



International Journal of Research in Agronomy

E-ISSN: 2618-0618
P-ISSN: 2618-060X
© Agronomy
NAAS Rating (2026): 5.20
www.agronomyjournals.com
2026; 9(8): 946-953
Received: 05-06-2026
Accepted: 09-07-2026

Gobikashri N
Department of Agronomy, Pandit
Jawaharlal Nehru College of Agriculture
and Research Institute (PAJANCOA &
RI), Karaikal, Puducherry, India

Mohan R
Department of Agronomy, Pandit
Jawaharlal Nehru College of Agriculture
and Research Institute (PAJANCOA &
RI), Karaikal, Puducherry, India

Poonguzhalan R
Department of Agronomy, Pandit
Jawaharlal Nehru College of Agriculture
and Research Institute (PAJANCOA &
RI), Karaikal, Puducherry, India

Aruna L
Department of Soil Science and
Agricultural Chemistry, Pandit
Jawaharlal Nehru College of Agriculture
and Research Institute (PAJANCOA &
RI), Karaikal, Puducherry, India

Nadarajan S
Department of Plant Breeding and
Genetics, Pandit Jawaharlal Nehru
College of Agriculture and Research
Institute (PAJANCOA & RI), Karaikal,
Puducherry, India

Thirumeninathan S
Department of Agronomy, Pandit
Jawaharlal Nehru College of Agriculture
and Research Institute (PAJANCOA &
RI), Karaikal, Puducherry, India

Nivetha C
Department of Agronomy, Pandit
Jawaharlal Nehru College of Agriculture
and Research Institute (PAJANCOA &
RI), Karaikal, Puducherry, India

Corresponding Author:
Gobikashri N
Department of Agronomy, Pandit
Jawaharlal Nehru College of Agriculture
and Research Institute (PAJANCOA &
RI), Karaikal, Puducherry, India

Identification of high performing rice cultivars under coastal delta conditions through Principal Component and Cluster Analysis

Gobikashri N, Mohan R, poonguzhalan R, Aruna L, Nadaradjan S, Thirumeninathan S and Nivetha C

DOI: <https://www.doi.org/10.33545/2618060X.2026.v9.i8l.6358>

Abstract

Aim: To compare the growth, yield attributes and performance of rice cultivars and determine superior cultivars for sustainable rice production in coastal delta environment by multivariate approach (PCA & Cluster analysis).

Methodology: The field study was carried out at Kharif 2025 at PAJANCOA & RI, East farm, Karaikal, India. The seven rice varieties (CO 51, ADT 46, KKL (R) 2, IR 20, BPT 5204, Karuppukavuni, CR 1009) were tested under Randomized Block Design with three replications. The growth and yield traits used were plant height, number of tillers, Leaf Area Index (LAI), number of productive tillers, dry matter production, panicle traits, test weight, grain production, straw production, crop duration, and were analyzed with Principal Component Analysis (PCA), hierarchical clustering, and K-means clustering.

Results: Significant differences were found between cultivars for growth and yield parameters. The first three principal components accounted for 92.68% of the total variation, LAI, panicle weight, dry matter production, grain yield, straw yield and crop duration were the most important for cultivar discrimination. The three most distinct cultivars were assigned to three clusters by cluster analysis and the elbow method determined that three was the optimum number of clusters. The higher grain yield (6298 kg ha⁻¹), dry matter production (15944 kg ha⁻¹), maximum LAI at flowering (7.27) and crop duration (149 days) were observed in CR 1009. KKL (R) 2 and ADT 46 also had desirable growth and yield characteristics.

Interpretation: Multiple agronomic traits were combined and PCA and cluster analysis were effective in identifying promising rice cultivars. CR 1009 has potential for production under coastal delta conditions and is a useful cultivar for sustainable rice production and improvement programmes.

Keywords: Rice cultivars, principal component analysis, cluster analysis, multivariate analysis, coastal delta, yield traits

Introduction

Stability of rice cultivars to the changing climatic conditions has become a key issue for sustainable crop production. CSA promotes the use of cultivars that are both resilient to environmental pressures and efficient in resource utilization, and are capable of sustaining high productivity (FAO, 2013). The ability to identify such cultivars involves assessing them across multiple agronomic traits and is beyond grain yield.

There is substantial variability in cultivar traits related to growth duration, biomass production, canopy development and yield between rice cultivars, leading to a range in their ability to adapt to specific agro-ecological conditions. Many agronomic characters are interrelated and the evaluation of these traits separately may not be sufficient to indicate the overall performance or the relationships between cultivars. With this in mind, multivariate statistical methods can be used to integrate more than one of these traits to assess cultivar performance (Hair *et al.*, 2019) [5].

Principal Component Analysis (PCA) is a multivariate powerful technique that allows the conversion of correlated variables into a reduced dimension of principal components that captures most of the variance (Yali and Nepir, 2026) [17]. This allows identification of specific traits that vary between cultivars and allows for the correct visualisation of genotype-trait relationships (Jolliffe, 2002) [7]. Hierarchical Cluster Analysis (HCA) is used to categorize

cultivars into groups of cultivars that are similar in overall distance from one another (Johnson & Wichern, 2007) [6], thus providing support for identification of superior cultivars for breeding and cultivation.

Multivariate statistical techniques have been widely applied in rice improvement programmes to characterize genetic diversity, classify cultivars and identify promising genotypes for sustainable production under diverse environmental conditions (Nachimuthu *et al.*, 2014; Hair *et al.*, 2019) [5, 11]. Such analyses provide valuable information for cultivar selection by integrating multiple growth and yield traits into a single analytical framework. Therefore, the present study employed Principal Component Analysis and Hierarchical Cluster Analysis to evaluate seven rice cultivars grown under the coastal agro-ecosystem of Karaikal.

