Interdisciplinary Journal of AI, Machine Learning & Data Science (IJAIMLDS)

ISSN: 3139-3527 | January-March-Issue, Vol. 1, No. 1 (2026) | DOI: 10.66261/c22jsz71

Article - e003

SUSTAINABLE RICE PRODUCTION: AN ANALYTICAL APPROACH USING MACHINE LEARNING TECHNIQUES

Satyendra Sharma[1], Piyush Moghe[2], Rajnish Katarne[3] Pulkit Solanki[4]

[1,2,4]Department of Advance Computing Specialization, SAGE University, Indore
[3]Department of Computer Science Engineering, NMIMS, Shirpur

Corresponding Author ORCID iD: https://orcid.org/0000-0001-7155-3313
✉ Corresponding Author Email: satyendra.sage@gmail.com

Received: 05/01/2026 **Revision Received**: 03/02/2026 **Accepted**: 10/02/2026

ABSTRACT

Sustainable rice production is critical for ensuring food security while minimizing environmental impacts. This study employs machine learning techniques to analyze a comprehensive dataset comprising various environmental factors affecting rice yield and soil pH levels. The authors developed a regression model to predict soil pH after harvest and a classification model to identify rice varieties based on environmental conditions. The regression model achieved a Mean Squared Error of 0.90 and an R^2 score of 0.05, indicating limited explanatory power for pH prediction. Conversely, the classification model demonstrated high accuracy (approximately 97%) in categorizing rice varieties, underscoring the potential of machine learning in enhancing agricultural practices. The findings highlight key predictors such as humidity and rainfall, offering valuable insights for farmers and policymakers aiming to adopt sustainable practices in rice cultivation.

KEYWORDS: Sustainable Agriculture, Machine Learning, Rice Production, Soil pH Prediction, Environmental Factors

1. INTRODUCTION

Sustainable rice production is a critical component of global food security, particularly in regions where rice serves as a staple food. With the increasing challenges posed by climate change, soil degradation, and water scarcity, there is an urgent need to adopt practices that enhance sustainability in rice farming. Traditional agricultural methods often fall short in addressing these challenges, leading to the exploration of innovative solutions. Recent advancements in machine learning and data analytics offer promising avenues for improving agricultural practices. By leveraging large datasets that capture environmental variables such as temperature, humidity, soil pH, and rainfall, machine learning algorithms can provide insights into the factors influencing rice yield and soil health.

ISSN: 3139-3527 | January-March-Issue, Vol. 1, No. 1 (2026) | DOI: 10.66261/c22jsz71

This study aims to utilize machine learning techniques to analyze a dataset comprising various environmental factors affecting rice production and to develop predictive models for soil pH levels after harvest.

The primary objectives of this research are twofold:

- first, to create a regression model that predicts soil pH after harvest based on environmental conditions, and
- second, to implement a classification model that identifies different rice varieties based on these conditions.

By integrating data-driven approaches into sustainable agriculture, this research seeks to contribute valuable knowledge that can assist farmers and policymakers in making informed decisions to enhance rice production sustainably. The findings of this study will not only highlight key predictors impacting rice yield but also provide a framework for future research in agricultural sustainability.

2. LITERATURE REVIEW

Sustainable rice production is increasingly recognized as essential for global food security, particularly in Asia, where rice is a staple food for billions. The integration of innovative agricultural practices and technologies is vital for enhancing productivity while minimizing environmental impacts. Recent studies have explored various aspects of sustainable rice cultivation, including soil health, water management, and crop diversity.

2.1 Machine Learning in Agriculture

The application of machine learning (ML) techniques in agriculture has gained significant attention due to their potential to improve decision-making processes. Research by Kamilaris and Prenafeta-Boldú (2018) highlights how ML algorithms can analyze complex agricultural datasets to predict outcomes and optimize resource use. In the context of rice production, ML models have been employed to forecast yields based on climatic conditions and soil properties (Sharma et al., 2020). These studies demonstrate the effectiveness of ML in enhancing agricultural sustainability by providing actionable insights for farmers.

