Machine Learning Approach Leveraging Crop Datasets for Optimizing Crop Yield Production

[1]Bulus Bali, [2]Zakawa N. Ngida, [1]Isacha Habila
[1]Department of Computer, Adamawa State University, Mubi Adamawa State, Nigeria.
[2]Department of Botany, Adamawa State University, Mubi Adamawa State, Nigeria.
bali930@asdu.ng
DOI: 10.56201/ijcsmt.v10.no6.2024.pg119.132

Abstract
Crop yield prediction based on environmental, soil, water, and weather parameters has become a vital area of research, addressing the growing need for sustainable agricultural practices and food security. This study adopts a machine learning approach to optimize crop yield production within the agricultural landscape of Michika Local Government Area, Adamawa State. Applying extensive crop datasets, machine learning techniques are utilized to analyze, interpret, and uncover critical factors and patterns influencing crop yields. The main objective is to develop a robust predictive model that empowers farmers and stakeholders with actionable insights, to enhance agricultural productivity. The results demonstrate a substantial improvement in yield predicting accuracy through machine learning-based methods compared to traditional approaches. ANN with the lowest RMSE (3136.8), the lowest MAE (2502.2), and higher R^2 (0.073255), indicates the most accurate predictions. These findings underscore the transformative potential of artificial intelligence in advancing precision agriculture, enabling resource-efficient farming, and bolstering food security. This study also highlights avenues for future research, including optimizing resource allocation strategies, identifying resilient crop varieties, predicting and mitigating crop diseases, and mapping soil suitability for diverse crops. Such efforts would further drive the adoption of smart agricultural systems, enhancing productivity and sustainability while supporting the transition to climate-resilient farming practices.

Keywords*: Crop datasets; Machine Learning; Predictive modeling; Crop yield; Weather parameters*

1. Introduction

Modern agriculture faces significant challenges, including meeting the rising demand for food production, addressing resource constraints, and ensuring environmental sustainability. As the population of Michika expands, the agricultural sector encounters multifaceted challenges such as climate change, soil degradation, and suboptimal resource utilization. Traditionally, farmers have primarily utilized conventional farming methods, which, unfortunately, lack precision, resulting in reduced productivity and prolonged time consumption. Precision farming, leveraging advanced technologies, enhances productivity by precisely identifying the essential practices for each season (Durai, & Shamili, 2022). Integrating machine learning techniques presents promising avenues for tackling these challenges by optimizing agricultural practices and elevating productivity. AI-driven technologies provide innovative solutions to refine agricultural processes, enhance yield predictions, optimize resource management, and mitigate the adverse effects of climate change on crop production. This study addresses the need to utilize machine learning algorithms in the agricultural domain, specifically focusing on their empirical effectiveness in optimizing crop yield production.

The agricultural sector in Michika, like other parts of Nigeria, faces challenges such as unpredictable weather conditions, inadequate knowledge of soil nutrient levels, pest and disease management, and inefficient resource use. These challenges contribute to low productivity and food insecurity. Conventional farming methods fail to address the complexities of modern agricultural systems, leaving farmers vulnerable to losses and inefficiencies. Accurate crop yield prediction remains a significant hurdle, as it requires integrating multiple factors, such as climate conditions, soil characteristics, and farming practices (Prabavathi & Chelliah, 2022). Addressing this gap demands advanced tools and methodologies that can provide actionable insights and empower farmers to make informed decisions (Mgendi, 2024).

This study aims to leverage machine learning techniques to optimize crop yield in Michika by developing predictive models that integrate environmental, soil, and management factors. Specifically, the objectives are to employ machine learning algorithms on historical data for crop yield prediction, compare various machine learning techniques for forecasting crop yields, and assess the effectiveness of these techniques in predicting yields to recommend the most suitable approach tailored to Michika's conditions.

There is a pressing need to enhance agricultural productivity in Michika Local Government Area Adamawa State, like other parts of Nigeria. Agriculture plays a crucial role in the economy of the region, and optimizing crop yield production can contribute to food security, poverty reduction, and overall economic development (Jiya et al., 2023). Leveraging machine learning techniques will harness the power of algorithms to analyze large agricultural datasets and uncover patterns that can significantly enhance crop yield production. Furthermore, the utilization of crop datasets that contain information related to climate conditions, soil quality, farming techniques, and crop performance, among others offers an opportunity to gather valuable insights into agricultural practices specific to the area. Leveraging machine learning algorithms for optimizing crop yield production based on historical data in the area holds great potential to address agricultural challenges, enhance food security, and contribute to the socio-economic development of the region

2. Related works

Crop yield prediction is a pivotal factor in agricultural success and efficiency, yet it poses significant challenges to farmers. Knowledge gaps regarding surplus harvests, unpredictable weather, seasonal rainfall, soil nutrients, fertilizer accessibility, pest management, and post-harvest losses contribute to declining crop production (Abdulraheem et al., 2022). Addressing these challenges requires a holistic approach that integrates advanced technologies and disseminates knowledge to empower farmers (Mgendi, 2024). In the quest to optimize crop yield, farmers require timely and insightful guidance for predicting productivity, conducting analyses, and unlocking the full potential of their crops. The real challenge lies in the accuracy of yield prediction, where tapping into a farmer's past experiences with a specific crop becomes a valuable resource for making informed predictions in subsequent crop cycles (Lata & Chaudhari, 2019).