Materials and methods

Seven rice cultivars, namely CO 51, ADT 46, KKL (R) 2, IR 20, BPT 5204, Karuppukavuni, and CR 1009, representing short, medium and long-duration groups, were evaluated under field conditions at PAJANCOA & RI, Karaikal, India, during Kharif 2025. The experiment was conducted in a randomized block design with three replications following the recommended package of practices (TNAU & DOA, 2020).

Multivariate statistical analysis was performed using R software (version 4.x.x) to evaluate the performance of seven rice cultivars based on eleven growth and yield traits. The dataset comprised plant height at harvest (PH @H), number of tillers at harvest (NT @H), leaf area index at flowering (LAI @F), number of productive tillers at harvest (NPT @H), dry matter production at harvest (DMP @H), panicle length (LP), panicle weight (WP), test weight (TW), grain yield (GY), straw yield (SY) and days to harvest (DAYH). Prior to analysis, all variables were standardized to zero mean and unit variance to eliminate differences in measurement scales. Principal Component Analysis (PCA) was performed using the FactoMineR package (Le *et al.*, 2008) [9], and the principal components were retained based on the Kaiser criterion (eigenvalue > 1). The amount of contribution of each variable to the principal components were examined using loading values and contribution percentages. The relationships between cultivars, agronomic traits were graphically

displayed using scree plots, variable correlation plots, treatment score plots and PCA biplots generated using the factoextra package (Kassambara & Mundt, 2020) [8]. Standardized trait data was used for Hierarchical Cluster Analysis (HCA). The dissimilarity matrix was generated according to the Euclidean distance and clusters were generated by Ward's minimum variance (Ward.D2) method (Ward, 1963; Teague and Awosile, 2026) [15, 16]. The dendrogram was used to classify the rice cultivars based on their growth and yield attributes.

The result of HCA was validated using K-means clustering. The within cluster sum of squares (WSS) elbow method was used to find the optimum number of clusters, and the cultivars were clustered in three clusters by using 25 random initializations (nstart = 25). Agronomic traits were clustered and the cluster-wise means were calculated for interpretation. The FactoMineR, factoextra, cluster and tidyverse packages were used to create all the graphical visualization and multivariate analyses in R.

Results and discussion

Principal Component Analysis

The eleven agronomic traits used here to measure the variance among the seven cultivars of rice were well captured by PCA. The first three principal components had eigenvalues exceeding one (PC1 = 4.79, PC2 = 4.11 and PC3 = 1.29) and accounted for 92.68% of the total variation between cultivars. PC1 explained 43.59% of the total variance, PC2 explained 37.33% and PC3 explained 11.76%, showing that these components captured a significant portion of the variance in the data. The contribution of all other principal components was less than 6% and so contributed very little to the cultivar discrimination. The sum of variances above the 90% threshold shows that the selected principal components were able to explain an adequate amount of the variation between rice cultivars, which is necessary for good interpretation of multivariate analysis (Abdi & Williams, 2010; Rencher & Christensen, 2012; Azimov *et al.*, 2026) [1, 2, 13] (Table 1; Figure 1). The scree plot also showed that the majority of information was contained within the first three principal components, with the eigenvalues dropping quickly after PC3, in line with the selection of these components for further interpretation (Figure 1).

Table 1: Eigenvalues, percentage of variance explained and cumulative variance of the principal components derived from growth and yield traits of rice cultivars

Principal Component	Eigenvalue	Variance Explained (%)	Cumulative Variance (%)
PC1	4.795	43.59	43.59
PC2	4.106	37.33	80.92
PC3	1.293	11.76	92.68
PC4	0.598	5.43	98.11
PC5	0.195	1.77	99.88
PC6	0.013	0.12	100.00

Contribution of Agronomic Traits to Principal Components

The loading matrix showed that some agronomic traits were highly involved in the separation of cultivars. Number of tillers at harvest had a strong negative correlation (−0.866) with PC1, while plant height at harvest (0.509), leaf area index at flowering (0.852), panicle weight (0.841), days to harvest (0.909), dry matter production (0.588), grain yield (0.589) and test weight (0.600) had positive correlations. Biomass accumulation, as reflected in dry matter production, grain yield and straw yield, was one of the most important factors that accounted for the observed variability and was an important component in differentiating the performance of cultivars (Table 2).

Days to harvest (17.22%), leaf area index (15.14%), number of

tillers (15.64%) and panicle weight (14.75%) were important contributors to PC1 and straw yield (20.48%), plant height (16.78%), dry matter production (14.60%), grain yield (14.87%) and test weight (13.80%) were important contributors to PC2 in the contribution analysis. The results suggest that vegetative growth, biomass accumulation and crop duration are important factors determining yield formation under coastal conditions and thus influence each other. Similar observations have been reported in multivariate analyses of crop plants, where principal component analysis effectively identifies the key traits contributing to phenotypic variation (Mohammadi & Prasanna, 2003; Yan & Kang, 2003) (Table 2; Figures 2, 5 and 6) [10, 18].

