Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX

DOI: 10.XXXXX/jossda.2026.1.1.007

JOURNAL OF STATISTICAL SCIENCES & DATA ANALYTICS
Department of Statistics, Ahmadu Bello University, Zaria, Nigeria | www.jossda.abu.edu.ng www.jossda.com.ng

Integrating MANOVA and Decision Tree Classification to Analyze Crop Responses to Fertilizer Nutrients and Soil Conditions

A. Usman[1], I. A. Sadiq[1], M. Tasiu[1], Y. Aminu[1], A. I. Ishaq[1],Y. Zakari[1], Y. Aliyu[1], R. O. David[1], and M. Sirajo[1]
[1]*Department of Statistics, Ahmadu Bello University, Zaria, Nigeria*

Corresponding Author:
A. Usman, Department of Statistics, Ahmadu Bello University, Zaria, Nigeria.
abubakarusman28@gmail.com

Abstract

This study examines the relationship between fertilizer nutrients (Nitrogen, Phosphorus, and Potassium), soil conditions, and environmental factors in relation to crop response using secondary data from the Kaggle Repository Crop Recommendation Dataset, comprising 2,200 observations across 22 crop classes. The study was guided by four objectives: to describe the characteristics of fertilizer nutrients and soil conditions, assess differences in nutrient and environmental variables across crop types using Multivariate Analysis of Variance (MANOVA), examine the relationships among environmental variables using Pearson correlation, and evaluate the ability of a Decision Tree model to classify crop types and identify important predictors. The findings showed substantial variability in fertilizer nutrients, particularly Potassium, while soil pH was comparatively stable. MANOVA indicated significant differences in the combined fertilizer nutrient and soil/environmental characteristics across crop types. The correlations among environmental variables were generally weak, suggesting that the variables provide distinct information about crop conditions. The Decision Tree model achieved an overall classification accuracy of 89.24% (Kappa = 0.8873), with Rainfall (19.87%), Potassium (19.20%), and Humidity (17.58%) emerging as the most influential predictors. By integrating MANOVA for assessing multivariate differences with Decision Tree classification for identifying important predictors, the study provides a complementary statistical and machine-learning approach to understanding crop responses. The findings demonstrate the potential of combining these methods to support precision agriculture through evidence-based, crop-specific fertilizer recommendations and improved management of soil and environmental conditions.

Keywords: MANOVA; decision tree; fertilizer nutrients; soil conditions; crop classification; precision agriculture.

1. Introduction

Agriculture remains central to the economies of developing nations, particularly across Sub-Saharan Africa, where it underpins rural livelihoods, employment, and food security (Food and Agriculture Organisation [FAO], 2020). In Nigeria, crop productivity depends on the interaction between fertilizer application and inherent soil chemical properties. The macronutrients Nitrogen (N), Phosphorus (P), and Potassium (K) must be supplied in balanced proportions to support plant metabolism and growth (Brady & Weil, 2016; Havlin et al., 2014), while soil pH, moisture, and climatic conditions determine how effectively these nutrients are absorbed (Tisdale et al., 2013).

Despite this, agricultural practice in many developing countries still relies on generalized “blanket” fertilizer recommendations that overlook spatial variation in soil chemistry and crop-specific nutrient needs (Aliyu et al., 2022; Shehu et al., 2019). This contributes to economic waste from over-application and environmental harm, including soil acidification and nutrient leaching (Adelekan et al., 2010; Lal, 2009). Addressing this requires a shift toward site-specific, data-driven fertilizer management.

Statistically, Multivariate Analysis of Variance (MANOVA) is well suited to testing whether nutrient and environmental mean vectors differ significantly across crop groups (Johnson & Wichern, 2019; Montgomery, 2019), but it cannot generate predictive decision rules. Machine learning methods such as Decision Tree classification (CART) complement MANOVA by partitioning the feature space into interpretable rules without requiring distributional assumptions (Breiman et al., 1984; Han et al., 2012). Few studies, however, integrate parametric hypothesis testing with predictive classification within a single framework (Shuaibu et al., 2024).

