

INNOVATION IN CROP NUTRITION PLANNING BASED ON RAINFALL PREDICTION USING SINGULAR SPECTRUM ANALYSIS AND BOOSTING TO OPTIMIZE AGRICULTURAL MANAGEMENT

Yuslena Sari^{1*}, Mambang², Muhammad Zulfadhilah³, Subhan Panji Cipta⁴, Muhammad Nursandi⁵, Finki Dona Marleny⁶, Ricardus Anggi Pramunendar⁷, Sunardi Sunardi⁸, Eka Setya Wijaya⁹, Andreyan Rizky Baskara¹⁰, Aurelia Monica Sari¹¹, Muhammad Alkaff¹²

Department of Informayion Technology, Universitas Lambung Mangkurat, Banjarmasin, 70123, Indonesia^{1,9,10,11,12}

Information Technology, Faculty of Science and Technology, Sari Mulia University, Banjarmasin, Indonesia^{2,3,4,5}

Informatics, Faculty of Engineering, University of Muhammadiyah, Banjarmasin, Indonesia⁶

Department of Informatics Engineering, Semarang, Faculty of Computer Science, Universitas Dian Nuswantoro, 50131, Indonesia⁷

Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Lambung Mangkurat, Banjarbaru, 70714, Indonesia⁸

Department of Computer Science, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia¹²

yuzlena@ulm.ac.id

Received: 08 March 2025, Revised: 25 January 2026, Accepted: 13 February 2026

*Corresponding Author

ABSTRACT

The high variability of rainfall in tropical climates presents a major challenge for agricultural management, as weather uncertainty often leads to inefficient fertilization practices due to nutrient loss. This study aims to develop a robust framework for rainfall prediction, which can inform a flexible and precise crop nutrient scheduling system. Utilizing an hourly rainfall dataset ($n=6,624$) obtained from IoT sensors, the research proposes an approach that integrates Singular Spectrum Analysis (SSA) for signal decomposition and noise reduction with Gradient Boosting algorithms (LightGBM and XGBoost). Spline interpolation was employed to handle missing data, while SSA served to disentangle deterministic trends from random noise, enabling the models to perform more effectively on the refined dataset. Empirical evaluation demonstrates that the SSA-XGBoost hybrid model achieves superior performance, with an RMSE of 0.0057 and an R^2 of 0.8278, significantly outperforming the SSA-LightGBM model (R^2 0.2879), which struggled to capture non-linear patterns within this dataset. The high predictive accuracy of the SSA-XGBoost model facilitates the implementation of responsive nutrient management strategies, wherein fertilizer application can be deferred during forecasted periods of high rainfall to prevent runoff and environmental pollution. This research contributes to the field of hydroinformatics by demonstrating the effectiveness of combining SSA and XGBoost as a cost-efficient yet high-performance solution for mitigating climate-related risks in tropical wetland agriculture.

Keywords : LightGBM, Prediction, Rainfall, Singular Spectrum Analysis, XGBoost

1. Introduction

Rainfall is one of the major factors that affect plant growth in various agricultural ecosystems. The availability of water from rainfall not only helps the photosynthesis process, but also affects the availability of nutrients in the soil. The interaction between rainfall and nutrients is critical as rainfall can accelerate the leaching of nutrients in the soil, causing plants to lack essential nutrients (Dupas et al., 2024; Elbasiouny et al., 2022b; Komatsu et al., 2026; J. Li et al., 2025; S. Li et al., 2025). On the other hand, too little rain can limit nutrient uptake by plant roots, resulting in slow growth and lower yields.

In the context of global climate change, rainfall patterns are becoming increasingly erratic, making it difficult for farmers to determine the right time to apply nutrients to crops. Therefore, it is important to develop nutrient management strategies that consider rainfall patterns. Providing the right nutrients at the right time can improve resource use efficiency and maintain farm sustainability.

Optimal nutrient delivery is key in increasing crop yields while maintaining ecosystem balance (Ding et al., 2025; Dong, 2024; Liu et al., 2025; Valladares-Castellanos et al., 2024; Y. Yang & Ma, 2025). Proper crop nutrient management is essential in modern agriculture, where excess or deficiency of nutrients can have a significant impact on crop health and yield quality. In the context of climate change that increases rainfall variability, the development of a weather prediction-based nutrient scheduling system is an urgent need. This research comes with a novelty through the development of a crop nutrient scheduling system that leverages rainfall prediction, aiming to maximize water and nutrient use efficiency and minimize the negative impacts of weather variability. This solution is designed to help farmers avoid potentially environmentally damaging over-fertilization and reduce the risk of nutrient deficiencies in crops, which often reduce the quality and quantity of crop yields (Murumkar et al., 2025).

The agricultural sector in tropical nations, particularly in archipelagic countries such as Indonesia, is currently under significant pressure from the combined effects of global climate variability and the imperative to ensure sustainable food security. In tropical farming systems, rainfall functions as the primary stochastic factor, governing not only water availability for photosynthesis but also fundamentally modulating the physico-chemical dynamics of soil nutrients. Rainfall plays a dual role; while it facilitates nutrient uptake through mass flow and diffusion, it can simultaneously trigger detrimental leaching and surface runoff (T. T. Cao et al., 2025; Mora-Motta et al., 2025; Paramesha et al., 2025; Singh et al., 2025; X. Yang et al., 2025). Consequently, balanced water management is paramount for enhancing input efficiency in agriculture.

However, alongside climate change, rainfall patterns are becoming increasingly erratic, evidenced by shifts in monsoon onset, heightened extreme weather events, and pronounced seasonal fluctuations. This instability renders traditional methods, such as fixed planting calendars or historically based local knowledge, less relevant for anticipating actual climatic conditions (Getachew et al., 2025; Olaleye et al., 2026). The inability to adapt to these forecasts adversely affects fertilizer-use efficiency. Heavy rainfall occurring shortly after fertilizer application (>50 mm/day) can accelerate the dissolution and leaching of nutrients, particularly mobile nitrate (NO_3^-), leading to economic losses and environmental pollution. Conversely, unanticipated rainfall deficits can impede nutrient uptake due to reduced soil water potential, potentially causing plant growth stagnation.

In response to this uncertainty, modernizing agricultural management necessitates a paradigm shift from reactive practices toward data-driven precision agriculture. The adoption of Machine Learning (ML) and the Internet of Things (IoT) offers a pathway to mitigate these risks. Previous research has employed ML to enhance the accuracy of rainfall prediction in tropical regions. Studies by (Baskara et al., 2024; Dong, 2024; Samad et al., 2020) utilized Long Short-Term Memory (LSTM) models to analyze temporal patterns, while proposed a hybrid methodology. In the Indonesian context, (Arassah et al., 2025; Puspasari et al., 2023) demonstrated the effectiveness of Random Forest and ensemble models in projecting extreme weather events.

Although the potential of these technologies is widely acknowledged, a significant gap persists between the capabilities of standard algorithms and the complexity of tropical rainfall-related data. The core challenge lies in the data's inherently noisy nature, stemming from localized convective storm phenomena. While powerful, boosting models such as Light Gradient Boosting Machine (LightGBM) and Extreme Gradient Boosting (XGBoost)—despite their efficacy in handling non-linear relationships—are highly sensitive to noise (Hameed et al., 2025; Pringandana & Kusnawi, 2025). Without appropriate preprocessing, these models are prone to overfitting, wherein the algorithm "memorizes" noise rather than learning the underlying subtle climatic trends. Consequently, the resulting predictions exhibit low accuracy during testing and prove unreliable for high-stakes decision-making, such as fertilization scheduling.

Therefore, this study posits that the path to accurate rainfall prediction lies not merely in algorithmic selection but necessitates advanced data processing to disentangle genuine signals from interference. This research proposes a novel hybrid framework integrating a signal decomposition method, Singular Spectrum Analysis (SSA), with ensemble boosting models

(comprising LightGBM and XGBoost). SSA functions as a critical preprocessing stage, deconstructing the rainfall time series into its constituent trend, seasonal, and noise components. By reconstructing the data using only the signal components and discarding the noise, this study provides "cleaned" input for the boosting models—an approach supported by recent findings from (Diniaty et al., 2024; Narimani et al., 2023; Tian et al., 2025) in hydrology.

