Journal of Science Research and Reviews

PRINT ISSN: 1595-9074
E-ISSN: 1595-8329
DOI: https://doi.org/10.70882/josrar.2025.v2i3.57
Homepage: https://josrar.esrgngr.org

Original Research Article

Smart Farming for Groundnut Yield Prediction Using IoT and Machine Learning

***[1]Bala, Abdullahi A., [1]Olanrewaju, Oyenike M. and [2]Echobu, Faith O.**

[1]Department of Computer Science, Federal University Dutsin-Ma.
[2]Department of Information Technology, Federal University Dutsin-Ma.
*Corresponding Author's email: alibala1898@fcatp.edu.ng; alibala1989@gmail.com

KEYWORDS
IoT,
Machine learning,
Parameters,
Model performance,
Agriculture.

ABSTRACT
The integration of IoT in agriculture has revolutionized crop production by enhancing productivity, quality, and efficiency while reducing labor costs and boosting farmer income. IoT sensors provide precise data on environmental, soil, and plant factors, critical for predicting crop yields. In this study, groundnut crops were cultivated in 20 pots and monitored using IoT devices over 120 days, generating 480 data instances. Parameters like temperature, soil moisture, and nutrients (nitrogen, phosphorus, potassium) were measured to track growth metrics. Machine learning models Multi-Layer Perceptron (MLP), K-Nearest Neighbors (KNN), and Random Forest (RF) were developed using bagging techniques to predict yield and model growth rates based on NPK levels. Model performance was evaluated using R-squared, MAE, RMSE, and RMSLE metrics. For yield prediction, KNN outperformed RF and MLP with the highest R-squared (0.87), lowest MAE (2.1033), and lowest RMSE (2.0119), while MLP performed worst. Conversely, in modeling growth rates influenced by NPK, MLP excelled with the highest R-squared (0.52), lowest MAE (1.3499), MSE (2.7220), RMSE (1.6498), and an exceptionally low RMSLE (0.0024). Overall, KNN was the top performer for yield prediction, followed by RF and MLP, whereas MLP was superior for growth rate predictions. This highlights the potential of IoT and machine learning in advancing agricultural intelligence.

CITATION
Bala, A. A., Olanrewaju, O. M., & Echobu, F. O. (2025). Smart Farming for Groundnut Yield Prediction Using IoT and Machine Learning. *Journal of Science Research and Reviews*, *2*(3), 1-7. https://doi.org/10.70882/josrar.2025.v2i3.57

INTRODUCTION

Groundnut cultivation is a cornerstone of Nigeria's agricultural sector, contributing significantly to the economy and food security. The northern regions, with favorable climates and soils, are central to groundnut farming, particularly during the rainy season. Groundnuts are a critical source of protein and oil, but challenges like pests, poor infrastructure, and market instability hinder productivity. Nigeria is the third largest producer of groundnut in 2019 with annual production of 4.4 million tonnes after China 17.1 million tonnes and India 6.7 million (FAO, 2021). Improving farming techniques, storage systems, and value-added products is essential for strengthening the sector.

Climate and soil conditions play a vital role in groundnut farming. The crop thrives in warm temperatures and requires 500–1,200 mm of well-distributed rainfall annually. Sandy loam soils are ideal for optimal nutrient absorption. However, population growth and limited arable land pressure farmers to maximize yields on suboptimal soils, often leading to inefficient fertilizer use due to imprecise application. Research has identified optimal nutrient levels for maximum yield, providing guidelines for farmers. However, these approaches are static and fail to account for dynamic changes in soil nutrients or environmental factors over time Ezihe et al (2017).

To address these limitations, IoT-based technologies offer real-time data collection and analysis. A research farm in Kano used IoT sensors to monitor environmental factors, plant growth, and soil nutrients, enabling continuous data acquisition. This data was leveraged to develop predictive machine learning models that optimize resource use and improve farming practices. IoT, which connects devices via sensors and cloud computing, facilitates precision agriculture by enabling targeted interventions like optimized irrigation and fertilization. Smart agriculture integrates IoT, AI, and data analytics to enhance efficiency, reduce waste, and increase yields, making farming more adaptive and sustainable. the Internet of Things that assists farmers in improving their agricultural operations and making the most use of their agricultural land for increased production and profitability Rekha et al (2017),

IoT frameworks form the backbone of modern precision agriculture. Jayaraman et al. (2016) demonstrated the feasibility of low-cost IoT systems using Raspberry Pi and soil sensors, achieving 85% accuracy in yield estimation for small farms. Similarly, Kodali et al. (2019) deployed LoRaWAN-based networks for paddy field monitoring, achieving 90% accuracy over a 2 km range. However, scalability remains a challenge, as highlighted by Tzounis et al. (2017), who noted that IoT systems often struggle with large-scale deployments due to fragmented communication protocols. Gavhane et al. (2018) addressed soil nutrient optimization using NPK sensors and fuzzy logic, reducing fertilizer waste by 35%, but emphasized calibration challenges in heterogeneous soil conditions.