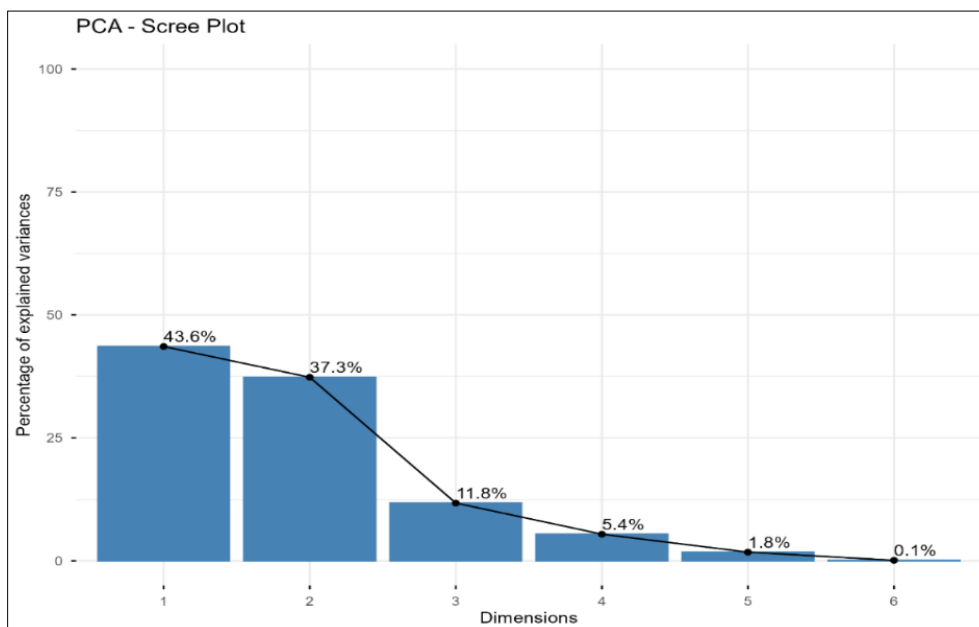


Fig 1: Eigenvalues and percentage of variance for the principal components are shown in a scree plot

Table 2: Principal component loadings for growth and yield traits of first three PCs

Trait	PC1	PC2	PC3
Plant height at harvest (PH @H)	0.509	-0.830	0.199
Number of tillers at harvest (NT @H)	-0.866	0.355	0.290
Leaf area index at flowering (LAI @F)	0.852	0.459	-0.153
Number of productive tillers at harvest (NPT @H)	-0.306	0.411	0.748
Dry matter production at harvest (DMP @H)	0.588	0.774	0.198
Panicle length (LP)	0.566	-0.261	0.622
Panicle weight (WP)	0.841	-0.386	0.119
Test weight (TW)	0.600	-0.753	0.237
Grain yield (GY)	0.589	0.781	0.037
Straw yield (SY)	0.242	0.917	0.116
Days to harvest (DAYH)	0.909	0.276	-0.276

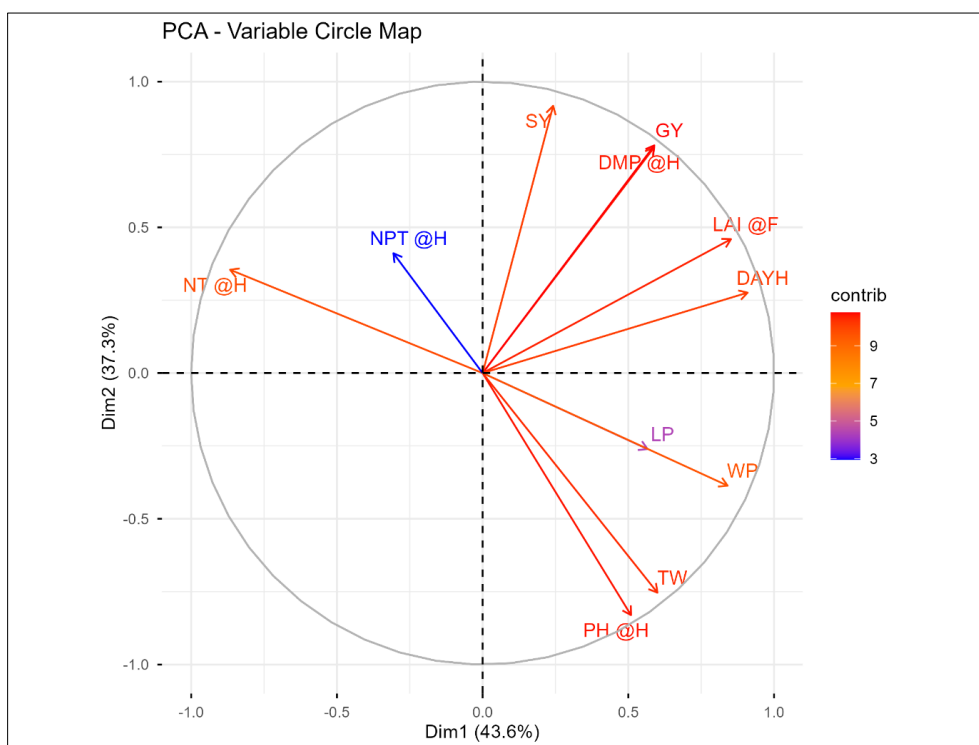


Fig 2: Variable correlation plot showing the contribution of the growth and yield traits to the principal components