This study makes significant contributions in three principal aspects. First, it offers a methodological innovation by implementing a Denoising-Ensemble technique specifically designed to address the limitation of boosting models when confronted with the volatile fluctuation characteristic of tropical data. Second, it advances the development of a Decision Support System (DSS) logic that translates rainfall prediction outputs into specific nutrient management actions—such as deferring fertilizer application during projected periods of high rainfall—to foster climate-resilient farming, aligning with insights from (Qin et al., 2023). Third, it provides contextual validation using a high-quality, high-frequency (hourly) dataset from in-situ sensors at Universitas Lambung Mangkurat, yielding a precise representation of climate dynamics in an often-underrepresented wet tropical region within the global literature.

The main innovation in this research lies in the adaptive approach that considers weather conditions and the specific needs of plants in each phase of their growth. The research not only focuses on analysing the relationship between rainfall and nutrient uptake by plants, but also develops an accurate rainfall prediction model to plan the optimal timing of nutrient application. By considering dynamic rainfall patterns, the research devised a more adaptive nutrient scheduling strategy, thereby improving resource use efficiency while maintaining ecosystem balance. Another novelty is the assessment of the impact of this scheduling system on crop yield and ecosystem balance. Through a comprehensive analysis, the research evaluates the potential for reducing the risk of nutrient leaching due to excessive rainfall, which often leads to reduced fertiliser use efficiency and environmental pollution.

By integrating more accurate weather prediction technology in the nutrient scheduling system, this research offers a practical solution that can be applied in the field, benefiting farmers in improving the sustainability of farming practices and resilience to climate variability. Achieving the goal of developing an optimal nutrient scheduling system based on rainfall prediction requires analytical methods that can capture complex and irregular rainfall data patterns due to climate variability. One method that is very useful in analysing time series data such as rainfall is Singular Spectrum Analysis (SSA). SSA is a technique capable of decomposing time series into trend, cycle and noise components, so that hidden patterns in rainfall data can be better identified. By utilising SSA, this research will identify the long-term pattern and seasonal variability of rainfall, which is then used for prediction (Tian et al., 2025)

In addition, in an effort to improve the accuracy of rainfall prediction and model nutrient scheduling, Boosting-based machine learning methods, such as LightGBM (Light Gradient Boosting Machine) and XGBoost (Extreme Gradient Boosting), will be used (Cui et al., 2021). Both algorithms are renowned for their ability to handle large and complex data and have the advantage of performing accurate predictions considering multiple variables. By combining the decomposition results from SSA with the Boosting model, a more adaptive and robust system to rainfall variations can be built.

LightGBM and XGBoost were chosen because they can overcome overfitting problems and can perform predictions quickly without sacrificing accuracy (Narimani et al., 2023). By using weather variables such as rainfall, temperature, humidity, as well as historical crop nutrient data, the model is expected to provide more precise nutrient scheduling recommendations in accordance with local climate conditions. The combination of SSA for rainfall pattern analysis and Boosting for prediction will be an innovative approach in this research, enabling more efficient crop nutrient scheduling and positively impacting agricultural yields.

2. Literature Review

a) Rainfall Rate

Rainfall in Indonesia is influenced by many meteorological and geographical factors. The vast area of Indonesia that lies below the equator experiences significant rainfall. Rainfall

patterns that vary between regions are influenced by the orography, topography and geography of the islands in Indonesia. Rainfall plays an important role in human life, affecting the agricultural sector, plantations, and causing the risk of floods and landslides (Diykh et al., 2026; Ghosh et al., 2025; Mashooq et al., 2025; Sutanto et al., 2025; R. Wang et al., 2026; Yamada et al., 2025)

b) Rainfall Prediction Studies

Several studies have applied machine learning techniques to improve rainfall prediction accuracy, particularly in tropical regions characterized by high variability, extreme rainfall intensity, and strong seasonal cycles. Samad et al. (Dong, 2024; Samad et al., 2020) utilized LSTM-based models that effectively captured rainfall patterns and improved predictive accuracy, while (Hu et al., 2025; Johny et al., 2022; G. D. Kumar et al., 2024) found that LSTM in rainfall forecasting. In India Patil and Kulkarni (2023) proposed a hybrid ML approach combining physics-based and data-driven principles, achieving up to 25% improvement over traditional models. Sawah (Sawah, 2026) further demonstrated that ML algorithms such as ANN, SVM, and ensemble learning outperform conventional statistical techniques in modeling nonlinear rainfall patterns. In Indonesia, Puspasari et al., (2023) applied a Random Forest model for rainfall and flood prediction with an accuracy of 94.93%, Arassah et al., (2025) combined SVC–RNN achieving 98.5% classification accuracy, and Y. Wang et al., (2026) developed an adaptive stacking ensemble model with Bayesian Optimization that produced low MSE across multiple regions. These studies confirm that machine learning approaches provide more reliable and accurate rainfall predictions in tropical climates compared to traditional models, effectively addressing the challenges posed by complex and highly variable weather conditions.

c) Time Series Prediction

Forecasting is the process of predicting future events based on past and present data. The methods are divided into qualitative, based on subjective factors and quantitative, using historical data. Quantitative techniques include time series and regression models, which focus on analyzing temporal patterns such as trend, seasonal, periodic, and irregular (Alkaff et al., 2019; Jiang et al., 2024; C. Li et al., 2026; Lin et al., 2026; Mehdizavareh et al., 2025; C. Wang et al., 2026; W. Wang et al., 2025; Wu et al., 2025).

d) Internet of Thing (IoT)

This research uses an Internet of Things (IoT)-based monitoring tool with an automatic type of tipping bucket. IoT connects sensors, actuators, and physical objects to the internet to collect data, including rainfall measurements for forecasting. Data from rainfall sensors are stored in the cloud (firebase) and can be accessed flexibly via the internet using SIM modules for sending and receiving data (Ahire et al., 2022; Mahdavejad et al., 2018; Zakaria and Michael, 2017).

e) Singular Spectrum Analysis (SSA)

Singular Spectrum Analysis (SSA) is a method rooted in Karhunen–Loève transformation theory, employed to extract underlying components—such as trend, seasonal patterns, and noise—from time series data. Through its three core stages of embedding, singular value decomposition (SVD), and reconstruction, SSA effectively filters out noise, thereby clarifying dominant patterns and enhancing predictive accuracy. Owing to its capacity to handle nonlinear and complex data structures, SSA has been widely adopted in climatological and hydrological analyses. For instance, Diniaty et al., (2024) integrated SSA with machine learning techniques to improve monthly rainfall forecasting accuracy. Similarly, Cui et al. (2021) demonstrated that applying SSA to river flow data effectively reduces noise prior to modeling with LightGBM, leading to superior prediction performance.

f) Light Gradient Boosting Machine (LightGBM)

The Light Gradient Boosting Machine (LightGBM) is a machine learning method developed by Microsoft Research Asia in 2016, based on the Gradient Boosting Decision Tree (GBDT) algorithm. LightGBM is known for its computational efficiency, high accuracy, low

memory consumption, and ability to handle large-scale datasets effectively (Bui et al., 2021; Cui et al., 2021; Narimani et al., 2023). Compared to other ensemble learning methods, each algorithm exhibits unique strengths and weaknesses when applied to time-series prediction. XGBoost, for instance, offers effective L1/L2 regularization and robust pruning mechanisms that enhance model stability for non-linear data, but it tends to be slower than LightGBM (Bentéjac et al., 2021). LightGBM, on the other hand, achieves faster training and better scalability for large datasets, although it can be prone to overfitting when handling imbalanced data (Q. Cao et al., 2023; Yao et al., 2022). Meanwhile, CatBoost and Random Forest (RF) models are generally more interpretable and resistant to overfitting, but they may be less efficient for continuous or high-dimensional time-series data (Sahin, 2020). An illustration of the LightGBM workflow is presented in Fig. 1.