Several studies have focused on developing predictive models for crop yield using statistical and machine learning approaches. Medar and Rajpurohit (2014) conducted a survey of data mining techniques for crop yield prediction, emphasizing the importance of environmental and soil attributes in influencing yield outcomes. In the context of groundnut yield prediction, Shah and Shah (2018) analyzed multiple machine learning algorithms, including Multiple Linear Regression, K-Nearest Neighbors (KNN), Artificial Neural Networks (ANN), and Regression Trees. Their study demonstrated that the KNN algorithm outperformed other methods in predicting groundnut yields based on soil, environmental, and abiotic factors. The aim of this research work is to develop groundnut yield production model using machine learning and smart farming tool.

MATERIALS AND METHODS

The methodology adopted in this research is shown in figure 1 which outlines a comprehensive workflow for the research work. It began with IoT device programming and calibration, where sensors and devices were correctly configured for data collection. The next step, experimental site setup and planting, involved preparing the field and planting crops, which was crucial for gathering relevant data. Dataset acquisition and data capturing followed, where IoT devices collected data on environmental and crop conditions. Data preprocessing was then performed to clean and prepare the data for analysis. The development of the model phase involved training machine learning algorithms using the processed data. Finally, testing and validation of the model and performance evaluation ensured the model's accuracy and effectiveness, providing a robust solution for the intended application.

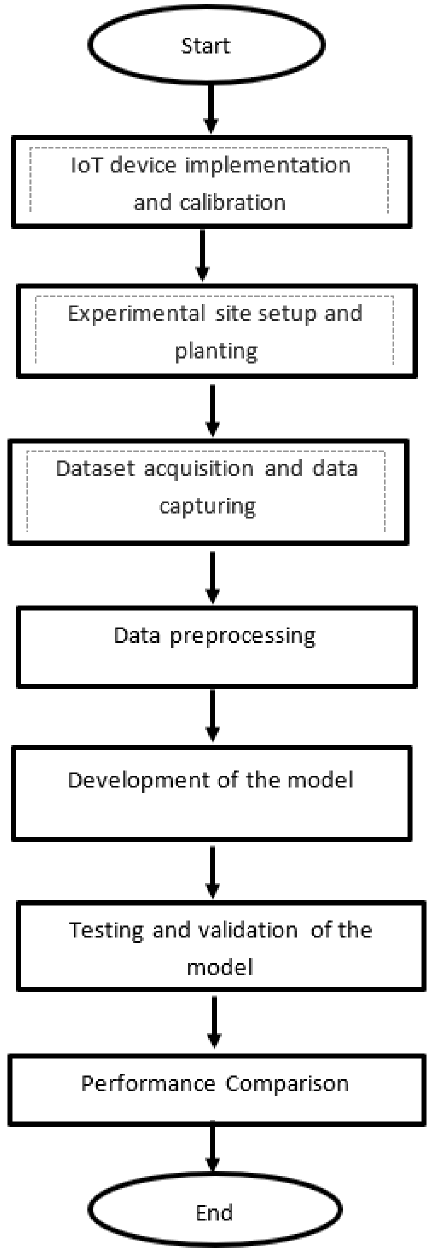


Figure 1: Workflow Adopted in the Research

IoT Device Implementation

Soil moisture sensors, temperature sensors, Modbus, NPK sensors, and an Arduino Uno were integrated to implement the IoT devices used for data capture in groundnut cultivation as shown in figure 2. The Arduino Uno served as the central microcontroller, interfacing with the sensors to collect real-time data on soil moisture, temperature, and nutrient levels (Nitrogen, Phosphorus, and Potassium). The Modbus protocol was utilized to enable communication between the Arduino Uno and other devices, ensuring seamless data transmission to the central system. The Arduino Uno processed the sensor data, which provided critical insights into the environmental and soil conditions, facilitating precise monitoring and management of the groundnut crop. This integration of IoT devices, powered by the Arduino Uno, significantly enhanced the accuracy and efficiency of data-driven decision-making in the cultivation process.

