Volume 2, Issue 1

Innovation Series Quarterly

https://journal.academicapress.org/aps/index.php/isq

Research Article

GR-AI-N: HYBRID MLR-LSTM MACHINE LEARNING-DRIVEN RICE YIELD FORECASTING SYSTEM WITH GENERATIVE-AI INTEGRATED PLATFORM FOR REAL-TIME OPTIMIZED PLANTING TIME PREDICTIONS AND RECOMMENDATIONS

Hayla D. Caitor[1, *], Il Nam O. Cho[2], *Ayman Yazeed S. Latip[3], Han Mark M. Lumayal[4], and Ava Claire T. Ponteras[5]

[1] Tupi National High School; haylagracecaitor0629@gmail.com
[2] Tupi National High School; choilnam2@gmail.com
[3] Tupi National High School; aymanyazeedlatip@gmail.com
[4] Tupi National High School; hanmarklumayal48@gmail.com
[5] Tupi National High School; avaponteras001@gmail.com
* Correspondence: aymanyazeedlatip@gmail.com; 09949435343 .

Keywords:
Multiple Linear Regression (MLR)
Machine Learning
Precision Agriculture
Decision-Support System
Rice-Yield Forecasting

Abstract: *Rice production in the Philippines is highly sensitive to climate variability, causing yield losses and inefficient resource use. This study presents GR-AI-N, a hybrid forecasting system combining Multiple Linear Regression (MLR), Long Short-Term Memory (LSTM) networks, and a Generative AI chatbot to predict rice yields and provide real-time recommendations. Historical geo-climatic data informed the hybrid MLR–LSTM model, which achieved high accuracy ($R^2 = 0.99$) and reliable planting forecasts through 2100. The chatbot demonstrated strong performance (F1 = 0.9766). User surveys on functionality, usability, adaptability, and acceptability showed excellent results. GR-AI-N thus delivers accurate, AI-driven decision support for sustainable rice production.*

1. INTRODUCTION

Self–related perceptions have turned into an essential subject in education, given its impacts on students' behavior and academic performance. These self-related psychological perceptions include self – concept, self – esteem, and self – efficacy. Education in the Philippines represents the ever-changing facets of life. Augmenting the current educational system through the implementation of the K to 12 curriculums. An additional year for primary education has been implemented, facing the reality that students' workload and financial problems caused most stress (Essel & Owusu, 2017).

Learned beliefs, attitudes, and impressions of an individual towards oneself best illustrate the idea of self- concept. According to the study of Emmanuel (2014), he looked into the correlation among the motivation, self- concept, and academic performance of the basic education students. Results showed that self-concept and academic performance were significantly correlated.

Furthermore, in the study entitled "Self-Concept, Study Habit and Academic Performance of Students" by Chamundeswari (2014), the results showed a significant correlation among the students' self-concept, study habits, and academic performance. Self – esteem is the totality of a person's evaluation of his worth. Mruk (2010) stated the various ways of defining self-esteem: 1) as an attitude whether it is positive, negative or behavioral reaction 2) based on discrepancy—on how the ideal-self differs from the perceived self, 3) a person's response towards himself, 4) a personality.

In the study between self-esteem and secondary school students' academic performance by Bhagat (2017), respondents studying in government and private schools show a positive but not significant relationship found between negative self-males and females with their academic performance. Self – efficacy is the perception to do an assigned task more effectively. It is a positive attitude toward the self that any tasks given could be accomplished through a person's capabilities.

In Bandura's Social Cognitive Theory of 1994, it was stated that four processes could be affected by self- efficacy: cognitive, motivational, affective, and selection processes. Self-efficacy significantly affects the cognitive processes of a person. The more that a person believes in himself, the more that the goal will be more comfortable for them to be achieved while those who have low self-efficacy focuses on the negative side, and it is more likely that they would experience difficulty and failure in achieving their goals. With self-efficacy, a person motivates himself through the goals they set for themselves and how they overcame the challenges while achieving those goals.

Balami (2015) studied the relationship between distance learner students' self-efficacy and academic performance; results showed that learners' self-efficacy has no significant relationship with their academic performance.

In this study, the researcher aims to contribute to Psychology and Education by finding the impact of self- concept, self–esteem, and self–efficacy on senior high school students' academic performance..

2. LITERATURE REVIEW

2.1 Geo-climatic Factors and Rice Yield

Geo-climatic factors such as temperature, rainfall, humidity, wind speed, and dew point strongly influence rice growth and development (Ahmed et al., 2019; Cruz et al., 2021). These variables affect critical stages including germination, tillering, flowering, and grain filling (IRRI, 2020). Small deviations in climate conditions can lead to measurable changes in yield output (Peng et al., 2014). Farmers rely heavily on stable weather patterns to plan planting schedules (Lansigan et al., 2000). As climate variability increases, rice production becomes more uncertain (IPCC, 2022).

Studies conducted in the Philippines indicate that temperature and rainfall are the most influential climatic variables affecting rice yield (Ahmed et al., 2019). High temperatures during flowering often result in spikelet sterility and reduced grain formation (Peng et al., 2014). Excessive or insufficient rainfall disrupts water availability and field preparation (Lansigan et al., 2000). These conditions increase production risks for smallholder farmers (FAO, 2021). Climate sensitivity remains a major concern in rice-growing regions (IPCC, 2022).

Research in South Cotabato showed strong statistical relationships between geo-climatic variables and rice yield (Dela Cruz et al., 2022). Temperature and humidity exhibited the highest influence on yield variability (Ahmed et al., 2019). Wind speed and precipitation also contributed to yield differences across seasons (Bian et al., 2025). These findings confirm that climate factors are significant predictors of production outcomes (IRRI, 2020). Accurate climate monitoring is therefore essential for yield forecasting (FAO, 2021).

Similar results have been reported in other Southeast Asian countries with comparable climates (Lobell et al., 2011). Studies link extreme climate events to delayed planting and reduced harvests (IPCC, 2022). El Niño and La Niña episodes significantly affect rice productivity (Lansigan et al., 2000). Yield losses during these events can persist for multiple seasons (FAO, 2021). Climate resilience is therefore a priority in agricultural planning (IRRI, 2020).

Overall, geo-climatic factors are fundamental inputs in rice yield prediction models (Ahmed et al., 2019). Their inclusion improves the explanatory power of forecasting systems (Peng et al., 2014). Continuous data collection enhances model reliability over time (FAO, 2021). Climate-based forecasting supports better timing of farming activities (IRRI, 2020). This approach contributes to more adaptive rice production systems (IPCC, 2022).

2.2 Non-geoclimatic Environmental Conditions

Non-geoclimatic factors also play a significant role in rice productivity (Bian et al., 2025). These include soil quality, crop variety, land management, and farming inputs (FAO, 2021). Such factors interact with climate conditions to shape yield outcomes (IRRI, 2020). Ignoring them can reduce the accuracy of predictive models (Lobell et al., 2011). Comprehensive analysis requires both climatic and non-climatic data (Peng et al., 2014).

Soil properties such as nutrient content, pH level, moisture retention, and salinity directly affect rice growth (Bian et al., 2025). Poor soil conditions restrict root development and nutrient uptake (FAO, 2021). This results in weaker plants and lower yields (IRRI, 2020). Soil degradation further increases vulnerability to climate stress (IPCC, 2022). Maintaining soil health is therefore critical for productivity (FAO, 2021).

Research shows that soil management practices can improve rice yields under variable conditions (Bian et al., 2025). Organic amendments enhance soil structure and microbial activity (FAO, 2021). These improvements increase nutrient availability to crops (IRRI, 2020). Better soil conditions also improve water efficiency (Lobell et al., 2011). As a result, yield stability is enhanced across seasons (Peng et al., 2014).

Crop variety selection also influences rice productivity (IRRI, 2020). Different varieties have varying tolerance to heat, drought, and pests (Peng et al., 2014). Selecting appropriate varieties improves adaptation to local conditions (FAO, 2021). Variety choice affects grain quality and maturity period (IRRI, 2020). This decision plays a major role in farm profitability (Lobell et al., 2011).

Including non-geoclimatic factors makes forecasting systems more location-specific (Ahmed et al., 2019). These variables allow models to reflect actual farm conditions (Peng et al., 2014). This improves the relevance of predictions for farmers (FAO, 2021). Integrated models support better decision-making at the field level (IRRI, 2020). Such systems enhance the practical value of yield forecasts (IPCC, 2022).

2.3 Rice Cultivation Practices and Yield Forecasting

Rice cultivation in the Philippines is closely tied to seasonal weather patterns and farming practices (Vicente et al., 2019). Most rice farmers follow fixed planting calendars based on traditional knowledge (Lansigan et al., 2000). While this approach has historical value, it often fails to reflect current climate variability (IPCC, 2022). As a result, planting schedules may not align with optimal environmental conditions (FAO, 2021). This misalignment increases the risk of yield loss and inefficient resource use (Qu et al., 2021).

Studies show that delayed or mistimed planting significantly reduces rice productivity (Vicente et al., 2019). Rainfed farms are especially vulnerable to irregular rainfall onset (Lansigan et al., 2000). Inconsistent water availability affects seedling establishment and early growth (IRRI, 2020). Yield reductions are commonly observed when planting does not match rainfall patterns (FAO, 2021). Accurate forecasting can help address these timing issues (Qu et al., 2021).

Callueng, Latip, and Rejas (2025) found that climate data is often available but underutilized by farmers. The lack of accessible forecasting tools limits practical application (Casinillo et al., 2023). Without decision-support systems, farmers rely on intuition rather than data (IRRI, 2020). This leads to inefficient use of seeds, water, and fertilizers (Qu et al., 2021). Yield variability therefore remains high across seasons (FAO, 2021).

Research in Southeast Asia confirms that farms using data-informed practices achieve higher efficiency (Casinillo et al., 2023). Improved scheduling enhances land and labor productivity (Darith et al., 2025). Timely planting also improves varietal performance (IRRI, 2020). Forecasting tools support better alignment between inputs and environmental conditions (FAO, 2021). This integration improves overall farm outcomes (IPCC, 2022).

Overall, rice cultivation outcomes depend on both agronomic practices and forecasting accuracy (Vicente et al., 2019). Yield forecasting supports better planning and risk reduction (Qu et al., 2021). Modern systems help farmers respond to changing conditions (IRRI, 2020). Data-driven cultivation improves efficiency and resilience (FAO, 2021). This approach strengthens food security in climate-sensitive regions (IPCC, 2022).

2.4 Mathematical Modeling in Agricultural Prediction

Mathematical models form the foundation of agricultural yield prediction (Peng et al., 2014). These models describe relationships between environmental variables and crop output (Bian et al., 2025). Regression techniques are commonly used due to their simplicity and interpretability (Yan et al., 2025). In the Philippines, multiple linear regression has been widely applied (Callueng et al., 2024). Such models provide baseline estimates for yield forecasting (FAO, 2021).

Callueng et al. (2024) developed a regression model using temperature, rainfall, humidity, wind speed, and dew point. The model achieved high R^2 values during both dry and wet seasons. These results indicate strong explanatory power (Peng et al., 2014). However, the model operated at a monthly resolution (Yan et al., 2025). This limited its usefulness for daily decision-making (FAO, 2021).

