Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 7, No.3, July 2025, pp. 02503023-01 ~ 02503023-10
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v7i3.2019

Artificial Neural Network-Based Forecasting of Rice Yield Using Environmental and Agricultural Data

Priyanto, Muhammad Faisal*, Mochamad Imamudin

Informatics Engineering, Universitas Islam Negeri Maulana Malik Ibrahim Jl. Gajayana No.50, Malang City, East Java, Indonesia

*mfaisal@ti.uin-malang.ac.id

Abstract. This study presents a high-accuracy predictive model for rice production in Indonesia using Artificial Neural Networks (ANN), achieving an R^2 of 98.11%, Mean Absolute Error (MAE) of 0.0966, and Mean Squared Error (MSE) of 0.0189. Climate variability remains a significant challenge to rice cultivation in regions like Malang City, where unpredictable environmental factors such as rainfall, temperature, and humidity hinder effective crop planning and yield estimation. To address this, we developed a Multilayer Perceptron (MLP)-based ANN model incorporating agro-environmental variables: rainfall, temperature, humidity, harvested area, and production quantity. Historical data from 2009 to 2024 were sourced from the Meteorology, Climatology, and Geophysics Agency (BMKG) and the Central Statistics Agency (BPS). The dataset underwent preprocessing, including cleaning, feature extraction, Z-Score normalization, and partitioning into training and testing sets. The proposed ANN architecture consists of an input layer, three hidden layers, and an output layer for regression tasks. Comparative evaluation against Random Forest, K-Nearest Neighbors, and Support Vector Regression demonstrated the ANN's superior ability to model complex nonlinear relationships in agricultural data. The results highlight the role of intelligent data-driven systems in enhancing the accuracy of yield forecasting, supporting sustainable agricultural practices, and informing national food security policy.

Keywords: Artificial Neural Network, Rice Yield Prediction, Agro-environmental, Climate-smart agriculture, Sustainable Farming

(Received 2025-05-26, Revised 2025-07-01, Accepted 2025-07-01, Available Online by 2025-07-17)

1. Introduction

Indonesia is one of the largest agrarian countries in the world, playing a crucial role in food supply, especially rice, which is the main staple for more than 270 million people [1]. As a country with diverse climates and geographical conditions, Indonesia faces significant challenges in rice production. Farmers often struggle to achieve optimal yields, especially amidst the uncertainty caused by climate change and varying soil conditions [2]. Increasingly erratic weather patterns, such as extreme rainfall and fluctuating temperatures, contribute to high uncertainty, making it difficult to predict and plan for optimal crop production[3]. This calls for serious attention from various stakeholders, including the government, researchers, and agricultural practitioners, to find effective solutions to ensure national food security[4].

Unpredictable weather, such as irregular rainfall, can lead to droughts or floods, both of which can damage rice crops [5]. Furthermore, soil conditions, including pH levels and moisture, also play a critical role in determining the quality and quantity of the harvest. This uncertainty creates difficulties for farmers in planning production, which ultimately threatens food security in the region. Artificial Neural Networks (ANN), a method in artificial intelligence (AI), have been proposed as a solution due to their ability to model nonlinear relationships between various environmental factors and crop yields [6]. Unlike traditional statistical methods, ANNs can model complex dependencies between variables like rainfall, temperature, humidity, and soil pH, offering more accurate predictions under uncertain conditions [7]. ANN has also proven capable of handling irregular and noisy data, which is typical in agricultural contexts [8], [9].

