

Article

Analysis of morphological characteristics of six varieties of cassava using discriminant analytical approach

Samuel U. Enogwe^{1,*}, Chinelo U. Chikwelu², Goodluck O. Merua³, and Ihuoma L. Onyeugo¹

¹ Department of Statistics, Michael Okpara University of Agriculture, Umudike, Nigeria

² Department of Statistics, Nnamdi Azikiwe University, Awka, Nigeria

³ Department of Statistics, Ignatius Ajuru University of Education, Rumuolumeni, Port Harcourt, Nigeria

* Correspondence: enogwe.samuel@mouau.edu.ng/senogwe@yahoo.com

Abstract: Cassava (*Manihot esculenta* Crantz) is a critical staple crop in sub-Saharan Africa, but its varietal identification is often challenged by phenotypic plasticity. This study employed Multiple Discriminant Analysis (MDA) to evaluate the efficacy of six morphological traits in distinguishing six cassava varieties obtained from the National Root Crops Research Institute (NRCRI), Nigeria. The traits measured were leaf length, leaf width, leaf area, leaf index, plant height, and mean angle. Data from 96 plants were subjected to descriptive statistics, multivariate analysis of variance (MANOVA), and canonical discriminant analysis. While leaf area and mean angle showed high variability, MANOVA indicated no significant difference in the group mean vectors, confirming substantial phenotypic overlap. The most discriminatory traits were leaf area, leaf index, and mean angle. Overall classification accuracy was low (45.8%), with an apparent error rate of 54.2%. Only two varieties, TMS98212 (75%) and NR100235 (56.7%), were classified with moderate accuracy; the remaining four were entirely misclassified. Canonical plots visually confirmed the distinct clustering of TMS98212 and NR100235 against the extensive overlap of the other varieties. We conclude that morphological traits alone are insufficient for reliable cassava varietal identification and recommend integrating molecular techniques with advanced statistical models for a more robust classification system.

Keywords: cassava, morphological traits, multiple discriminant analysis, varietal classification, Nigeria, phenotypic overlap.

Received: 22 September 2025; Revised: 24 December 2025; Accepted: 18 January 2026; Published: 3 March 2026



Copyright: ©2026 the Author(s). Published by JSSCI. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0).

Journal Abbreviation: J. Stat. Sci. Comput. Intell.

1. Introduction

Cassava (*Manihot esculenta* Crantz) stands as a fundamental pillar of food security and economic stability across sub-Saharan Africa, serving as a primary dietary staple for over 800 million people and a crucial source of income for smallholder farmers. In Nigeria, the world's largest producer, cassava's role extends beyond direct consumption to a rapidly expanding industrial sector, where it is processed into starch, high-quality flour, sweeteners, and bioethanol, thereby driving rural economies and national agricultural policy. The crop's

renowned resilience—thriving in marginal soils, tolerating intermittent drought, and yielding reliably under suboptimal conditions—cements its status as an indispensable crop for climate adaptation strategies. Despite this profound significance, the entire cassava value chain is persistently hampered by a critical, unresolved issue: the accurate and reliable identification of distinct varieties. Farmers, extension agents, and even breeders frequently rely on visual morphological characteristics such as leaf shape, stem coloration, and plant architecture.