Regression-based models assume linear relationships among variables (Bian et al., 2025). In real farming systems, interactions are often nonlinear (Peng et al., 2014). This reduces prediction accuracy under complex conditions (Yan et al., 2025). Static coefficients also limit adaptability to changing environments (IPCC, 2022). These limitations highlight the need for more flexible methods (FAO, 2021).

Time-series regression has been used to improve forecasting performance (Bian et al., 2025). These models account for temporal patterns in yield data (Yan et al., 2025). While effective, they still require predefined assumptions (Peng et al., 2014). Real-time adaptation remains challenging (FAO, 2021). As a result, model responsiveness is constrained (IPCC, 2022).

Despite limitations, mathematical models remain essential in agricultural research (Peng et al., 2014). They provide transparency and interpretability (Yan et al., 2025). These models serve as benchmarks for advanced systems (FAO, 2021). Integrating them with adaptive techniques improves performance (Bian et al., 2025). This combination supports more accurate forecasting frameworks (IPCC, 2022).

2.5 Machine Learning and Web-Based Decision Support Systems

Machine learning offers advanced capabilities for analyzing complex agricultural data (Bushara et al., 2025). These algorithms can model nonlinear relationships among variables (Yan et al., 2025). Machine learning systems improve prediction accuracy compared to traditional models (Yang et al., 2025). They adapt as new data becomes available (FAO, 2021). This makes them suitable for dynamic farming environments (IPCC, 2022).

Random Forest and XGBoost have shown strong performance in crop yield prediction (Bushara et al., 2025). These models handle high-dimensional data effectively (Yan et al., 2025). Ensemble methods reduce overfitting and improve generalization (Yang et al., 2025). Studies report accuracy levels exceeding traditional regression (Bushara et al., 2025). These findings support their use in rice yield forecasting (FAO, 2021).

Machine learning models perform best when integrated with real-time data (Yan et al., 2025). Daily weather updates improve responsiveness (IPCC, 2022). Soil and varietal inputs further enhance precision (IRRI, 2020). This integration allows localized predictions (FAO, 2021). Such systems outperform static forecasting tools (Peng et al., 2014).

Web-based platforms enable practical deployment of machine learning models (De Jesus et al., 2025). These systems allow farmers to access predictions easily (Damba, 2025). Interactive interfaces improve understanding and usability (MacDonald et al., 2025). Farmers can input site-specific data (Yang, 2025). This increases relevance and adoption (FAO, 2021).

Overall, machine learning combined with web-based systems transforms agricultural decision-making (Bushara et al., 2025). These platforms deliver real-time, location-specific guidance (Damba, 2025). They bridge the gap between research and practice (IRRI, 2020). Farmers benefit from timely and actionable insights (FAO, 2021). This approach supports sustainable and resilient rice production (IPCC, 2022).

3. METHODOLOGY

This chapter describes and discusses the results of the entire study, providing a comprehensive analysis of the collected data. It presents the research procedure in detail and the statistical treatment.

3.1 Research Design

The study utilized a quantitative, developmental, and descriptive–analytical research design aimed at developing and evaluating a hybrid rice yield forecasting and decision-support system. Quantitative methods were employed to analyze the statistical relationships between geo-climatic variables—such as temperature, precipitation, humidity, wind speed, and dew point—and rice yield in various regions of the Philippines. Descriptive and analytical techniques were implemented to interpret historical data trends and evaluate the predictive performance of various models, including Multiple Linear Regression (MLR), Long Short-Term Memory (LSTM), and the hybrid MLR–LSTM approach.

The developmental component of the research involved the design, construction, and integration of the GR-AI-N web-based platform. This encompassed the creation of the forecasting core, the generative AI chatbot (GRAINKEEPER), and the geospatial visualization dashboard. The system underwent iterative testing and refinement using historical datasets from 2010 to 2024 to ensure its accuracy, reliability, and adaptability across different regional contexts. Model validation was performed using statistical performance metrics such as R^2, RMSE, MAE, MAPE, and paired t-tests.

Additionally, the study employed a system evaluation design to assess the efficacy of the developed platform as a decision-support tool. A user-based evaluation was conducted through a Likert-scale survey to measure the functionality, usability, adaptability, and acceptability of the GR-AI-N system. This comprehensive application of statistical modeling, machine learning evaluation, and user assessment ensured that the research design effectively addressed both the technical accuracy and practical usability of the proposed system..

3.2 Respondent /Participants

The study's respondents comprised thirty (30) individuals, encompassing agricultural research personnel and farming enthusiasts, who took part in the system evaluation phase. These participants were chosen to evaluate the functionality, usability, adaptability, and acceptability of the GR-AI-N web-based platform. Their feedback was collected via a Likert-scale survey following their interaction with the system, offering insights into the platform's effectiveness as a real-time decision-support tool for rice farming applications.

3.3 Instruments of the Study

The primary instruments utilized in this study comprised secondary datasets and computational tools for model development and validation. Historical geo-climatic data, which included temperature, precipitation, humidity, wind speed, and dew point from the years 2010 to 2024, were sourced from PAGASA Synoptic Stations and reputable weather data websites. Additionally, rice yield data were obtained from official agricultural records. Statistical analysis and Multiple Linear Regression modeling were conducted using the Statistical Package for the Social Sciences (SPSS), while machine learning and hybrid model development were implemented through Python-based frameworks.

Furthermore, a researcher-developed Likert-scale questionnaire was employed as the principal instrument for system evaluation. This survey was designed to measure four dimensions: functionality, usability, adaptability, and acceptability of the GR-AI-N platform. For the evaluation of the generative AI chatbot, a confusion matrix was utilized to assess accuracy, precision, recall, and F1-score across predefined user intentions. These instruments ensured both quantitative rigor in model assessment and structured user feedback for evaluating system performance and practicality.

3.4 Procedure

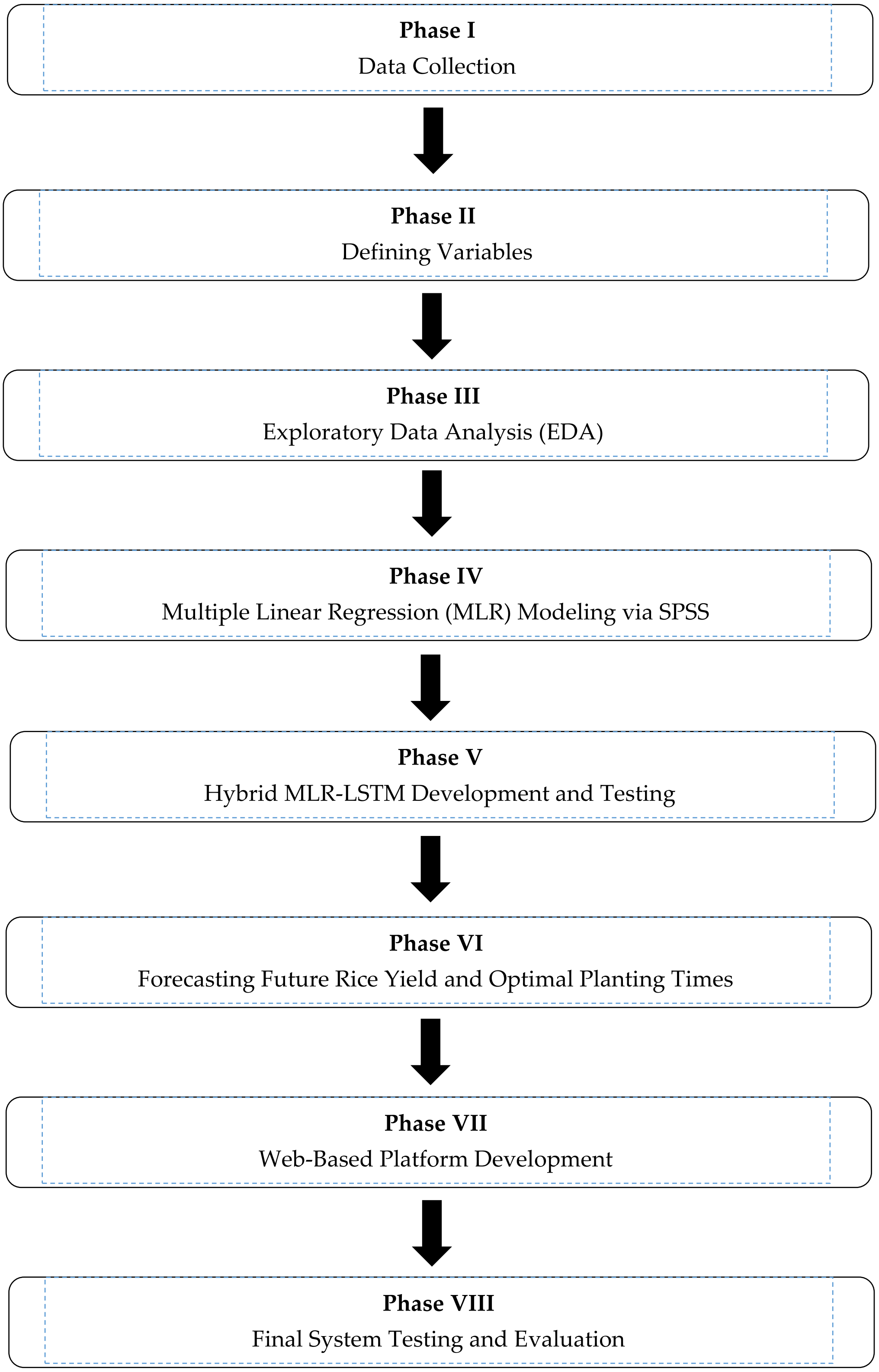


Figure 1. Research Procedure

3.5 Data Collection

To ensure data quality and model reliability, all raw geo-climatic and rice yield data underwent a systematic pre-processing flow before being used for model training and validation. Data pre-processing is a critical stage in machine learning, as it removes noise, handles missing values, and transforms heterogeneous data into structured, standardized inputs (Kotsiantis et al., 2006). In this study, the pre-processing pipeline involved several steps, including data cleaning, transformation, feature selection, scaling, and dateset partitioning. These processes aimed to optimize data consistency and improve model convergence during the Multiple Linear Regression (MLR) and Long Short-Term Memory (LSTM) modeling phases.