Several recent studies have explored the application of advanced deep learning models, particularly Long Short-Term Memory (LSTM), in crop yield prediction tasks due to their effectiveness in handling time-series data. For instance, [10] developed an LSTM model optimized with an Improved Optimization Function (IOF) and demonstrated that it outperformed other architectures such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Gated Recurrent Units (GRU), achieving an RMSE of 2.19 and MAE of 25.4 Similarly, [11] compared LSTM, GRU, and CNN models for rice crop detection using time-series satellite data and found that temporal models like LSTM provided more stable and accurate results across seasonal datasets. Despite these advancements, most studies have focused on LSTM or CNN, while Artificial Neural Networks (ANN) remain relatively underutilized in agricultural prediction tasks. [12] acknowledged this imbalance and emphasized the potential of ANN in sustainable agrarian applications, especially where data availability is limited or interpretability is required. Moreover, [13] found that ANN models, although simpler, achieved competitive accuracy compared to LSTM for crop yield prediction based on climatic parameters. These findings reveal a significant research gap regarding the comparative effectiveness and applicability of ANN, particularly in modeling nonlinear agro-environmental relationships under climate uncertainty.

Therefore, this study will focus solely on using ANN as the primary method for predicting rice yields. The ANN model in this research adopts a fixed Multilayer Perceptron (MLP) architecture with an input layer, three hidden layers (64, 32, and 16 neurons), and one output layer, using ReLU activation in hidden layers and a linear function in the output. The model inputs include weather-related variables (rainfall, temperature, humidity) and agricultural factors (soil pH, harvest area, and nutrients), with data spanning from 2009 to 2024. To ensure effective learning and generalization, preprocessing steps include Z-Score normalization and data cleaning. The model is trained using Stochastic Gradient Descent (SGD) with 200 epochs and a batch size of 32. A 5-fold cross-validation strategy is applied to evaluate model robustness and prevent overfitting.

The use of a fixed architecture is intentional, as it provides a balance between performance, simplicity, and reproducibility—important considerations for practical implementation in sustainable agricultural systems. While LSTM-based models are often more complex and resource-intensive, this research demonstrates that a properly designed ANN can offer comparable predictive accuracy with greater interpretability[14], [15]. This study contributes to sustainability science by providing an accessible and scalable approach to intelligent crop yield forecasting. The findings can support agricultural planning, mitigate climate-related risks, and ultimately improve national food security.

2. **Methods**

This study proposes the use of Artificial Neural Networks (ANN), specifically the Multilayer Perceptron (MLP) architecture, to predict rice yields based on weather and soil condition data. The superior performance of the Artificial Neural Network (ANN) model in this study can be attributed to its ability to model complex, non-linear relationships between agro-environmental variables. Unlike traditional models such as K-Nearest Neighbors (KNN) or Support Vector Regression (SVR), which depend on relatively simple approximation strategies, ANN utilizes multiple layers and nonlinear activation functions to capture intricate interactions in data with high variability and noise. These characteristics are especially crucial for modeling agricultural data, where environmental conditions do not follow linear trends. For instance, a recent study [16] in the Chhattisgarh region of India demonstrated that ANN models outperformed traditional regression techniques including SMLR, LASSO, and Ridge

Regression—in predicting rice yield, owing to their ability to handle nonlinear and high-dimensional climate-agriculture relationships

This methodology is further illustrated in the flow chart below, which provides a visual representation of the steps involved, from data collection and preprocessing to model development, training, and evaluation.

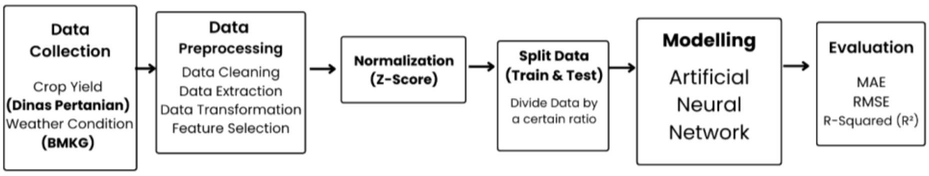

Figure 1. Research flow diagram

2.1. Data Collection

The dataset used in this study consists of weather and agricultural data relevant to rice production in Malang City. Weather data comprising rainfall, temperature, and humidity were sourced from the Meteorology, Climatology, and Geophysics Agency (BMKG), while agricultural data, such as harvested area and rice yield from 2009 to 2024, were obtained from the Central Statistics Agency (BPS). These two data sources provide essential input features for the predictive model, capturing the influence of environmental and soil-related conditions on rice yields.