Crop yield prediction ventures into the realm of advanced algorithms, forecasting crop production by harnessing a diverse array of data points. These include temperature, rainfall, pH levels, pesticide and fertilizer application, and various meteorological variables. The intricacies of crop yield forecasting hinge on the application of predefined criteria, creating a pathway to anticipate and enhance the productivity of each crop cycle (Oluwole et al., 2022). Conventionally, farmers have predominantly employed conventional farming methods. However, these methods lack precision, leading to decreased productivity and extensive time consumption. Precision farming, on the other hand, enhances productivity by identifying the necessary practices for each season. Elements of precision farming encompass forecasting weather conditions, soil analysis, recommending suitable crops for cultivation, and determining precise quantities of fertilizers and pesticides required (Abdulraheem et al., 2022; Durai & Shamili, 2022). Several researchers have leveraged machine learning methodologies in many parts of the world, including regression trees, random forests, multivariate regression, association rule mining, and artificial neural networks, to predict crop yields. Utilizing machine-learning (ML) models, crop-yield is taken as an implicit function of input variables, encompassing factors such as weather components, and soil conditions, thereby presenting a potentially complex and nonlinear function. According to (Durai & Shamili, 2022; Mgendi, 2024) in the context of Precision Farming (PF), innovative technologies such as ML, Data Mining, Data Analytics, and Internet of a Thing (IoT) converge to usher in a new era of data-driven agricultural efficiency. These innovative tools are not just data collectors; they are orchestrators of insights, contributing to a reduction in manual labor while concurrently amplifying productivity. At the heart of this technological symphony, ML techniques take center stage, autonomously detecting, identifying, and predicting outcomes by extracting knowledge, and relationships from the vast datasets under scrutiny (Domingues et al., 2022). In the agricultural landscape of Nigeria, as delineated by (Jiya et al., 2023), crop cultivation emerges as the predominant pursuit, carrying significant implications for families and the broader economy.

The accurate prediction of crop yields stands as a critical challenge in the agricultural domain, necessitating advanced AI-driven solutions (Mgendi, 2024). Farmers confront a plethora of uncertainties, including limited knowledge about potential harvest excess, unpredictable weather-conditions, and dynamic seasonal rainfall policies. The intricacies extend to the depletion of soil nutrition levels, the volatile availability and cost of fertilizers, challenges in pest control, and the ever-present specter of post-harvest losses (Oluwole et al., 2022). These multifaceted factors collectively contribute to a decline in crop production (Durai & Shamili, 2022). The integration of AI technologies can provide a sophisticated framework for addressing these challenges comprehensively. AI algorithms can analyze vast datasets, assimilating information on weather patterns, soil health, fertilizer dynamics, and pest management. This data-driven approach enables precise predictions and recommendations, empowering farmers with actionable insights to optimize crop yield, resource allocation, and risk mitigation. Exploiting AI holds the promise of enhancing decision-making processes, promoting sustainable practices, and fostering resilience in the face of diverse challenges (Durai & Shamili, 2022; Mgendi, 2024).

Traditionally, farmers have adhered to conventional farming practices, which lack precision and result in reduced productivity and time consumption, particularly in the face of unpredictable climate conditions (Zhou & Ismaeel, 2021; Prabavathi & Chelliah, 2022). Exploiting the use of variable farming resources like soil, fertilizer, and weather data, including crop information, is of utmost importance. Achieving high yields hinges on the efficient deployment of these resources (Kwaghtyo & Eke, 2023). In AI-driven agriculture, forecasting weather conditions, scrutinizing soil properties, suggesting optimal crops for cultivation, and determining precise quantities of fertilizers and pesticides constitute key facets of precision farming (Zhou & Ismaeel, 2021). Exploiting cutting-edge technologies such as IoT, data mining, data analytics, and machine-learning, precise farming systematically gathers data, trains sophisticated systems, and forecasts outcomes (Mgendi, 2024). This technological integration not only minimizes the need for manual labor but also enhances overall productivity in the agricultural landscape (Durai & Shamili, 2022; Mgendi, 2024). Predicting crop yields stands out as a formidable challenge in the domain of smart agriculture, prompting the proposal and validation of numerous models (Kuradusenge et al., 2023; Mgendi, 2024). Given that crop production is influenced by various factors like weather, climate, seed quality, fertilization, and soil composition, addressing this challenge requires the integration of diverse datasets (Oluwole et al., 2022). Studies by Bali & Garba (2021), Bali et al. (2021) and Bali (2024) demonstrated that neural network-based methods can effectively learn from environmental data, self-organize their structures, and deliver strong prediction performance.