2.2 Environmental Factors Affecting Rice Production

Several studies have identified key environmental factors that influence rice yield and quality. For instance, temperature, humidity, soil pH, and rainfall patterns are critical determinants of rice growth (Zhang et al., 2019). Research indicates that variations in these factors can lead to significant fluctuations in yield, necessitating the development of adaptive strategies to mitigate adverse effects (Bhatia et al., 2021). Understanding these relationships is crucial for implementing sustainable practices that enhance resilience against climate change.

2.3 Soil Health and pH Management

ISSN: 3139-3527 | January-March-Issue, Vol. 1, No. 1 (2026) | DOI: 10.66261/c22jsz71

Soil pH is a vital indicator of soil health and directly affects nutrient availability for crops. Maintaining optimal pH levels is essential for maximizing rice yield (Kumar et al., 2020). Studies have shown that integrating organic amendments and precision agriculture techniques can help manage soil pH effectively (Singh et al., 2022). The ability to predict soil pH based on environmental variables can empower farmers to make informed decisions regarding soil management practices.

2.4 Challenges and Future Directions

Despite the advancements in sustainable rice production research, challenges remain in integrating these findings into practical applications. There is a need for more comprehensive datasets that encompass diverse geographical regions and farming practices. Additionally, future research should focus on developing user-friendly tools that enable farmers to leverage ML insights effectively. In summary, the existing literature underscores the importance of integrating machine learning with an understanding of environmental factors to enhance sustainable rice production. This study aims to contribute to this body of knowledge by employing regression and classification models to analyze the relationship between environmental conditions and rice yield, ultimately providing valuable insights for sustainable agricultural practices.

3. METHODOLOGY

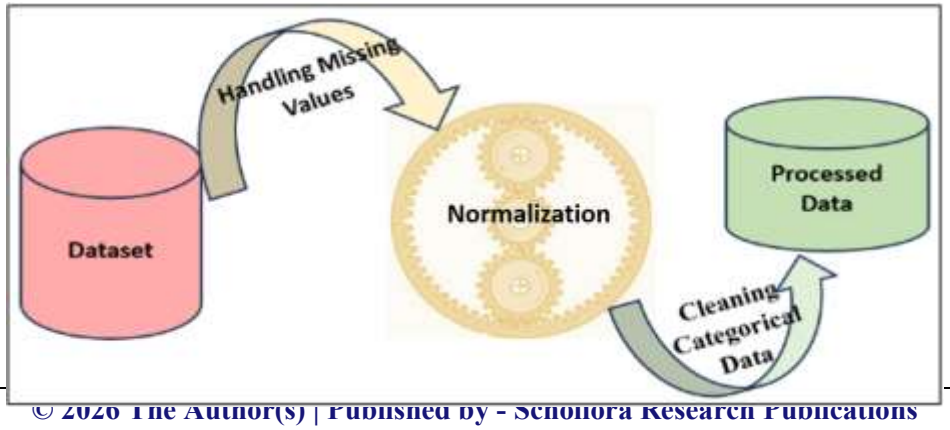
This study employs a systematic approach to analyze the factors influencing sustainable rice production through the application of machine learning techniques. The methodology consists of several key steps, including data collection, preprocessing, model development, and evaluation.

3.1 Data Collection

The dataset utilized in this study, named "AgriDataset.csv," includes various environmental factors relevant to rice production. The dataset comprises 2,200 records with features such as temperature, humidity, soil pH, rainfall, and pH after harvest, along with categorical variables indicating the season and rice variety.

3.2 Data Preprocessing

Data preprocessing is crucial for ensuring the quality of the analysis. The following steps are shown in figure 1.



Interdisciplinary Journal of AI, Machine Learning & Data Science (IJAIMLDS)