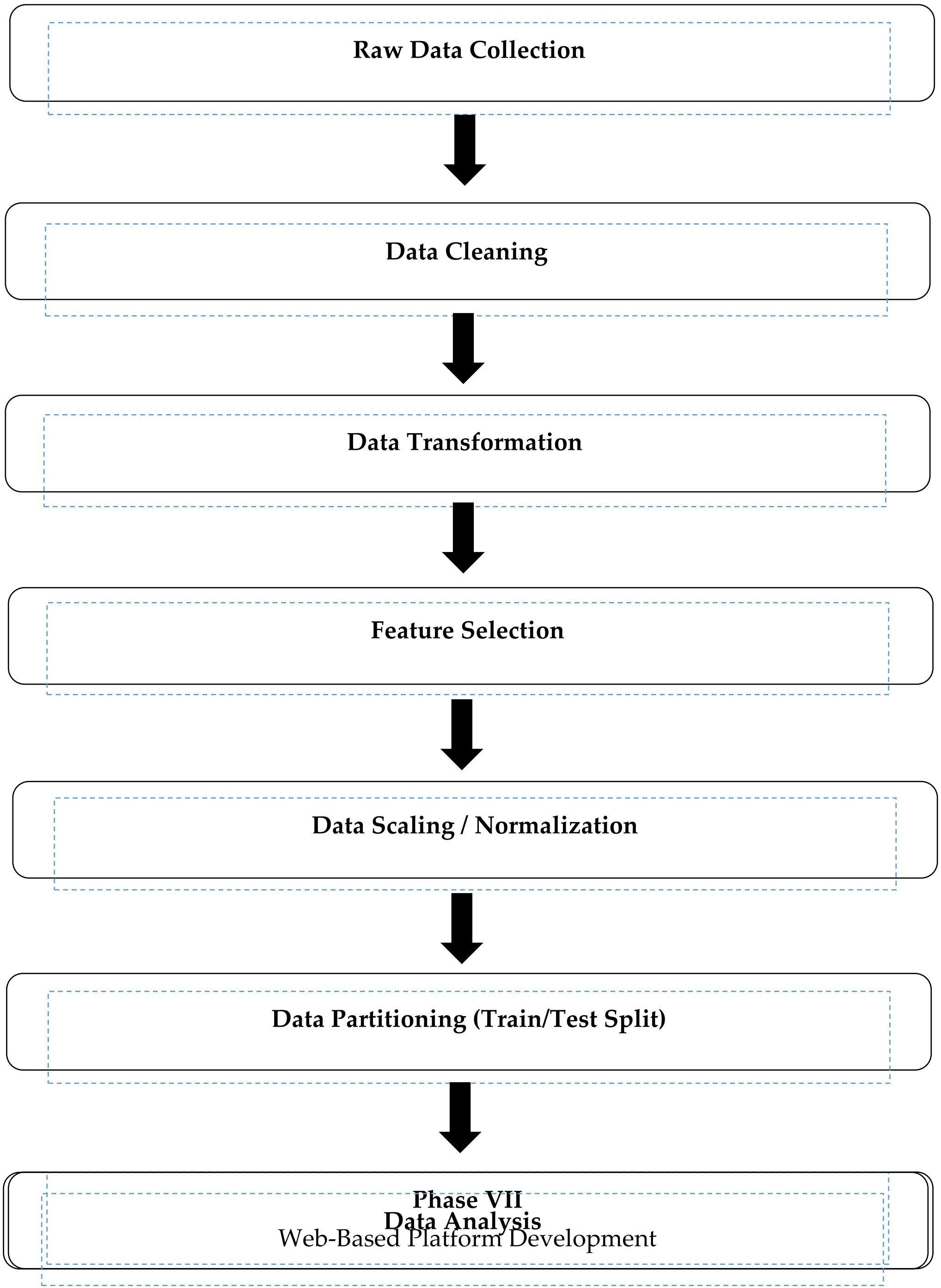


Figure 2. Data Collection

3.6 Data Analysis

The data will first be subjected to the Kolmogorov–Smirnov and Shapiro–Wilk normality tests to determine the distribution of variables using the Statistical Package For Social Sciences software. To assess the relationships between geo-climatic factors and rice yields, the Pearson Product-Moment Correlation Coefficient (Pearson r) will be employed.

$$r = \frac{(n\sum xy) - (\sum x)(\sum y)}{\sqrt{(n\sum x^2 - \sum x^2)\,(n\sum x^2 - \sum x^2)}}$$

Where;

r as the pearson-r correlation coefficient

x as the first variable

y as the second variable

n as the number of samples

The regional Multivariate Linear Regression (MLR) models will then be validated using a t-test at a 0.05 level of significance to check whether the predicted values significantly differ from the actual yields across the 16 regions.

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S^2{}_1}{N_1} + \frac{S^2{}_2}{N_2}}}$$

Where;

x_1 *as the mean of the predicted yield*

x_1 *as the mean of the actual yield*

s_1 *as the standard deviation of the predicted yield*

s_2 *as the standard deviation of the actual yield*

Model accuracy will also be assessed using performance metrics such as the coefficient of determination (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).

$$R^2 = 1 - \frac{\sum_i (y_i - \widehat{y_i})^2}{\sum_i (y_i - \bar{y})^2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (X_i - x_i)^2}{N}}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |x_i - x|$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y_\iota}}{y_i}\right|$$

For the Hybrid MLR–LSTM model, the same evaluation metrics (R^2, RMSE, MAE, and MAPE) will be used to allow for a direct comparison with the traditional MLR models. Meanwhile, the Generative AI-powered chat-bot will be evaluated through a confusion matrix analysis, which will provide values for Accuracy, Precision, Recall, and F1 Score.

$$Accuracy\ = \frac{\sum TP + TN}{\sum TP + FP + FN + TN}$$

$$Precision\ = \frac{\sum TP}{\sum TP + FP}$$

$$Recall\ = \frac{\sum TP}{\sum TP + FN}$$

$$f1 = \frac{2(Recall)(Precision}{Recall + Precision}$$

Where;

TP as the number of True Positives

FP as the number of False Positives

TN as the number of True Negatives

FN as the number of False Negatives

These combined statistical and computational techniques will determine the effectiveness, accuracy, and reliability of the forecasting models, the hybrid framework, and the chatbot in supporting rice farmers' decision-making processes

3.7 Ethical Considerations

This study was carried out with careful attention to ethical responsibility in both data use and system development. All geo-climatic and rice yield data used in the GR-AI-N system were obtained from publicly available and credible sources, such as government and institutional databases. No personal, confidential, or farmer-identifiable information was collected or analyzed at any stage of the research. Throughout the data preprocessing and modeling process, efforts were made to ensure data accuracy, fairness, and transparency, reducing the risk of bias or misleading results. The forecasting models were designed to provide decision support rather than absolute predictions, recognizing that agricultural outcomes are influenced by factors beyond climate data alone.

The Generative AI chat-bot was intended to assist users by offering guidance and explanations, not to replace farmers' judgment or local expertise. Model limitations and uncertainties were clearly acknowledged to prevent overdependent on automated outputs. Overall, the study aimed to promote responsible use of artificial intelligence in agriculture while prioritizing farmer welfare, sustainability, and food security.To ensure proper handling of data with ethically sound means, this study employed an anonymous respondent system which resulted into survey responses remaining confidential.

3.8 System Flow

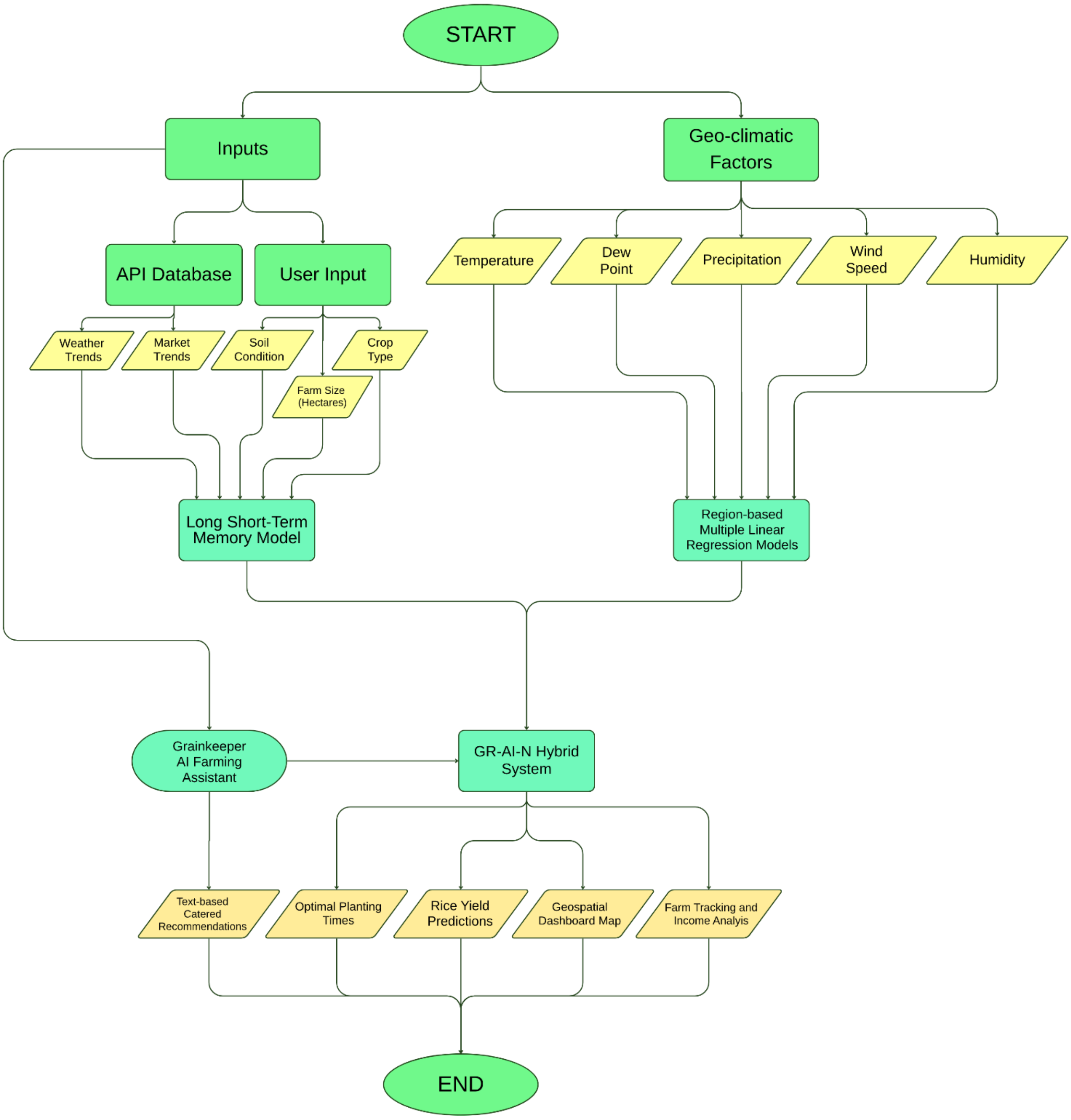


Figure 3. System Flow

4. FINDINGS

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam non pretium arcu. Nullar euismod est nec mauris placerat, id imperdiet sem tincidunt. Sed erat nulla, bibendum vitae tincidunt quis, porta eget dui. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Suspendisse at sollicitudin metus. Donec egestas nibh eget ante aliquam tempor. Donec sagittis lorem vitae nibh accumsan, id porta risus feugiat. Nunc a sagittis mauris. Aliquam ut venenatis neque. Morbi rutrum rutrum nisl, eget vestibulum enim volutpat ut (Ahmad, 2022; Mohamad & Ali, 2019).