Awe and Dias (2022) developed a deep-learning model, specifically a generative Artificial Neural Networks (ANN), to effectively capture and estimate the complex and non-parametric features of agricultural output. The study was conducted in Nigeria with data covering forty years (1980-2019). They compared the results with the autoregressive integrated moving average model (ARIMA). The empirical findings demonstrate the superiority of the hybrid ANN model over the traditional Box-Jenkins ARIMA methodology for forecasting non-stationary time series. The proposed ANN model significantly improves forecasting accuracy in this context. A fuzzy inference system (FIS) was developed by (Shuaibu et al., 2021) to forecast rice yield in Jigawa state, Nigeria, for the period 2021-2030. This forecasting model incorporated variables such as "rainfall, land availability, and historical rice production data". The findings of this study indicate that in the year 2022, a total land area of 177,950 hectares will be demanded, along with a rainfall of 873 millimeters, to achieve a rice production of 1,070,000 metric tons. Utilizing an ANN multilayer perceptron (MLP), Alves et al. (2018) predicted soybean productivity by considering growth habits, sowing density, and agronomic characteristics. The outcomes reveal a commendable 98% success rate on the training dataset and a noteworthy 72% accuracy on the validation dataset, demonstrating the network's capability to estimate soybean productivity based on agronomic features, growth habits, and population density. Recognizing the profound impact of crop production setbacks, this research pivots toward a comparative exploration of diverse techniques for predicting crop yields.

3. Methodology

3.1.Study Area

This exploration is slated to unfold within the Michika Local Government Area of Adamawa State, strategically situated in the far northeast, marked by geographic coordinates covering latitude 11°8' South and longitude 15°13' E. This locality shares a northern boundary with Madagali LGA and extends further to form an international boundary with the Republic of Cameroon to the northeast. Borno State bounds it to the West, while Mubi and Hong LGAs define its southern boundary. The Michika LGA, a mosaic of diversity, encompasses four distinctive development areas: Michika Metropolitan, Garta, Bazza, and Madzi, sprawled across a land mass measuring 961 km2. The dynamic shade of this region is woven together by an estimated population of 211,124 occupants. Embarking on a temporal line, this study is projected for a duration of seven months, from May to November during the Wet Season. This strategic time frame aims to encapsulate and scrutinize the nuances of environmental dynamics and agricultural practices, offering a comprehensive insight into the intricacies of the Michika Local Government Area.

3.2.Crop yield parameters

In this area, groundnut, maize, and rice are the predominant crops. The essential weather factors impacting crop yield include precipitation, air temperature, air moisture, solar radiation, and wind speed. An AI-driven analysis, guided by, was employed to examine climate and weather parameters. Additionally, historical crop yield data linked to soil and weather conditions, encompassing temperature, humidity, rainfall, and soil pH, were employed for the nominated crops in the region. The equipment used for data collection is shown in Table 1.

Table 1 Showing equipment used for data collection

S/N	Equipment	Use
1	Soil pH Tester (Excellent Corrosion Proof Soil Fertility Nutrient Meter)	Measures soil pH
2	Digital thermometer	Measures temperature
3	Hygrometer (wet and dry bulb)	Measures humidity
4	Rain gauge (Programmable Water Timer)	Measures rainfall

3.3.Selected Crops

Groundnut *(Arachis hypogaea)*, Maize *(Zea mays)*, and Rice *(Oryza sativa)* datasets were used. The accession of crop yield data involved collating information from the agricultural department and various farmer cooperatives within the study area during the Wet season. The major factors of the dataset include District, farm size (ha), Fertilizers (Nitrogen, Phosphorus, Potash) (Kg), temperature (^{o}C), rainfall (mm), humidity (RH), yield (kg/ha). May to November was regarded as the Wet Season. The crop yield in kg/ha was taken as the target variable. The visualization of the crop yield dataset is shown in Figure 1.