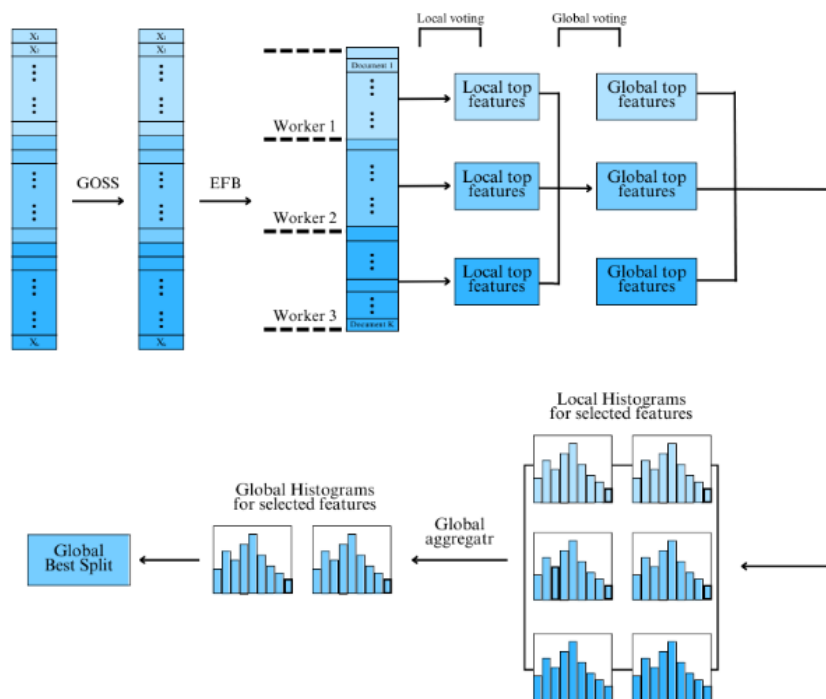


Fig. 1. Illustration of LightGBM Method Flow

Figure 1 illustrates the core operational mechanism of the LightGBM algorithm, highlighting its Gradient-Based One-Sided Sampling (GOSS) approach and leaf-wise tree growth strategy. In contrast to classical level-wise tree development methods, this figure demonstrates how LightGBM prioritizes splits at leaves with the greatest gradients or prediction errors, represented by darker or more dominant data subsets. The process underscores computational efficiency, as data with small gradients (already well-predicted instances) are either down-sampled or randomly selected, while the model focuses more intensively on learning from complex, high-gradient data. This elucidates why LightGBM can achieve substantially faster model training compared to other ensemble techniques.

LightGBM introduces Gradient-Based One-Sided Sampling (GOSS) and Exclusive Feature Bundling (EFB) methods. GOSS organises learning based on decreasing absolute value of gradient, retains subset 'A' with top- $a \times 100\%$ highest gradient cases, and randomly selects subset B with $(1-a) \times 100\%$ from size of $b \times |A|$. Consistent objects are partitioned into vectors based on the estimated variance of subsets A and B shown in the following equation (V. Kumar et al., 2023).

(1)

$$\tilde{V}_j(d) = \frac{1}{n} \left(\frac{\left(\sum_{x_i \in A_\tau} g_i + \frac{1-a}{b} \sum_{x_i \in B_\tau} g_i \right)^2}{n_\tau^j(d)} + \frac{\left(\sum_{x_i \in A_\gamma} g_i + \frac{1-a}{b} \sum_{x_i \in B_\gamma} g_i \right)^s}{n_\gamma^j(d)} \right)$$

Description:

$\tilde{V}_j(d)$ = Function value \tilde{V} at the j-th iteration and parameter d

n = total value of data or sample

A_τ = A subset of the data that fulfils the conditions or criteria τ

B_τ = Another subset of the data that fulfils the conditions or criteria τ

g_i = Weight or function value of the i-th data

a = A constant that governs the proportion of data in the subset A_τ

b = A constant that governs the proportion of data in the subset B_τ

$n_\tau^j(d)$ = Total data in subset A_τ on the j-th iteration and parameter d

A_γ = A subset of the data that fulfils the conditions or criteria γ

B_γ = Another subset of the data that fulfils the conditions or criteria γ

$n_\gamma^j(d)$ = Total data in subset A_γ on the j-th iteration and parameter d

s = Exponent or power applied to the sum in a subset A_γ and B_γ

Where $A_\tau = \{x_i \in A: x_{ij} \leq d\}$, $A_\gamma = \{x_i \in A: x_{ij} > d\}$, $B_\tau = \{x_i \in B: x_{ij} \leq d\}$, $B_\gamma = \{x_i \in B: x_{ij} > d\}$ and to normalize B value to size of A_c , coefficient $(1 - a)/b$ is used.

Exclusive Feature Bundling (EFB) in LightGBM is an automatic feature selection that combines separate features, such as categorical inputs, to reduce the dimensionality of the data without losing information. It changes the complexity from $O(\text{data} \times \text{features})$ to $O(\text{data} \times \text{bundles})$, where the number of bundles is much smaller than the original number of features. This change speeds up the learning time by 20 times, making LightGBM a gradient-enhanced decision tree with GOSS and EFB implementations (V. Kumar et al., 2023).

g) Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) method, proposed by Tianqi Chen in 2016, is an ensemble learning algorithm based on boosting method and has attracted great attention in machine learning applications. The main formula of this method is as follows (Bentéjac et al., 2021).

$$\hat{y}_i = f_1(x_i) + f_2(x_i) + \dots + f_k(x_i) \quad (2)$$

Description:

\hat{y}_i = Predicted or estimated target value for i-th data

$f_1(x_i), f_2(x_i), \dots, f_k(x_i)$ = Predictor functions applied to the data of x_i

x_i = Data atau fitur-fitur dari data ke-i

An illustration of the flow of the XGBoost method can be seen in Fig. 2

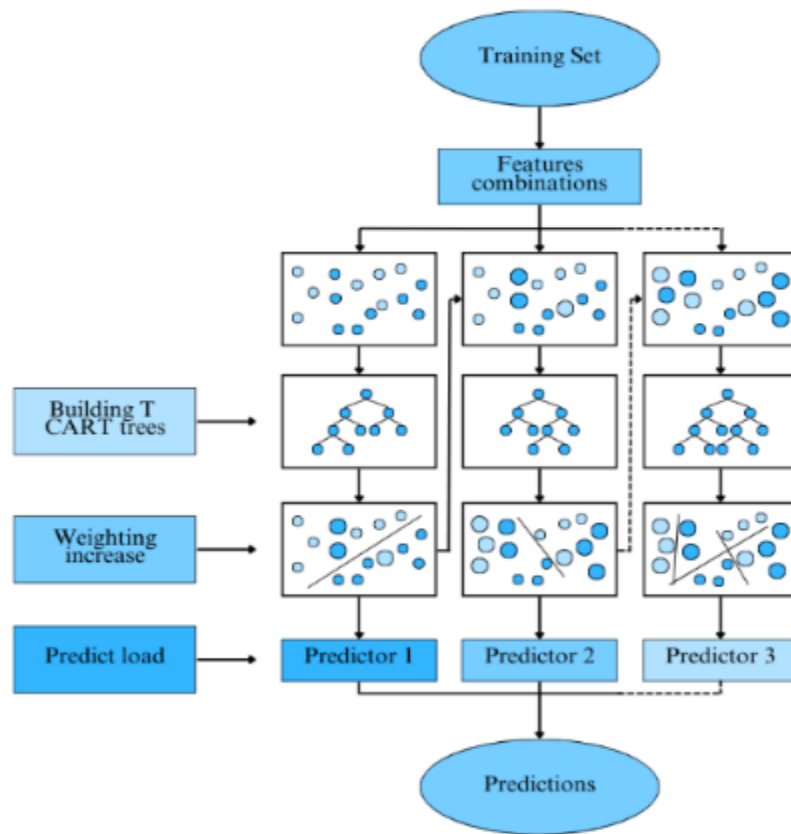


Fig. 2. Illustration of XGBoost Method Flow

Figure 2 represents the principle of Additive Training in XGBoost. The diagram depicts the sequential process wherein each new decision tree (Tree1, Tree2, ..., TreeK) is added not independently, but to correct the residuals (errors) from the preceding tree. The final predicted value (\hat{y}_i) is the sum of scores from all trees in the sequence ($\sum f_k(x_i)$). This visualization reinforces XGBoost's strength in incrementally reducing bias through a gradient descent mechanism on the objective function, which is further augmented by regularization terms to prevent overfitting on highly variable data such as rainfall.