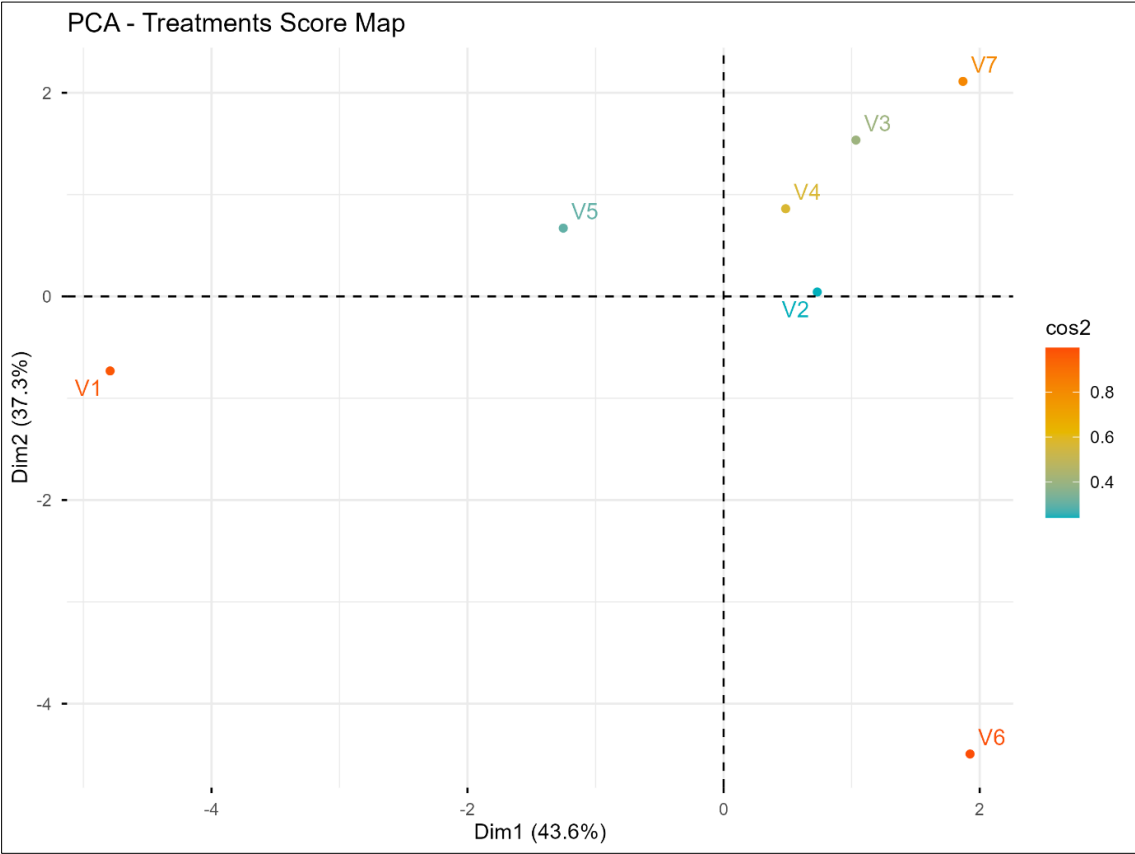


Fig 3: Principal component score plot of the distribution of rice cultivars

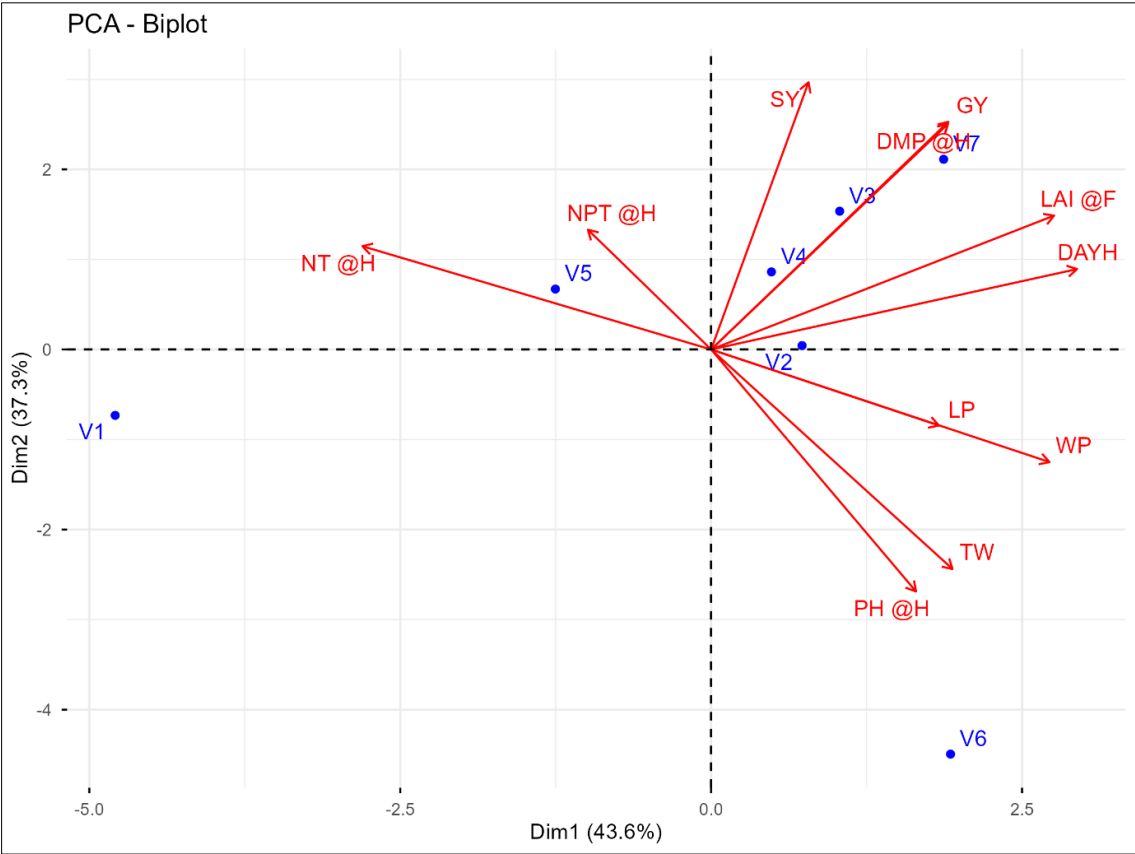


Fig 4: Principal component biplot of association between the rice cultivars, growth and yield traits