District	Crop	Temperature	Rainfall	pH	Humidity	FarmSize	N	P	K	yield_value
Categorical	Categorical	Number	Number	Number	Number	Number	Number	Number	Number	Number
District	Crop	Temperature	Rainfall	pH	Humidity	Farm size	N	P	K	yield_value
Central	G/nut	28	778.04	5.8	65.3	4	0	0	0	3000
Central	G/nut	28.5	696.12	6.3	74.6	2	0	27	0	780
Central	G/nut	28	730.32	5.5	78.5	2	0	54	0	896
Central	G/nut	34.5	908.02	6.2	46.74	3	0	81	0	992
Central	G/nut	36	908.02	5.2	46.7	0.5	0	27	0	431
Central	G/nut	35.5	908.03	6	46.72	1	0	0	0	486
Central	G/nut	30	908.02	7	46.71	2	0	54	0	1120
Central	G/nut	31.5	908.02	5.6	46.7	4	0	108	0	5687
Central	G/nut	31	908.01	6	46.74	6	0	450	0	5830
Central	G/nut	31	908.02	6.5	46.76	7	0	0	0	6982
Central	G/nut	28	945.22	6	79.57	3	0	0	0	3980
Central	G/nut	27	815.23	5	78.8	0.7	0	0	0	458
Central	G/nut	27	809.03	5.4	78.5	1.8	0	0	0	703
Central	G/nut	28.5	780.24	6.5	78.5	2.4	0	0	0	1323
Central	G/nut	27	730.09	6	78.5	1	0	0	0	501
Central	G/nut	26.5	730.1	6.5	73.41	2.5	0	0	0	1502
Central	G/nut	24	730.16	5.1	73.41	3	0	0	0	4002
Central	G/nut	26.5	730.07	5.5	73.41	0.7	0	0	0	461
Central	G/nut	27	730.2	5.4	50.13	2	0	54	0	1201

Figure 1 presents the visualization of the crop yield dataset.

3.4.DATA PRE-PROCESSING

The data for this exploration passed a series of preprocessing steps aimed at optimizing the accuracy of crop yield production predictions. MATLAB (2021a) served as the primary development platform for creating and assessing the proposed model. Initially, a rigorous data cleaning process was executed, involving transformative activities to exclude any duplicate or irrelevant data points within the dataset. Following this, feature selection was performed to assess the correlation between features and the target variable, crop yield. This analysis aids in detecting and retaining the most pertinent features. Subsequently, feature scaling was employed to standardize numerical features, employing techniques such as min-max scaling or standardization. This normalization ensures that all numerical features are on a comparable scale. Concurrently, categorical variables were converted into numerical representations to enhance compatibility with machine learning algorithms. In the final stages of preprocessing, the dataset was partitioned into distinct sets for training, validation, and testing. The training set was employed for model training, the validation set contributed to hyperparameter tuning, and the testing set served as the benchmark for estimating the model's performance.

This research explores four categories of machine learning algorithms: ANN, decision trees (DT), Random Forest (RF), and support vector machines (SVM). The motivation for exploring machine learning algorithms such as ANN, DT, RF, and SVM stems from their ability to handle complex, nonlinear connections within data and give accurate predictions for diverse agricultural applications. These algorithms are well-suited to optimize crop yield predictions due to their inflexibility, scalability, and rigidity in modeling intricate patterns and relations between various environmental, soil, and management factors. The activity diagram for the model is shown in Figure 2.