h) Min-Max Normalization

Min-Max Normalisation is a simple data scaling method that converts data values into the range [0, 1]. In addition to facilitating computation, this method also helps find the global optimal solution. The equation to perform min-max normalisation is as follows (Sahin, 2020):

$$x_i^{new} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (3)$$

Description:

x_i^{new} = normalized data

x_i = data that will be normalized

x_{min} = minimal value on dataset

x_{max} = maximal value on dataset

i) Model Evaluation

1) Root Mean Squared Error (RMSE)

Root Mean Squared Error (RMSE) calculates the error between the predicted value and the actual

value, and then finds its root. RMSE is a popular continuous variable performance metric, providing an objective picture of prediction performance. The lower the RMSE value, the more accurate the method. The basic assumption is that the error value should be unbiased and follow a normal distribution. The equation for calculating RMSE is as follows (Diykh et al., 2026):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

Description:

n = total number of data

y_i = actual value

\hat{y}_i = result of prediction value

2) Mean Absolute Error (MAE)

Mean Squared Error (MAE) calculates the error between the predicted value and the actual value, then it is converted into absolute value. MAE is used for model evaluation and is useful if outliers represent corrupted data. The equation to calculate MAE is as follows (Luo et al., 2021):

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

Description:

n = total number of data

y_i = actual value

\hat{y}_i = prediction value

3) R-Squared (R^2)

R-squared (R^2) is the proportion of variance in the dependent variable that can be predicted from the independent variables. As a coefficient of determination with values between 0 and 1, R-squared indicates how much variation in the independent variable affects the dependent variable. The closer to a value of 1, the better the model. The equation to calculate R^2 is as follows (Sahin, 2020):

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (6)$$

Description:

n = total number of data

y_i = actual value

\hat{y}_i = prediction value

\bar{y} = averaged value

3. Research Method

a. Rainfall Prediction Model

Rainfall prediction is performed using the Singular Spectrum Analysis (SSA) method which is able to separate data components into trend, seasonality, and noise. Then, two Boosting models, namely LightGBM and XGBoost, are used to make predictions based on the decomposed data. The steps performed in this method are as follows:

- 1) Rainfall Data Collection: The historical rainfall data used in this study was taken from local weather stations or open weather datasets, such as BMKG or global meteorological data sources. The data used includes several historical years to capture seasonal variations and long-term trends.

- 2) Data Pre-processing: The rainfall data obtained was pre-processed to overcome the problem of missing data or outlier data. Interpolation and normalisation techniques are used to ensure that the data is ready for further processing by the prediction algorithm.
- 3) Data Decomposition with Singular Spectrum Analysis (SSA): After data pre-processing, SSA was applied to decompose the rainfall time series into principal components, namely:
 - Trend: Indicates long-term patterns that occur in rainfall data.
 - Seasonal: Identifying seasonal fluctuations, such as higher rainfall in the wet season.
 - Noise: It is a component that has no particular pattern and can be ignored for predictive analysis.
- 4) Application of Boosting Models: After decomposing the data, the LightGBM and XGBoost models were applied to perform rainfall prediction based on the trend and seasonal components generated by SSA. These two models were chosen for their advantages in handling large data, their ability to manage complex variables, and their efficiency in prediction.
- 5) The LightGBM model is used to perform rainfall prediction based on weather variables such as temperature, humidity, and relevant air pressure. The algorithm is fast in model training and has the ability to handle large-scale data with high efficiency.
- 6) XGBoost was used to compare the prediction results of LightGBM. XGBoost also handles outliers and noise in the data well, and is well known for minimising overfitting.
- 7) Model Performance Evaluation: The performance of the rainfall prediction model was evaluated using several evaluation metrics, such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R-squared. (R^2). These metrics are used to measure how well the model is able to predict rainfall based on existing data.
- 8) Model Validation: To ensure the accuracy of the predictions, cross-validation was performed by dividing the data into subsets. The most accurate prediction model was further tested using separate test data to measure the generalisability of the model in predicting future rainfall.

With this method, it is expected that the resulting prediction model will have high accuracy in forecasting rainfall based on the historical patterns that have been analysed. The model can then be integrated into a crop nutrient scheduling system to improve the efficiency of water and nutrient use.

b. Nutrient Scheduling

Nutrient application is designed to be adjusted according to the resulting rainfall prediction. When high rainfall is predicted, nutrient application is reduced to prevent nutrient leaching, while at low rainfall predictions, nutrient application is increased. The steps involved in scheduling nutrient delivery are as follows:

- 1) Rainfall Classification Based on Prediction: The rainfall predictions generated by the models (LightGBM and XGBoost) are classified into several categories, such as (Murumkar et al., 2025):
 - High rainfall (> 50 mm/day): High nutrient leaching potential.
 - Moderate rainfall (20-50 mm/day): Moderate risk of nutrient leaching.
 - Low rainfall (< 20 mm/day): Nutrient delivery is more effective due to minimal leaching.
- 2) Adjustment of Nutrient Feeding Schedule: Based on the rainfall prediction results from the above categories, adjustments to the nutrient feeding schedule are made as follows.
 - High rainfall: Nutrient application is reduced or temporarily delayed until rainfall is reduced, to avoid leaching of nutrients. If nutrients are applied, a slow-release form of fertilizer is used which reduces the risk of leaching.
 - Rainfall is moderate: Nutrient application is done with adjusted doses, considering the risk of partial leaching. The use of quick-acting fertilizers is preferred.
 - Low rainfall: Nutrient delivery can be increased to ensure crops get enough nutrients as the risk of leaching is minimal. Liquid fertilizer or nutrients delivered through irrigation can be applied to ensure absorption efficiency.
- 1) Integration with Plant Growth Phase: Nutrient schedules are also customized according

to the growth phase of the crop. For example, in the early vegetative phase, plants require more nitrogen for leaf growth, while in the generative phase, the focus is more on phosphorus and potassium to support fruit or seed formation. Rainfall predictions will determine the intensity and type of nutrients applied in each phase

- 2) Use of Weather and Crop Monitoring Technology: To ensure proper scheduling, weather and soil sensors are used to monitor field conditions in real-time. Actual rainfall data, soil moisture and crop nutrient requirements are synchronized with model predictions to provide more accurate information regarding the timing and amount of nutrient application.
- 3) Dynamic Evaluation and Adjustment: This nutrient delivery scheduling system is evaluated periodically. Based on crop growth results and actual weather conditions, adjustments are made to ensure nutrient efficiency remains optimal. The rainfall prediction model will be continuously updated with new data, so that the nutrient delivery schedule can be adjusted more dynamically following changes in rainfall patterns.

This approach is expected to make nutrient delivery more efficient, reduce the risk of nutrient leaching, increase crop yields, and preserve the environment.

c. Data and Implementation

Hourly rainfall data from Lambung Mangkurat University sensor is used as dataset. The prediction process is performed using SSA method for data decomposition, and Boosting for rainfall prediction.