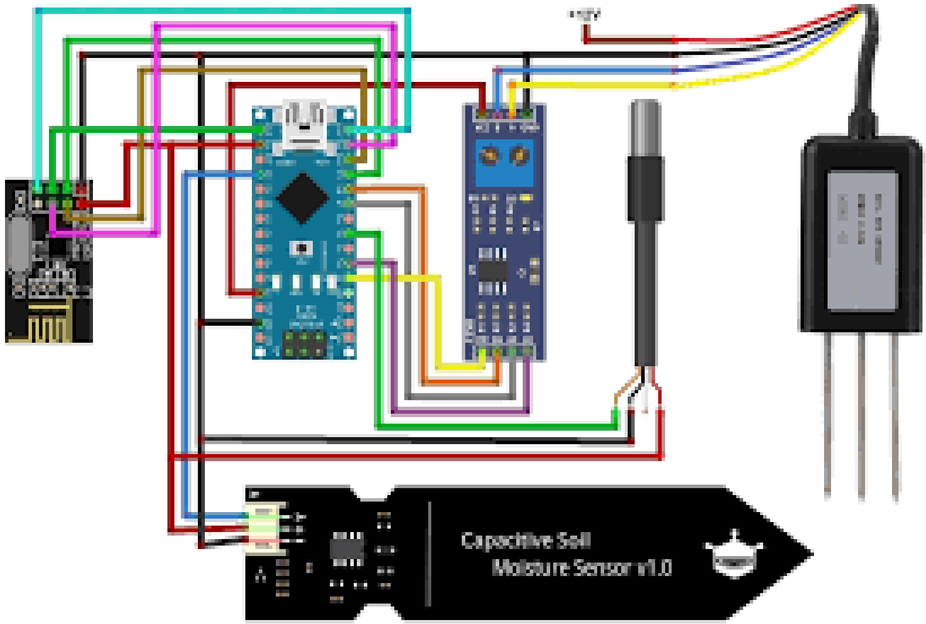


Figure 2: IoT Implementation Block

IoT Device Calibration

The test was carried out by immersing the temperature sensor's probe into wet soil at regular intervals and comparing its readings with those of a thermometer for calibration. Similarly, the soil nutrient sensor readings were compared against laboratory soil nutrient analyses to ensure accuracy. The results demonstrated high precision, with an average accuracy of 93.91% for temperature, 96.5% for Nitrogen, 93.3% for Potassium, and 97.2% for Phosphorus. These findings suggest the sensors are reliable for monitoring soil conditions effectively, as shown in the figure 3.