This study examines how soil fertility and environmental conditions influence crop performance by considering fertilizer nutrients, including Nitrogen, Phosphorus, and Potassium, together with rainfall, temperature, humidity, and soil pH. Because these variables differ in their levels and variability, descriptive statistics are first used to summarise their distributions and provide an understanding of the general characteristics of the dataset. Since crop performance is influenced by several soil and environmental factors simultaneously, MANOVA is employed to determine whether the combined mean vectors of these variables differ significantly across crop groups. In addition, the relationships between these factors and crop types may involve complex patterns that are not easily explained using conventional statistical methods alone. Therefore, Decision Tree classification is applied to identify the most influential predictors and generate interpretable classification rules for distinguishing among crop types. The integration of these methods provides both statistical evidence of differences among crop groups and a practical approach for identifying the factors most useful for crop classification.

2. Literature Review

Soil fertility reflects the soil's capacity to supply balanced nutrients for plant growth (Brady & Weil, 2016). Nitrogen drives vegetative growth and chlorophyll synthesis; Phosphorus supports energy transfer and root development; Potassium regulates osmotic balance and disease resistance (Mengel & Kirkby, 2001; Tisdale *et al*., 2013). Nutrient uptake, however, depends on soil pH and climatic factors, with optimal availability generally occurring between pH 6.0 and 7.5 (Tisdale *et al*., 2013); outside this range, nutrients such as phosphorus become chemically fixed and unavailable to roots (Havlin *et al*., 2014). Temperature, humidity, and rainfall further govern soil moisture, microbial activity, and transpiration, shaping crop response (Draper & Smith, 1998; Troeh & Thompson, 2005).

Because agricultural datasets typically involve correlated response variables, MANOVA is preferred over multiple univariate ANOVAs, as it evaluates group mean differences while controlling the family-wise Type I error rate and preserving covariance structure (Johnson & Wichern, 2019; Rencher & Christensen, 2012). Test statistics such as Wilks' Lambda allow researchers to determine whether crop types differ in their multivariate nutrient and environmental profiles (Gomez & Gomez, 1984; Montgomery, 2019).

Machine learning has increasingly complemented classical inference in agricultural research (Kamilaris & Prenafeta-Boldú, 2018). Decision Tree algorithms, particularly Classification and Regression Trees (CART), are valued for interpretability, requiring no assumptions of normality or linearity (Breiman et al., 1984; Han et al., 2012). By recursively splitting the feature space using impurity measures such as the Gini Index, CART produces transparent if-then rules linking soil and climate variables to crop outcomes.

This study is grounded in three theoretical pillars: The Law of the Minimum/Soil Fertility Theory, which holds that crop productivity is constrained by the scarcest resource (Brady & Weil, 2016); Multivariate Statistical Theory, governing hypothesis testing under correlated variables (Johnson & Wichern, 2019); and Information Theory, underpinning entropy-based recursive partitioning in Decision Trees (Breiman et al., 1984). While prior studies (Shehu et al., 2019; Aliyu et al., 2022; Shuaibu et al., 2024) have examined soil fertility or crop classification in isolation, this study integrates both, applying MANOVA to validate group separation before Decision Tree modelling for predictive classification.

3. Methodology

3.1 Research Design and Data Description

This study utilises a quantitative, secondary analytical research design. The dataset analysed was extracted from the Kaggle repository ("Crop Recommendation Dataset"). It comprises N=2,200 structured observations across 22 distinct agricultural crop categories (100 observations per crop class). The continuous predictor variables and target categorical response are summarised below:

1. Primary Soil Nutrients:
   i. Nitrogen (N) content in soil (kg/ha)
   ii. Phosphorus (P) content in soil (kg/ha)
   iii. Potassium (K) content in soil (kg/ha)
2. Climatic and Soil Conditions:

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
Usman *et al.* 2026 **DOI**: 10.XXXXX/jossda.2026.1.1.007

i. Temperature (°C)
ii. Relative Humidity (%)
iii. Soil pH level (0-14 scale)
iv. Rainfall (mm)

3. Target Categorical Class (Y): Crop type label (e.g., rice, maize, chickpea, kidney beans, banana, mango, coffee, etc.).

3.2 Exploratory and Descriptive Statistical Analysis

Descriptive statistical parameters were calculated across all numerical variables to establish overall baseline distributions and dispersion characteristics. Evaluated parameters include: sample mean ($\overline{X}$), sample variance (s^2), standard deviation (s), minimum (Min), maximum (Max), coefficient of variation, skewness, kurtosis, and interquartile range (IQR), which were computed for all numerical variables to establish baseline distributional characteristics.