- 1) Data Collection: This study utilized critical hydrological data collected from an Internet of Things (IoT)-enabled automatic weather station (AWS) deployed at Universitas Lambung Mangkurat in South Kalimantan. The instrumentation featured a tipping-bucket rain gauge, calibrated for precise rainfall intensity measurement. A GSM module facilitated automated data transmission by sending telemetry packets to a cloud-based Firebase database, enabling continuous, unmanned data logging. The dataset comprises 6,624 hourly rainfall entries, capturing the characteristic diurnal and seasonal fluctuations of the tropical monsoon climate. This detailed temporal dataset was essential for training the proposed machine learning models to identify rapid shifts in weather patterns.
- 2) Implementation: Implementation of LightGBM and XGBoost methods with data decomposition techniques using SSA for rainfall prediction. The implementation stage will be carried out using Jupyter Notebook tools. The programming language used is python.
- 3) Dataset Input: The dataset used in this study is hourly rainfall data with a total of 6624 data in the form of files with the extension .csv and imported using the pandas library.
- 4) Preprocessing: Preprocessing is done to prepare the data before it is processed at the next stage. In this research, data with empty values will be processed using traditional techniques and interpolation, then data normalization is performed using min-max normalization.
- 5) Data Decomposition: In the data decomposition stage, the main components of the dataset are separated into trend, seasonal, and noise so that they are easier to interpret. The data decomposition process in this study uses the SSA technique.
- 6) Data division: The dataset used will be divided into training data and test data for benchmarking. The training data will be used to train the model, while the test data will be used to test the model.
- 7) Model Building: LightGBM and XGBoost models were built using several parameters namely, num_iterations, learning_rate, and max_depth parameters in lightGBM and n_estimators, eta, and max_depth parameters in XGBoost.
- 8) Testing: Model testing is done by utilising test data to test the pre-trained LightGBM and XGBoost models. This test is conducted to obtain model performance using regression loss in the form of RMSE and MAE, and regression score in the form of R2.

4. Results and Discussions

This section presents the results of rainfall prediction using the LightGBM and XGBoost models, and their impact on nutrient scheduling.

a. Prediction Model Performance

The models compared were the LightGBM and XGBoost models. The results showed that the XGBoost model with SSA resulted in lower RMSE and higher R2 compared to LightGBM. This shows that XGBoost is more accurate in predicting rainfall and is more suitable in the context of nutrient scheduling that is sensitive to changes in rainfall. Results can be seen in Table I.

Table 1 - Comparison of Model Prediction Results

Model	RMSE	MAE	R2
LightGBM	0.0115858097	0.0042129124	0.2879378609
XGBoost	0.0056971936	0.0034495675	0.8278182711

Based on table I, it can be seen that the lowest RMSE and MAE values are in using the XGBoost model, and the highest R2 value is in using the LightGBM model. A comparison of the RMSE, MAE, and R2 values based on the data division ratio can be seen in Fig. 3.

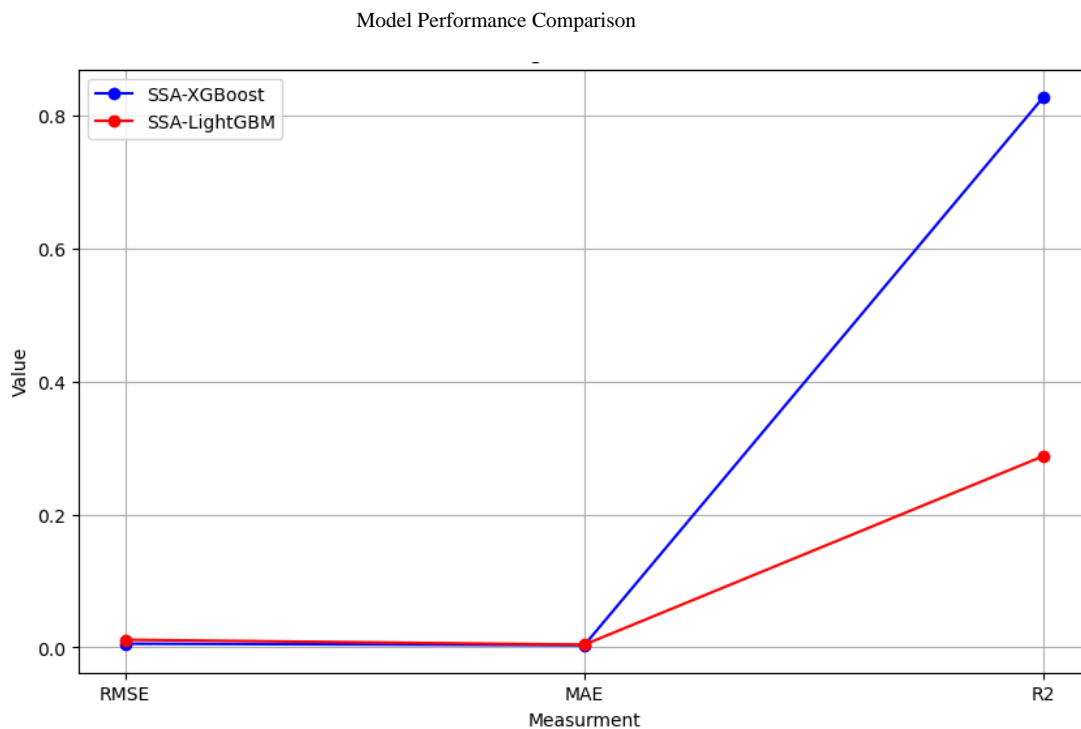


Fig. 3. Performance Comparison of LightGBM and XGBoost Models

b. Nutrient Utilization Efficiency

By utilizing rainfall prediction, the efficiency of nutrient use is increased, as nutrients are applied according to crop needs and anticipated rainfall. This reduces wastage and increases crop yield.

c. Comparative Performance Analysis

A practical evaluation of the proposed hybrid framework was conducted using an hourly rainfall dataset (6,624 observations) obtained from the weather monitoring station of Lambung Mangkurat University. The experimental process began with time series analysis using Singular Spectrum Analysis (SSA) to decompose the data into trend, seasonal, and random noise

components, followed by training ensemble models based on XGBoost and LightGBM on the reconstructed data. The results indicate a clear performance difference between the two algorithms; the SSA-XGBoost model achieved significantly better results with an RMSE of 0.0057 and an MAE of 0.0034, and attained the highest accuracy with an R^2 value of 0.8278. In contrast, the SSA-LightGBM model, despite using the same input, struggled to capture data variability, as reflected by a higher error (RMSE of 0.0116) and a low R^2 value of only 0.2879. The comparative visualization of these results (see Figure 3 and Table 2) clearly demonstrates that the level-wise tree-growing architecture of XGBoost provides substantially better stability than the leaf-wise approach of LightGBM for the characteristics of this local rainfall dataset.

Table 2 - Performance Comparison of the Proposed Models with Related Studies.

Evaluation Metric	SSA-XGBoost (This Study)	SSA-LightGBM (This Study)	XGBoost Runoff (Hameed et al., 2025)	LightGBM Humidity (Pringandana & Kusnawi, 2025)	Technical Interpretation
RMSE	0,0057	0,0116	1,554 (m ³ /s)	0,0786	The low RMSE achieved by XGBoost indicates a very small error deviation, which is crucial for avoiding critical mistakes in fertilization decision-making.
MAE	0,0034	0,0042	-	-	The low mean absolute error reflects strong consistency in daily predictions. XGBoost successfully explains the data variance to a very high degree, comparable to international benchmarks (Hameed), whereas LightGBM performs poorly in the context of this dataset.
R2	0,8278	0,2879	0,904	0,7981	

The analysis presented in Table 2 supports the argument that the failure of LightGBM is data-dependent rather than systemic. In the study by Wibawa et al., (2025) LightGBM demonstrated strong performance with an R^2 value of 0.79; however, in the present study, its performance is unsatisfactory, achieving an R^2 of only 0.28. This discrepancy highlights the critical importance of benchmarking algorithms for each specific geographical location and dataset type. No single algorithm can be expected to perform optimally across all scenarios in the field of hydrology.