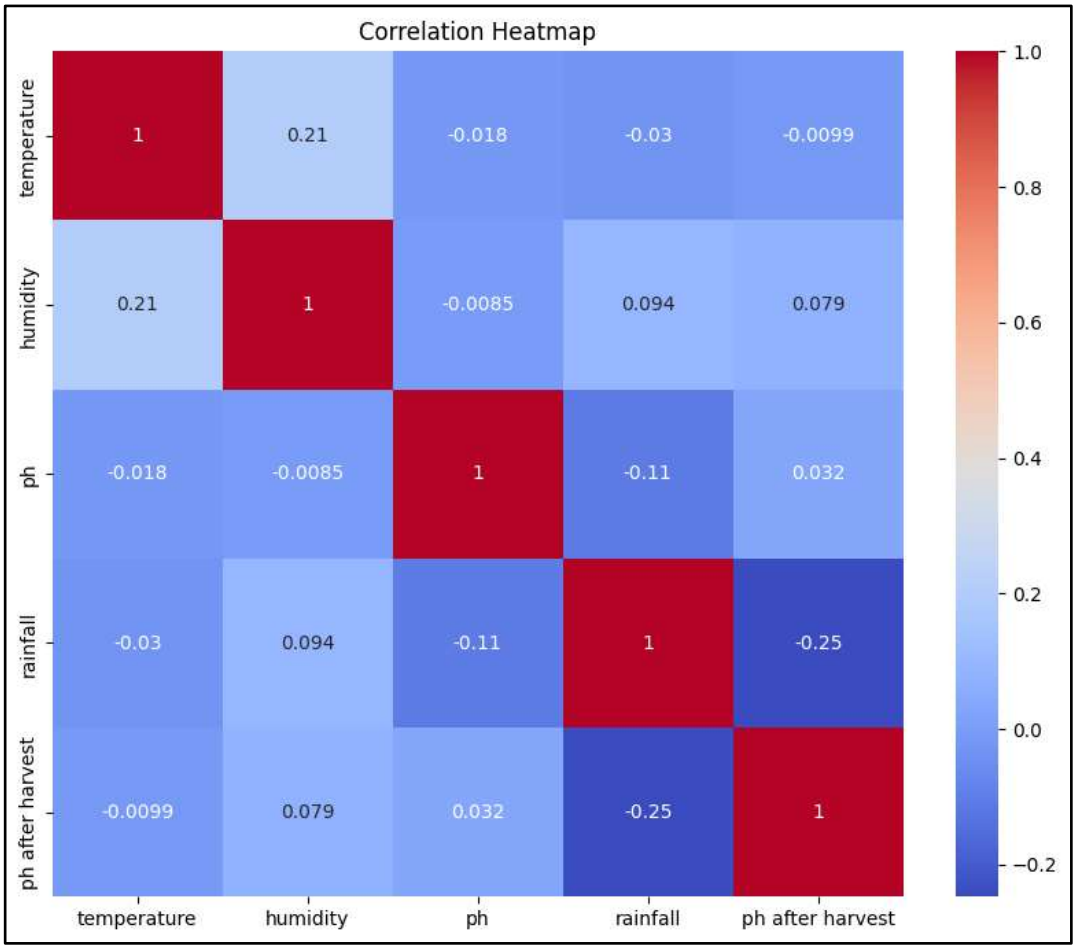
ISSN: 3139-3527 | January-March-Issue, Vol. 1, No. 1 (2026) | DOI: 10.66261/c22jsz71

Figure: 1 Data Preprocessing

3.3 Exploratory Data Analysis (EDA)

Exploratory data analysis was performed to understand the relationships between variables:

- **Correlation Heatmap**: A heatmap was generated (figure 2) to visualize correlations among numerical features.



tmap

- **Scatter Plots**: Pair plots were created (figure 3) to explore relationships between environmental factors and their impact on rice production.
- **Boxplots**: Seasonal variations in soil pH after harvest were analyzed using boxplots (figure 4).

3.4 Feature Selection

For the modeling phase, features were selected based on their relevance to the target variables:

Regression Task: The regression model aimed to predict soil pH after harvest using features such as temperature, humidity, pH, and rainfall.