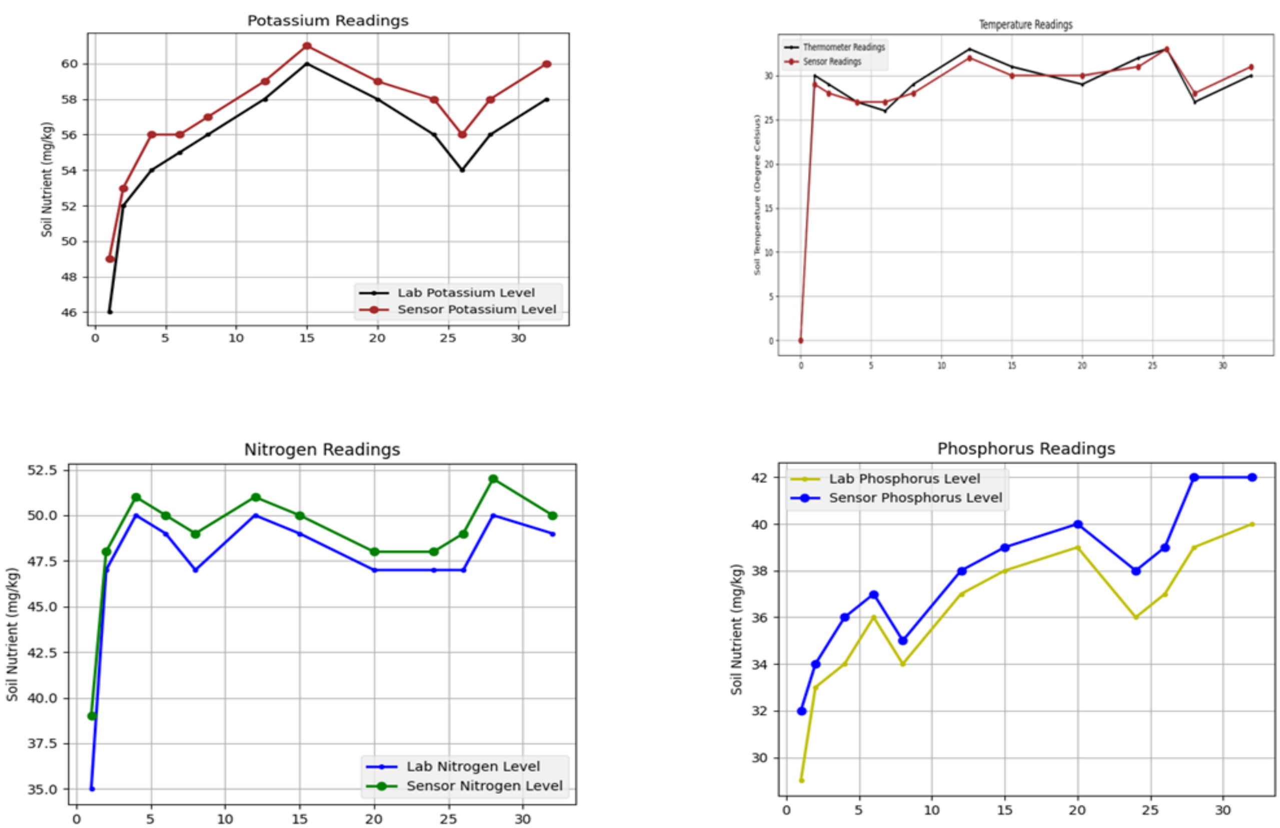


Figure 3: Device Calibration

Experimental Site

The experiment was carried out during the dry season of 2024, in the Department of Soil Science, Federal College of Agricultural Produce Technology Kano (11° 11' N, 07° 38' E) and 686 m above sea level in Nigeria's Northern Guinea savannah ecological zone. There are two seasons in the area: dry season (November-April) and the rainy season (May-October). It also exhibits a mono-modal rainfall pattern, with an annual mean of 1110 mm ranging from 950 to 1270 mm. The average yearly temperature is 25 degrees Celsius.

Feature Selection

In building a crop yield prediction model, soil temperature, soil moisture, soil Nitrogen (N), Phosphorus (P), and Potassium (K) are critical features due to their fundamental roles in plant growth and crop productivity. These factors collectively determine the soil's capacity to support crop health and development, and they significantly influence yield outcomes.

Temperature is a pivotal factor in crop yield prediction models. By reflecting its influence on crop physiology and growth stages, ML models provide insights into agricultural productivity under varying climatic conditions. Research underscores its importance, with studies demonstrating that integrating temperature data significantly enhances prediction accuracy across regions and crops. Soil moisture determines the water accessible to plants, influencing physiological processes like photosynthesis and transpiration. Water stress due to insufficient soil moisture often results in reduced yields Qinqing et al (2022).

Data Collection

The data was collected using various IoT sensors such as soil nutrient monitors, soil moisture, soil temperature sensors, and from weather stations in addition to the plant growth record (Plant height, leaves, flowers and yield). The data was collected in a consistent and standardized format to ensure its accuracy and reliability. Temperature, relative humidity, wind speed, wind direction and rainfall data were collected from climate data record of Agro-climatological station. Altogether, the dataset contained records of soil and climatic condition for the growing period.

All the collected data were structured, numerical, and continuous, enabling seamless integration into analytical models. These datasets formed a robust foundation for understanding the interactions between environmental conditions, soil properties, and plant responses.

Soil temperature, moisture, nitrogen, potassium, phosphorus was recorded in addition to environmental data from weather station, as well as crop data, consisting of the height of the plants, the count number of leaves per stand, count number of tillers, and number of flowers.

Harvesting was done when almost all plants had reached the end of active vegetative growth. Two plants were selected randomly and tagged in each net pot, and the total dry matter, total fresh weight and number of pod were recorded.

Data Pre-processing

The collected data was pre-processed to remove any missing values, outliers, and irrelevant information. Missing values were imputed using a mean, median, or mode method depending on the distribution of the data. Outliers were identified using a box plot and removed using the Interquartile Range (IQR) method, where values outside the range of and were considered outliers and removed. The removal of outliers was based on the principle that they can significantly affect the accuracy of the prediction model. Data was later transformed to scale down to have a mean of 0 and a standard deviation of 1 using standardization formula as follows

$$\text{Xscaled} = \frac{X-\mu}{\sigma} \quad (1)$$

Where μ is the mean of the data and σ is the standard deviation of the date. The remaining data was then used for further analysis. Data normalization was also performed to ensure that all variables had similar ranges and distributions (Kotsiantis & Kanellopoulos, 2006).

Model development environment

The model was developed using Weka (Waikato Environment for Knowledge Analysis), a comprehensive data mining suite that offers various machine learning algorithms for classification, clustering, and association tasks. This software also provides tools for data preprocessing, visualization, and other functionalities to transform data into an appropriate format for mining.

Three algorithms—K-Nearest Neighbors (KNN), Multi-Layer Perceptron (MLP), and Random Forest—were chosen due to their complementary strengths and alignment with the dataset's characteristics. KNN is valued for its simplicity and flexibility, making it effective for smaller datasets and scenarios where interpretability is important. MLP, a type of neural network, excels in capturing complex, non-linear relationships, making it suitable for datasets with intricate patterns. Random Forest stands out for its robustness in handling high-dimensional data and its resistance to overfitting, thanks to ensemble learning. Each of these algorithms supports both classification and regression tasks, providing versatile tools for analyzing varied dataset structures.