However, these traits exhibit high phenotypic plasticity, meaning they are profoundly influenced by environmental factors like soil fertility, rainfall patterns, and pest pressure, leading to widespread misidentification. This problem has tangible consequences, including the erroneous distribution of planting materials, contamination of breeder seed stocks, inaccurate yield assessments, and the hampered adoption of improved, disease-resistant varieties, ultimately constraining crop productivity and genetic conservation efforts. Morphological characterization, due to its accessibility and low cost, remains the most widely used method for varietal distinction in resource-limited settings. Yet, its limitations are well-documented in the scientific literature. For example, Nwakuya et al. (2022) applied Multiple Discriminant Analysis (MDA) to five cassava varieties and reported a high classification accuracy, but notably identified a different set of key discriminatory traits than other studies, highlighting the context-specific nature of morphological expression. Similarly, Kouakou et al. (2023) found that branching architecture was the primary differentiating trait among hundreds of accessions in Côte d'Ivoire, further underscoring this variability. Crucially, seminal works by Asare et al. (2011) in Ghana and Derso and Mohamud (2018) in Ethiopia, which combined morphological descriptors with molecular markers, converged on a singular conclusion: while morphological traits offer a valuable first-level screening tool, they are inherently insufficient for conclusive identification and must be complemented by molecular genetic analysis to achieve precision. This collective body of evidence illuminates a critical research gap: a comprehensive, quantitative benchmark is urgently needed to define the precise limitations and intrinsic discriminatory power of a standardized set of quantitative morphological traits for a nationally relevant cohort of Nigerian cassava varieties. It is this specific gap that the present study seeks to fill.

This research is designed to apply Multiple Discriminant Analysis (MDA)—a robust multivariate statistical technique selected for its specific proficiency in maximizing separation between pre-defined groups and identifying the linear combinations of variables that best differentiate them—to six prominent cassava varieties obtained from the National Root Crops Research Institute (NRCRI). The scope of this work is a rigorous evaluation of six key quantitative morphological traits: leaf length, leaf width, leaf area, leaf index, plant height, and mean angle. We aim to solve the specific problems of the undefined empirical rate of varietal misclassification and the unclear statistical hierarchy of trait importance for these Nigerian varieties. The significance of this research is its potential to provide national breeding programs, seed certification agencies, and extension services with an evidence-based, statistically validated framework for preliminary varietal sorting and management. The novel contribution to knowledge of this study is its definitive quantification of the extensive phenotypic overlap among these commercially significant varieties, precisely identifying leaf area, leaf index, and mean angle as the most influential—yet ultimately inadequate—traits for reliable standalone classification. By doing so, it provides an empirical foundation that strongly justifies the necessary integration of molecular tools with advanced statistical models to develop a robust and reliable identification system for the Nigerian cassava sector.

2. Methodology

2.1. Source of Data

The study was conducted using primary data collected from the National Root Crops Research Institute (NRCRI), Umudike, Nigeria. A total of 100 cassava plants were sampled across ten labeled varieties. However, four of these—NR8082, AR9-46, TMEB419, and NR100325—appeared only once and were excluded from discriminant analysis since robust covariance estimation requires multiple samples per group. The final analysis focused on six main varieties: NR100235 ($n = 30$), TMS98212 ($n = 36$), B515 ($n = 12$), COB61 ($n = 7$), IITA-TMS-IBA980581 ($n = 6$), and NWAGERI ($n = 5$), resulting in a total sample size (N) of 96 plants. Morphological traits measured included leaf width (cm), leaf length (cm), leaf area (cm²), leaf index (ratio of length to width), plant height (cm), and mean angle (degrees between petiole and stem).

2.2. Morphological Predictor Variables

Six quantitative morphological traits were measured for each plant. The selection of these particular traits is supported by their established use in the discrimination of cassava and other crop germplasm. For instance, Ezekwe (2017) successfully used leaf length, leaf width, and petiole length in a discriminant analysis of fluted pumpkin morphotypes, while Githiri et al. (2006) highlighted the importance of leaf morphology in characterizing cassava genotypes. The traits measured were:

X_1 : Leaf Length (L , in cm)

X_2 : Leaf Width (W , in cm)

X_3 : Leaf Area (A , in cm²), calculated as $A = L \times W \times 0.75$

X_4 : Leaf Index (LI), calculated as $LI = \frac{L}{W}$

X_5 : Plant Height (H , in cm)

X_6 : Mean Angle (θ , in degrees between the petiole and the stem)

2.3. Multiple Discriminant Analysis (MDA) Framework

Multiple Discriminant Analysis is a multivariate technique used to describe the relationship between a single categorical dependent variable (cassava variety) and multiple continuous predictor variables (morphological traits). This method has a strong precedent in agricultural research. Montalvo et al. (2014) effectively used discriminant analysis to classify cocoa clones in Peru based on pod and bean characteristics, demonstrating its utility for perennial crops. Similarly, Olayiwola et al. (2022) applied the technique to differentiate sesame accessions using morphological descriptors. The analysis in this study proceeds in two main stages, following this established framework.