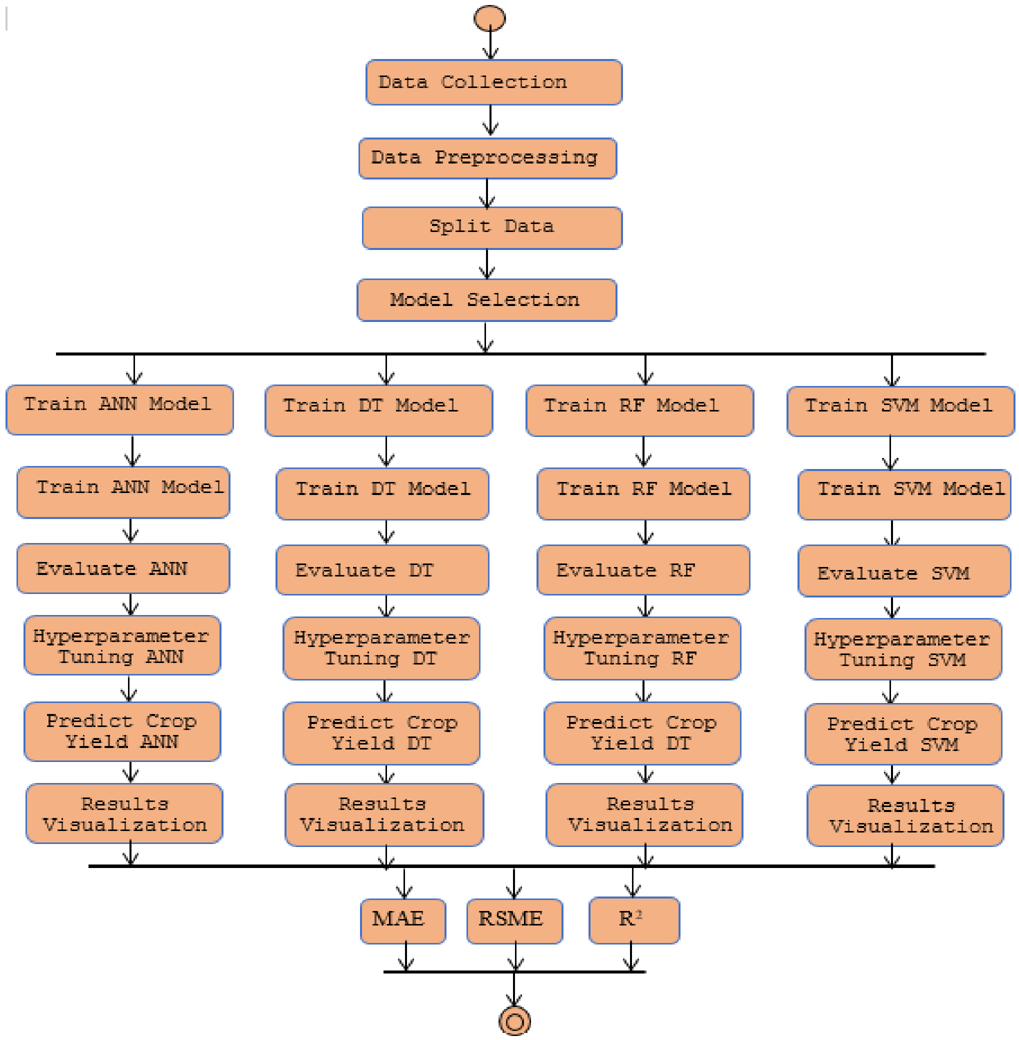

Figure 2: The activity diagram depicting the flow of tasks and decisions in the crop yield prediction process.

i) *Data Collection*: Relevant historical data on crop yields, climate, soil conditions, and other relevant factors were collected.

ii) *Data Preprocessing*: The data was prepared and standardized, handling missing values, scaling, and encoding categorical variables. The data from inputs and target datasets were standardized between zero and one using (1) to avoid large numeric ranges from the values of the predictor variables.

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

where x is the respective variable, $xmin$ is the minimum value, $xmax$ is the maximum, and $xnormal$ is the standardized value.

iii) *Split Data*: The data was divided into training and testing datasets.
iv) *Model Selection*: Each of the machine learning models ANN, DT, RF, and SVM was designated for training.
v) *Model Training*: Each model was trained on the training dataset.
vi) *Model Evaluation*: The performance of each model using appropriate metrics was estimated. The statistical metrics employed to estimate the performance of artificial intelligence-driven predictive models are RMSE (Root Mean Squared Error), R-squared, and Mean Absolute Error (MAE). The motivation for using RMSE, R-squared, and MAE to evaluate models in crop yield prediction is to give complementary insights into accuracy and reliability. RMSE measures the average squared error, lower values indicate better model performance, correct large errors, and assess model accuracy and sensitivity to variability. MAE measures the average absolute error, lower values indicate better performance, treat all deviations equally, and balance average accuracy while being robust to outliers. R^2 indicates the proportion of variance explained by the model; values closer to 1 indicate better performance, while negative values suggest the model performs worse than a simple mean predictor. The training errors were tracked using the RMSE function, which is specified as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=n}^{n}(y_{targeti} - y_{net}}\ )^2$$

(2)

R^2, or the relationship between the modeled data value and the observed data value is examined via the coefficient of determination given by:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(x - x_i)^2}{\sum_{i=1}^{n}(x - \bar{x})^2} \quad (3)$$

Where
x_i = the observed value
x = the modeled value
n = the number of errors,
Σ = summation symbol,
$(x_i - x)^2$ = the square of absolute errors.
R^2 = the correlation coefficient
$\bar{x}$ = the average value

The Mean Absolute Error (MAE) represents the average of the absolute differences between the actual target values and the predicted values. It is mathematically defined as shown in (4).

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \bar{y}_i\right|$$

(4)

vii) *Hyperparameter Tuning*: The model hyperparameters were optimized for improved model performance.
viii) *Prediction:* The trained models were used to predict crop yield for new data.

ix) *Results Visualization*: The output was visualized and the predictions were compared with actual values using charts or plots.

This entire process was implemented in MATLAB using built-in functions; fitcensemble, fitctree, svm, and trainNetwork functions for each machine learning model, along with built-in evaluation functions; crossval, cvLoss, and predict.

4. Interpretation of Results and Discussion

The results from the four machine learning models; ANN, DT, RF, and SVM as shown in Table 2 provide valuable insights into the performance of the models in predicting crop yield. The performance of each model was discussed based on the three metrics: RMSE, MAE, and R^2.

Table 2 present a clear comparison of how each model performs evaluated using RMSE, MAE, and R^2.

Model	RMSE	MAE	R2
DT	5992.0	3841.8	-2.3817
SVM	3315.6	2593.9	-0.035425
RF	3430.6	2733.0	-0.10846
ANN	3136.8	2502.2	0.073255

DT: The high RMSE (5992) value for the DT model indicates that its predictions diverge significantly from the actual yield values on average. This suggests that the DT could not capture the complexity of the relationship between the predictors and the target variable well. MAE: Similarly, the MAE (3841.8) is relatively high, meaning that the average absolute error between the predicted and actual values is large. This further supports the conclusion that the model is struggling to predict crop yields accurately. The negative R^2 (-2.3817) value indicates that the DT model is performing worse than a simple mean-based model (i.e., always predicting the average yield). This is a clear sign of poor model performance and suggests that the DT is overfitting or capturing very little of the variance in the target variable. This shows that the DT model is underperforming, likely due to overfitting or the inability to capture the complex patterns in the data. The negative R^2 and high RMSE/MAE suggest that this model is not suitable for crop yield prediction in this case.