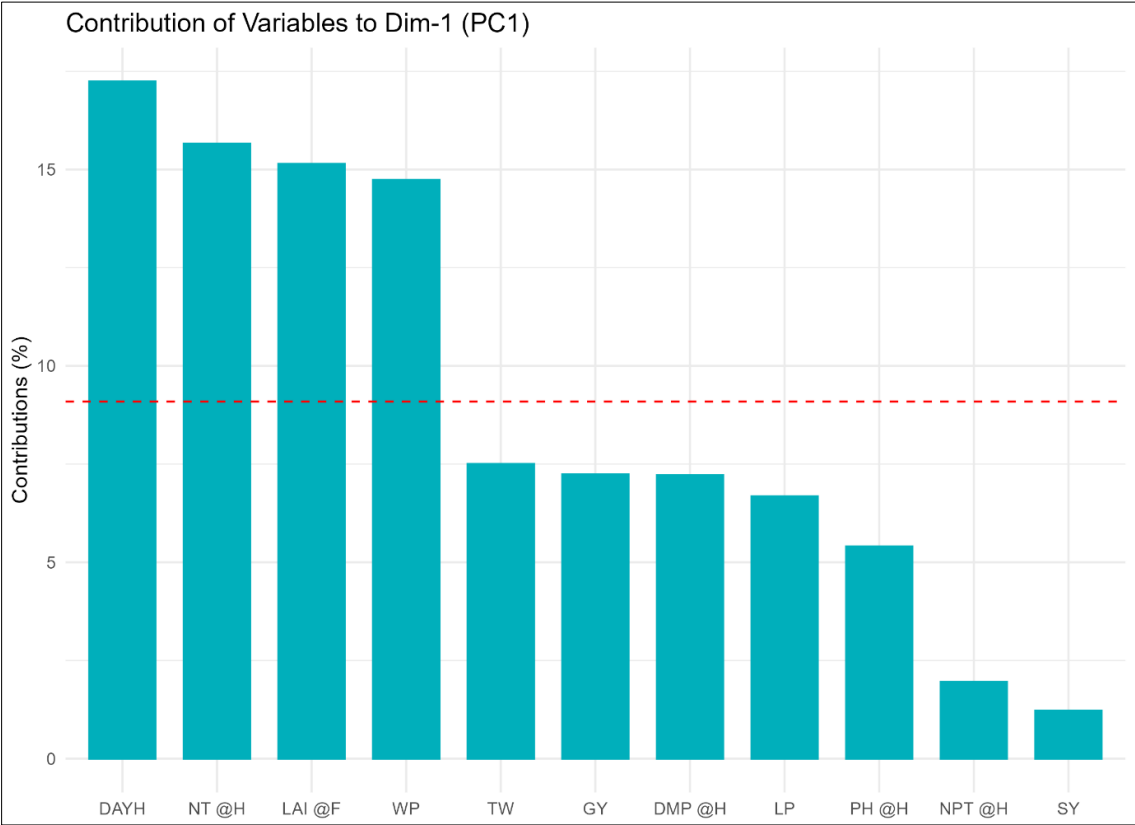


Fig 5: Contribution of growth and yield traits to the first principal component (PC1)

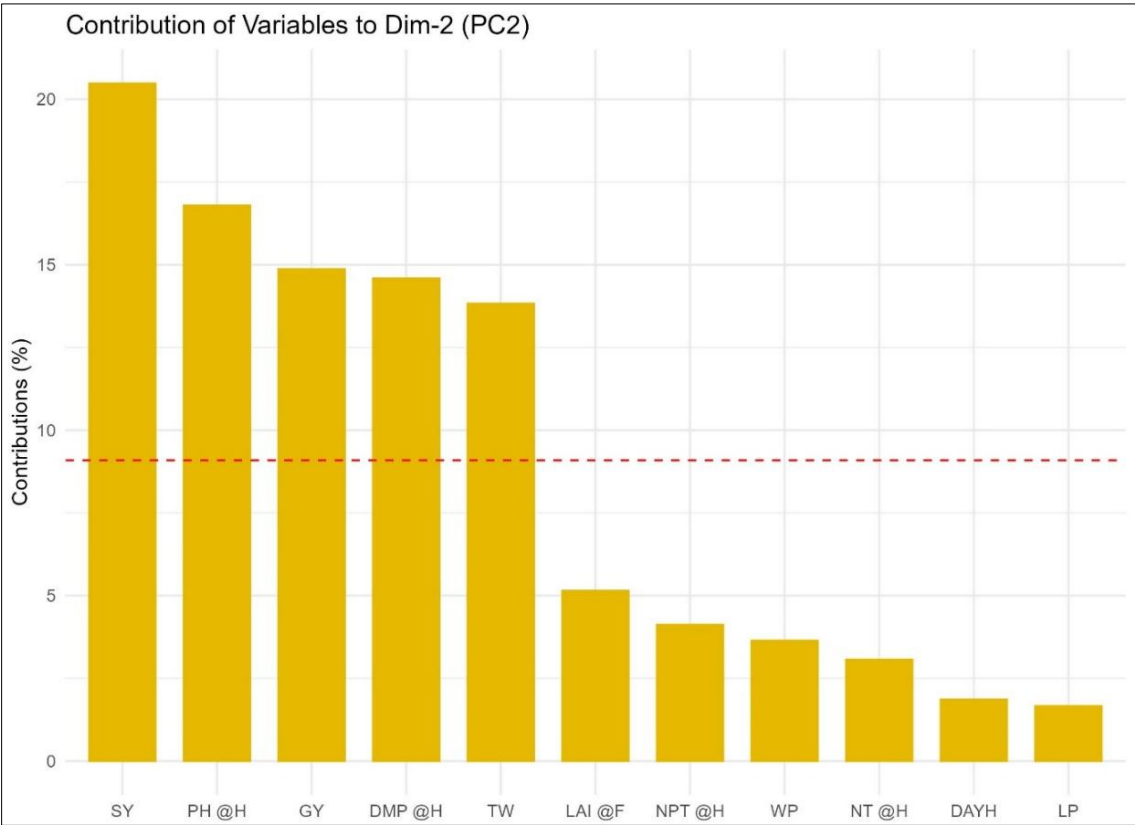


Fig 6: Contribution to second principal component (PC2) of growth and yield traits