Nearest Neighbors (KNN) is widely used for prediction because of its simplicity, interpretability, and effectiveness in various scenarios. KNN operates by identifying the closest data points in the feature space and assigning the predicted output based on their values, making it intuitive and easy to understand. It is particularly effective when the dataset is small, as it does not require a complex training phase, and it adapts well to changing data distributions because it relies solely on local information. Furthermore, KNN can model non-linear decision boundaries effectively, especially when combined with an appropriate distance metric such as Euclidean or Manhattan distance. However, its performance can degrade with high-dimensional data (the "curse of dimensionality"), which requires careful preprocessing, such as feature selection or dimensionality reduction Han et al (2012).

Multi-Layer Perceptron (MLP) is a powerful algorithm for prediction due to its ability to model complex, non-linear relationships in data. As a type of feedforward neural network, MLP consists of multiple layers of nodes, each employing activation functions to capture intricate patterns that traditional linear models might miss. This capability makes MLP particularly effective for tasks with high-dimensional or structured data, such as image recognition or natural language processing. Additionally, MLP's flexibility allows it to adapt to diverse prediction tasks, including classification and regression, by adjusting the network architecture and hyperparameters. However, MLP requires careful tuning and sufficient training data to avoid overfitting and to achieve optimal generalization. Goodfellow et al (2016).

Random Forest is widely used for prediction due to its robustness, accuracy, and versatility. It is an ensemble learning method that combines multiple decision trees, each trained on random subsets of the data and features, to make predictions. This process reduces overfitting, which is a common issue with individual decision trees, and enhances generalization to unseen data. Random Forest performs well with high-dimensional datasets and can handle both classification and regression tasks effectively. Furthermore, it offers built-in mechanisms for estimating feature importance, making it a valuable tool for interpretability and feature selection in complex datasets. Hastie et al (2009).

Performance evaluation

The Mean Absolute Error (MAE) measures the average magnitude of errors, giving a clear understanding of the model's accuracy without accounting for the direction of the errors. The Root Mean Squared Error (RMSE), which is derived by taking the square root of the Mean Squared Error (MSE), provides an interpretable metric that aligns with the original unit of measurement and heavily penalizes larger errors. Additionally, Relative Absolute Error (RAE) and Root Relative Absolute Error (RRAE) and RSquare (R^2), or Coefficient of Determination, is essentially the squared value of the sample correlation coefficient (denoted as r) between the observed outcomes and their predicted values. This coefficient ranges from 0 to 1 were employed to compare the performance of the model relative to a baseline, offering insights into its effectiveness in minimizing errors.

R^2 (Coefficient of determination):

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

RMLE (Root Mean Log Error):

$$\text{RMLE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\ln(\hat{y}_i + 1) - \ln(y_i + 1))^2} \quad (3)$$

RMSLE (Root Mean Squared Log Error):

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

MAE (Mean Absolute Error):

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

RAE (Relative Absolute Error):

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

RRSE (Root Relative Squared Error):

$$\text{RRSE} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(\bar{y} - y_i)^2}} \quad (7)$$

Where: $\hat{y}_i$ = predicted value for the i-th observation, $y_i =$ actual value for the i-th observation, $\bar{y} =$ mean of the actual values, $\text{n} =$ number of observations, $\ln =$ natural logarithm, $\log =$ logarithm (base 10 or natural, depending on context)

RESULTS AND DISCUSSION

Data Description

The table labelled Table 1 Statistical description of the data provides a summary of five variables: Temp (temperature), N (nitrogen), P (phosphorus), K (potassium), Moisture, and Yield pod plant^{-1}.

Mean temperature of 29.57°C with low std of 0.48, Indicates a narrow temperature range due to controlled condition. A small variation of Nitrogen (N) and Phosphorus (P) show minimal variability across samples, however Potassium (K) shows slightly higher variability compared to nitrogen and phosphorus. The yield data shows the largest variability (std = 2.79), reflecting a broader range of outcomes in the research. The relatively small standard deviations across Temp, N, P, and K indicate consistency in soil conditions. Moisture and Yield exhibit slightly greater variability, suggesting these are influenced by other factors or show natural variability. The quartiles give a good sense of distribution for each variable, with most variables showing small interquartile ranges, highlighting uniformity.

Table 1: Statistical description of the data

	Temp	N	P	K	Moisture	Yield pod plant^{-1}
Count	20.00000	20.00000	20.00000	20.00000	20.00000	20.00000
Mean	29.57470	18.01600	22.523750	54.82750	34.94950	21.00000
Std	0.48291	0.540929	0.250015	0.599042	1.983677	2.79096
Min	28.78000	17.44000	22.00000	53.75000	31.20000	13.00000
25%	29.32750	17.57500	22.37750	54.43250	34.04000	20.00000
50%	29.54000	17.85000	22.55500	54.80500	34.96500	22.00000
75%	29.81500	18.37500	22.67500	55.09250	36.34250	23.00000

Data Visualization

The data were graphically represented using stack bar charts making it easier to identify patterns, trends, outliers, and relationships. The stacked bar chart in figure 4, presents the proportional contributions of various parameters—yield, moisture, potassium, phosphorus, nitrogen, temperature, fresh weight, dry weight, number of flowers, and number of leaves for different pots. Each bar represents a pot, and the parameters are color-coded, showing their relative percentages within the total measurement. Yield and moisture occupy significant portions across pots, indicating their high contribution to overall plant development. Nitrogen, phosphorus, and potassium show consistency across pots, reflecting similar nutrient availability. The number of leaves and flowers forms a smaller proportion compared to parameters like temperature or fresh weight, but these are essential for assessing vegetative and reproductive growth. Overall, the chart effectively illustrates the distribution of multiple factors influencing plant performance, highlighting both consistency and variation across pots.