The main finding of this study—namely, the substantially superior performance of XGBoost compared to LightGBM in predicting tropical rainfall after SSA-based denoising—provides new insights into the field of hydro-informatics. These results are consistent with recent studies by (Hameed et al., 2025) which reported that XGBoost consistently outperforms other ensemble models as well as deep learning approaches such as LSTM in capturing nonlinear patterns in environmental datasets of medium sample size.

Although LightGBM is often praised for its computational efficiency in handling large-scale data, as reported by (Pringandana & Kusnawi, 2025) in humidity prediction. From an agricultural perspective, the high accuracy achieved by the SSA-XGBoost model ($R^2 > 0.8$) has significant implications for crop nutrient management. (Elbasiouny et al., 2022a; Qin et al., 2023) emphasize that synchronizing fertilization timing with soil moisture conditions is an effective strategy for minimizing nutrient losses. With the ability of the SSA-XGBoost model to accurately predict rainfall, nutrient scheduling systems can proactively delay fertilization when heavy

rainfall events (exceeding 50 mm/day) are anticipated—a strategy that, according to (Margenot & Lee, 2023), can reduce phosphorus and nitrogen losses to water bodies by up to 80%. Furthermore, the integration of such predictive capabilities supports a more precise implementation of the 4R Nutrient Stewardship principle (Right Time) compared to calendar-based methods, as recommended by (Jayarani et al., 2024) in the development of advanced agricultural systems.

Overall, this study confirms that in the context of rainfall prediction in tropical regions with limited yet highly variable sensor data availability, the combination of XGBoost with SSA-based decomposition represents the most robust and reliable solution compared to other boosting methods.

These findings indicate that computational speed (as offered by LightGBM) should not come at the expense of predictive accuracy, particularly when model outputs are used for critical decision-making processes such as nutrient management. The widespread adoption of this hybrid model has the potential to transform agricultural practices from reactive to predictive, thereby improving fertilizer use efficiency and reducing environmental impacts caused by nutrient leaching in tropical areas.

The implications of high predictive accuracy are particularly significant for the transition toward climate-resilient precision agriculture. Elbasiouny et al. (2022) and Dong et al. (2023) (Elbasiouny et al., 2022a; Qin et al., 2023) emphasize that synchronizing fertilization timing with soil moisture conditions is essential for enhancing nutrient use efficiency (NUE). With the capability of the SSA-XGBoost model to accurately predict heavy rainfall events, this system enables the proactive implementation of fertilization delay strategies, which, according to Skidmore et al. (2023), are crucial for reducing the risk of nutrient losses and phosphorus runoff into water bodies commonly associated with excessive rainfall.

In general, this study concludes that the SSA-XGBoost hybrid model provides a robust, efficient, and scalable solution for managing water availability and crop nutrient dynamics in tropical regions, while effectively bridging the gap between weather prediction and real-world agricultural applications.

5. Conclusion

This study demonstrates that integrating Singular Spectrum Analysis (SSA) with XGBoost constitutes a highly effective methodology for addressing the challenge of rainfall prediction in highly volatile tropical regions. The hybrid SSA-XGBoost model achieved superior predictive performance, evidenced by a high accuracy ($R^2 = 0.8278$) and a low error rate (RMSE = 0.0057), significantly outperforming comparable algorithms such as LightGBM. Furthermore, the model's outputs are directly applicable to developing adaptive cultivation systems. The practical utility of this research is substantial, as it provides a data-driven foundation for optimizing fertilizer scheduling. By enabling farmers to avoid applications during forecasted periods of heavy rainfall, the framework helps prevent nutrient leaching, thereby enhancing cost efficiency and environmental protection. Consequently, this study recommends the broader adoption of the SSA-XGBoost methodology within intelligent agricultural development initiatives in Indonesia, supporting the advancement of more sustainable food security.

Acknowledgement

This research is funded by the DRTPM KATALIS scheme with Agreement/Contract Number 066/E5/PG.02.00/PL/BATCH.2/2024

References

- Alkaff, M., Khatimi, H., Puspita, W., & Sari, Y. (2019). Modelling and predicting wetland rice production using support vector regression. *Telkomnika (Telecommunication Computing Electronics and Control)*, *17*(2), 819–825. <https://doi.org/10.12928/TELKOMNIKA.V17I2.10145>
- Arassah, F. I., Sadik, K., Sartono, B., & Sofan, P. (2025). Optimizing Machine Learning for Daily Rainfall Prediction in Bogor: A Statistical Downscaling Approach. *Eduvest - Journal of*

- Universal Studies*, 5(6), 7006–7018. <https://doi.org/10.59188/EDUVEST.V5I6.51307>
- Baskara, A. R., Maulida, M., Lestiyanto, M. T. M., Sari, Y., Mustamin, N. F., & Wijaya, E. S. (2024). Explicit Content Classification in Indonesian Song Lyrics Using the LSTM-CNN Method. *2024 9th International Conference on Informatics and Computing, ICIC 2024*, 1–6. <https://doi.org/10.1109/ICIC64337.2024.10956794>
- Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54(3). <https://doi.org/10.1007/s10462-020-09896-5>
- Bui, Q. T., Chou, T. Y., Hoang, T. Van, Fang, Y. M., Mu, C. Y., Huang, P. H., Pham, V. D., Nguyen, Q. H., Anh, D. T. N., Pham, V. M., & Meadows, M. E. (2021). Gradient boosting machine and object-based cnn for land cover classification. *Remote Sensing*, 13(14). <https://doi.org/10.3390/rs13142709>
- Cao, Q., Wu, Y., Yang, J., & Yin, J. (2023). Greenhouse Temperature Prediction Based on Time-Series Features and LightGBM. *Applied Sciences (Switzerland)*, 13(3). <https://doi.org/10.3390/app13031610>
- Cao, T. T., Anh Le, H., & Eppe, G. (2025). Nutrient dynamics, environmental impacts, and feed efficiency in intensive whiteleg shrimp (*Litopenaeus vannamei*) farming on sandy soils in Ninh Thuan, Vietnam. *Aquaculture Reports*, 44, 103050. <https://doi.org/10.1016/J.AQREP.2025.103050>
- Cui, Z., Qing, X., Chai, H., Yang, S., Zhu, Y., & Wang, F. (2021). Real-time rainfall-runoff prediction using light gradient boosting machine coupled with singular spectrum analysis. *Journal of Hydrology*, 603.
- Ding, X., Tao, Y., Luo, J., Xi, H., Tao, Q., Li, J., Huang, C., Du, F., & Ou, W. (2025). How do water quality requirements and spatial flows shape the supply-demand balance of water provision ecosystem services? Evidence from the Taihu Lake Basin. *Applied Geography*, 183, 103723. <https://doi.org/10.1016/J.APGEOG.2025.103723>
- Diniaty, D., Fauzi, A. M., Sunarti, T. C., Raharja, S., & Helmi, F. (2024). Determination of Superior Commodities For The Development of Small and Medium Industries in Kampar Regency. *Journal of Applied Engineering and Technological Science (JAETS)*, 5(2), 995–1010. <https://doi.org/10.37385/JAETS.V5I2.4727>
- Diykh, M., Ali, M., Farouque, A. A., Aldhafeeri, A. A., Jamei, M., & Labban, A. (2026). A robust artificial intelligence informed over complete rational dilation wavelet transform technique coupled with deep learning for long-term rainfall prediction. *Engineering Applications of Artificial Intelligence*, 165, 113426. <https://doi.org/10.1016/J.ENGAPPAI.2025.113426>
- Dong, S. (2024). Precipitation Prediction Using Long Short-Term Memory Networks: Improving Seasonal Rainfall Forecast Accuracy for Flood and Drought Prevention. *Proceedings - 2024 International Conference on Advances in Electrical Engineering and Computer Applications, AEECA 2024*, 345–349. <https://doi.org/10.1109/AEECA62331.2024.00067>
- Dupas, R., Faucheux, M., Senga Kiessé, T., Casanova, A., Brekenfeld, N., & Fovet, O. (2024). High-intensity rainfall following drought triggers extreme nutrient concentrations in a small agricultural catchment. *Water Research*, 264, 122108. <https://doi.org/10.1016/J.WATRES.2024.122108>
- Elbasiouny, H., El-Ramady, H., Elbehiry, F., Rajput, V. D., Minkina, T., & Mandzhieva, S. (2022a). Plant Nutrition under Climate Change and Soil Carbon Sequestration. *Sustainability (Switzerland)*, 14(2), 1–20. <https://doi.org/10.3390/su14020914>
- Elbasiouny, H., El-Ramady, H., Elbehiry, F., Rajput, V. D., Minkina, T., & Mandzhieva, S. (2022b). Plant Nutrition under Climate Change and Soil Carbon Sequestration. *Sustainability 2022, Vol. 14*, 14(2). <https://doi.org/10.3390/SU14020914>
- Getachew, B., Kewessa, G., Hailu, W., & Girma, G. (2025). Observed climate trends and farmers' adaptation strategies in Dendi District, West Shewa Zone, Ethiopia. *Climate Services*, 38, 100548. <https://doi.org/10.1016/J.CLISER.2025.100548>
- Ghosh, S., Mukhoti, S., & Sharma, P. (2025). Quantifying rainfall-induced climate risk in rainfed agriculture: A volatility-based time series study from semi-arid India. *Agricultural Water Management*, 319, 109775. <https://doi.org/10.1016/J.AGWAT.2025.109775>