Distribution of Rice Cultivars Based on Principal Components
The score plot (PCA) and biplot clearly distinguished the seven rice cultivars based on their agronomic performance. Cultivars falling in the positive quadrant (PC1 and PC2) were found to have

better growth and yield characteristics while those in the other quadrants had relatively poor growth and yield characteristics. The biplot also showed the relationships among the cultivars and measured variables and cultivar combinations with desirable growth and productivity characteristics were identified. The

distinct separation of cultivars shows a high amount of phenotypic variation and the PCA was useful for selecting cultivars that were superior in several agronomic traits at the same time (Hair *et al.*, 2019; Yan & Kang, 2003) (Figures 3 and 4) [5, 18].

Cluster Analysis

The rice cultivars were clustered into three clusters based on their similarity in growth and yield performance using the hierarchical cluster analysis. The dendrogram showed good discernment of the cultivars into groups that have different agronomic characteristics, reflecting high phenotypic variability of the materials evaluated (Figure 7). The elbow method was used to identify the best number of clusters for the K-means analysis and the within cluster sum of squares (WCSS) values were plotted against k values. The elbow plot was very clearly convex at k = 3, suggesting three clusters were the most suitable number for

partitioning the rice cultivars (Figure 8). The seven cultivars of rice were divided into three groups by K-means clustering that was based on optimal cluster number obtained by the hierarchical clustering. Cluster I comprised ADT 46 (V2), KKL (R) 2 (V3), IR 20 (V4), BPT 5204 (V5) and CR 1009 (V7), indicating relatively similar agronomic performance. The trait combination of Karuppukavuni (V6) differed from the other varieties and it was grouped separately due to its higher plant height, panicle weight and lower grain yield than the other varieties (Table 3; Figure 9).

The similarity of the hierarchical and K-means cluster shows that the cultivar classification is reliable and there is a high level of phenotypic diversity among the rice cultivars evaluated. Other clustering methods are extensively adopted for the classification of genotypes according to several quantitative traits (Rao, 1952; Johnson & Wichern, 2007) [6, 12].

Table 3: Rice cultivars were classified using hierarchical and K-means cluster analysis with growth and yield traits

Cultivar	Hierarchical Cluster	K-means Cluster
CO 51 (V1)	1	2
ADT 46 (V2)	2	1
KKL (R) 2 (V3)	2	1
IR 20 (V4)	2	1
BPT 5204 (V5)	2	1
Karuppukavuni (V6)	3	3
CR 1009 (V7)	2	1

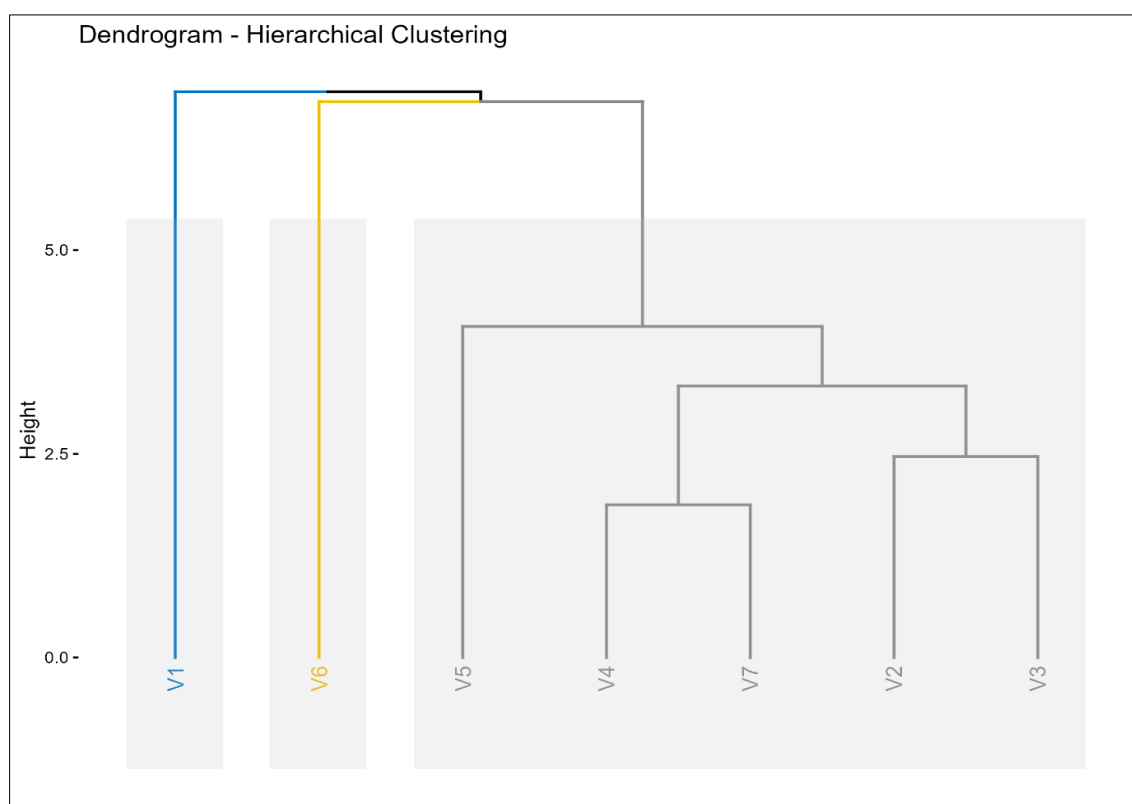


Fig 7: Hierarchical cluster dendrogram based on growth and yield traits of rice cultivars.