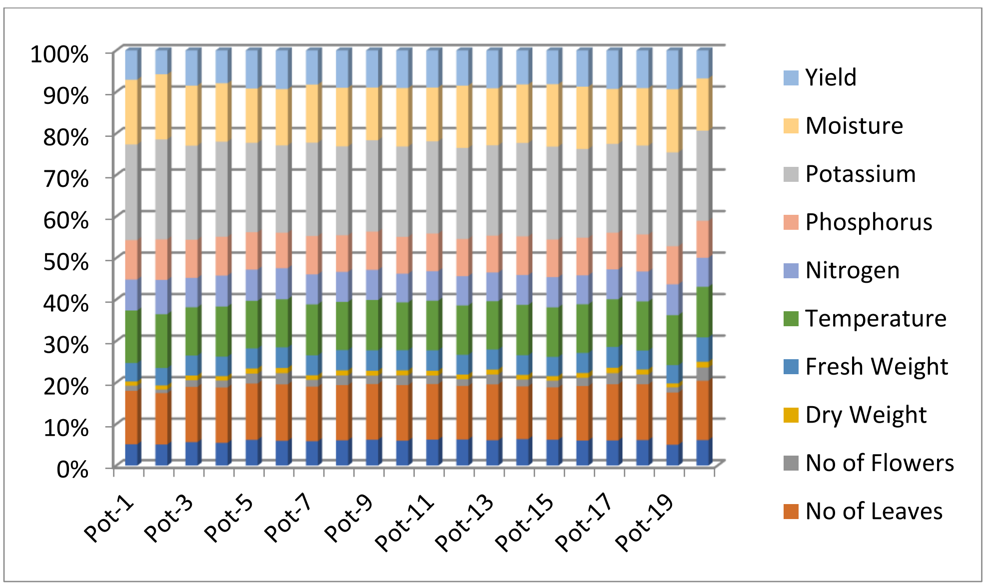


Figure 4: Data Visualization

Yield Prediction Model Result

Table 2 presents the performance measures of individual models, specifically focusing on MLP, RF, and KNN algorithms. The table provides various metrics such as R-squared, Mean Square Logarithmic Error (MASLE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Relative Absolute Error (RAE), and Root Relative Absolute Error (RRSE) to evaluate these models comprehensively.

Among the three models, KNN demonstrates the highest R-squared value at 0.87, indicating it explains 87% of the variance in the data, followed by RF with 0.85 and MLP with 0.49. When examining the MAE metric, RF performs slightly better than KNN, with values of 2.2334 and 2.1033 respectively, while MLP shows the highest error at 2.3087. The RMSE values reveal a similar trend where KNN and RF are comparable at 2.9837 and 2.0119, whereas MLP has a significantly higher value at 3.2268.

The RMSE penalizes larger errors, and RF's lower value suggests it mitigated outlier effects better. Relative Absolute Error (RAE): KNN had the smallest RAE (97%), highlighting its accuracy compared to the baseline mean predictor. RF (103%) was slightly better than MLP (107%). Root Relative Squared Error (RRSE): Both RF and KNN achieved similar values (105%), slightly outperforming MLP (113%).

Overall, this table suggests that KNN and RF outperform MLP across most metrics, making them preferable choices for the given task. The differences in performance among the models highlight the importance of selecting appropriate algorithms based on specific evaluation criteria. This comparative analysis offers valuable insights into the strengths and weaknesses of each model when applied to the dataset.

Table 2: The error measure of individual models

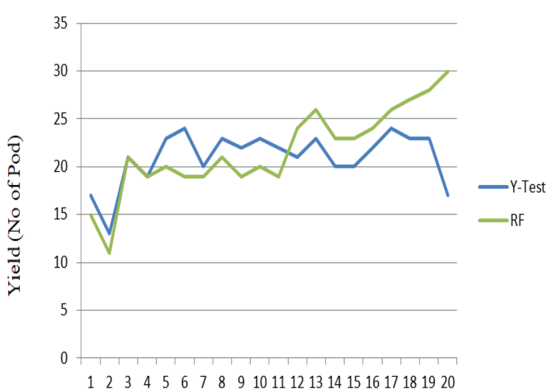
	R^2	MALE	MASLE	MAE	RMSE	RAE	RRSE
MLP	0.49	0.1085	0.163	2.3087	3.2268	107%	113%
RF	0.85	0.1056	0.1509	2.2334	2.9837	103%	105%
KNN	0.87	0.0998	0.1531	2.1033	2.0119	97%	105%