- Hameed, M. M., Masood, A., Hamid, A., Elbeltagi, A., Razali, S. F. M., & Salem, A. (2025). Forecasting monthly runoff in a glacierized catchment: A comparison of extreme gradient boosting (XGBoost) and deep learning models. *PLOS ONE*, *20*(5), e0321008. <https://doi.org/10.1371/JOURNAL.PONE.0321008>
- Hu, F., Yang, Q., Yang, J., Shao, J., & Wang, G. (2025). An adaptive rainfall-runoff model for daily runoff prediction under the changing environment: Stream-LSTM. *Environmental Modelling & Software*, *191*, 106524. <https://doi.org/10.1016/J.ENVSOFT.2025.106524>
- Jayarani, P., Channa Basava, P., Nagaraj, P., & Kusuma Latha, P. (2024). *Optimizing fertilizer usage in agriculture with AI Driven Recommendations*. *21*(1), 628. <https://doi.org/10.0805/Jbse.2024577123>
- Jiang, H., Zhang, Y., Qian, C., & Wang, X. (2024). Comment on papers using machine learning for significant wave height time series prediction: Complex models do not outperform auto-regression. *Ocean Modelling*, *189*, 102364. <https://doi.org/10.1016/J.OCEMOD.2024.102364>
- Johny, K., Pai, M. L., & S., A. (2022). A multivariate EMD-LSTM model aided with Time Dependent Intrinsic Cross-Correlation for monthly rainfall prediction. *Applied Soft Computing*, *123*, 108941. <https://doi.org/10.1016/J.ASOC.2022.108941>
- Komatsu, K. J., Reinhart, K., Alley, S., Porensky, L. M., Wilcox, K. R., & Koerner, S. E. (2026). Sensitivity of soil nutrient pools, but stability of microbial processes, under reduced rainfall and altered grazing management in northern mixed-grass prairie. *Soil Biology and Biochemistry*, *214*, 110071. <https://doi.org/10.1016/J.SOILBIO.2025.110071>
- Kumar, G. D., Pradhan, K. C., & Tyagi, S. (2024). Deep Learning Forecasting: An LSTM Neural Architecture based Approach to Rainfall and Flood Impact Predictions in Bihar. *Procedia Computer Science*, *235*, 1455–1466. <https://doi.org/10.1016/J.PROCS.2024.04.137>
- Kumar, V., Kedam, N., Sharma, K. V., Khedher, K. M., & Alluqmani, A. E. (2023). A Comparison of Machine Learning Models for Predicting Rainfall in Urban Metropolitan Cities. *Sustainability (Switzerland)*, *15*(18). <https://doi.org/10.3390/su151813724>
- Li, C., Jiang, H., Zhou, R., Dou, Y., Shen, Z., Zhang, L., Qian, X., & Mao, J. (2026). Hypergraph attention and periodic fusion learning for enhanced flight delay prediction. *Information Fusion*, *129*, 104076. <https://doi.org/10.1016/J.INFFUS.2025.104076>
- Li, J., Zhu, X., Xu, Q., Ao, L., Yu, X., & Zhang, S. (2025). Coupling of vegetation filter strips and biochar reduced the loss of nitrogen and phosphorus nutrients in subtropical riparian areas under continuous rainfall conditions. *Environmental Technology & Innovation*, *40*, 104473. <https://doi.org/10.1016/J.ETI.2025.104473>
- Li, S., Wei, J., Xu, H., Lu, L., Gao, A., Jiang, J., Yu, Z., & Feng, H. (2025). Rainfall-derived dissolved organic matter and nutrients in the Chaohu Lake watershed: Urban source tracing and ecological risk insight. *Journal of Environmental Chemical Engineering*, *13*(5), 119020. <https://doi.org/10.1016/J.JECE.2025.119020>
- Lin, X., Yang, Z., Lin, X., Zhong, W., & Zhang, H. (2026). A non-intrusive load monitoring method for commercial buildings based on time-series periodicity adaptive fusion. *Energy and Buildings*, *354*, 117031. <https://doi.org/10.1016/J.ENBUILD.2026.117031>
- Liu, N., Guo, T., Zhang, H., Yang, G., Wei, B., Xu, H., Ren, H., Badgery, W., Kemp, D., Nie, Z., Lee, M. R. F., Rillig, M. C., & Zhang, Y. (2025). Optimized grazing management enhances multiple ecosystem services by maintaining plant diversity and dominance in grasslands. *One Earth*, *8*(7), 101319. <https://doi.org/10.1016/J.ONEEAR.2025.101319>
- Luo, J., Zhang, Z., Fu, Y., & Rao, F. (2021). Time series prediction of COVID-19 transmission in America using LSTM and XGBoost algorithms. *Results in Physics*, *27*. <https://doi.org/10.1016/j.rinp.2021.104462>
- Margenot, A. J., & Lee, J. (2023). The fate of nitrogen of ammonium phosphate fertilizers: A blind spot. *Agricultural and Environmental Letters*, *8*(2). <https://doi.org/10.1002/AEL2.20116>
- Mashooq, B., Singh, Y., & Manzoor, S. I. (2025). Federated Learning-based Framework for Rainfall Prediction in Smart Agriculture. *Procedia Computer Science*, *259*, 2004–2013. <https://doi.org/10.1016/J.PROCS.2025.04.156>