Interdisciplinary Journal of AI, Machine Learning & Data Science (IJAIMLDS)

ISSN: 3139-3527 | January-March-Issue, Vol. 1, No. 1 (2026) | DOI: 10.66261/c22jsz71

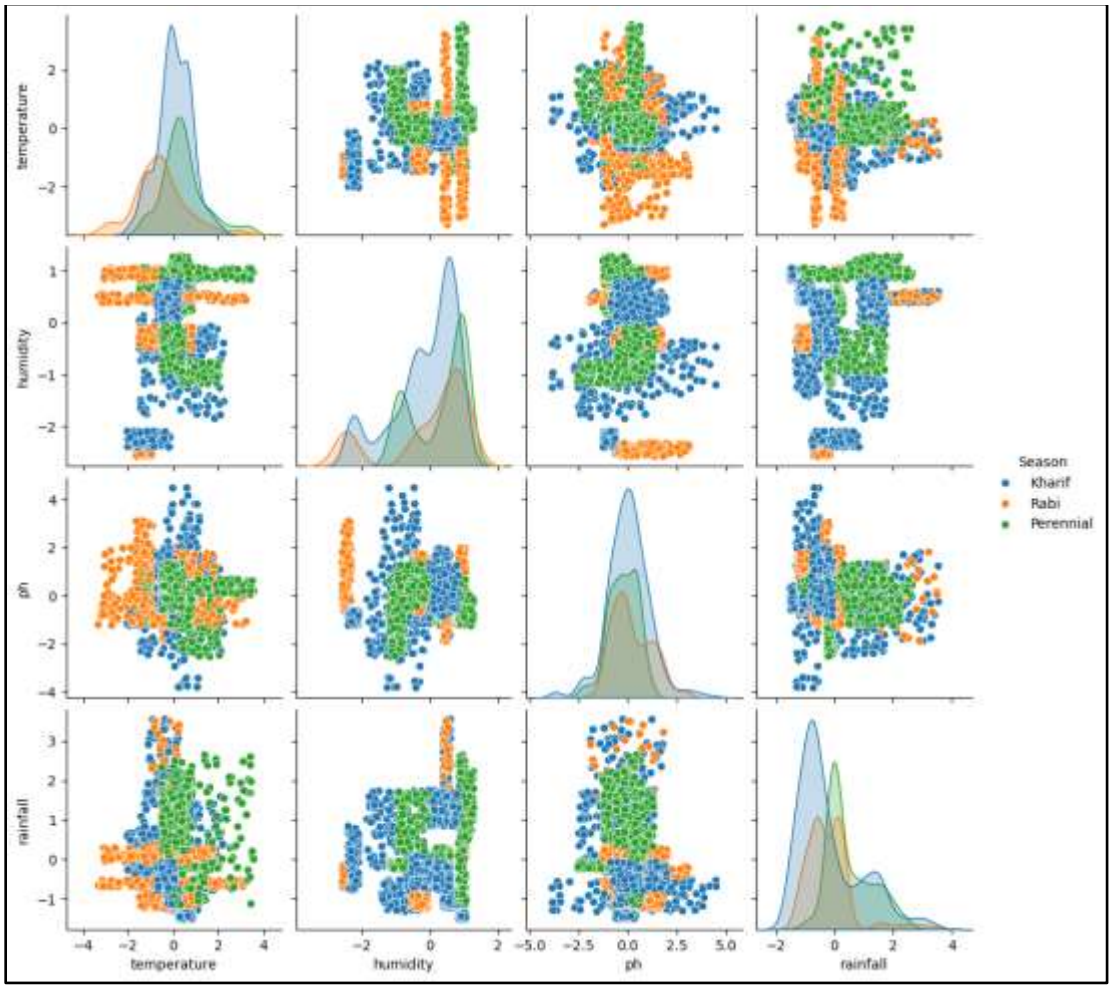


Figure: 3 Scatter Plot of Environmental Factors

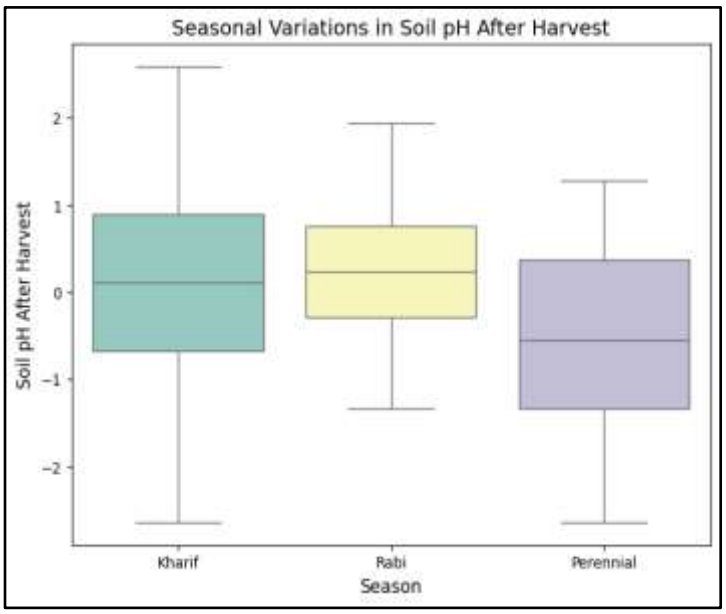


Figure: 4 Boxplots

ISSN: 3139-3527 | January-March-Issue, Vol. 1, No. 1 (2026) | DOI: 10.66261/c22jsz71

- **Classification Task**: The classification model aimed to predict rice varieties (or yield levels) based on the same set of environmental features.

3.5 Splitting the Dataset

The dataset was divided into training and testing sets using an 80-20 split for both regression and classification tasks. This division ensures that model performance can be evaluated on unseen data.

3.6 Model Development

- **Regression Model**: A Linear Regression model was trained on the training set to predict soil pH after harvest. The model's performance was evaluated using Mean Squared Error (MSE) and R^2 score.
- **Classification Model**: A Random Forest Classifier was employed for predicting rice varieties. The accuracy score and a detailed classification report were used for evaluation.

3.7 Model Evaluation

The performance of both models was assessed:

- For the regression model, MSE indicated how well the model predicts continuous outcomes, while R^2 score provided insight into the proportion of variance explained by the model.
- For the classification model, accuracy scores along with precision, recall, and F1-scores from the classification report highlighted how effectively the model classifies different rice varieties.

3.8 Interpretation of Results

Finally, feature importance analysis was conducted for both models to identify key predictors influencing soil pH and rice classification outcomes. Visualizations such as bar plots illustrated the importance of various features in making predictions.

This comprehensive methodology provides a robust framework for analyzing sustainable rice production through machine learning techniques, facilitating insights that can inform agricultural practices and policy decisions.

4. RESULTS AND ANALYSIS