2.3.1. Stage 1: Deriving Discriminant Functions

The goal of this stage is to find linear combinations of the predictor variables that maximize the separation between the six cassava varieties. This is achieved by solving the eigenvalue problem, a standard approach in

multivariate statistics:

$$(W^{-1}B - \lambda I)v = 0 \quad (1)$$

where

B is the between-groups sum of squares and cross-products (SSCP) matrix.

W is the within-groups SSCP matrix.

λ is an eigenvalue, representing the ratio of between-group to within-group variance for a discriminant function.

v is the corresponding eigenvector of coefficients.

The number of significant discriminant functions extracted (ss) is the smaller of p (number of variables) and $g - 1$ (number of groups minus one). For this study with $p = 6$ and $g = 6$, up to $s = 5$ functions could be derived. The relative importance of each function is indicated by its eigenvalue and the proportion of variance it explains.

The canonical discriminant function for dimension k is:

$$Y_k = v_{k1}X_1 + v_{k2}X_2 + \dots + v_{kp}X_p \quad (2)$$

where Y_k is the score on the k -th discriminant function for a given plant. To aid interpretation, standardized discriminant coefficients were computed by multiplying the raw coefficients by the pooled within-group standard deviation of the corresponding variable. This allows for the assessment of the relative contribution of each morphological trait.

2.3.2. Stage 2: Developing the Classification Rule

The second stage of MDA involves classifying the cassava plants into one of the six varieties based on their morphological measurements. This was done using a Bayesian classification procedure that minimizes the total probability of misclassification. The linear classification function for variety i is given by:

$$CF_i(\mathbf{x}) = c_{i0} + c_{i1}X_1 + c_{i2}X_2 + \dots + c_{ip}X_p \quad (3)$$

where the coefficients c_{ij} are calculated as:

$$c_{ij} = (N - g) \sum_{k=1}^p \bar{x}_{ik} s^{jk} \quad (4)$$

and the constant c_{i0} is:

$$c_{i0} = -\frac{1}{2} \sum_{j=1}^p c_{ij} \bar{x}_{ik} \quad (5)$$

where

$\bar{x}_{ik}$ is the mean of variable k in group i

s^{jk} are the elements of the inverse of the pooled within-group covariance matrix

Prior probabilities were incorporated using the group proportions $p_i = n_i / N$.

Classification Rule

For a plant with morphological measurement vector $\mathbf{x}$, we compute $CF_i(\mathbf{x})$ for each variety $i = 1, 2, 3, 4, 5, 6$. The plant is classified into the variety k that yields the highest classification score:

$$\text{Assign to variety } k \text{ where } k = \arg \max_i [CF_i(\mathbf{x})] \quad (6)$$

2.4. Evaluation of Discriminatory Power and Classification Accuracy

2.4.1. Multivariate Tests

Before proceeding with MDA, we tested the fundamental assumption of equal group mean vectors using Multivariate Analysis of Variance (MANOVA) with Wilks' Lambda, a standard preliminary test is given by:

$$\Lambda = \frac{|W|}{|B + W|} \quad (7)$$

The parameters of the Wilks' Λ distribution are p = number of variables, v_B = degrees of freedom for hypothesis and v_W = degrees of freedom for error. The Likelihood ratio test rejects H_0 if $\Lambda > \Lambda_{\alpha, p, v_B, v_W}$.