2.2. Data Preprocessing

To ensure data quality and suitability for ANN modeling, several preprocessing steps were conducted. These include data cleaning to address missing values and outliers, feature extraction to reduce redundancy, and transformation of categorical variables into numerical formats[17]. The selected features rainfall, temperature, humidity, soil pH, moisture, and nutrients were then standardized using Z-Score normalization, ensuring all features have a mean of 0 and standard deviation of 1.

.

2.3. Normalization (Z-Score)

To ensure that all the features contribute equally to the model, normalization is applied to the dataset. The Z-Score normalization method is used to transform the data so that each feature has a mean of 0 and a standard deviation of 1. This is especially important in ANN models, as the scaling of the data impacts the performance and convergence speed of the network.

2.4. Splitting Data (Train & Test)

The dataset is split into two subsets: one for training the model and the other for testing its performance. Typically, around 70-80% of the data is used for training, and the remaining 20-30% is reserved for testing. This division helps evaluate how well the model generalizes to unseen data.

2.5. ANN Model Development

This study adopts the Multilayer Perceptron (MLP) architecture of Artificial Neural Networks (ANN) for rice yield prediction. The model consists of an input layer with five features—rainfall, temperature, humidity, harvest amount, and harvest area—which represent key environmental and agricultural factors.

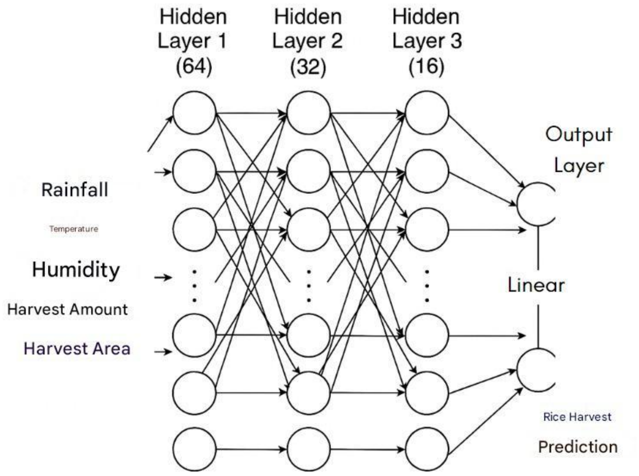

Figure 2. ANN architecture

As shown in Figure 2, the ANN architecture includes three hidden layers with 64, 32, and 16 neurons, respectively. Each layer uses the ReLU activation function to capture non-linear relationships within the data. These hidden layers enable the model to learn complex interactions, such as how temperature and humidity jointly influence rice growth. The final output layer contains a single neuron with a linear activation function, suitable for generating continuous output values in regression tasks.

2.6. Model Training with K-Fold Cross Validation

To evaluate the robustness and generalization of the model, a 5-fold cross-validation strategy was employed. The dataset was split into five equal parts; in each iteration, four parts were used for training and one part for testing. This process was repeated five times so that each subset served once as the testing set. The final performance metrics were averaged across all folds to assess model stability and reliability under varying data partitions.

The training process utilized a feedforward propagation mechanism, where the input data flows through multiple layers to compute predictions, followed by backpropagation for updating the model's weights. The optimization was carried out using Stochastic Gradient Descent (SGD) as the learning algorithm. The learning rate was set to 0.01, and the momentum was fixed at 0.9 to accelerate convergence and prevent local minima. The model was trained for 1,000 epochs per fold, with a batch size of 32 samples per iteration. Early stopping was not used, as the cross-validation strategy already provided a form of regularization and robustness check. The relatively high number of epochs ensured that the model had sufficient opportunity to converge, especially given the non-linear and multi-dimensional nature of agro-environmental data.

No hyperparameter tuning was applied in this study. Instead, the architecture and training parameters were predefined based on insights from prior empirical research and commonly adopted best practices for applying ANN in agricultural yield forecasting. This deliberate use of a fixed configuration was intended to balance model performance with computational efficiency, making the approach more accessible and reproducible especially in scenarios with limited computing resources or in field-based agricultural settings.