Multiple Linear Regression (MLR) is a statistical method that models the relationship between two or more independent variables and a single dependent variable by fitting a linear equation to observed data. The theoretical assumption is that each unit change in an independent variable produces a consistent change in the dependent variable while holding other predictors constant. This technique has been widely applied in agricultural research because it allows not only the quantification of individual effects of climatic factors but also their combined influence on crop yields. Similarly, Yan et al. (2025) emphasized that integrating multiple predictors in regression enhances the robustness of crop yield forecasts compared to single-variable approaches.

The general form of a multiple linear regression equation is expressed as:

$$y$$

Where:

“y” is the output variable, also called dependent variable “T” is the Temperature

“D” is the Dew Point

“H” is the Humidity

“W” is the Wind Speed

“P” is the Precipitation

“a” is the estimated regression coefficients

Area of Coverage	Constant	Temperature	Dew Point	Precipitation	Wind Speed	Humidity
Region 1	-1037950.07	44830.5115	21917.63769	7039.900704	-18765.532	-1770.69628
Region 2	-170415.323	-383.614732	45198.123	-5770.19564	-51440.8258	-2384.66355
Region 3	2441515.061	245611.6906	-296632.343	-10775.5759	24416.92098	37579.99277
Region 4-A	174703.6232	-2513.540898	-2316.10549	-574.599114	1220.703298	444.819668
Region 4-B	2462696.68	-7727.476056	-85345.6814	1319.553168	7552.726695	-1893.08943
Region 5	-1088431.23	56816.44592	-23635.0436	-565.266929	11163.58297	3078.610271
Region 6	-2282654.8	28665.02941	39439.85311	9865.761607	14974.7581	10348.35713
Region 7	1265295.898	-27552.1949	-16438.4877	-2461.81571	-827.002089	-247.733141
Region 8	156704.5721	-9727.507485	12445.77482	756.703129	2646.832458	269.258349
Region 9	-1698181.04	-16084.12925	94152.10214	4434.446643	50325.78834	-1586.53123
Region 10	358634.5322	-1476.718307	-2507.2149	153.106817	-3460.6665	-1443.86259

Region 11	-242052.608	9666.202531	6119.261278	4730.314147	8512.536544	-1767.15139
Region 12	-170415.323	-383.614732	45198.123	5770.195649	-51440.8258	-2384.66355
Region 13	743292.0608	-4011.03759	-1976.28725	4855.152645	-3268.16435	-5434.78269
CAR	947101.6095	-16394.57511	-43532.9522	5910.98534	4644.365482	6123.945553
BARMM	-980922.938	43919.76003	-1643.64681	2318.57697	-10601.396	1742.901708

Table 1. MLR Model Coefficients .

After the computation of the MLR Model Coefficients, the researchers moved on to construct the 16 regional MLR models, excluding NCR because of no recorded rice production. Taking into account 5 independent variables for the factors; temperature, dew point, precipitation, wind speed and humidity, and substituting the computed coefficients to complete the baseline models. The difference in coefficient values ensures the different conditions of each region is properly modeled for optimal accuracy in rice yield forecasting. This aligns with the study of Yan et.al. (2025), that tailoring regression models to localized data improves accuracy and adaptability.

Region	MLR Models
Region 1	$\hat{Y} = -1037950.07 + 44830.51T + 21917.64D + 7039.91P - 18765.53W - 1770.70H$
Region 2	$\hat{Y} = -170415.32 - 383.61T + 45198.12D - 5770.20P - 51440.83W - 2384.66H$
Region 3	$\hat{Y} = -2441515.06 + 245611.70T - 296632.33D - 10775.58P + 24416.92W + 37579.99H$
Region 4-A	$\hat{Y} = 174703.62 - 2513.54T - 2316.11D - 574.60P + 1220.70W + 444.82H$
Region 4-B	$\hat{Y} = 2462696.68 - 7727.47T - 85345.64W + 1319.55P + 7552.72W - 1893.09H$
Region 5	$\hat{Y} = -1088431.20 + 56816.45T - 23635.04D - 565.27P + 11163.59W + 3078.61H$
Region 6	$\hat{Y} = -2282654.81 + 28665.03T + 39439.85D + 9865.76P + 14974.76W + 10348.36H$
Region 7	$\hat{Y} = 1265295.90 - 27552.19T - 16438.49D - 2461.82P - 827W - 247.73H$
Region 8	$\hat{Y} = 156704.57 - 9727.51T + 12445.77D + 756.7P + 2646.83W + 269.26H$
Region 9	$\hat{Y} = -1698181.04 - 16084.13T + 94152.10D + 4434.45P + 50325.79W - 1586.53H$
Region 10	$\hat{Y} = 358634.53 - 1476.72T - 2507.21D + 153.11P - 3460.66W - 1443.86H$
Region 11	$\hat{Y} = -242052.60 + 9666.20T + 6119.26D + 4730.31P + 8512.54W - 1767.15H$
Region 12	$\hat{Y} = -170415.32 - 383.61T + 45198.12D - 5770.20P - 51440.83W - 2384.66H$
Region 13	$\hat{Y} = 743292.06 - 4011.04T - 1976.29D + 4855.15P - 3268.16W - 5434.78H$
CAR	$\hat{Y} = 947101.61 - 16394.58T - 43532.95D + 5910.99P + 4644.37W + 6123.95H$
BARMM	$\hat{Y} = -980922.94 + 43919.76T - 1643.65D + 2318.58P - 10601.40W + 1742.90H$

Table 2. Region-based MLR Models .

The results of the 16 Multiple Linear Regression (MLR) models, representing each region of the Philippines except NCR, shows the variability of climatic factors on how it affects rice yield across different regions. While temperature, precipitation, and wind speed had positive effects on rice yield, the actual effect of this climate variables differ from region to region, this further supports the researchers aim of modelling different climatic conditions as accurate as possible. The predicted rice yields from 2014 to 2024 were generated by replacing the computed coefficients of each region-based Multiple Linear Regression (MLR) model into the general regression equation. The independent variables acts as inputs, while the coefficients acts as the one being able to successfully produce accurate forecasts. This successful modelling of relationships provides the foundation of the project to interpret climatic data into yield estimates, By embedding the coefficients into the MLR formula, the study produced annual yield predictions that could then be directly compared with observed data for each region.

After calculating the coefficients for each region-based Multiple Linear Regression (MLR) model, the researchers used these values in the general regression equation to predict rice yields from 2014 to 2024. The model used temperature, precipitation, humidity, wind speed, and dew point as input variables to measure how each climate factor affects rice yield. The coefficients show the expected change in yield for every unit increase in each variable, as explained by Field (2018). This method allowed the researchers to turn climate data into yield estimates, with each region's equation reflecting its own geo-climatic conditions. This approach is consistent with Enovejas et al. (2020), who found that including climate variables in regression models gives reliable rice yield projections in Nueva Ecija. The researchers then compared the yearly predictions with the actual observed yields for each region to check how well the model performed.

To build the hybrid system, the researchers first used MLR to create a strong baseline for rice yield prediction based on geo-climatic factors such as temperature, rainfall, humidity, wind speed, and dew point. These MLR predictions were then combined with the original climate data and sent into the LSTM model. Here, MLR works as a guiding map. LSTM then steps in to adjust for changes over time and nonlinear patterns that MLR cannot catch. This setup keeps the model easy to understand while making it more powerful. Sun et al. (2022) showed that mixing regression with LSTM can greatly reduce prediction errors, while Lagrazon and Tan (2023) described hybrid models as a practical balance between accuracy and real use for farmers in the Philippines. After setting up the hybrid model, the researchers used it to predict rice yields from 2014 to 2024. These results were compared directly with actual production values in metric tons. This comparison anchors the model in real data. By blending statistical structure with time-based learning, the hybrid approach aims to reduce the gaps seen in standalone models and deliver reliable and egion-specific forecasts for rice farming decisions.

	REGION 1		REGION 2	
Year	ACTUAL	PREDICTED	ACTUAL	PREDICTED
2014	1548373	1548158.84	1185196	1185196
2015	1602814	1602823.29	1244148	1244105
2016	1707695	1707120.88	1270901	1270504
2017	1750057	1750122.81	1347646	1347691
2018	1796229	1793214.97	1364849	1364123

2019	1777121	1777001.96	1291644	1291043
2020	1805126	18054642.3	1200642	1200642
2021	1865052	1863121.39	1319505	1319512
2022	1719990	1719104.42	1343125	1343183
2023	1851265	1851263.25	1269594	1269293
2024	1877113	1877113.89	1264117	1264122

	REGION 3		REGION 4-A	
Year	ACTUAL	PREDICTED	ACTUAL	PREDICTED
2014	2958415	2958231	390189	390101
2015	2616083	2616002	399193	399827
2016	3220607	3220523	389272	389127
2017	3409468	3409128	411785	411343
2018	3765150	3765107	405582	405091
2019	3304310	3304653	392907	392565
2020	3342883	3342001	407124	407994
2021	3634807	3634285	410833	410284
2022	3615115	3615192	420233	420545
2023	3730180	3730043	380782	380003
2024	3635148	3635121	386920	386785

	REGION 4-B		REGION 5	
Year	ACTUAL	PREDICTED	ACTUAL	PREDICTED
2014	857532	874658.2	1081078	1086479.56
2015	981669	983173.52	1070917	1075905.14
2016	1030626	1002007.37	1173307	1177182.79
2017	1033942	1052106.73	1243241	1240257
2018	1081878	1074267.73	1258147	1250988.96
2019	1081833	1067129.67	1264448	1266285.71
2020	1080408	1087951.45	1275492	1276531.37
2021	1159830	1159107.04	1335077	1335229.02
2022	1093168	1093998.32	1350438	1352282.45
2023	1407016	1409906.6	1192903	1191677.91
2024	1163101	1166696.36	1294992	1294220.09

	REGION 6		REGION 7	
Year	ACTUAL	PREDICTED	ACTUAL	PREDICTED
2014	1789693	1781874.66	270449	280824.47
2015	2245038	2219576.43	322862	324576.44
2016	2292201	2258584.36	327120	313183.83
2017	2090790	2053872.16	347697	283298.28
2018	2052574	2005650.71	338822	323732.58
2019	2056824	2052057.13	336194	328350.02

2020	1916230	1944496.97	231982	235036.5
2021	2230837	2232160.65	325209	323254.04
2022	2232253	2233388.63	309459	305818.34
2023	2077791	2070776.66	223087	220660.05
2024	2290580	2242372.65	260524	264670.43

	REGION 8		REGION 9	
Year	ACTUAL	PREDICTED	ACTUAL	PREDICTED
2014	964145	964854	552645	552612
2015	979648	979243	622203	622003
2016	994972	990052	618811	618121
2017	989794	989105	639113	639033
2018	982596	982332	656777	651212
2019	955709	955099	661775	661988
2020	954844	954553	581111	581434
2021	945565	945221	700590	701003
2022	946877	946102	728673	728612
2023	900245	900231	767437	767332
2024	927095	927123	645407	645001

	REGION 10		REGION 11	
Year	ACTUAL	PREDICTED	ACTUAL	PREDICTED
2014	402811	402563	403811	403984
2015	416543	416533	416543	416234
2016	448703	448034	448703	448002
2017	421692	421967	421495	421998
2018	452267	452032	487367	487212
2019	441868	441232	456368	456231
2020	417948	411003	413438	413894
2021	433665	433023	433435	433231
2022	488105	488452	485976	485281
2023	450368	450895	467953	467231
2024	466764	466799	451563	451421

	REGION 12		REGION 13	
Year	ACTUAL	PREDICTED	ACTUAL	PREDICTED
2014	1185196	1185643	405871	405197.15
2015	1244148	1242391	416997	416632.56
2016	1270901	1270943	469205	469415.18
2017	1347646	1347932	583838	583804.42
2018	1364849	1364454	574499	577516.5
2019	1291644	1291012	481312	489399.52
2020	1200642	1201045	462182	461024.75

2021	1319505	1319656	482675	483365.69
2022	1343125	1343100	510071	517203.49
2023	1269594	1263434	449369	449069.33
2024	1264117	1264609	503332	503722.42

	CAR		BARMM	
Year	ACTUAL	PREDICTED	ACTUAL	PREDICTED
2014	400415	400540.39	623343	623347.43
2015	428949	428250.78	576873	576795.22
2016	453461	453939.89	542760	529223.93
2017	460170	460176.63	612083	569684.12
2018	452609	452131.09	557149	584976
2019	400911	405270.25	488215	482105.21
2020	382848	389190.92	544486	541457.73
2021	445006	445313.74	577984	577083.29
2022	391105	391465.58	657933	654348.72
2023	485156	485038.25	659586	649541.26
2024	333972	331284.49	720731	726580.08

Table 3. Hybrid Model Rice Yield Predictions (Quarterly 2014-2024).