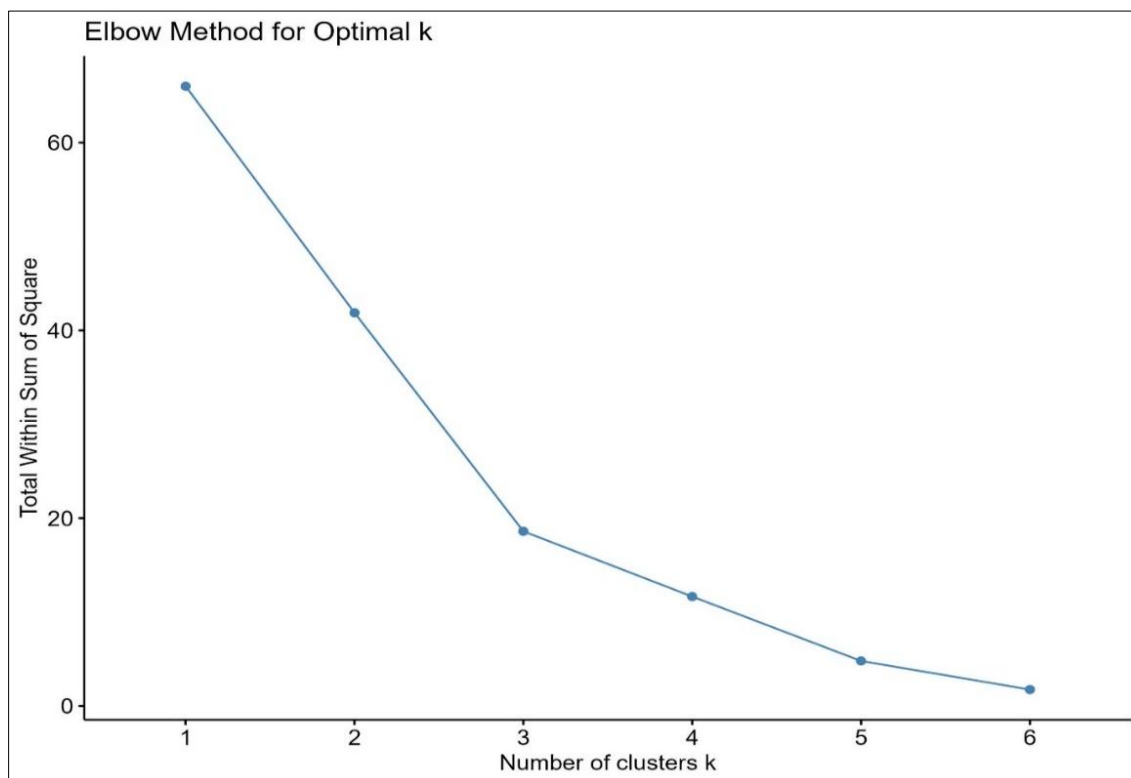


Fig 8: Elbow plot to determine the optimal number of clusters (k) for the K-means clustering