2.7 Evaluation Matrix

The performance of the ANN model is evaluated using three standard regression metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2). MAE measures the average magnitude of prediction errors, providing a straightforward indication of accuracy[18]. RMSE penalizes larger errors more heavily, offering insight into prediction consistency[19]. R^2 evaluates how well the model explains the variance in the actual data, with values closer to 1 indicating stronger predictive power[20]. Together, these metrics offer a comprehensive assessment of model accuracy and generalization ability.

3. Results and Discussion

3.1. Data Collection

The dataset used in this study was sourced from BMKG and BPS, covering the years 2009–2024. Weather data include average temperature (°C), humidity (%), and rainfall (mm), while agricultural data consist of harvested area and rice yield (tons). These features serve as key inputs for predicting rice production using the ANN model.

Table 1. Used Variables

Category	**Variable Name**	**Unit**
Input Features	Average Temperature	°C
	Humidity	%
	Rainfall	mm
	Harvest Area	Hectares
Output(Target)	Rice Yield	Tons

These variables capture essential agro-environmental conditions that influence rice yield and form the basis of model training and evaluation.

3.2. Data Preprocessing

Proper data preprocessing is essential to ensure that the data used in the Artificial Neural Networks (ANN) model is of high quality and suitable for analysis. The steps involved in data preprocessing are as follows:

3.2.1. Data Cleaning

Before training the ANN model, data cleaning was conducted to ensure quality and reliability. The raw dataset contained several missing values, duplicate entries, and potential outliers. Missing values were addressed using mean or median imputation based on the distribution of each variable. Duplicate records were removed to prevent redundancy, and outliers were either corrected or excluded to avoid skewing the model. Table 2 presents examples of how missing or invalid values were handled during this process.

Table 2. Data cleaning process

Variable	**Original Value**	**Imputed Value (Mean/Median)**
Temperature	22.5	23.0
Rainfall	NULL (8888)	120.0
Humidity	78.2	78.2

To further support data validation, Figure 1 illustrates the distribution and trend of rice production against key environmental variables such as average temperature (TAVG), rainfall (RR), and relative humidity (RH_AVG). These visualizations reveal the influence of each variable on rice yield over time and help identify extreme values or anomalies. For example, significant dips or spikes in rainfall or temperature often correspond with changes in production levels. Such visual checks are essential for confirming data integrity prior to model training.

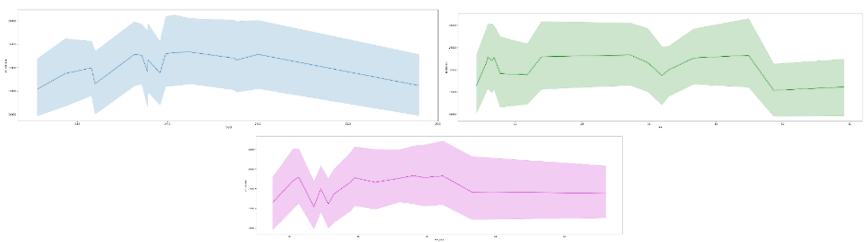

Figure 3. Production distribution to environmental variables

3.2.2. *Data Extraction and Transformation*

Data transformation and extraction are critical steps in preparing the dataset for the Artificial Neural Networks (ANN) model. During data extraction, only the relevant features from both the weather and agricultural datasets are selected, ensuring that unnecessary variables are excluded, which reduces redundancy and enhances model efficiency. This includes key environmental factors such as temperature, rainfall, and humidity, as well as agricultural variables like harvest area.

Table 3. Selected Features for Model

Feature	Description
Temperature (°C)	Average temperature for the season
Rainfall (mm)	Total rainfall during the growing season
Humidity (%)	Average humidity during the season
Rice Yield (tons)	Total rice yield for the season (target variable)
Harvest Area (hectares)	Area of land used for rice cultivation

The data transformation process follows, which standardizes the format of the dataset. This involves converting categorical variables into numerical values through encoding and aligning measurement units for consistency across the dataset, such as ensuring rainfall is in millimeters and temperature in degrees Celsius. These combined steps ensure that the data is both streamlined and consistent, making it ready for model training.