3.3 Multivariate Analysis of Variance (MANOVA) Model Specification

To test the null hypothesis (H_0) that mean vectors of soil nutrients and environmental conditions do not differ significantly across crop types, two distinct MANOVA models were specified:

Model 1 (Soil Macronutrients):

$$Y_1 = (N, P, K)^T \tag{1}$$

Model 2 (Environmental Parameters):

$$Y_2 = (Temperature, Humidity, pH, Ra\inf all)^T \tag{2}$$

Using the same grouping variables, suppose there are g crop types and n_i observations in the i^{th} crop type. The multivariate response vector for the j^{th} observation in the i^{th} crop type is expressed as

$$Y_{ij} = \begin{bmatrix} Y_{1ij} \\ Y_{2ij} \\ . \\ . \\ . \\ Y_{pij} \end{bmatrix} \tag{3}$$

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX
 DOI: 10.XXXXX/jossda.2026.1.1.007

where p denotes the number of response variables. For fertilizer nutrients, $p=3$, whereas for soil conditions, $p=4$.

The general linear model matrix equation for MANOVA is expressed as:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \tag{4}$$

where Y_{ij} is the $p\times1$ multivariate response vector for j^{th} observation in the i^{th} crop group $(i=1,2,...,22;\ j=1,2...,100)$, μ is the $p\times1$ grand mean vector across all observations, τ_i is the $p\times1$ treatment effect vector corresponding to the i^{th} crop group, and ε_{ij} is the px1 random error vector assumed to follow an independent and identically distributed multivariate normal distribution:

$$\varepsilon_{ij} \sim N_p(0,\Sigma) \tag{5}$$

For the fertilizer nutrients, the model can be expressed explicitly as

$$\begin{bmatrix} N_{ij} \\ P_{ij} \\ K_{ij} \end{bmatrix} = \begin{bmatrix} \mu N \\ \mu P \\ \mu K \end{bmatrix} + \begin{bmatrix} \tau N_i \\ \tau P_i \\ \tau K_i \end{bmatrix} + \begin{bmatrix} \varepsilon N_{ij} \\ \varepsilon P_{ij} \\ \varepsilon K_{ij} \end{bmatrix} \tag{6}$$

Similarly, for the soil condition variables,

$$\begin{bmatrix} Temperature_{ij} \\ Humidity_{ij} \\ PH_{ij} \\ Rain\_fall_{ij} \end{bmatrix} = \begin{bmatrix} \mu T \\ \mu H \\ \mu PH \\ \mu R \end{bmatrix} + \begin{bmatrix} \tau T_i \\ \tau H_i \\ \tau PH_i \\ \tau R_i \end{bmatrix} + \begin{bmatrix} \varepsilon T_{ij} \\ \varepsilon H_{ij} \\ \varepsilon PH_{ij} \\ \varepsilon R_{ij} \end{bmatrix} \tag{7}$$

The hypotheses tested for each MANOVA model were

$$H_0 : \mu_1 = \mu_2 = ... = \mu_g, \tag{8}$$

indicating that the mean vectors of the response variables are equal across all crop types, against

$$H_1 : \exists i \neq j \text{ such that } \mu_i \neq \mu_j \tag{9}$$

indicating that at least one crop type has a different mean vector.

The test statistic used was Wilks' Lambda, defined as:

$$\Lambda = \frac{|W|}{|T|} = \frac{|W|}{|W+B|} \tag{10}$$

Where: W is the within-group sums of squares and cross-products (SSCP) matrix, B is the between-group SSCP matrix, and T is the total SSCP matrix.

3.4 Pearson Correlation Analysis

Pearson correlation coefficients were computed among the environmental variables (Temperature, Humidity, pH, Rainfall) to assess the strength and direction of linear relationships between them.

3.5 Decision Tree Classification

A Decision Tree (CART) model was fitted using all seven predictors to classify crop type. The model recursively partitions the predictor space using an impurity-reduction criterion, producing interpretable decision rules. Model performance was evaluated using overall classification accuracy and Cohen's Kappa statistic, and predictor importance was ranked according to each variable's contribution to reducing classification error.