2.4.2. Classification Matrix and Error Rates

The performance of the MDA classification rule was evaluated using a classification (confusion) matrix. Let n_{ij} represent the number of plants from actual variety j that were classified into predicted variety i .

The apparent error rate (APER) was calculated as:

$$APER = \frac{\text{Total misclassified plants}}{N} = \frac{N - \sum_{i=1}^6 n_{ii}}{96} \times 100 \quad (8)$$

The overall classification accuracy is simply:

$$\text{Accuracy} = 100\% - APER \quad (9)$$

2.4.3. Mahalanobis Distance

To quantify the separation between varieties, we computed the squared Mahalanobis distance between variety i and variety j , a robust measure of multivariate distance commonly used in such analyses:

$$D^2(i, j) = (\bar{x}_i - \bar{x}_j)^T S_{pooled}^{-1} (\bar{x}_i - \bar{x}_j) \quad (10)$$

All analyses were performed using R statistical software.

3. Results and Discussion

The results of the Multiple Discriminant Analysis (MDA) are presented sequentially, from the initial

description of the data to the final classification performance, with each output directly corresponding to the methodological steps outlined.

3.1. Descriptive Statistics and Trait Variability

Prior to multivariate analysis, the distribution of each morphological trait was examined. Table 1 presents the descriptive statistics for the six traits across all 96 samples. Leaf Area exhibited the highest variability (Standard Deviation = 35.91, range: 72.71–273.20 cm²), indicating substantial differences in leaf size across the sampled plants. Plant Height also showed considerable variation (SD = 44.26 cm). In contrast, Leaf Width and Length were more consistent. This high variability in Leaf Area and Plant Height suggested their potential, yet uncertain, role in discriminating between varieties, as they could be influenced by both genetic and environmental factors.

Table 1: Descriptive Statistics of Morphological Traits

Trait	Mean	Std. Dev.	Min	Max
Leaf Width	19.32	2.37	13.50	23.70
Leaf Length	13.07	1.74	8.62	17.33
Leaf Area	127.50	35.91	72.71	273.20
Leaf Index	2.43	1.19	1.03	7.42
Plant Height	146.07	44.26	54.00	251.00
Mean Angle	14.18	4.35	6.20	24.60

3.2. Inter-Varietal Mean Comparisons

Table 2 displays the mean values for each trait by variety, allowing for a preliminary visual assessment of differences. For instance, TMS98212 had the highest mean Leaf Area (133.83 cm²) and Plant Height (161.50 cm), while NWAGERI had the lowest mean Leaf Area (116.21 cm²) but the highest Mean Angle (17.5°). COB61 showed a notably lower mean Plant Height (114.86 cm). These differences, however, are univariate and do not account for the correlation between traits or the within-group variance, which is why multivariate tests are essential.

Table 2: Mean Morphological Traits by Variety

Variety	Leaf Width	Leaf Length	Leaf Area	Leaf Index	Plant Height	Mean Angle
NR100235	19.68	13.23	131.49	2.58	140.27	13.59
TMS98212	19.33	13.17	133.83	2.52	161.50	14.97
B515	18.46	12.51	126.55	2.05	160.50	14.93
COB61	19.14	12.90	121.06	2.87	114.86	12.34
IITA-	19.10	13.05	118.99	2.18	160.17	14.18
IBA980581						
NWAGERI	18.61	12.62	116.21	2.06	150.00	17.50

3.3. Multivariate Analysis of Variance (MANOVA)

The MANOVA test, using Wilks' Lambda, was conducted to test the null hypothesis that the population mean vectors of the six morphological traits are equal across the six cassava varieties. The result was not statistically significant ($p > 0.05$). This indicates a failure to reject the null hypothesis, providing statistical

evidence that there are no significant differences in the overall morphological profiles of the six varieties when all traits are considered simultaneously. This foundational result underscores the high degree of phenotypic overlap and sets the stage for the challenging classification task.