3.2.3. *Z Score Normalization*

Z-Score normalization was applied to standardize all features, ensuring they have a mean of 0 and a standard deviation of 1. This step is crucial for ANN models, which are sensitive to differences in feature scales.

Table 4. Z-Score Normalization

Feature	Original Value	Z-Score Normalized Value
Temperature	22.5	-0.75
Rainfall	120	1.25
Humidity	78.2	-0.20
Soil pH	5.8	0.10

As shown in Table 4, Z-Score normalization brings all features onto a comparable scale, helping the model learn more effectively. Without normalization, features like rainfall—due to their larger numerical range—could disproportionately influence the learning process. Standardizing the input ensures balanced contributions from all variables and improves the model's prediction accuracy.

3.2.4. *Split Data*

To evaluate model performance and generalization, the dataset was divided into training and testing subsets. Approximately 70–80% of the data was used for training, enabling the ANN to learn patterns and minimize prediction errors. The remaining 20–30% served as test data to assess the model's ability to make accurate predictions on unseen inputs. This approach helps prevent overfitting and ensures reliable evaluation on real-world scenarios.

3.3. *Modelling and Evaluation*

This section evaluates the performance of four machine learning algorithms: Artificial Neural Networks (ANN), Random Forest (RF), K-Nearest Neighbors (KNN), and Support Vector Regression (SVR), all selected for their effectiveness in handling complex regression tasks.To ensure robustness and reduce the risk of overfitting, model evaluation was conducted using a 5-fold cross-validation strategy. This iterative process provided a more reliable estimate of model performance across different data partitions.

The evaluation used R^2-Score, Mean Absolute Error (MAE), and Mean Squared Error (MSE) to measure model accuracy and error levels. ANN was chosen as the primary model due to its ability to capture non-linear relationships in agricultural data. RF, KNN, and SVR served as comparative models, each offering different strengths—such as RF's handling of feature interactions, KNN's local proximity modeling, and SVR's kernel-based decision boundaries. Among the four, the ANN model achieved the best performance with an R^2 score of 98.11%, indicating it could explain nearly all the variance in rice yield. Despite a slightly higher MAE (0.0966) compared to RF, its overall prediction accuracy was superior. The ANN's MSE was also low at 0.0189, showing minimal error between predicted and actual values. These results demonstrate that ANN effectively learns the complex interactions between input features like weather conditions, soil characteristics, and irrigation practices. Its strong performance reaffirms its suitability for non-linear forecasting tasks in agriculture, particularly in rice yield prediction. The internal structure of the proposed Artificial Neural Network, including the number of hidden layers and activation functions used, is visually depicted in Figure 2.

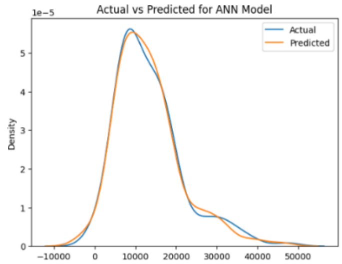

Figure 4. Actual Vs Predicted for ANN Model

The plot demonstrates that the ANN model produces predictions that are nearly identical to the actual observed values. This suggests a strong relationship between the predicted and actual outcomes and highlights the effectiveness of ANN in capturing the underlying patterns in rice yield prediction. Also, the Learning Curve for ANN in Figure 3 shows that as the model gets more data, its performance stabilizes and remains consistent across both training and cross-validation sets, suggesting a well-trained model with good generalization. This matches the Actual vs Predicted plot, where the model's predicted values are very close to the actual values, confirming the model's robustness in making accurate predictions.