.

With the accuracy of the Hybrid MLR-LSTM model validated, the researchers now proceed to the forecasting of Rice Yield values upto the year of 2100, and with this values, the researchers can now easily determine the optimal planting times for a specific region in a specific time. Regional forecasts was done to maintain the accuracy given by the region-based MLR models, combining its capabilities with the temporal-focused machine learning LSTM model.

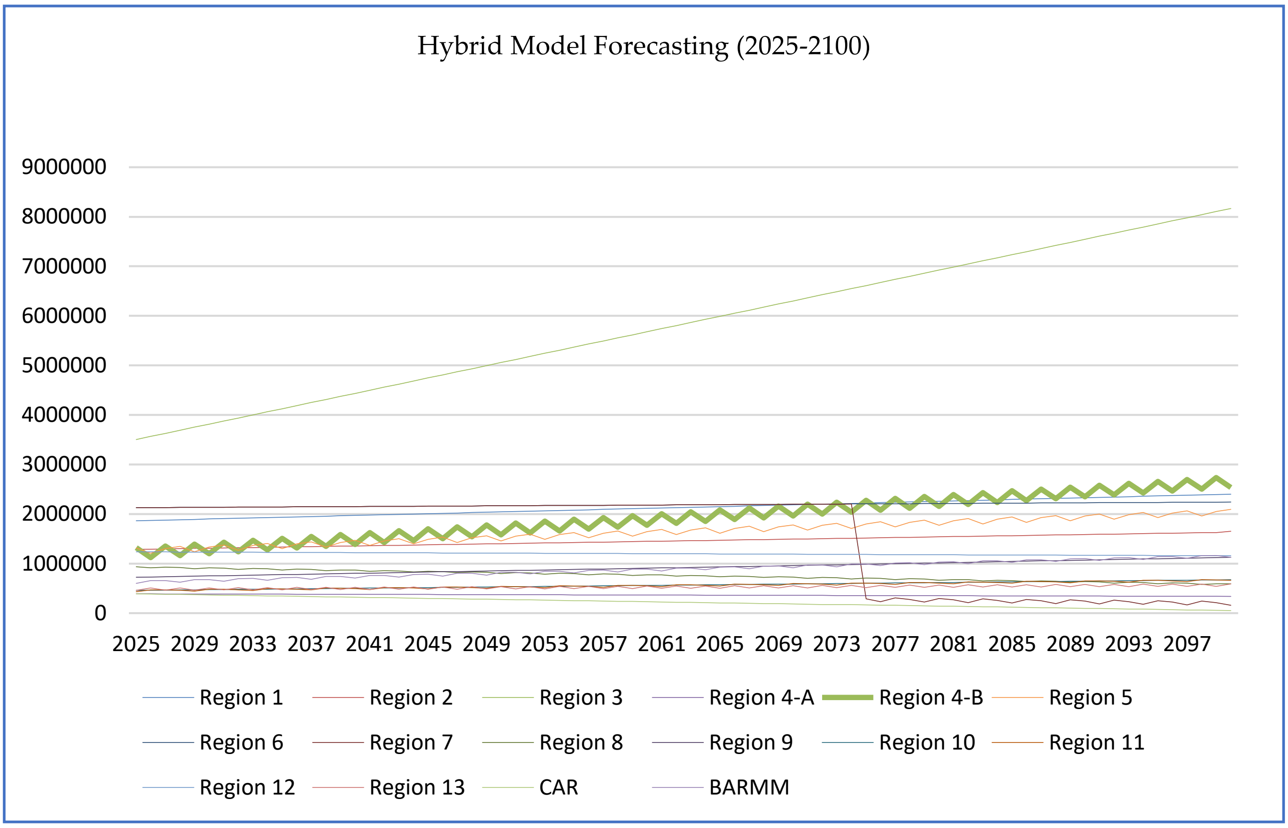


Figure 4. Rice Yield Forecasting (2025-2100)

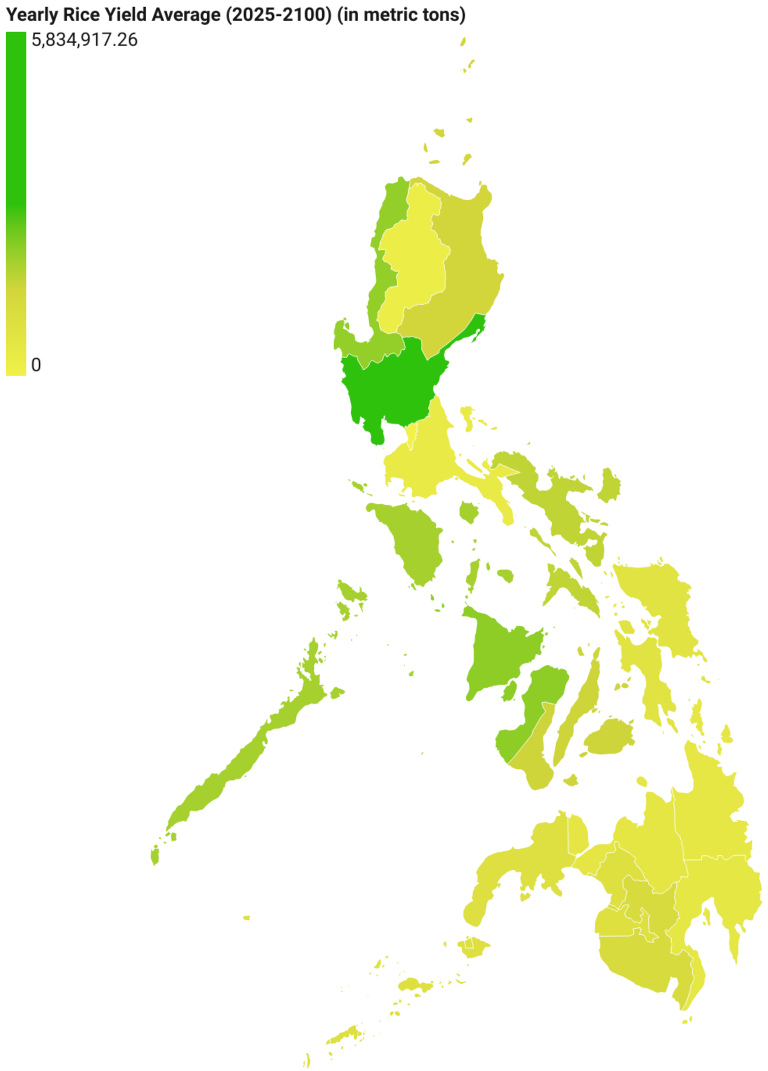
All regions of the Philippines from 2025-2100 was successfully projected using the Hybrid model. Region 3 reveal steady growth of rice yield production, while region 7 from 2073-2077 will experience a drop in rice production. Other Regions have no sudden change in its trendline and displayed steady progressions

2025-2100 Mean of Annual Forecasted Yields

REGION	Yield (metric tons)
Ilocos Region	2,114,584
Cagayan Valley	1,459,839
Central Luzon	5,834,917
CALABARZON	363,725
MIMAROPA	1,928,275
Bicol Region	1,655,570
Western Visayas	2,185,387

Central Visayas	1,506,177
Eastern Visayas	757,116
Zamboanga Peninsula	922,350
Northern Mindanao	567,179
Davao Region	560,212
SOCCSKSARGEN	1,197,377
Caraga	524,779
CAR	219,982
BARMM	893,332
NCR	No production

Table 4. Annual Rice Yield Mean From Forecast (2025-2100)



Map: Ayman Yazeed S. Latip • Source: From Hybrid Model Rice Yield Forecasting • Map data: © OSM • Created with Datawrapper

Figure 5. Choropleth Map of Forecasted Rice Yield.

Using the hybrid MLR–LSTM forecasting model, rice yield projection from 2025-2100 reveal strong regional variations across the Philippines. Central Luzon is predicted to remain the highest-yielding region, maintaining consistent productivity throughout the forecast period. Western Visayas and Ilocos Region follow with stable yield growth, while Bicol and MIMAROPA show moderate upward trends. However, due to fluctuations, Eastern Visayas and CAR are projected to experience slower yield recovery. Overall, with no abrupt declines most regions show steady progress in long-term production. These predictions will now be used as the basis for the websites Optimal Planting Time recommendations for a specific location and time.

Aside from statistical and machine learning models, the GR-AI-N platform integrates Generative Artificial Intelligence as an added innovation. This component is designed to improve user interaction and information delivery. Generative AI produces new outputs, such as text or responses, by learning patterns from existing data (Goodfellow et al., 2014). In agriculture, this technology is increasingly used to support advisory and decision-making systems. Tripathi et al. (2023) showed that generative AI helps farmers understand complex forecasts by translating them into natural, conversational language. Jain et al. (2021) also highlighted its use in adaptive chatbots that deliver real-time and personalized recommendations.