3.4. Multiple Discriminant Analysis and Classification

Despite the non-significant MANOVA, MDA was performed to explore the maximal separation possible. Table 3 presents the structure matrix, which shows the pooled within-groups correlations between the discriminating variables and the standardized canonical discriminant functions. These correlations are the most valid measure for assessing the relative contribution of each variable to the discriminant functions. The results identify Leaf Area and Leaf Index as having the highest correlations with the first function, confirming their role as the primary discriminators. Mean Angle loaded strongly on a subsequent function, highlighting its unique role in separating varieties like NWAGERI.

Table 3: Structure Matrix: Pooled Within-Groups Correlations

Trait	Function 1	Function 2
Leaf Area	0.756	0.102
Leaf Index	0.682	-0.215
Plant Height	0.451	-0.389
Leaf Length	0.350	0.118
Leaf Width	0.301	0.055
Mean Angle	0.150	0.894

The ultimate test of the MDA model is its classification accuracy. Table 4, the confusion matrix, shows the classification results based on the derived discriminant functions. The overall classification accuracy was 45.8%, meaning the APER was 54.2%. TMS98212 and NR100235 were classified with 75% and 56.7% accuracy, respectively. Critically, the four other varieties (B515, COB61, IITA-IBA980581, and NWAGERI) had a 0% correct classification rate, with all their members misclassified into either TMS98212 or NR100235. This indicates that the morphological profiles of these four varieties are not distinct enough to be differentiated from the two dominant groups.

Table 4: Confusion Matrix (Rows = Actual, Columns = Predicted)

Actual	Predicted					
	NR100235	TMS98212	B515	COB61	IITA- IBA980581	NWAGERI
NR100235	17	13	0	0	0	0
TMS98212	9	27	0	0	0	0
B515	5	7	0	0	0	0
COB61	5	2	0	0	0	0
IITA- IBA980581	2	4	0	0	0	0
NWAGERI	1	4	0	0	0	0

4. Discussion of Results

The results demonstrated that morphological traits such as leaf area, leaf index, and mean angle showed the greatest variability across cassava varieties, suggesting their potential as discriminating factors. However, the MANOVA results indicated no statistically significant differences in group mean vectors, reflecting the strong influence of phenotypic plasticity. This means that while traits varied visually, environmental effects masked true genetic differences.

The discriminant analysis achieved a modest classification accuracy of 45.8% (APER = 54.2%). Varieties TMS98212 (75%) and NR100235 (56.7%) were the most accurately classified, while B515, COB61, IITA-IBA980581, and NWAGERI were completely misclassified. Canonical plots confirmed these findings, showing distinct clustering for TMS98212 and NR100235, whereas other varieties overlapped heavily. The high misclassification rates, particularly among varieties with small sample sizes, underscore both data imbalance and genuine morphological similarity.

From a breeding perspective, the results highlight the limitations of relying solely on morphological descriptors for cassava varietal identification. Traits like plant height and leaf area, though visually distinct, are highly influenced by environmental conditions, making them unreliable as standalone classifiers. The prominence of leaf index and mean angle in discriminant functions suggests they could serve as quick field indicators, particularly for distinguishing varieties such as NWAGERI. Nevertheless, robust varietal identification requires integration with molecular markers such as SSRs and SNPs, which capture genetic diversity more effectively. Combining morphological and molecular approaches, supported by advanced statistical and machine learning models, will enhance cassava breeding, conservation, and extension services.

5. Conclusion and Recommendations

This study evaluated the effectiveness of morphological traits in cassava varietal classification using Multiple Discriminant Analysis. While some separation was observed—particularly between TMS98212 and NR100235—the overall classification accuracy was low (45.8%, APER = 54.2%), reflecting extensive phenotypic overlap. Leaf index, leaf area, and mean angle emerged as the most informative discriminators, but their predictive power was insufficient for consistent varietal identification.