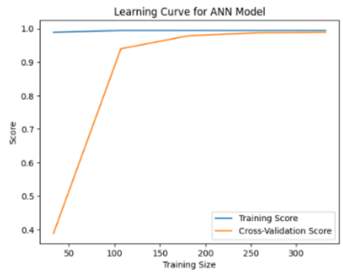

Figure 5. Cross Validation and Training ANN

For a more comprehensive evaluation, we also compared the ANN model with other machine learning models commonly used in regression tasks. The results of these models are summarized below:

Table 5. results of these models

Model	R2	MAE	MSE
Artificial Neural Network	**98.107207**	**0.096600**	**0.018928**
Random Forest	97.834867	0.1611	0.181447
K-Nearest Neighbors	78.730988	0.2655	0.286234
Support Vector Regression	76.990749	0.2910	0,455139

The ANN model's ability to capture complex, non-linear relationships in the data makes it particularly suitable for agricultural predictions, where the interactions between variables like soil moisture, temperature, and humidity can be highly intricate. ANNs can model these interactions better than traditional linear models or tree-based methods, making them highly effective for forecasting rice yields in varying climatic and environmental conditions. This results is clearly reflected in the performance metrics shown in Figure 4, where the ANN outperforms other models across all evaluation criteria (R^2, MAE, and MSE).

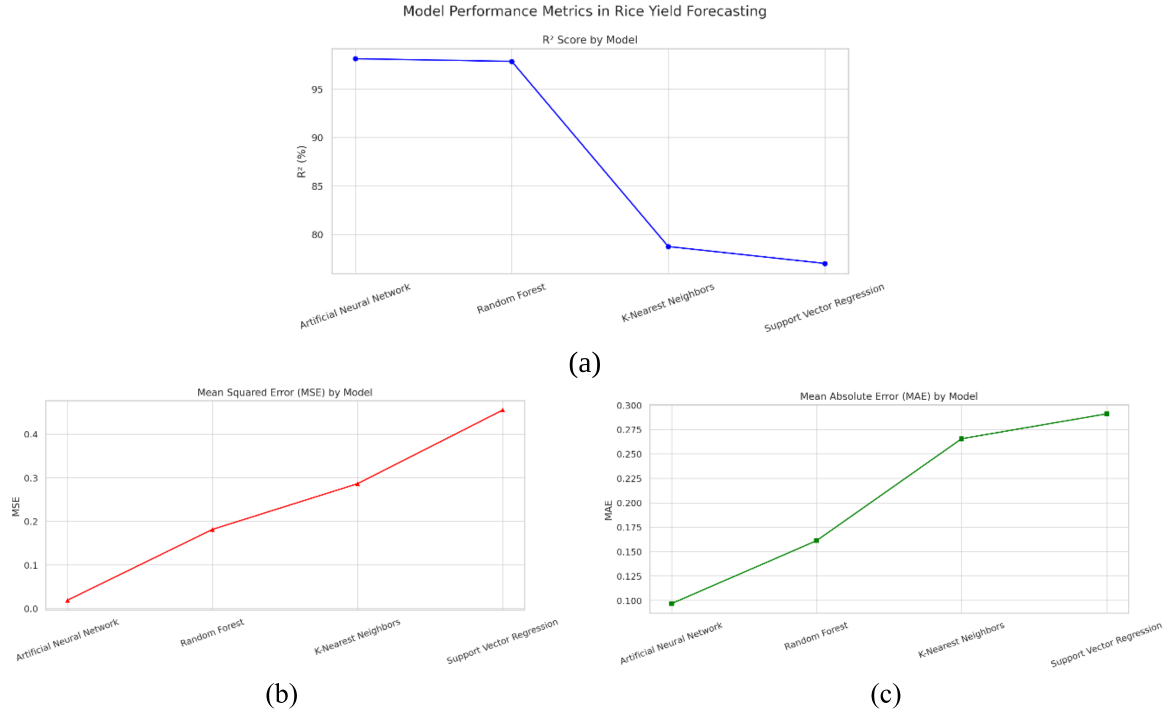

Figure 6. Performance comparison of rice yield forecasting models:
(a) R^2 scores,(b) Mean Absolute Error,(c) Mean Squared Error

Given the non-linear nature of agricultural data, ANNs are well-suited for modeling rice yield predictions. They are capable of handling the complex relationships that exist between input features (weather data, soil type, irrigation, etc.) and the output (rice yield). The ability of ANNs to model such complex interactions makes them a preferred choice in this context.In comparison, Random Forest and K-Neighbors were chosen for their flexibility in capturing feature interactions and local relationships, respectively. However, their performance was not as strong as ANN, indicating that ANNs provide a more accurate and robust approach for rice yield prediction.