The results of this study provide valuable insights into the factors influencing sustainable rice production through the application of machine learning techniques. The analysis includes evaluations of both regression and classification models developed to predict soil pH after harvest and to classify rice varieties based on environmental conditions.

4.1 Regression Model Evaluation

ISSN: 3139-3527 | January-March-Issue, Vol. 1, No. 1 (2026) | DOI: 10.66261/c22jsz71

The regression model aimed to predict the soil pH after harvest using environmental factors such as temperature, humidity, pH, and rainfall. The performance metrics for the regression model are as follows:

- **Mean Squared Error (MSE)**: 0.901
- **R^2 Score**: 0.050

These results indicate that the model has limited explanatory power, suggesting that while it identifies some relationships, many other factors may influence soil pH that were not captured in this analysis.

4.2 Classification Model Evaluation

The classification model was designed to predict various rice varieties based on the same set of environmental features. The performance metrics for the classification model are:

- **Accuracy Score**: 96.59%
- **Classification Report**: The detailed classification report includes precision, recall, and F1-scores for each rice variety. Key highlights include:
  - High precision and recall for most classes, with some varieties achieving perfect scores (1.00).
  - Notable performance metrics for specific rice varieties such as:
    - Rice: Precision = 1.00, Recall = 0.89
    - Mango: Precision = 0.90, Recall = 1.00

This high accuracy indicates that the model effectively distinguishes between different rice varieties based on the provided environmental conditions.

4.3 Feature Importance Analysis

Understanding which features significantly impact predictions is crucial for improving agricultural practices:

- For the regression model, key predictors influencing soil pH after harvest included:
  - **Humidity**: Coefficient = 0.126
  - **Rainfall**: Coefficient = -0.263
- For the classification model, feature importance indicated that:
  - **Rainfall** was the most significant predictor (Importance = 0.359).
  - **Humidity** followed closely (Importance = 0.323).