To strengthen cassava classification and breeding efforts, the following are recommended:

- (i) Increase sample sizes across multiple planting cycles to reduce bias and improve model robustness.
- (ii) Integrate morphological descriptors with molecular markers (e.g., SSR, SNP) for reliable identification.
- (iii) Apply advanced statistical and machine learning approaches such as random forests or support vector machines to enhance predictive accuracy.
- (iv) Develop practical, farmer-oriented field guides highlighting leaf index and petiole angle as quick reference traits.

Overall, morphological traits remain valuable as low-cost, field-level tools, but their limitations necessitate complementary molecular and computational approaches for effective varietal identification.

Author contributions: Conceptualization, S.U.E. and I.L.U.; Methodology, S.U.E. and I.L. O.; Software, S.U.E., C.U.C. and I.L. O.; Validation, S.U.E. and I.L.O.; Formal Analysis, S.U.E., G.O.M., C.U.C. and I.L.O.; Investigation, S.U.E., G.O.M., C.U.C. and I.L.O.; Resources, S.U.E; Data Curation, I.L.O.; Writing—original draft preparation, I.L.O.; Writing—review and editing, S.U.E, and C.U.C.; Visualization, S.U.E.; Supervision, S.U.E.; Project Administration, I.L.O.; Funding Acquisition, None. All authors have read and agreed to the published version of the manuscript.

Funding Statement: This research received no external funding.

Data Availability: Not applicable.

Acknowledgments: The authors thank the anonymous reviewers for their knowledgeable suggestions, which improved the quality of this paper.

Conflict of interest: The authors declare no conflict of interest.

References

- [1] Asare, P. A., Galyuon, I. K. A., Sarfo, J. K., & Tetteh, J. P. (2011). Morphological and molecular diversity studies of cassava germplasm in Ghana. *African Journal of Biotechnology*, 10(63), 13900–13908.
- [2] Derso, E., & Mohamud, S. (2018). Morphological characterization of cassava varieties in Ethiopia. *International Journal of Plant Research*, 8(2), 45–53.
- [3] Ezekwe, A. G. (2017). Application of discriminant analysis in the morphometric Githiri, S. M., et al. (2006). Morphological variation of cassava clones in the coastal region of Kenya. *Journal of Plant Breeding and Crop Science*, 1(2), 21-27.
- [4] FAO. (2013). *Cassava: A vital food security crop in Africa*. Food and Agriculture Organization of the United Nations, Rome.
- [5] Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis* (6th ed.). Pearson Prentice Hall.
- [6] Kouakou, D. A., Kouadio, Y. J., Kouassi, K. E., Koffi, K. F., & Akaffou, D. S. (2023). Agro-morphological variability of cassava in Côte d'Ivoire. *Journal of Applied Biosciences*, 181, 18962–18973.
- [7] Montalvo, C., et al. (2014). Multivariate analysis of morphological characteristics of Peruvian cocoa beans (*Theobroma cacao* L.). *Scientia Agropecuaria*, 5(3), 123-135.
- [8] Nwakuya, M. T., Biu, E. O., & Merua, G. O. (2022). Multiple Discriminant Analysis of five cassava varieties. *International Journal of Innovative Mathematics, Statistics & Energy Policies*, 10(4), 71–83.
- [9] Olayiwola, M. O., et al. (2022). Morphological characterization and discriminant analysis of sesame (*Sesamum indicum* L.) accessions. *Journal of Agricultural Science and Practice*, 7(2), 45-53.
- [10] Zhang, D., Liu, X., & Yang, C. (2019). Genetic diversity of cassava germplasm revealed by morphological traits and SSR markers. *PLOS ONE*, 14(3), e0212875.



Disclaimer/Publisher's Note: The views, opinions, and content expressed in all articles are solely those of the respective author(s) and contributor(s) and do not necessarily reflect those of the JSSCI, its editors, or the publisher. JSSCI and its editorial team assume no responsibility for any harm or damage resulting from the use of information, methods, or products mentioned in the publication.