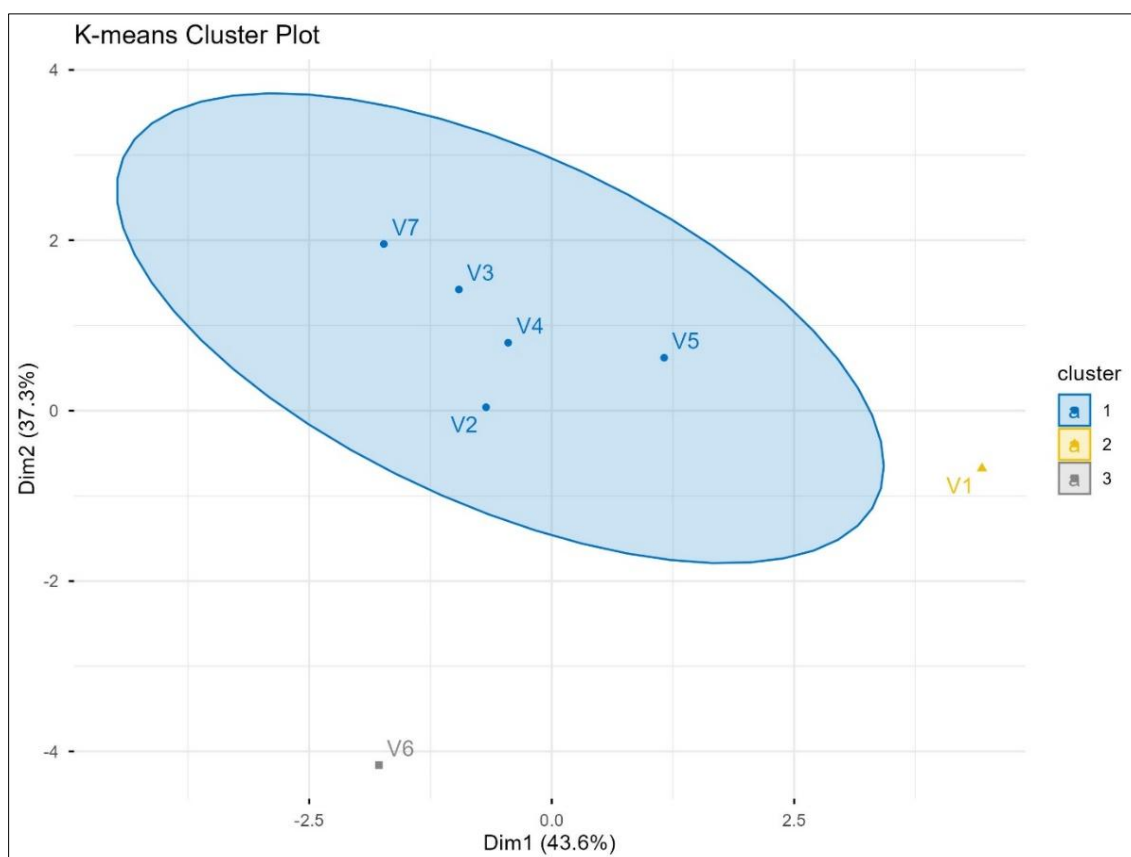


Fig 9: Rice cultivars classified by multivariate traits using K-means cluster plot

Identification of High-Performing Rice Cultivars

Combination of this interpretation with that of cluster analysis showed the cultivar CR 1009 (V7) to be the most superior, due to its high grain yield (6,298 kg ha⁻¹), high dry matter yield (15,944 kg ha⁻¹), highest LAI at flowering (7.27) and the longest crop duration (149 days), which makes this cultivar a good prospect

for coastal delta conditions. KKL (R) 2 (V3) and ADT 46 (V2) also performed well and had higher LAI (6.31 and 6.54 respectively) and a balance in growth and yield attributes. Karuppukavuni (V6), on the other hand, had an unusually high plant height (180.96 cm) and panicle weight (4.15 g), but grain yield (2,892 kg ha⁻¹) was relatively low, which suggests that a

higher vegetative growth and biomass accumulation alone does not necessarily translate into higher grain productivity and differences in source-sink balance and assimilate partitioning efficiency between the two varieties (Smith et al., 2018) [14]. CO 51 (V1) had low productivity as compared with the rest, lower LAI at flowering (2.74) and shorter crop duration (106 days) making it distinct (Table 3; Figures 4, 7 and 8). This information demonstrates the value of multivariate analysis to the identification of superior rice cultivars, based on multiple growth and yield traits instead of grain yield. Other applications of multivariate techniques for identification of superior crop genotypes by combining several yield components are reported by Mohammadi & Prasanna (2003) [10].

Conclusion

The diversity among cultivars of rice was found using multivariate analysis, which was effective. PCA identified the major traits contributing to cultivar variation, while hierarchical and K-means clustering grouped cultivars into distinct clusters with three optimal groups confirmed by the elbow method. These findings demonstrate the usefulness of multivariate approaches for cultivar characterization, selection of promising genotypes and supporting future rice improvement programmes.

References

1. Abdi H, Williams LJ. Principal component analysis. Wiley Interdiscip Rev Comput Stat. 2010;2(4):433-459.
2. Azimov A, Shavkiev J, Koraev A. Genetic diversity, population structure, and salinity tolerance of upland cotton (*Gossypium hirsutum* L.) germplasm revealed by SSR markers. 2026.
3. Cruz CD, Regazzi AJ, Carneiro PCS. Modelos biométricos aplicados ao melhoramento genético. 4th ed. Viçosa: UFV; 2012.
4. Food and Agriculture Organization of the United Nations. Climate-smart agriculture sourcebook. Rome: Food and Agriculture Organization of the United Nations; 2013.
5. Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate data analysis. 8th ed. Cengage Learning; 2019.
6. Johnson RA, Wichern DW. Applied multivariate statistical analysis. 6th ed. Pearson Prentice Hall; 2007.
7. Jolliffe IT. Principal component analysis. 2nd ed. Springer; 2002.
8. Kassambara A, Mundt F. factoextra: Extract and visualize the results of multivariate data analyses. R package version 1.0.7; 2020.
9. Le S, Josse J, Husson F. FactoMineR: An R package for multivariate analysis. J Stat Softw. 2008;25(1):1-18.
10. Mohammadi SA, Prasanna BM. Analysis of genetic diversity in crop plants salient statistical tools and considerations. Crop Sci. 2003;43(4):1235-1248.
11. Nachimuthu VV, Robin S, Sudhakar D, Raveendran M, Rajeswari S, Manonmani S. Evaluation of rice genetic diversity and variability in a population panel by principal component analysis. Indian J Sci Technol. 2014;7(10):1555-1562.
12. Rao CR. Advanced statistical methods in biometric research. New York: John Wiley & Sons; 1952.
13. Rencher AC, Christensen WF. Methods of multivariate analysis. 3rd ed. John Wiley & Sons; 2012.
14. Smith MR, Rao IM, Merchant A. Source-sink relationships in crop plants and their influence on yield development and nutritional quality. Front Plant Sci. 2018;9:1889.
15. Teague E, Awosile B. Distribution and diversity of *Salmonella* serovars from beef cattle before, at, and after processing in relation to human-derived serovars. Front Vet Sci. 2026;13:1897484.
16. Ward JH Jr. Hierarchical grouping to optimize an objective function. J Am Stat Assoc. 1963;58(301):236-244.
17. Yali W, Nepir G. Assessment of genetic diversity among elite sorghum [*Sorghum bicolor* (L.) Moench] genotypes based on agro-morphological traits using principal component and cluster analyses under drought-stressed environments. Ecol Genet Genomics. 2026;100503.
18. Yan W, Kang MS. GGE biplot analysis: A graphical tool for breeders, geneticists, and agronomists. Boca Raton: CRC Press; 2003.