The Decision Tree seeks a prediction function

$$Y = f(\boldsymbol{X}) \tag{11}$$

Where $f(\bullet)$ is obtained by recursively partitioning the predictor space into homogeneous subsets.

At each node, the impurity is measured using the Gini Index:

Where: K is the number of crop classes and p_i is the proportion of observations belonging to class i.

The reduction in impurity after splitting is:

$$\Delta G = G_{parent} - \sum_{j=1}^{m} \frac{n_j}{n} G_j \tag{12}$$

where G_{parent} is the impurity before splitting, G_j is the impurity of the j^{th} child node, n_j is the number of observations in child node j and n is the total number of observations in the parent node.

The split with the largest impurity reduction was selected.

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX

DOI: 10.XXXXX/jossda.2026.1.1.007

The model performance was evaluated using:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \text{ X } 100\% \tag{13}$$

where TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative.

After fitting the Decision Tree model, the importance of each predictor variable was assessed based on the total reduction in Gini impurity contributed by that variable throughout the tree.

Let X_j denote the j^{th} predictor variable.

The importance score is defined as

$$VI(X_{(j)}) = \sum_{t \in X_j} \Delta G_t \tag{14}$$

where $VI(X_j)$ is the importance score of Variable X_j and ΔG_t is the decrease in Gini impurity at node t.

4. Results

Secondary data representing fertilizer nutrient application and soil/environmental conditions across 22 crop types were analysed, drawn from the Kaggle repository titled "Crop Recommendation" Dataset (N=2200; 100 observations per crop, no missing values). Descriptive statistics, MANOVA, Pearson correlation, and a Decision Tree model were applied in sequence to characterise the variables, test for group differences, examine inter-variable relationships, and evaluate predictive performance.

4.1 Descriptive Statistics

Table 1 presents the descriptive statistics of the fertilizer nutrients and soil variables. The mean Nitrogen, Phosphorus, and Potassium contents were 50.55, 53.36, and 48.15 kg/ha, respectively. Potassium exhibited the greatest variability (CV = 105.19%), followed by Nitrogen (CV = 73.03%), while Phosphorus showed comparatively lower variability (CV = 61.81%). Among the environmental variables, mean Temperature (25.62°C) and soil pH (6.47) were relatively stable (CV = 19.77% and 11.96%, respectively), with pH the most stable variable overall. Mean Humidity (71.48%) and Rainfall (103.46 mm) showed moderate to substantial variability (CV = 31.15% and 53.12%, respectively). Overall, fertilizer nutrients, particularly potassium, displayed greater variability than most soil condition variables, indicating the dataset captures diverse nutrient and environmental conditions suitable for multivariate analysis and classification.

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX

DOI: 10.XXXXX/jossda.2026.1.1.007

Table 1. Descriptive Statistics of Fertilizer Nutrients and Soil Variables

Variable	n	Mean	SD	CV (%)
Nitrogen (N)	2200	50.55	36.92	73.03
Phosphorus (P)	2200	53.36	32.99	61.81
Potassium (K)	2200	48.15	50.65	105.19
Temperature (°C)	2200	25.62	5.06	19.77
Humidity (%)	2200	71.48	22.27	31.15
Soil pH	2200	6.47	0.77	11.96
Rainfall (mm)	2200	103.46	54.96	53.12

To further examine the distribution and variability of the fertilizer nutrients, boxplots were used to visually present the distribution of Nitrogen (N), Phosphorus (P), and Potassium (K). The boxplots provide information on the median, interquartile range, overall spread, and potential outliers of the nutrient variables, thereby complementing the descriptive statistics presented in Table 1.