- Mehdizavareh, H., Khan, A., & Cichosz, S. L. (2025). Enhancing glucose level prediction of ICU patients through hierarchical modeling of irregular time-series. *Computational and Structural Biotechnology Journal*, 27, 2898–2914. <https://doi.org/10.1016/J.CSBJ.2025.06.039>
- Mora-Motta, D., Llanos-Cabrera, M. P., Chavarro-Bermeo, J. P., Estrada-Bonilla, G. A., & Silva-Olaya, A. M. (2025). Soil quality responses of a tropical Andisol to contrasting land-use systems: Insights from multiple indexing approaches. *Science of The Total Environment*, 1002, 180648. <https://doi.org/10.1016/J.SCITOTENV.2025.180648>
- Murumkar, A., Tapas, M., Martin, J., Kalcic, M., Shedekar, V., Goering, D., Thorstensen, A., Boles, C., Redder, T., & Confesor, R. (2025). Advancing SWAT modeling with rainfall risk-based fertilizer timing to improve nutrient management and crop yields. *Agricultural Water Management*, 316, 109555. <https://doi.org/10.1016/J.AGWAT.2025.109555>
- Narimani, R., Jun, C., Saedi, A., Bateni, S. M., & Oh, J. (2023). A multivariate decomposition–ensemble model for estimating long-term rainfall dynamics. *Climate Dynamics*, 61(3–4), 1625–1641. <https://doi.org/10.1007/s00382-022-06646-x>
- Olaleye, O. S., Faloye, O. T., Akintola, O. A., Babalola, T. E., Ogunrinde, A. T., Akinremi, J. O., Jimoh, R. A., Alonge, A. E., & Sodiq, K. A. (2026). Assessment of rainfall Spatio-temporal trends and climate variability for sustainable agriculture and environmental resilience in parts of Ekiti state, south West Nigeria. *Ecological Frontiers*. <https://doi.org/10.1016/J.ECOFRO.2025.12.020>
- Paramesha, V., Kumar, P., Prabhakar, M., Gopinath, K. A., Ravisankar, N., Mohan Kumar, R., Jyoti Nath, A., Jinger, D., & Bhattacharjee, S. (2025). Assessing the impact of nutrient management on productivity, economics, soil quality, energy efficiency, and life cycle assessment in rice-based farming systems. *Journal of Agriculture and Food Research*, 23, 102278. <https://doi.org/10.1016/J.JAFR.2025.102278>
- Pringandana, C. G. L., & Kusnawi, K. (2025). A Comparative Analysis of Hyperparameter-Tuned XGBoost and LightGBM for Multiclass Rainfall Classification in Jakarta. *Jurnal Teknik Informatika (Jutif)*, 6(4), 2467–2483. <https://doi.org/10.52436/1.jutif.2025.6.4.4965>
- Puspasari, R. L., Yoon, D., Kim, H., & Kim, K. W. (2023). Machine Learning for Flood Prediction in Indonesia: Providing Online Access for Disaster Management Control. *Economic and Environmental Geology*, 56(1), 65–73. <https://doi.org/10.9719/EEG.2023.56.1.65>
- Qin, D., Tominaga, R., & Saneoka, H. (2023). Uptake and Use Efficiency of Major Plant Nutrients for Climate-Resilient Agriculture. *Climate-Resilient Agriculture*, 2, 35–50. https://doi.org/10.1007/978-3-031-37428-9_2
- Sahin, E. K. (2020). Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. *SN Applied Sciences*, 2(7), 1–17. <https://doi.org/10.1007/s42452-020-3060-1>
- Samad, A., Bhagyanidhi, Gautam, V., Jain, P., Sangeeta, & Sarkar, K. (2020). An Approach for Rainfall Prediction Using Long Short Term Memory Neural Network. *2020 IEEE 5th International Conference on Computing Communication and Automation, ICCCA 2020*, 190–195. <https://doi.org/10.1109/ICCCA49541.2020.9250809>
- Sawah, M. S. (2026). A novel feature selection-driven framework for sustainable and efficient rainfall classification using machine learning models. *Unconventional Resources*, 9, 100286. <https://doi.org/10.1016/J.UNCRESES.2025.100286>
- Singh, A., Kiboi, M., Bhullar, G. S., Sisodia, B. S., Bautze, D., Patidar, I., Konde, N., Ledroit, C., & Riar, A. (2025). Long-term impacts of farming systems and preceding crops on soil organic carbon and physiochemical properties in Vertisols of subtropical India. *Soil Advances*, 4, 100083. <https://doi.org/10.1016/J.SOILAD.2025.100083>
- Sutanto, S. J., Paparrizos, S., Nauta, L., Supit, I., Lefèvre, V., Kranjac-Berisavljevic, G., Gandaa, B. Z., Dogbey, R., Jamaldeen, B. M., & Ludwig, F. (2025). DROP app: A hydroclimate information service to deliver scientific rainfall, local rainfall, and soil moisture forecasts for agricultural decision-making. *Heliyon*, 11(4), e42740. <https://doi.org/10.1016/J.HELIIYON.2025.E42740>
- Tian, Y., Fu, W., Xiang, Y., Xiong, Q., Cui, R., Ao, Z., & Lei, X. (2025). Urban real-time rainfall-

- runoff prediction using adaptive SSA-decomposition with dual attention. *Journal of Hydrology*, 653, 132701. <https://doi.org/10.1016/J.JHYDROL.2025.132701>
- Valladares-Castellanos, M., de Jesús Crespo, R., Xu, Y. J., & Douthat, T. H. (2024). A framework for validating watershed ecosystem service models in the United States using long-term water quality data: Applications with the InVEST Nutrient Delivery (NDR) model in Puerto Rico. *Science of The Total Environment*, 949, 175111. <https://doi.org/10.1016/J.SCITOTENV.2024.175111>
- Wang, C., Zhang, Q., Tao, M., Hu, H., Xue, C., Xue, F., & Dong, Z. (2026). Estimation of seasonal ecological water demand in arid zone of Northwest China: An approach using the LSTM-random forest regression model. *Journal of Environmental Management*, 397, 128240. <https://doi.org/10.1016/J.JENVMAN.2025.128240>
- Wang, R., Chen, Y., Wu, H., Liu, J., Wang, M., & Duan, J. (2026). A flood susceptibility prediction method for climate change scenarios driven by coupled land simulation and spatiotemporal dual convolution synergy. *Journal of Hydrology*, 664, 134366. <https://doi.org/10.1016/J.JHYDROL.2025.134366>
- Wang, W., Wen, X., Zhang, M., Wang, Y., Zheng, Y., Gong, M., & Liu, D. (2025). A gate-aware GRU model with trend-residual decomposition and quantile regression for remaining useful life prediction of IGBT. *Microelectronics Journal*, 165, 106852. <https://doi.org/10.1016/J.MEJO.2025.106852>
- Wang, Y., Su, Y., Zheng, Z., Zhou, Z., & Wang, X. (2026). Bayesian-optimized CNN-LSTM neural network for predicting road construction dust concentrations. *Developments in the Built Environment*, 25, 100843. <https://doi.org/10.1016/J.DIBE.2026.100843>
- Wibawa, T. S., Ningrum, N. K., & Syahreza, A. (2025). Comparison of CatBoost and LightGBM Models for Air Humidity Prediction. *Journal of Applied Informatics and Computing*, 9(3), 803-809.
- Wu, S., Kong, L., Wang, A., & Feng, X. (2025). Thermal error prediction of CNC Swiss-type lathe under variable operating conditions based on symbolic regression and time series mixup enhancement. *CIRP Journal of Manufacturing Science and Technology*, 63, 362–374. <https://doi.org/10.1016/J.CIRPJ.2025.09.018>
- Yamada, H., Yamasaki, T., Kanuma, N., & Nishimura, T. (2025). Application of GeoWEPP to a cabbage monocropping region: developing agricultural strategies to mitigate water erosion under short-duration intense rainfall. *CATENA*, 258, 109242. <https://doi.org/10.1016/J.CATENA.2025.109242>
- Yang, X., Wong, S. C., Ge, L. Y., Chan, W. P., & Lisak, G. (2025). Enhancing vegetable yield and quality in urban farming under soil and soil-less conditions: the synergistic role of biochar in food security and waste management. *Journal of Cleaner Production*, 513, 145724. <https://doi.org/10.1016/J.JCLEPRO.2025.145724>
- Yang, Y., & Ma, Q. (2025). From deficit to balance: Identifying blue-green infrastructure networks based on trade-offs and synergies between water and terrestrial ecosystem services in a water sensitive region. *Ecohydrology & Hydrobiology*, 25(4), 100656. <https://doi.org/10.1016/J.ECOHYD.2025.100656>
- Yao, X., Fu, X., & Zong, C. (2022). Short-Term Load Forecasting Method Based on Feature Preference Strategy and LightGBM-XGboost. *IEEE Access*, 10(June), 75257–75268. <https://doi.org/10.1109/ACCESS.2022.3192011>