The high accuracy of ANN enables predictive tools that support farmers' decisions (planting, irrigation, harvesting) and government strategies (logistics, subsidies, resource allocation). For sustainable agriculture, such models aid in reducing inputs, improving land use, and supporting food security.

However, limitations exist, the model used fixed parameters (no hyperparameter tuning), relies on region-specific data (Malang), and lacks interpretability. To enhance generalizability and transparency, future research should explore explainable AI (XAI), larger multi-regional datasets, and sensor-based inputs. In conclusion, the ANN model shows strong promise for real-world agricultural applications, particularly in supporting climate-resilient farming and food system planning at regional and national levels.

4. Conclusion

This study confirms the effectiveness of Artificial Neural Networks (ANN), particularly the Multilayer Perceptron (MLP) architecture, in predicting rice yields based on agro-environmental variables. By integrating weather and agricultural data specifically rainfall, temperature, humidity, harvest area, and harvest quantity the ANN model successfully captured complex nonlinear relationships that influence rice production. The model achieved high prediction accuracy, with an R^2 score exceeding 98%, outperforming other machine learning models such as Random Forest, K-Nearest Neighbors, and Support Vector Regression. Beyond model accuracy, the findings highlight the potential of ANN to support data-driven decision-making in agriculture. The predictive insights from the model can assist policymakers, agricultural agencies, and farmers in planning adaptive strategies for planting schedules, irrigation management, and resource allocation particularly under climate uncertainty. However, this study also acknowledges certain limitations. The fixed ANN architecture, while interpretable and efficient, may not generalize optimally across different crops or regions without further tuning. Moreover, the model relies heavily on historical environmental data; future work should consider integrating remote sensing data and dynamic weather forecasts to enhance predictive robustness. Future research should also explore hybrid models that combine ANN with domain-specific knowledge or optimization frameworks to further improve interpretability and scalability in diverse agricultural settings. Overall, the study contributes to the advancement of sustainable precision agriculture by offering a practical, scalable AI-based tool for yield forecasting and adaptive planning, supporting national efforts toward food security and climate resilience.

References

[1] S. Park, “World Bank,” in *Essential Concepts of Global Environmental Governance*, Second edition. | Abingdon, Oxon; New York: Routledge, 2021.: Routledge, 2020, pp. 289–290. doi: 10.4324/9780367816681-118.

[2] A. Dhamira and I. Irham, “THE IMPACT OF CLIMATIC FACTORS TOWARDS RICE PRODUCTION IN INDONESIA,” *Agro Ekonomi*, vol. 31, no. 1, Sep. 2020, doi: 10.22146/ae.55153.

[3] K. E. Ukhurebor *et al.*, “Precision agriculture: Weather forecasting for future farming,” in *AI, Edge and IoT-based Smart Agriculture*, Elsevier, 2022, pp. 101–121. doi: 10.1016/B978-0-12-823694-9.00008-6.

[4] A. A. Mulema, L. Cramer, and S. Huyer, “Stakeholder engagement in gender and climate change policy processes: Lessons from the climate change, agriculture and food security research program,” *Front Sustain Food Syst*, vol. 6, Nov. 2022, doi: 10.3389/fsufs.2022.862654.

[5] S. Saud *et al.*, “Comprehensive Impacts of Climate Change on Rice Production and Adaptive Strategies in China,” *Front Microbiol*, vol. 13, Jun. 2022, doi: 10.3389/fmicb.2022.926059.