RF: with RMSE (3430.6) is slightly worse than SVM, indicating higher error. It is the third best, slightly worse than SVM with MAE (2733). R^2 (-0.10846) is negative, meaning the model also performs worse than the baseline mean predictor. This indicates that RF struggles to generalize well and has suboptimal performance compared to SVM and ANN.

SVM: This is the second best after ANN with RMSE (3315.6), indicating decent predictive accuracy. Second best after ANN with MAE (2593.9). The R^2 (0.035425) is close to zero, showing the model explains very little of the variance and marginally outperforms the baseline. This shows that SVM performs moderately well but has limited predictive power.

ANN: the lowest RMSE (3136.8), indicates the most accurate predictions. The lowest MAE (2502.2), is suggesting the smallest average prediction errors. R^2 (0.073255) is slightly positive, indicating the model explains a small amount of variance and outperforms the baseline predictor. This indicates that ANN outperforms the other models in all metrics, making it the best-suited model for this dataset and problem.

5. The Comparison of Models

The bar chart in Figure 3 illustrates the comparison of RMSE values for four different machine-learning models: DT, SVM, RF, and ANN.

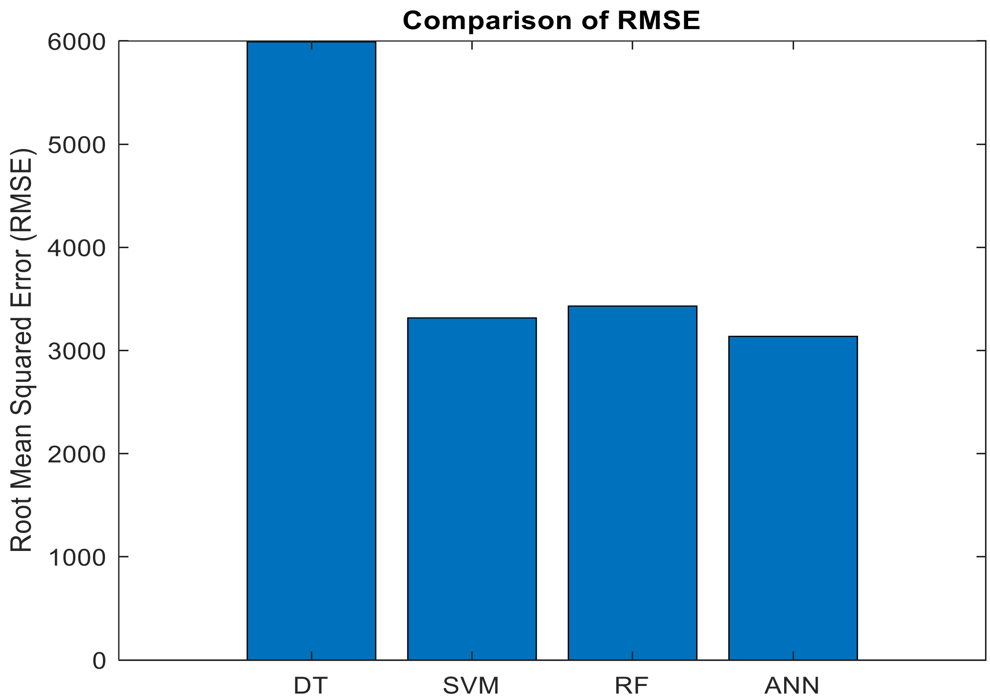

Figure 3 visualizes a comparison of the RSME of the models

DT is the worst-performing model, with high RMSE and MAE values, and a very negative R^2, indicating that the model is poorly fitted to the data and might be overfitting or not capturing the relevant features of the dataset.

The bar chart in Figure 4 presents a comparison of the MAE values for four different machine-learning models: DT, SVM, RF, and ANN.