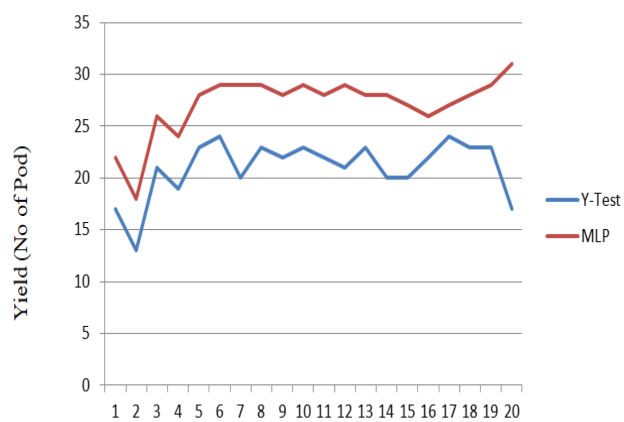
The graph below in figure 5 compares the performance of three machine learning models MLP (Multi-Layer Perceptron), RF (Random Forest), and KNN (K-Nearest Neighbors) in predicting yield (number of pods) against actual test values (Y-Test). The MLP model (red line) consistently overestimates the yield compared to the actual Y-Test values, showing a tendency toward higher predictions. The RF model (green line) generally follows the Y-Test values more closely, but it diverges in certain instances, particularly around instances 16–20, where it shows a steeper increase.

The KNN model (purple line) tracks the Y-Test values well, with occasional minor deviations, such as in instances 6 and 18, where it slightly overestimates or underestimates the yield. The Y-Test values (blue line) represent the actual observed yield, serving as the benchmark for evaluating the models.

MLP appears to have the least alignment with the actual data but maintains a smooth trend, which may indicate overfitting. RF strikes a balance between prediction accuracy and capturing trends, though it slightly exaggerates yield for higher values. KNN is the closest to Y-Test for most instances, particularly in middle-range values, suggesting its suitability for capturing localized patterns.

The combined analysis highlights the strengths and weaknesses of each model, with KNN and RF showing greater consistency with the actual test data. Overall, KNN and RF outperform MLP in terms of tracking the actual yield trend accurately.





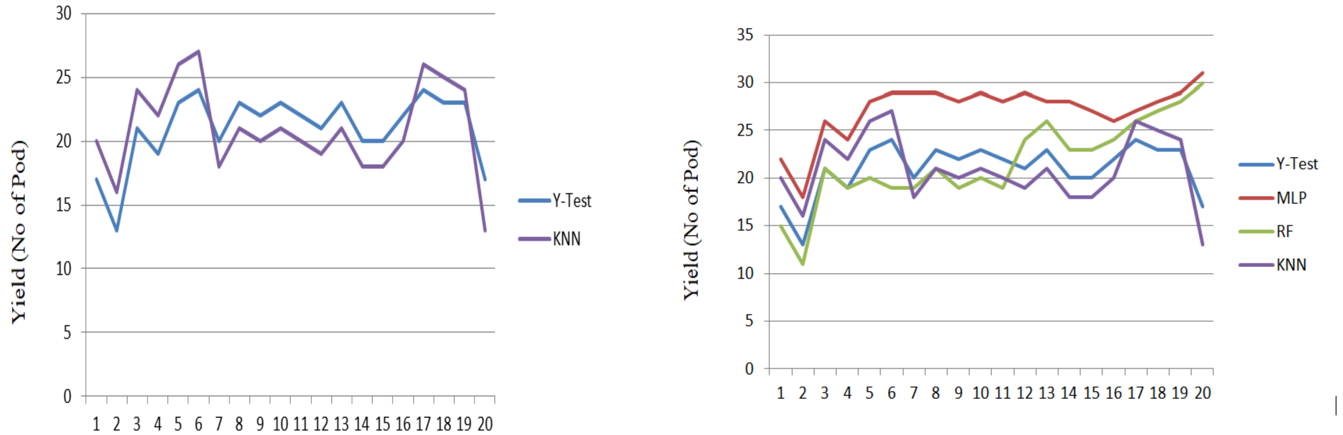


Figure 5: Performance Graphs

Comparative performance

The performance of the K-Nearest Neighbors (KNN) model in the 2018 paper by Shah and Shah, titled "Groundnut Crop Yield Prediction Using Machine Learning Techniques," demonstrates a strong ability to predict groundnut yield with an RMSE value of 1.2343, which is notably lower than the RMSE values achieved by KNN in the current study. In contrast, the KNN model developed in this research recorded an RMSE of 2.0119, indicating a comparatively lower accuracy when applied to the dataset collected from Kano State, Nigeria.