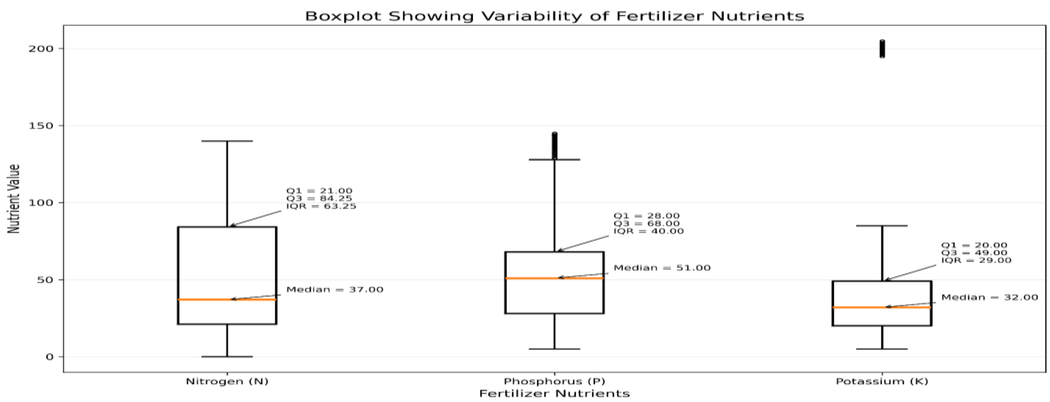

Figure 1. Boxplots showing the distribution and variability of Nitrogen, phosphorus, and potassium.

Figure 1 shows that Nitrogen had the widest interquartile range (IQR = 63.25), indicating greater variability within the middle 50% of observations compared with Phosphorus (IQR = 40.00) and Potassium (IQR = 29.00). The median values were 37.00, 51.00, and 32.00 for Nitrogen, Phosphorus, and Potassium, respectively. Potassium exhibited several high-value outliers, indicating the presence of unusually high observations. These outliers contribute to the relatively high variability observed for Potassium in the descriptive statistics.

4.2 Skewness and Kurtosis

Table 2 shows that Nitrogen, Phosphorus, Potassium, Temperature, pH, and Rainfall were all positively skewed, with Potassium showing the strongest skewness (2.37), suggesting relatively high extreme values. Humidity alone was negatively skewed (−1.09). Potassium also had the highest kurtosis (4.43), indicating a sharply peaked, heavy-tailed distribution, while Nitrogen had negative kurtosis (−1.06), indicating a flatter distribution. Temperature and pH were comparatively closer to symmetric, near-normal distributions.

Table 2. Skewness and Kurtosis of Fertilizer Nutrients and Soil Variables

Variable	N	Skewness	Kurtosis
N	2200	0.51	-1.06
P	2200	1.01	0.85
K	2200	2.37	4.43
Temperature	2200	0.18	1.22
Humidity	2200	-1.09	0.30
pH	2200	0.28	1.64
Rainfall	2200	0.96	0.60

4.3 Multivariate Analysis of Variance (MANOVA)

Table 3 presents the MANOVA results for fertilizer nutrients and soil condition variables across crop types. Fertilizer nutrients (Nitrogen, Phosphorus, and Potassium) differed significantly among crop types (Wilks' Lambda = 0.0000, F(2, 7) = 2796.14 [approx.], p < .001). Soil condition variables (Temperature, Humidity, pH, and Rainfall) likewise exhibited significant joint differences across crop types (Wilks' Lambda = 0.0016, F(4, –) = 419.32, p < .001). These results indicate that both fertilizer nutrients and soil conditions jointly and significantly influence crop classification, implying that different crop types are associated with distinct nutrient requirements and environmental profiles.

Table 3. MANOVA Results for Fertilizer Nutrients and Soil Conditions

Variable Group	Wilks' Λ	F (approx.)	p
Fertilizer nutrients (N, P, K)	0.0000	2796.14	< .001
Soil conditions (Temp., Humidity, pH, Rainfall)	0.0016	419.32	< .001

4.4 Pearson Correlation Among Environmental Variables

Table 4 presents the Pearson correlation coefficients among Temperature, Humidity, pH, and Rainfall. Correlations were generally weak, indicating little or no strong linear relationship among these variables. Temperature showed a weak positive correlation with Humidity (r = 0.205) and very weak negative correlations with pH (r = −0.018) and Rainfall (r = −0.030). Humidity showed a very weak positive relationship with Rainfall (r = 0.094) and a very weak negative relationship with pH (r = −0.008), while pH and Rainfall were weakly negatively correlated (r = −0.109). The weak inter-correlations suggest each variable contributes largely independent information to crop response, supporting their joint inclusion in the classification model.