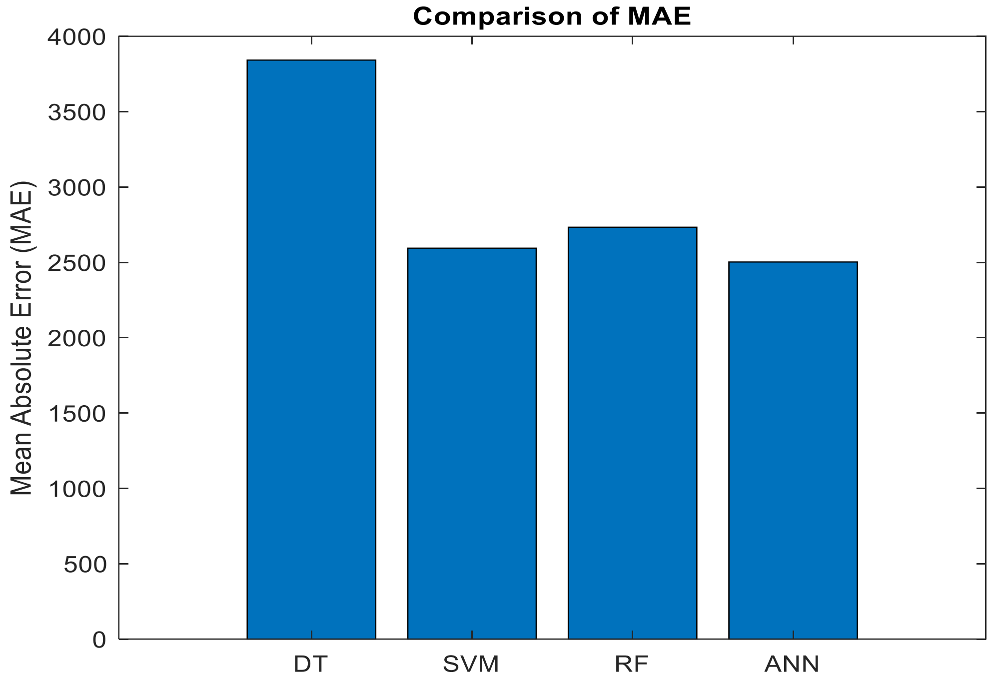

Figure 4 visualizes a comparison of the MAE of the models

SVM and RF: Both perform similarly, with SVM slightly outperforming RF in terms of RMSE and MAE. Both models show negative R^2 values, indicating they are not suitable for explaining the variance in crop yields. They are more accurate than DT, but not as effective as ANN. The bar chart in Figure 5 compares the MAE values across four machine-learning models: DT, SVM, RF, and ANN.

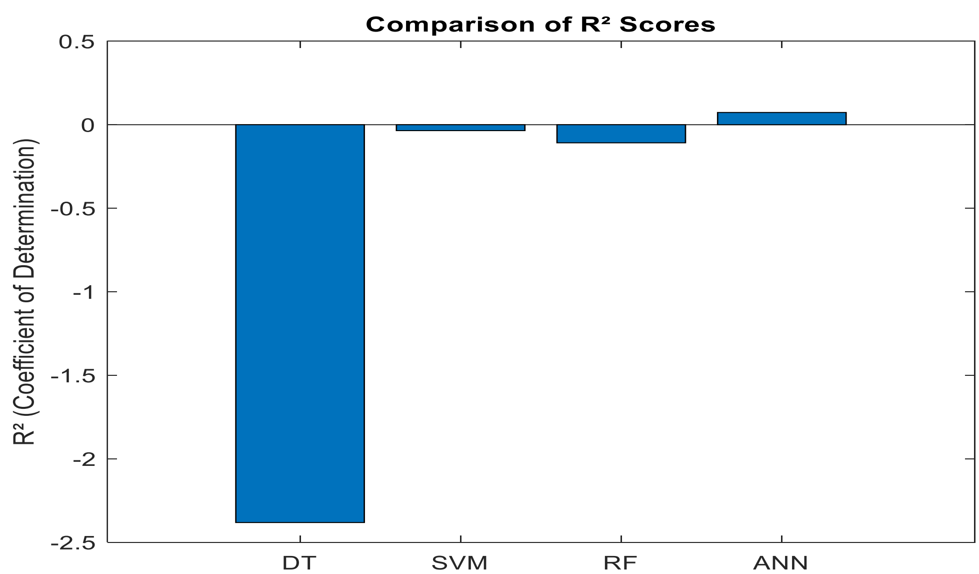

Figure 5 visualizes a comparison of the R^2 of the models

Best Performing Model: ANN consistently outperforms the other models in all metrics, followed by SVM. RF and DT perform poorly, with DT being the worst. Among the models, ANN shows the best performance in terms of RMSE, MAE, and R^2. The generally low or negative R^2 values indicate that all models, except ANN, struggle to capture the underlying patterns in the data. This suggests that the available features (Temperature, Rainfall, Humidity, and District) might not fully capture the complexity of the factors influencing crop yield, or that further feature engineering, data preprocessing, or model tuning may be required for better predictive accuracy. However, ANN's positive R^2 suggests it is slightly capable of explaining the variance, and the most reliable model for predicting crop yield.