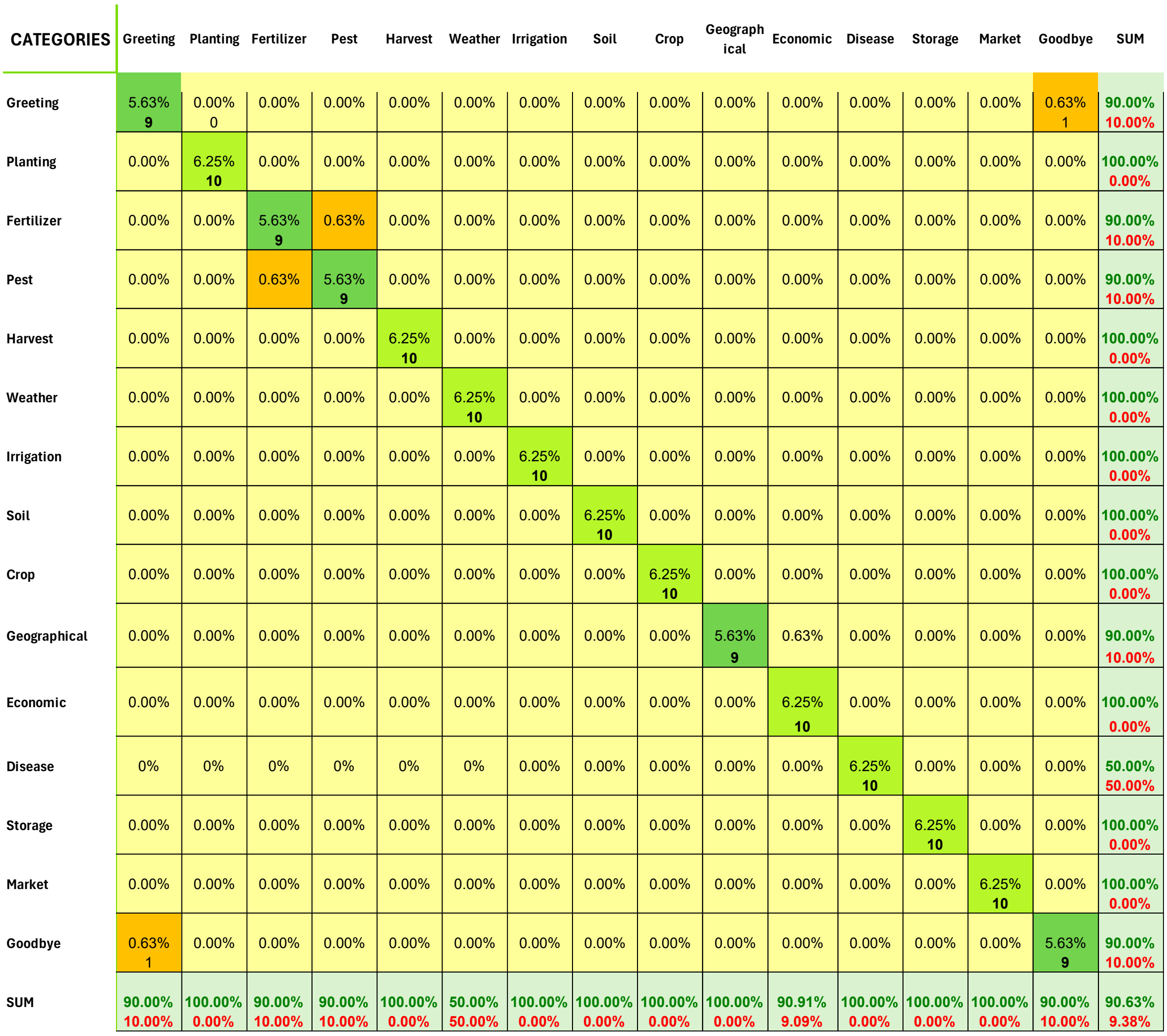
In this study, the generative AI module is embedded into the GR-AI-N website under the codename "GRAINKEEPER." This feature transforms the platform from a static forecasting tool into an interactive system. It acts as a bridge, connecting predictive analytics with practical, field-level decisions. Model performance was evaluated using a Confusion Matrix. The dataset consisted of 15 distinct user intentions, with 50 utterances per intention. An 80–20 data split was applied, where 80% of the utterances were used for training and 20% for testing. This setup enabled a clear assessment of the model's accuracy in c;lassifying user questions.

Intent	Description	Utterances (Training Data)	Utterances (Testing Data)
1. Planting Time Advice	Best planting dates & schedules	40	10
2. Fertilizer Recommendation	Type & amount of fertilizer	40	10
3. Pest Control Guidance	Identify & manage pests	40	10
4. Irrigation / Watering Advice	Water schedules & needs	40	10
5. Weather / Climate Inquiry	Weather & climate patterns	40	10
6. Soil Condition Guidance	Soil health & treatments	40	10
7. Rice Crop Type Info	Choosing & managing varieties	40	10
8. Farm Info Storage	Save farmer's details	40	10
9. Farm Info Recall	Retrieve stored details	40	10
10. Harvesting Tips	Harvest timing & methods	40	10
11. Yield Estimation	Estimate rice yield		10

12. Storage & post-harvest	Storage methods	40	10
13. Geographical Tips	Advice depending on location	40	10
14. Economic Advice	Market prices, profitability	40	10
15. General Help / Greetings	Greetings & FAQs	40	10

Table 5 Intentions For Confusion Matrix.

Following a 80-20 split, 40 utterances will be used for the initial training of the AI model, while 10 utterances for testing its responds. The intentions is designed to cover all basic topics about rice farming, this ensures the most important information about rice farming is successfully given by model, the model will be tested in terms of Accuracy, Precision, Recall, and f-1 Score. The use of an 80–20 train–test split in this study follows standard machine learning practice. This approach ensures that the model is trained on a sufficient portion of the data while still being evaluated on unseen samples (Goodfellow, Bengio, &



CATEGORIES	Greeting	Planting	Fertilizer	Pest	Harvest	Weather	Irrigation	Soil	Crop	Geographical	Economic	Disease	Storage	Market	Goodbye	SUM
Greeting	5.63% 9	0.00% 0	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.63% 1	90.00% 10.00%
Planting	0.00%	6.25% 10	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00% 0.00%
Fertilizer	0.00%	0.00%	5.63% 9	0.63%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	90.00% 10.00%
Pest	0.00%	0.00%	0.63%	5.63% 9	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	90.00% 10.00%
Harvest	0.00%	0.00%	0.00%	0.00%	6.25% 10	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00% 0.00%
Weather	0.00%	0.00%	0.00%	0.00%	0.00%	6.25% 10	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00% 0.00%
Irrigation	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	6.25% 10	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00% 0.00%
Soil	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	6.25% 10	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00% 0.00%
Crop	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	6.25% 10	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00% 0.00%
Geographical	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	5.63% 9	0.63%	0.00%	0.00%	0.00%	0.00%	90.00% 10.00%
Economic	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	6.25% 10	0.00%	0.00%	0.00%	0.00%	100.00% 0.00%
Disease	0%	0%	0%	0%	0%	0%	0.00%	0.00%	0.00%	0.00%	0.00%	6.25% 10	0.00%	0.00%	0.00%	50.00% 50.00%
Storage	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	6.25% 10	0.00%	0.00%	100.00% 0.00%
Market	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	6.25% 10	0.00%	100.00% 0.00%
Goodbye	0.63% 1	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	5.63% 9	90.00% 10.00%
SUM	90.00% 10.00%	100.00% 0.00%	90.00% 10.00%	90.00% 10.00%	100.00% 0.00%	50.00% 50.00%	100.00% 0.00%	100.00% 0.00%	100.00% 0.00%	100.00% 0.00%	90.91% 9.09%	100.00% 0.00%	100.00% 0.00%	100.00% 0.00%	90.00% 10.00%	90.63% 9.38%

Courville, 2016). Similar studies in conversational AI have shown that combining this data partitioning strategy with confusion matrix analysis provides a clear and reliable measure of model performance and helps identify areas for further improvement (Chen et al., 2021; Zhang et al., 2022).

Figure6 . Confusion Matrix Chart

The confusion matrix shows how well the model classified sample questions across 15 user intentions. The results indicate strong performance. Ten intentions were classified with perfect accuracy, while the remaining five intentions each had only one misclassified instance. Confusion matrices are widely used to evaluate classification models because they clearly display both correct predictions and errors. They provide insight into true positives, false positives, true negatives, and false negatives, making them useful for identifying specific patterns of misclassification (Provost & Fawcett, 2013). In natural language processing and chatbot systems, this method is especially important for multi-class tasks, where accurate intent recognition directly affects system reliability and user experience (Jurafsky & Martin, 2023).

Generative AI Assessment Summary

Metric	Value
Accuracy	0.98
Precision	0.9733
Recall	0.98
F-1 Score	**0.9766**

Table 6. Generative AI Assessment Summary.

Table 12 presents the performance of GRAINKEEPER model. The model achieved a high accuracy rate of 98%, a precision of 97.33%, and a recall of 98%. GRAINKEEPER scored an F1-score of 0.9766, demonstrating strong overall performance in accurately classifying user inputs while maintaining high precision and recall. This results is directly computed from the previously mentioned Confusion Matrix Test. GRAINKEEPER will serve as the window for farmers to have quick access to farming tips and information, completing the platform's purpose in becoming a all-in-one rice farming tool for farmers.

4.5 GR-AI-N Website Development

The GR-AI-N web platform acts as the backbone of the entire system. It brings together prediction models, data analysis, and AI-driven features in one space. This section explains how the platform was designed and built, with focus on the interface, system integration, and overall operation. The goal is simple: turn complex forecasts into a tool that farmers and researchers can easily use and understand. A system architecture diagram is used to show how the system works as a whole. For agricultural decision-support systems, these diagrams are especially valuable. They help illustrate how data sources, prediction models, and user interfaces work together to produce usable results for end-users (Mandal et al., 2023). In the case of GR-AI-N, the diagram is critical because the system combines MLR, LSTM, and a generative AI chatbot into one framework.