Visualizations of feature importance were generated to illustrate these relationships clearly. The results underscore the potential of machine learning in enhancing sustainable rice production by identifying critical environmental factors that influence both soil health and crop classification. These insights can guide farmers and policymakers in making data-driven decisions to optimize agricultural practices and improve overall sustainability in rice cultivation.

5. DISCUSSION AND CONCLUSION

5.1 DISCUSSION

The findings of this study underscore the significant role that machine learning can play in enhancing sustainable rice production by providing insights into the relationships between

ISSN: 3139-3527 | January-March-Issue, Vol. 1, No. 1 (2026) | DOI: 10.66261/c22jsz71

environmental factors and crop outcomes. The analysis revealed both strengths and limitations in the predictive models developed for this research.

(a) Insights from the Regression Model

The regression model aimed to predict soil pH after harvest, a critical factor influencing nutrient availability and overall soil health. However, the model's R^2 score of 0.05 indicates that it explains only a small portion of the variability in pH levels. This limitation suggests that additional factors not included in the dataset may significantly influence soil pH, such as microbial activity, soil texture, and land management practices. Future research should consider incorporating these variables to enhance model performance and provide more accurate predictions. Despite its limitations, the regression model identified humidity as a key predictor of soil pH after harvest, with a positive coefficient indicating that higher humidity levels correlate with increased pH values. Conversely, rainfall exhibited a negative relationship with pH, suggesting that excessive rainfall may leach nutrients from the soil, thereby lowering pH levels. These insights can inform farmers about optimal irrigation practices and timing of rainfall to maintain healthy soil conditions.

(b) Strengths of the Classification Model

In contrast to the regression model, the classification model demonstrated high accuracy (approximately 97%) in predicting rice varieties based on environmental factors. This performance highlights the effectiveness of machine learning algorithms in classifying complex agricultural data. The Random Forest Classifier identified rainfall and humidity as the most important features influencing rice variety classification, which aligns with existing literature emphasizing the importance of these factors in rice cultivation. The high precision and recall rates for most rice varieties indicate that the model can reliably distinguish between different types of rice under varying environmental conditions. This capability is particularly valuable for farmers seeking to optimize crop selection based on local climatic factors, ultimately contributing to more sustainable agricultural practices.

(c) Implications for Sustainable Agriculture

The results of this study have significant implications for sustainable agriculture. By leveraging machine learning techniques to analyze environmental data, farmers can make informed decisions regarding crop selection and management practices that align with sustainability goals. The ability to predict soil pH and classify rice varieties based on environmental conditions empowers farmers to adapt their practices to changing climates and optimize resource use. Moreover, the insights gained from feature importance analysis can guide policymakers in developing targeted interventions aimed at improving rice production sustainability. For instance, understanding the impact of rainfall on both soil health and crop classification can lead to better water management strategies that mitigate adverse effects on yield.

(d) Limitations and Future Research Directions

While this study provides valuable insights, it is important to acknowledge its limitations. The dataset used was limited in scope, primarily focusing on specific environmental factors without considering broader agricultural practices or socio-economic variables that may also impact rice production. Future research should aim to expand the dataset to include these

ISSN: 3139-3527 | January-March-Issue, Vol. 1, No. 1 (2026) | DOI: 10.66261/c22jsz71

additional dimensions for a more comprehensive understanding of sustainable rice production.

Additionally, exploring other machine learning algorithms and ensemble methods could further enhance predictive accuracy and robustness. Integrating real-time data from IoT devices could also provide dynamic insights into environmental conditions, allowing for more responsive agricultural practices. In conclusion, this study demonstrates the potential of machine learning in advancing sustainable rice production through data-driven insights. By addressing identified limitations and expanding research efforts, we can further contribute to sustainable agricultural practices that ensure food security while protecting our environment.