6. Conclusion

This research evaluates various yield prediction techniques, analyzing their effectiveness to recommend strategies based on weather parameters and nutrient requirements. The study developed a predictive model capable of accurately forecasting crop yields using features deduced from comprehensive datasets. This model offers farmers in the study area actionable insights into optimal crop varieties, fertilizer application, and prevailing environmental conditions, thereby enhancing decision-making and reducing potential losses. The findings emphasize the transformative potential of machine learning (DT, RF, and SVM) in optimizing agricultural productivity. While the ANN model demonstrated superior performance by minimizing prediction errors, the overall predictive accuracy was constrained, as indicated by low R^2 values. Addressing these limitations presents an important opportunity for future research, which could incorporate real-time data from IoT sensors, scale the model to different crops and regions, and develop more robust, user-friendly tools to promote widespread adoption. Additionally, this study highlights the importance of optimizing hyperparameters, such as learning rate, number of neurons, and activation functions, to improve predictive accuracy. Exploring hybrid modeling combining ANN with other techniques is also recommended, as they have the potential to leverage complementary strengths of various approaches, thereby enhancing overall model performance.

Acknowledgments

We sincerely appreciate Adamawa State University, Mubi, for providing financial support for this research through the TETFUND grant. Additionally, we are grateful to the authors of the valuable resources referenced, the domain experts who contributed their insights, and the agricultural department alongside the farmer cooperatives within the study area for their collaboration and support.

Conflict of Interest

The authors affirm that there are no conflicts of interest associated with this study.

Funding

This research was supported by funding from the Tertiary Education Trust Fund (TETFund) under the Institution-Based Research (IBR) program, with reference number TEFT/DR&D/UNI/MUBI/RG/2024/VOL.1, facilitated through Adamawa State University, Mubi.

References

Abdulraheem, M., Awotunde, J. B., Abidemi, E. A., Idowu, D. O., & Adekola, S. O. (2022). Weather prediction performance evaluation on selected machine learning algorithms. *IAES International Journal of Artificial Intelligence*, *11*(4), 1535.

Alves, G. R., Teixeira, I. R., Melo, F. R., Souza, R. T. G., & Silva, A. G. (2018). Estimating soybean yields with artificial neural networks. *Acta Scientiarum. Agronomy*, *40*.

Awe, O. O., & Dias, R. (2022). Comparative analysis of ARIMA and artificial neural network techniques for forecasting non-stationary agricultural output time series. *AGRIS on-line Papers in Economics and Informatics*, *14*(665-2023-114), 3-9.

Bali, B. (2024). Analysis of Emerging Trends in Artificial Intelligence in Education in Nigeria.

Bali, B., & Garba, E. J. (2021). Neuro-fuzzy approach for prediction of neurological disorders: a systematic review. *SN Computer Science*, *2*(4), 307.

Bali, B., Garba, E. J., & Ahmadu, A. S. (2021). Adaptive Neuro Fuzzy Inference System for Diagnosis of Stimulant Use Disorders.

Domingues, T., Brandão, T., & Ferreira, J. C. (2022). Machine learning for detection and prediction of crop diseases and pests: A comprehensive survey. *Agriculture*, *12*(9), 1350.

Durai, S. K. S., & Shamili, M. D. (2022). Smart farming using machine learning and deep learning techniques. *Decision Analytics Journal*, *3*, 100041.

Jiya, E. A., Illiyasu, U., & Akinyemi, M. (2023). Rice Yield Forecasting: A Comparative Analysis of Multiple Machine Learning Algorithms. *Journal of Information Systems and Informatics*, *5*(2), 785-799.

Kuradusenge, M., Hitimana, E., Hanyurwimfura, D., Rukundo, P., Mtonga, K., Mukasine, A., ... & Uwamahoro, A. (2023). Crop yield prediction using machine learning models: case of Irish potato and maize. *Agriculture*, *13*(1), 225.

Kwaghtyo, D. K., & Eke, C. I. (2023). Smart farming prediction models for precision agriculture: a comprehensive survey. *Artificial Intelligence Review*, *56*(6), 5729-5772.

Lata, K., & Chaudhari, B. (2019). Crop yield prediction using data mining techniques and machine learning models for decision support system. *Journal of Emerging Technologies and Innovative Research (JETIR)*.

Mgendi, G. (2024). Unlocking the potential of precision agriculture for sustainable farming. *Discover Agriculture*, *2*(1), 87

Oluwole, O. E., Osaghae, E. O., & Basaky, F. D. (2022). Machine learning Solution for Prediction of Soil Nutrients for Crop Yield: A Survey. *Machine learning*, *9*(9).

Prabavathi, R., & Chelliah, B. J. (2022). A comprehensive review on machine learning approaches for yield prediction using essential soil nutrients. *Universal Journal of Agricultural Research*, *10*(3), 288-303.

Shuaibu, A. M., Muhammad, M. N., & Abu-Safyan, Y. (2021). Forecasting Rice Production in Jigawa State, Nigeria using Fuzzy Inference System. *Dutse J. Pure Appl. Sci*, *7*(4), 203-213.

Zhou, Q., & Ismaeel, A. (2021). Integration of maximum crop response with machine learning regression model to timely estimate crop yield. *Geo-Spatial Information Science*, *24*(3), 474-483.