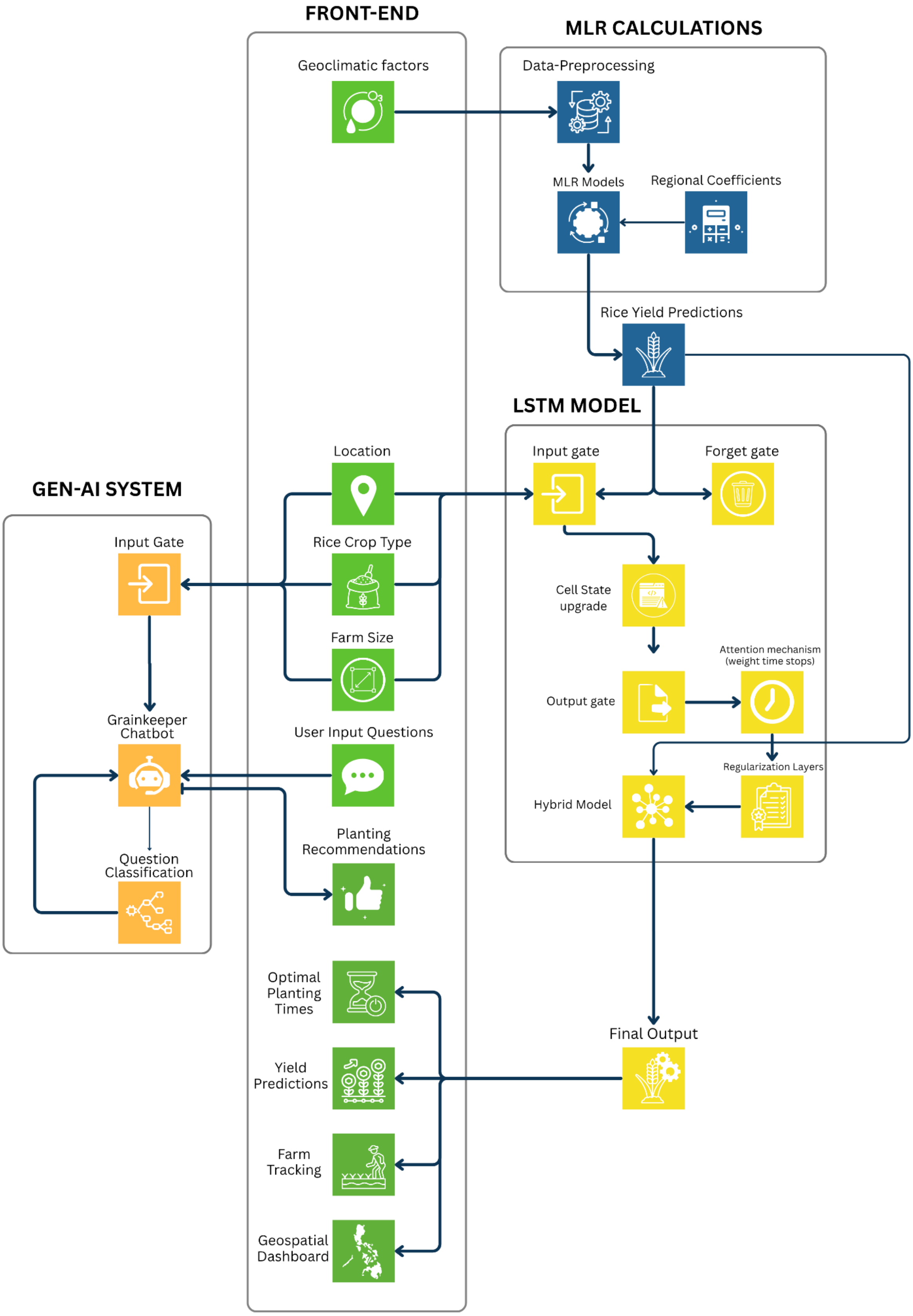
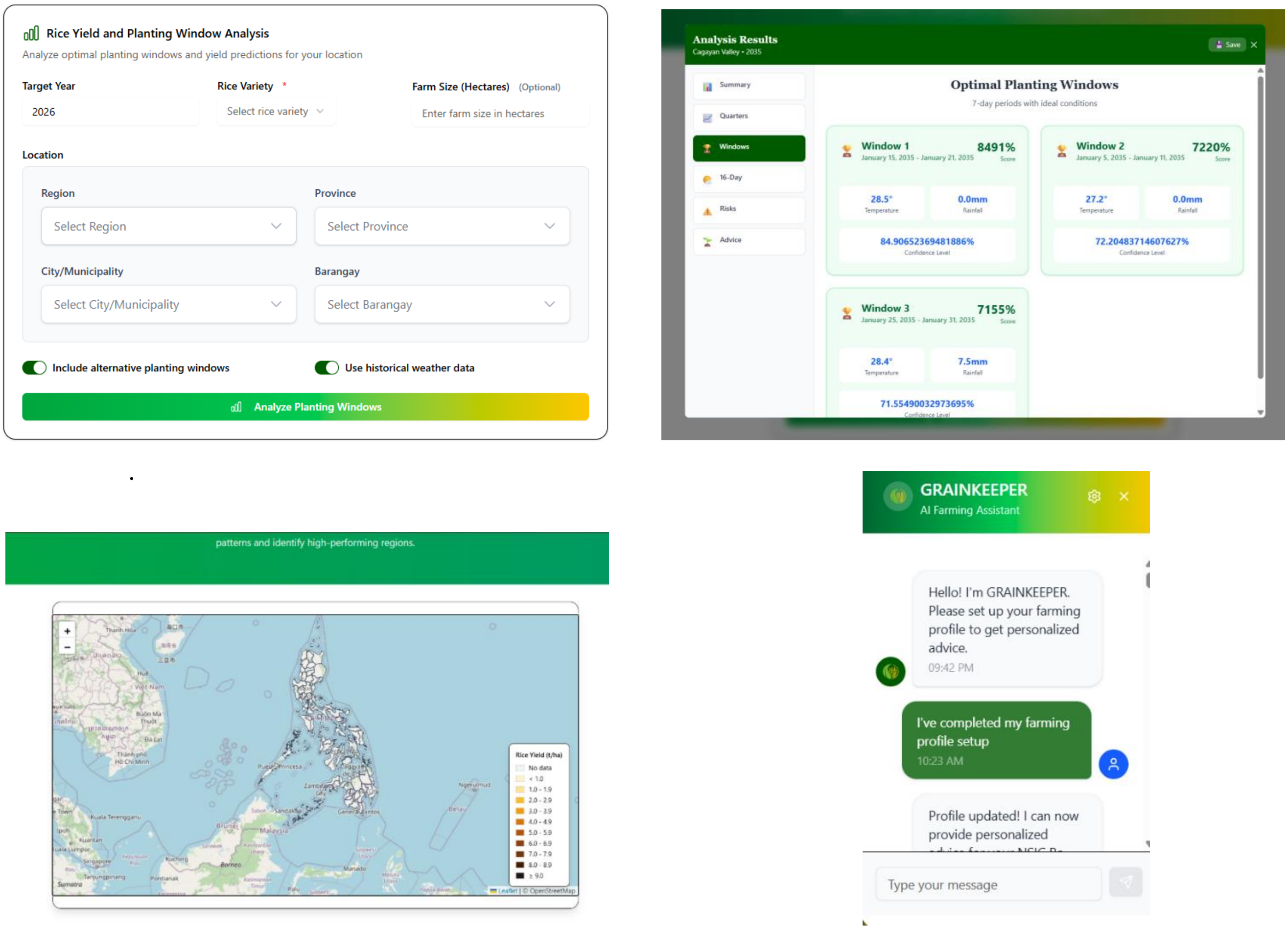

FRONT-END
MLR CALCULATIONS
Geoclimatic factors
Data-Preprocessing
MLR Models
Regional Coefficients
Rice Yield Predictions
LSTM MODEL
GEN-AI SYSTEM
Location
Input gate
Forget gate
Input Gate
Rice Crop Type
Cell State upgrade
Farm Size
Attention mechanism (weight time stops)
Output gate
Grainkeeper Chatbot
User Input Questions
Regularization Layers
Hybrid Model
Question Classification
Planting Recommendations
Optimal Planting Times
Final Output
Yield Predictions
Farm Tracking
Geospatial Dashboard

Figure 7 . System Architectural Diagram.

The GR-AI-N system architecture shows how data, models, and AI components are connected. It explains how information moves from input to output in a clear and structured way. The diagram is divided into three main parts: the Front-End, the MLR Calculations, and the LSTM Model. All of these components are linked to the Generative AI system. In the Front-End, geo-climatic variables such as temperature, precipitation, humidity, and wind speed are entered as raw inputs. These data first pass through a pre-processing stage to ensure consistency and quality. Once prepared, the data are sent to the MLR module. The MLR Calculations block estimates regional coefficients and produces initial rice yield predictions. These outputs are based on the linear relationships between climate variables and historical yield data. The results then serve as a foundation for further modeling and AI-driven analysis within the system



Rice Yield and Planting Window Analysis
Analyze optimal planting windows and yield predictions for your location

Target Year: 2026
Rice Variety *: Select rice variety
Farm Size (Hectares) (Optional): Enter farm size in hectares

Location
Region: Select Region
Province: Select Province
City/Municipality: Select City/Municipality
Barangay: Select Barangay

● Include alternative planting windows
● Use historical weather data

Analysis Results
Cagayan Valley • 2035

Optimal Planting Windows
7-day periods with ideal conditions

Window 1
January 15, 2035 - January 21, 2035
8491% Score
28.5° Temperature
0.0mm Rainfall
84.90652369481886% Confidence Level

Window 2
January 5, 2035 - January 11, 2035
7220% Score
27.2° Temperature
0.0mm Rainfall
72.20483714607627% Confidence Level

Window 3
January 25, 2035 - January 31, 2035
7155% Score
28.4° Temperature
7.5mm Rainfall
71.55490032973695% Confidence Level

GRAINKEEPER
AI Farming Assistant

Hello! I'm GRAINKEEPER. Please set up your farming profile to get personalized advice.
09:42 PM

I've completed my farming profile setup
10:23 AM

Profile updated! I can now provide personalized

patterns and identify high-performing regions.

Note: GR-AI-N Website is now accessable through the link http://grain-nine.vercel.app

Figure 8 . GR-AI-N System User Interfaces.

The GR-AI-N web platform provides an integrated and user-friendly environment that connects data input, forecasting, visualization, and AI-driven assistance into a single system. Users enter basic farm and location details through a simple Input Tab, which triggers the hybrid MLR–LSTM model to generate planting window predictions and yield estimates. Results are displayed in a structured format, including optimal planting periods, predicted yields, and climate-related risk assessments. These outputs are further visualized through an interactive geospatial dashboard that highlights regional trends and supports long-term planning. To enhance accessibility, the

GRAINKEEPER chatbot delivers data-driven recommendations in natural language, allowing users to easily understand and apply complex analytics in practical planting decisions.

To assess on-field performance, a Likert-scale survey will be conducted with 30 respondents, including agricultural researchers and farming enthusiasts. The survey evaluates four system dimensions: functionality, usability, adaptability, and acceptability. Participation is voluntary and anonymous, with no collection of personally identifiable information, and all responses are used solely for research purposes.

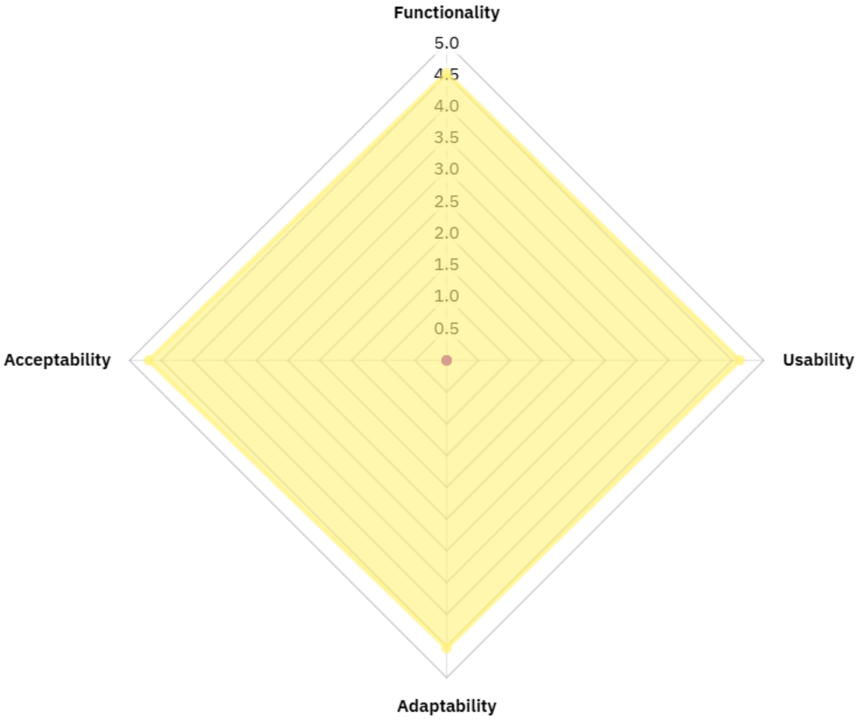


Figure 9. Survey Radar Chart.