[6] V. Sharma, A. K. Tripathi, and H. Mittal, “Technological revolutions in smart farming: Current trends, challenges & future directions,” *Comput Electron Agric*, vol. 201, p. 107217, Oct. 2022, doi: 10.1016/j.compag.2022.107217.

[7] P. Hara, M. Piekutowska, and G. Niedbała, “Selection of Independent Variables for Crop Yield Prediction Using Artificial Neural Network Models with Remote Sensing Data,” *Land (Basel)*, vol. 10, no. 6, p. 609, Jun. 2021, doi: 10.3390/land10060609.

[8] D. Elavarasan and P. M. D. Vincent, “Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications,” *IEEE Access*, vol. 8, pp. 86886–86901, 2020, doi: 10.1109/ACCESS.2020.2992480.

[9] M. G. M. Abdolrasol *et al.*, “Artificial Neural Networks Based Optimization Techniques: A Review,” *Electronics (Basel)*, vol. 10, no. 21, p. 2689, Nov. 2021, doi: 10.3390/electronics10212689.

[10] U. Bhimavarapu, G. Battineni, and N. Chintalapudi, “Improved Optimization Algorithm in LSTM to Predict Crop Yield,” *Computers*, vol. 12, no. 1, p. 10, Jan. 2023, doi: 10.3390/computers12010010.

[11] H. Crisóstomo de Castro Filho *et al.*, “Rice Crop Detection Using LSTM, Bi-LSTM, and Machine Learning Models from Sentinel-1 Time Series,” *Remote Sens (Basel)*, vol. 12, no. 16, p. 2655, Aug. 2020, doi: 10.3390/rs12162655.

[12] D. Elavarasan and P. M. D. Vincent, “Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications,” *IEEE Access*, vol. 8, pp. 86886–86901, 2020, doi: 10.1109/ACCESS.2020.2992480.

[13] J. Patel, B. Vala, and M. Saiyad, "LSTM-RNN Combined Approach for Crop Yield Prediction On Climatic Constraints," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, IEEE, Apr. 2021, pp. 1477–1483. doi: 10.1109/ICCMC51019.2021.9418231.

[14] A. M. Alqudah and Z. Moussavi, "A Review of Deep Learning for Biomedical Signals: Current Applications, Advancements, Future Prospects, Interpretation, and Challenges," *Computers, Materials & Continua*, vol. 83, no. 3, pp. 3753–3841, 2025, doi: 10.32604/cmc.2025.063643.

[15] I. Malashin, V. Tynchenko, A. Gantimurov, V. Nelyub, and A. Borodulin, "Applications of Long Short-Term Memory (LSTM) Networks in Polymeric Sciences: A Review," *Polymers (Basel)*, vol. 16, no. 18, p. 2607, Sep. 2024, doi: 10.3390/polym16182607.

[16] A. Satpathi *et al.*, "Comparative Analysis of Statistical and Machine Learning Techniques for Rice Yield Forecasting for Chhattisgarh, India," *Sustainability*, vol. 15, no. 3, p. 2786, Feb. 2023, doi: 10.3390/su15032786.

[17] P. Dhawas, A. Dhore, D. Bhagat, R. D. Pawar, A. Kukade, and K. Kalbande, "Big Data Preprocessing, Techniques, Integration, Transformation, Normalisation, Cleaning, Discretization, and Binning," 2024, pp. 159–182. doi: 10.4018/979-8-3693-0413-6.ch006.

[18] J. Qi, J. Du, S. M. Siniscalchi, X. Ma, and C.-H. Lee, "On Mean Absolute Error for Deep Neural Network Based Vector-to-Vector Regression," *IEEE Signal Process Lett*, vol. 27, pp. 1485–1489, 2020, doi: 10.1109/LSP.2020.3016837.

[19] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput Sci*, vol. 7, p. e623, Jul. 2021, doi: 10.7717/peerj-cs.623.

[20] D. S. K. Karunasingha, "Root mean square error or mean absolute error? Use their ratio as well," *Inf Sci (N Y)*, vol. 585, pp. 609–629, Mar. 2022, doi: 10.1016/j.ins.2021.11.036.