CONCLUSION

This research focused on using machine learning and IoT-based systems to predict groundnut yield and growth rates in response to NPK fertilizer application. The study implemented IoT devices for real-time data collection, developed predictive models using regression techniques, and evaluated their performance using metrics such as R^2, RMSE, MAE, and MSLE. The results demonstrated the effectiveness of Random Forest (RF) and Multi-Layer Perceptron (MLP) in predicting yield and growth rates of groundnut crop. The study contributes to smart farming practices by providing tools for data-driven decision-making in agriculture. Among the models, KNN achieved superior performance in yield prediction, particularly in reducing absolute and logarithmic errors, while RF demonstrated a strong balance between reliability and error minimization, ranking second. MLP trailed behind in yield forecasting. However, for modeling vegetative growth under NPK conditions, MLP outperformed KNN and RF despite challenges of underfitting caused by limited dataset size. The findings suggest KNN is optimal for yield prediction, whereas MLP shows greater promise for analyzing growth responses to nutrients. In light of the weak responses observed under the growth models, improving model performance requires expanding and diversifying the training dataset. Additionally, deep learning technique can also be explored in future research.

REFERENCES

Bachuwar, S., Kokate, P., Kadu, A., & Devendrachari, M. C. (2018). IoT-based soil nutrient monitoring using NPK sensors. International Journal of Advance Research in Science and Engineering, 7 (9), 1–7.

Chen, H., Wu, W., & Liu, H.-B. (2015). Assessing the relative importance of climate variables to rice yield variation using support vector machines. Springer. https://doi.org/10.1007/s11119-015-9406-6

Ezihe, J.A.C. & Agbugba, Ikechi & Idang, C.. (2017). Effect of climatic change and variability on groundnut (Arachis hypogea, L.) production in Nigeria. Bulgarian Journal of Agricultural Science. 23. 906-914.

Farooq, M., Khan, M. U., & Rehman, A. (2019). IoT-driven reductions in water and fertilizer use by 20–30%. In IoT applications in agriculture: Challenges and opportunities (pp. 45–67). Springer. https://doi.org/10.1007/978-3-030-12345-6_3

FAO (2021). Faostat. Retrieved May 13, 2021 from http://www.fao.org/faostat/en/#data/QC

Gavhane, A., Kokate, P., Kadu, A., & Devendrachari, M. C. (2018). IoT-based soil nutrient monitoring using NPK sensors. International Journal of Advance Research in Science and Engineering, 7 (9), 1–7.

Goap, A., Sharma, D., Shukla, A. K., & Krishna, C. R. (2018). An IoT-based smart agriculture system using machine learning. In 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI) (pp. 2372–2376). IEEE. https://doi.org/10.1109/ICACCI.2018.8554915

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

Gupta, V., Yadav, S., & Kumar, R. (2021). Crop yield prediction from meteorological parameters using artificial intelligence. AI in Agriculture. https://doi.org/10.1016/j.aiag.2021.100045

Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concepts and techniques . Morgan Kaufmann.

Khanna, A., & Kaur, S. (2019). Optimizing potato irrigation with ANNs, reducing water usage by 22%. Journal of Agricultural Informatics, 10 (2), 45–58. https://doi.org/10.17700/jai.2019.10.2.512

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. Computers and Electronics in Agriculture, 147, 70–90. https://doi.org/10.1016/j.compag.2018.02.016

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

Madhumathi, R., Arumuganathan, T., & Shruthi, R. (2020). Soil NPK and moisture analysis using wireless sensor networks. In 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1–6). IEEE. https://doi.org/10.1109/ICCCNT49239.2020.9225511

Mahmud, A., Esakki, B., & Seshathiri, S. (2020). Quantification of groundnut leaf defects using image processing algorithms. In Proceedings of International Conference on Trends in Computational and Cognitive Engineering: Proceedings of TCCE 2020 (pp. 649–658). Springer Singapore. https://doi.org/10.1007/978-981-15-9712-1_55

Rehman, A., Safdar, R., & Farooq, M. (2021). Gradient boosting and IoT sensors reduce wheat farming water use by 18%. Sustainability, 13 (12), 6789. https://doi.org/10.3390/su13126789

Rekha, P., Rangan, V. P., Ramesh, M. V., & Nibi, K. V. (2017, October). High yield groundnut agronomy: An IoT based precision farming framework. In *2017 IEEE Global Humanitarian Technology Conference (GHTC)* (pp. 1-5). IEEE.

Sharma, D., Goap, A., Shukla, A. K., & Krishna, C. R. (2020). IoT-based smart agriculture system for crop yield prediction and disease detection. IEEE Internet of Things Journal, 7 (8), 7123–7134. https://doi.org/10.1109/JIOT.2020.2985567

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and Computing, 14 (3), 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88