The radar chart explain that the GR-AI-N system show a well-rounded and balanced performance across the four evaluated dimensions: Functionality = 4.52, Usability = 4.62, Adaptability = 4.54, and Acceptability = 4.7. All dimension received a grand mean score greater than 4.0, indicating strong positive evaluations from respondents.

5. DISCUSSION.

5.1 Statistical Treatment

Normality tests and correlation analysis confirmed that the geo-climatic datasets largely met the assumptions for parametric testing. Temperature, wind speed, and precipitation showed strong positive correlations with rice yield, while humidity was negatively correlated. These results highlight the critical role of stable climate conditions in crop productivity. Even small shifts in these factors can significantly influence yields. The correlations also guided the selection of key variables for predictive modeling, ensuring the most influential factors were prioritized.

5.2 Multiple Linear Regression (MLR) Modeling

MLR effectively captured the linear relationships between geo-climatic variables and rice yields across regions. High R^2 values in regions like 4-A and 4-B indicate strong model performance under stable climatic conditions. However, deviations were observed during extreme weather, revealing limitations in capturing nonlinear or time-dependent dynamics. Regional differences in coefficients emphasize the importance of localized models. While MLR provides a reliable baseline, additional adaptive modeling approaches are necessary for comprehensive forecasting.

5.3 Hybrid MLR–LSTM Integration

The hybrid MLR–LSTM model outperformed the standalone MLR by combining interpretability with temporal flexibility. It captured nonlinear patterns and long-term trends that regression alone missed. Performance metrics, including R^2, RMSE, MAE, and MAPE, showed significant improvement, confirming the model's robustness. Prediction errors during extreme climatic events were reduced, enhancing reliability across regions. This demonstrates that hybrid models effectively balance statistical clarity with machine learning adaptability for agricultural forecasting.

5.4 Forecasting

Long-term forecasts up to 2100 revealed distinct regional variations in rice productivity. Central Luzon consistently exhibited the highest yields, while regions like Region 7 were more vulnerable to environmental fluctuations. These trends highlight areas that may require adaptive management strategies. Combining historical data with machine learning allows for more accurate anticipation of future yields. The forecasts provide practical insights to guide planting schedules and long-term agricultural planning.

5.5 Web-Based Platform and Generative AI Chatbot

The GR-AI-N platform, integrated with the GRAINKEEPER chatbot, demonstrated usability and real-world applicability. Survey results indicated that users found the system functional, adaptable, and easy to navigate. The chatbot provides natural-language, data-driven recommendations based on the hybrid model outputs. It translates complex predictive analytics into actionable advice for farmers. This AI-driven interface supports precision agriculture, improving decision-making and promoting sustainable crop management.

6. CONCLUSION

This study was designed to develop and test GR-AI-N, a hybrid forecasting system that combines Multiple Linear Regression (MLR), Long Short-Term Memory (LSTM) networks, and a Generative AI chatbot to improve rice yield predictions and planting time recommendations in the Philippines. The analysis showed that key geo-climatic factors—such as temperature, precipitation, wind speed, and dew point—had strong positive relationships with rice yield, while humidity exhibited a notable negative association. These findings confirmed that including these variables is important for accurate forecasting.

The region-based MLR models performed well as a baseline forecasting method, with an average R^2 of 0.87 and error rates (MAPE) below 10%, indicating reliable performance under typical conditions. However, their ability to capture complex patterns diminished during years with extreme climate events, where the assumption of linear relationships was insufficient.

To address this limitation, the study incorporated LSTM models, which are capable of learning sequential and nonlinear patterns within climatic data. The LSTM models significantly improved forecasting accuracy compared to regression alone, producing much lower error rates and maintaining a consistently high explanatory power, with an average R^2 of 0.9977 across 576 test cycles. The average deviation of LSTM predictions from actual yields, measured by MAPE, was approximately 0.46%, highlighting the model's exceptional precision and reliability.

By integrating MLR and LSTM into a hybrid framework, the system combined the interpretability and stability of regression with the flexibility and strength of deep learning. This hybrid model outperformed both standalone approaches, delivering near-perfect alignment with historical yield data and achieving the lowest overall error rates in the study.

Finally, embedding these models into a web-based platform with a Generative AI chatbot and geospatial dashboard ensured that complex scientific outputs were translated into clear, practical recommendations for farmers. Evaluation of the chatbot using confusion matrix testing over 15 iterations showed very high scores in accuracy, recall, and precision, with an overall F-1 score of 0.9766, demonstrating its reliability as an interactive advisory tool. Positive ratings for usability, acceptability, adaptability, and functionality further supported GR-AI-N's potential as a scalable, climate-smart decision support system that can enhance productivity and resilience in Philippine rice farming.

Acknowledgments: The researchers wish to extend their sincere gratitude to all individuals and institutions who contributed to the successful completion of this study. Special appreciation is due to Sir Nelson Malificiado, the research advisor, for his invaluable guidance and academic support, and to Ma'am Fretzelaine Dasas, the MCS coach, for her mentorship and continuous encouragement throughout the research process. The researchers also acknowledge Mr. Selwyn Uy for his consultancy and assistance in website design, as well as the Municipal Agricultural Office of the Local Government Unit of Tupi for providing valuable support and domain insights. Heartfelt thanks are extended to the survey respondents for their time and cooperation, and to the researchers' parents for their unwavering support and motivation. The researchers extend their appreciation to Archie Sarto, a classmate and comrade in MCS research, for his collaboration and shared efforts throughout the study, special recognition is also given to Ayeeshah Latip for her assistance in coding and technical support. The researchers also acknowledge Febellen Rejas and Alexander Callueng as project pioneers and seniors whose work and insights served as inspiration for this study. Above all, the researchers offer thanks to God for the wisdom, strength, and guidance that made this research possible.

References

Ahmed, A., et al. (2019). Seasonal climate patterns and ideal planting times for rice in the Philippines. Philippine Journal of Crop Science, 12(4), 211–225

Adisa, O. M., Botai, J. O., Adeola, A. M., Hassen, A., & Botai, C. M. (2022). Machine learning-based predictive modeling of maize yield using climate variables in South Africa. Agricultural and Forest Meteorology, 312, 108715. https://doi.org/10.1016/j.agrformet.2021.108715

Bian, L., Zhang, Y., & Chen, H. (2025). Effects of organic amendments on paddy field productivity. International Journal of Agronomy, 15(2), 45–60.

Callueng, A. E., Latip, A. Y. S., & Rejas, F. E. (2025). Multiple linear regression analysis of geo-climatic variables on rice yield in South Cotabato.

Casinillo, F., et al. (2023). Data-informed cultivation strategies and yield efficiency in Philippine rice farms. Philippine Journal of Agricultural Economics, 45(1), 55–67.

Chauhan, B. S., Jabran, K., & Mahajan, G. (Eds.). (2017). Rice production worldwide. Springer International Publishing. https://doi.org/10.1007/978-3-319-47516-5

Dubey, S., et al. (2018). Climate anomalies and rice yield disruptions in Southeast Asia. Climate Research, 76(3), 215–228.

Food and Agriculture Organization. (2013). Crop yield forecasting methods and early warning systems: A literature review. FAO. https://www.fao.org

Food and Agriculture Organization. (2024). FAOSTAT. https://www.fao.org/faostat/en

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., … Zaks, D. P. M. (2020). Solutions for a cultivated planet. Nature, 478(7369), 337–342. https://doi.org/10.1038/nature10452

Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concepts and techniques (3rd ed.). Elsevier.

Hasan, M., et al. (2020). Mechanized rice farming: Improving yield and labor efficiency. Journal of Smart Agriculture, 11(2), 89–101.

Hatfield, J. L., & Beresford, N. A. (2016). Effects of heat stress on spikelet sterility in rice. Agricultural Meteorology, 218, 15–23.

Jin, Y., et al. (2025). Climate-induced salt stress on hybrid rice varieties. Plant Science Journal, 212(5), 88–103.

Khaki, S., & Wang, L. (2019). Crop yield prediction using deep neural networks. Frontiers in Plant Science, 10, 621. https://doi.org/10.3389/fpls.2019.00621

Kotsiantis, S. B., Kanellopoulos, D., & Pintelas, P. E. (2006). Data preprocessing for supervised learning. International Journal of Computer Science, 1(2), 111–117.

Kumar, S., et al. (2021). Climate adaptation strategies for rice farming. Journal of Environmental Management, 283, 111–123.

Lasalita-Zapico, R., et al. (2008). Traditional rice landraces and biodiversity conservation in Lake Sebu, Philippines. Philippine Journal of Crop Science, 33(1), 77–88.

Nagai, T., & Makino, A. (2019). Optimal temperature ranges for rice growth. Plant Physiology Journal, 12(3), 140–150.

Ocampo, O., Makahiya, H. A., Pamplona, M., Motilla, J. K., Meynard, P., Bueno, C., & Sta Cruz, P. (2023). Variability of cost and return of dry direct-seeded lowland rice in major rainfed growing areas in the Philippines. Philippine Agricultural Scientist, 106(2), 208–216. https://ovcre.uplb.edu.ph/journals-uplb/index.php/PAS/article/view/

Philippine News Agency. (2024). Rice farmers in South Cotabato.

Phuong, T. T., Pham, H. T., & Nguyen, T. N. (2022). Data preprocessing impact on LSTM-based crop yield forecasting using climatic time-series data. Computers and Electronics in Agriculture, 202, 107353. https://doi.org/10.1016/j.compag.2022.107353

Qu, X., Kojima, D., Wu, L., & Ando, M. (2021). Losses in the rice harvest process: A review. Sustainability, 13(17), 9627. https://doi.org/10.3390/su13179627

Remshard, M., & Queenborough, S. A. (2023). Design of tables for the presentation and communication of data in ecological and evolutionary biology. Ecology and Evolution, 13(7), e10062. https://doi.org/10.1002/ece3.10062

Sattar, R., et al. (2023). Relative humidity effects on rice transpiration and nutrient uptake. Agricultural Water Management, 285, 104118.

Stuecker, C., et al. (2018). ENSO-related rainfall disruptions and rice yield losses. Environmental Research Letters, 13(9), 095103. https://doi.org/10.1088/1748-9326/aadce6

Vicente, J., et al. (2019). Adoption of modern planting schedules among Filipino rice farmers. Philippine Journal of Agricultural Extension, 22(1), 33–45.

Wang, L., Feng, H., & Li, Z. (2018). Maize yield forecasting by linear regression and artificial neural networks in Jilin, China. The Journal of Agricultural Science, 156(9), 1125–1135. https://doi.org/10.1017/S0021859618000581

Wassmann, R., et al. (2009). Breeding heat-resistant rice varieties. Field Crops Research, 110(3), 85–98.

Weather and Climate. (2015). Historical climate data for South Cotabato. https://www.weather-and-climate.com

Yan, H., et al. (2025). Integrating time-series data and regression for crop yield forecasting. International Journal of Computational Agriculture, 11(1), 15–29.