Table 4. Pearson Correlation Matrix of Environmental Variables

Variable	Temperature	Humidity	pH	Rainfall
Temperature	1.000	0.205	-0.018	-0.030
Humidity	0.205	1.000	-0.008	0.094
pH	-0.018	-0.008	1.000	-0.109
Rainfall	-0.030	0.094	-0.109	1.000

4.5 Classification Performance

Table 5 presents the classification performance of the Decision Tree model. The model achieved an overall accuracy of 89.24%, indicating that approximately 89 of every 100 crop observations were correctly classified. The Kappa statistic of 0.8873 indicates almost perfect agreement between predicted and actual crop types after accounting for chance agreement, demonstrating that fertilizer nutrients and soil variables provide a reliable basis for crop classification.

Table 5: Classification Performance of the Decision Tree Model

Measure	Value
Overall Accuracy	89.24%
Kappa Statistic	0.8873

4.6 Predictor Importance

Table 6 ranks the predictor variables by their contribution to crop classification. Rainfall was the most influential variable (19.87%), followed by Potassium (19.20%) and Humidity (17.58%). Phosphorus (12.83%), Temperature (12.15%), and Nitrogen (10.10%) made moderate contributions, while soil pH contributed the least (8.26%). These results highlight the combined importance of environmental conditions and fertilizer nutrients in determining crop response, with Rainfall, Potassium, and Humidity emerging as the key drivers of crop classification.

Table 6. Contribution of Predictor Variables to Crop Classification

Variable	Importance (%)
Rainfall	19.87
Potassium (K)	19.20
Humidity	17.58
Phosphorus (P)	12.83
Temperature	12.15
Nitrogen (N)	10.10
Soil pH	8.26

4.7 Discussion

This study examined the joint influence of fertilizer nutrients and soil/environmental conditions on crop response using MANOVA, Pearson correlation, and Decision Tree classification. The findings align with previous evidence that fertilizer nutrients and soil conditions vary systematically across crop types, supporting the view that crop-specific nutrient and environmental requirements should guide fertilizer application rather than blanket recommendations (Aliyu *et al.*, 2022; Shehu *et al.*, 2019). The high classification accuracy obtained (89.24%) is also consistent with Shuaibu *et al*. (2024), who reported strong predictive performance for machine learning models using similar soil and climatic predictors, reinforcing the value of non-parametric methods in agricultural classification tasks.

The predictor importance ranking, with Rainfall, Potassium, and Humidity emerging as the strongest drivers of crop classification, corroborates Shahhosseini *et al*. (2014), who similarly found that climatic variables contributed substantially to crop yield prediction alongside soil nutrients. However, soil pH contributed comparatively little in this study, despite its established theoretical role in nutrient availability (Havlin *et al*., 2014; Tisdale *et al*., 2013). This suggests that, within the range of pH values observed in this dataset, other environmental and nutrient factors exerted a stronger influence on crop differentiation than pH variation itself.

From a practical standpoint, the results offer applied value for agricultural extension services and policymakers. By demonstrating that fertilizer nutrients and soil/environmental conditions can jointly and reliably classify crop type, this framework supports a shift toward precision, site-specific fertilizer recommendations rather than generalised ones, potentially reducing both economic waste and environmental harm associated with over-application (Adelekan & Olawale, 2010; Lal, 2009).

Nevertheless, some limitations should be noted. First, the dataset used represents curated, cross-sectional observations rather than field-collected Nigerian agricultural data, which may limit direct generalizability to local farming conditions. Second, the Decision Tree model, while interpretable, may be prone to overfitting if not validated through techniques such as cross-validation or pruning. Third, the analysis did not account for potential interaction effects between nutrients and environmental variables, which could further explain crop-specific responses.

Future research should validate this framework using field-collected data from Nigerian agro-ecological zones, incorporate interaction effects between nutrient and environmental variables, and compare Decision Tree performance against ensemble methods such as Random Forest or Gradient Boosting to assess potential gains in predictive accuracy and robustness.

5. Conclusion

This study demonstrates that fertilizer nutrients (N, P, K) and soil/environmental conditions (temperature, humidity, pH, rainfall) both vary significantly across crop types, as confirmed by MANOVA ($p < .001$ for both models). Descriptive analysis revealed considerable variability in Potassium relative to the comparatively stable soil pH, while correlation analysis showed weak linear associations among environmental variables, suggesting each contributes largely independent information to crop response. The Decision Tree classifier achieved a high overall accuracy of 89.24% (Kappa = 0.8873), confirming that fertilizer nutrients and soil variables provide a reliable basis for crop classification. Rainfall, Potassium, and Humidity emerged as the most influential predictors, while soil pH contributed least.

