Mobile Programmer by BNSP. He can be contacted at email: farhanfathur73@gmail.com.