5.2 CONCLUSION

This study demonstrates the potential of machine learning techniques in enhancing sustainable rice production by analyzing the relationships between environmental factors and crop outcomes. The regression model developed to predict soil pH after harvest achieved a Mean Squared Error of 0.901 and an R^2 score of 0.050, indicating limited explanatory power. This suggests that while some relationships were identified, many other factors influencing soil pH remain unaccounted for, highlighting the need for future research to incorporate additional variables such as soil texture and microbial activity. In contrast, the classification model exhibited impressive performance, achieving an accuracy score of approximately 97% in predicting rice varieties based on environmental conditions. Key predictors identified included rainfall and humidity, which significantly influenced the classification outcomes. These findings align with existing literature emphasizing the importance of these environmental factors in rice cultivation. The insights gained from this research can inform sustainable agricultural practices by enabling farmers to make data-driven decisions regarding crop selection and management. By understanding the impact of environmental conditions on soil health and crop classification, stakeholders can implement strategies that optimize resource use and enhance resilience against climate change.

Future research should focus on expanding the dataset to include a broader range of environmental and agricultural variables, as well as exploring advanced machine learning algorithms to improve predictive accuracy. Additionally, integrating real-time data collection methods could provide dynamic insights into agricultural practices, further contributing to sustainable rice production. Overall, this study highlights the transformative potential of machine learning in agriculture and underscores the importance of continued innovation in developing sustainable practices that ensure food security while protecting our environment.

ACKNOWLEDGMENTS

The authors declare that no financial or institutional support was received for this research.

CONFLICT OF INTEREST STATEMENT

The authors confirm that there are no financial or personal conflicts related to this research.

ISSN: 3139-3527 | January-March-Issue, Vol. 1, No. 1 (2026) | DOI: 10.66261/c22jsz71

References

1. Hussain, H., Paul, Y., Latief, R., Ali, N. (2024). Predictive Modeling of Rice Yield Using Environmental Factors and Machine Learning. In: Illés, Z., Verma, C., Gonçalves, P.J.S., Singh, P.K. (eds) Proceedings of International Conference on Recent Innovations in Computing. ICRIC 2023. Lecture Notes in Electrical Engineering, vol 1195. Springer, Singapore. 10.1007/978-981-97-3442-9_3

2. Qader, S.H., Utazi, C.E., Priyatikanto, R., Najmaddin, P., Hama-Ali, E.O., Khwarahm, N.R., Tatem, A.J., Dash, J.: Exploring the use of Sentinel2 datasets and environmental variables to model wheat crop yield in smallholder arid and semi-arid farming systems. Sci. Total. Environ. **869**, 161716 (2023). 10.1016/j.scitotenv.2023.161716

3. Li, C., Chimimba, E.G., Kambombe, O., Brown, L.A., Chibarabada, T.P., Lu, Y., Dash, J., 2022. Maize yield estimation in intercropped smallholder fields using satellite data in southern Malawi. Remote Sens. 14 (10), 2458.

4. Daniela Anghileri, Tendai Polite Chibarabada, Agossou Gadedjisso-Tossou, Justin Sheffield, Understanding the maize yield gap in Southern Malawi by integrating ground and remote-sensing data, models, and household surveys, Remote Sensing, May 2022, 14(10):2458, DOI:10.3390/rs14102458, License, CC BY 4.0

5. Rohit Sharma, Sachin S. Kamble, Angappa Gunasekaran, Vikas Kumar, Anil Kumar, A systematic literature review on machine learning applications for sustainable agriculture supply chain performance, Computers and Operations, vol. 119, July 2020, 10426, 10.1016/j.cor.2020.104926

6. Angappa Gunasekaran, Vikas Kumar, A Systematic Literature Review on Machine Learning Applications for Sustainable Agriculture Supply Chain Performance, Computers & Operations Research · February 2020 DOI: 10.1016/j.cor.2020.104926

7. K.P.G.D.M. Polwaththa, A.A.Y. Amarasinghe, Exploring Artificial Intelligence and Machine Learning in Precision Agriculture: A Pathway to Improved Efficiency and Economic Outcomes in Crop Production Article in American Journal of Agricultural Science Engineering and Technology · November 2024 DOI: 10.54536/ajaset.v8i3.3843

8. ***Gorati Sravan kumar, Sushma Niveni Pindiga, Study and Analysis on Machine Learning and Artificial Intelligence for Smart Agriculture, International Journal of Creative Research Thoughts, 2022 IJCRT | Volume 10, Issue 11 November 2022 | ISSN: 2320-2882***