These findings underscore the value of integrating parametric multivariate testing with non-parametric machine learning for a more complete understanding of crop-environment relationships. For agricultural policy and practice, the results support a shift away from blanket fertilizer recommendations toward site-specific, data-driven strategies that account for the differential importance of rainfall and potassium management. Future research should extend this framework using longitudinal or field-experimental data and compare Decision Tree performance against ensemble methods such as Random Forest or Gradient Boosting.

Journal of Statistical Sciences & Data Analytics (JOSSDA)
Volume 1, Issue 2, December 2026 | Print ISSN XXXX-XXXX | Online ISSN XXXX-XXXX

DOI: 10.XXXXX/jossda.2026.1.1.007

References

Adelekan, B. A., & Olawale, S. O. (2010). Development of localised crop response models to fertilizer application. *Journal of Agricultural Science and Technology*, *12*(3), 45–56.

Aliyu, K. T., Kamara, A. Y., Huising, E. J., Jibrin, J. M., Shehu, B. M., Rurinda, J., Adam, A. M., Mohammed, I. B., & Vanlauwe, B. (2022). Maize nutrient yield response and requirement in the maize belt of Nigeria. Environmental Research Letters, 17(6), 064025. https://doi.org/10.1088/1748-9326/ac5bb1

Aliyu, K. T., Peter-Jerome, H., Adewopo, J., Kamara, A., & Dawaki, M. U. (2022). Assessing the spatial variability of soil properties to delineate nutrient management zones in smallholder maize-based systems in Nigeria. *Applied and Environmental Soil Science*, *2022*, 1–14. https://doi.org/10.1155/2022/5111635

Brady, N. C., & Weil, R. R. (2016). *The nature and properties of soils* (15th ed.). Pearson.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Wadsworth International Group.

Draper, N. R., & Smith, H. (1998). Applied Regression Analysis (3rd ed.). Wiley.

Food and Agriculture Organisation. (2020). *World food and agriculture: Statistical yearbook 2020*. FAO.

Gomez, K. A., & Gomez, A. A. (1984). *Statistical procedures for agricultural research* (2nd ed.). John Wiley & Sons.

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

Havlin, J. L., Tisdale, S. L., Nelson, W. L., & Beaton, J. D. (2014). *Soil fertility and fertilizers: An introduction to nutrient management* (8th ed.). Pearson.

Johnson, R. A., & Wichern, D. W. (2019). *Applied multivariate statistical analysis* (6th ed.). Pearson.

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

Mengel, K., & Kirkby, E. A. (2001). *Principles of plant nutrition* (5th ed.). Springer.

Montgomery, D. C. (2019). Design and Analysis of Experiments (10th ed.). Wiley.

Rencher, A. C., & Christensen, W. F. (2012). *Methods of multivariate analysis* (3rd ed.). John Wiley & Sons.

Shahhosseini, M., Martinez-Feria, R. A., Harrison, M., & Archontoulis, S. V. (2014). Using machine learning algorithms to improve crop yield prediction in complex agricultural systems. *Agronomy Journal*, *106*(6), 2111–2121.

Shehu, B. M., Lawley, M. A., & Merckx, R. (2019). Nutrient use efficiency and economic optimisation of fertilizer options for maize in northern Nigeria. *Agronomy for Sustainable Development*, *39*(4), 37–49. https://doi.org/10.1007/s13593-019-0582-4

Shehu, B. M., Lawan, B. A., Jibrin, J. M., Kamara, A. Y., Mohammed, I. B., Rurinda, J., Zingore, S., Craufurd, P. Q., Vanlauwe, B., Adam, A. M., & Merckx, R. (2019). Balanced nutrient requirements for maize in the Northern Nigerian Savanna: Parameterisation and validation of QUEFTS model. Field Crops Research, 241, 107585. https://doi.org/10.1016/j.fcr.2019.107585

Tisdale, S. L., Nelson, W. L., & Beaton, J. D. (2013). Soil Fertility and Fertilizers. Pearson Education

Troeh, F. R., & Thompson, L. M. (2005). Soils and Soil Fertility (6th ed.). Blackwell Pub.