IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

International Journal of Sustainable Development and Planning
Vol. 21, No. 6, June, 2026, pp. 2547-2561
Journal homepage: http://iieta.org/journals/ijsdp

Hybrid PCA-FCM Clustering for Provincial Rice Distribution Classification in Indonesia

Muhammad Faisal[1*], Muhammad Syafaat S. Kuba[2], Abd Rakhim Nanda[2], Udom A. L. Ewon[3], Eny Syatriana[4], Desi Anggreani[1], Rizki Yusliana Bakti[1], Muhammad Risal Haris[1], Intan Izyan Roslan[5]

[1] Department of Informatics, Universitas Muhammadiyah Makassar, Makassar 90221, Indonesia
[2] Department of Water Resources Engineering, Universitas Muhammadiyah Makassar, Makassar 90221, Indonesia
[3] Diploma in Design Craft, Politeknik Besut, Mukim Keluang 22200, Malaysia
[4] Department of English Education, Universitas Muhammadiyah Makassar, Makassar 90221, Indonesia
[5] Department of Information Technology & Communication, Politeknik Kuala Terengganu, Kuala Terengganu 20200, Malaysia

Corresponding Author Email: muhfaisal@unismuh.ac.id

https://doi.org/10.18280/ijsdp.210613

Received: 2 April 2026
Revised: 10 May 2026
Accepted: 22 May 2026
Available online: 30 June 2026

Keywords:
Principal Component Analysis, Fuzzy C-Means, hybrid clustering, rice distribution, agricultural logistics

ABSTRACT

Indonesia's rice distribution system exhibits substantial variation in production capacity, consumption pressure, logistics readiness, and price stability across provinces, creating persistent disparities in provincial rice-distribution characteristics. This study implements a hybrid analytical framework that integrates Principal Component Analysis (PCA) and Fuzzy C-Means (FCM) to evaluate provincial rice-distribution characteristics using multivariate indicators sourced from national institutions. PCA is employed to extract the dominant structural dimensions of the distribution system-production-surplus strength, logistics capacity, and price-demand variability-while reducing multicollinearity and improving interpretability. The resulting PCA components are clustered using FCM to capture transitional and overlapping provincial characteristics. Cluster validity is assessed through the Xie-Beni (XB) index and benchmarked against baseline models, showing that the Principal Component Analysis-Fuzzy C-Means (PCA-FCM) configuration provides a balanced trade-off between clustering compactness and fuzzy interpretability for capturing transitional provincial characteristics. The analysis identifies four effectiveness categories: Highly Effective, Moderately Effective, Surplus but Logistically Weak, and Deficit & Logistically Vulnerable. Spatial analysis reveals pronounced disparities between western and eastern regions, largely driven by infrastructure inequality. The findings provide a systematic, data-driven approach to classifying provincial distribution profiles and offer actionable insights for policymakers to design targeted interventions, strengthen logistics corridors, and enhance national food security resilience.

1. INTRODUCTION

Indonesia's rice-distribution system is shaped by persistent structural disparities arising from uneven production output, demographic concentration, consumption pressure, and logistics performance across provinces. As rice remains the nation's primary staple commodity, the resilience of the distribution network is essential for maintaining food security and socio-economic stability [1]. Significant gaps continue to separate high-surplus regions from deficit-dependent provinces, where distribution flow is further influenced by variations in road conditions, transport accessibility, warehousing capacity, and market connectivity [2-4]. Previous rice-distribution and agricultural logistics studies generally evaluate production capacity, transportation readiness, or market performance independently, limiting integrated understanding of provincial rice-distribution structures [5, 6]. Existing national assessments remain descriptive and single-indicator-based, limiting the ability to understand how production strength, consumption intensity, logistics readiness, and price behaviour operate as a unified system. This analytical gap restricts policymakers from identifying priority regions and designing targeted logistical interventions [7]. Rice-distribution evaluation increasingly requires multivariate analytical approaches to capture structural interactions among production, logistics readiness, consumption pressure, and market dynamics. Studies on dimensionality reduction demonstrate Principal Component Analysis (PCA)'s capability to reveal hidden structural patterns in high-dimensional datasets, including transportation behaviour, logistics readiness, and distribution variability [8, 9]. Previous studies further demonstrate that dimensionality-reduction techniques improve interpretability in agricultural distribution and regional performance evaluation systems [10]. These methodological advances reinforce PCA's role in summarizing interconnected indicators into interpretable components that support large-scale classification efforts [11, 12]. Clustering approaches likewise play a central role in distribution-performance analysis. Fuzzy C-Means (FCM) is widely applied because it accommodates gradual transitions between categories-an essential feature when evaluating provinces whose characteristics do not fit rigid boundaries [10, 13, 14]. In agricultural value chains, fuzzy clustering has been used to detect structural heterogeneity, variation in sustainability, and performance asymmetry across dynamic supply-chain environments [15-17]. Recent methodological studies emphasize that integrating PCA with FCM enhances cluster separability, stability, and interpretability in high-dimensional agricultural, environmental, and socio-economic datasets [18, 19]. Despite these developments, hybrid Principal Component Analysis-Fuzzy C-Means (PCA-FCM) approaches remain underutilised in national rice distribution research, leaving a gap in Indonesia's evidence-based policy framework [20].

Previous analytical frameworks emphasize that multivariate modelling improves regional distribution-system evaluation under heterogeneous provincial conditions [21]. Such insights strengthen the rationale for adopting a hybrid PCA-FCM model, which can capture both structural disparities and transitional characteristics in Indonesia's provincial distribution system. In the Indonesian context, hybrid analytical approaches such as multi-objective optimization and clustering have proven effective for regional prioritization and policy formulation, particularly in settings characterized by socio-economic and infrastructural heterogeneity [22, 23]. These applications illustrate how integrated quantitative frameworks outperform single-indicator assessments by identifying hidden structural constraints and mapping multi-dimensional development trajectories. Their relevance to rice-distribution studies underscores the suitability of PCA-FCM for classifying provinces into meaningful categories that reflect both logistical capacity and distribution vulnerability. Robust analytical frameworks are increasingly required to improve interpretability and classification reliability in high-dimensional distribution evaluation systems [24, 25]. Applying these principles to national rice-distribution analysis ensures that the resulting cluster structure is not only statistically valid but also operationally relevant for strategic decision-making and policy implementation. To address this limitation, the present study implements a hybrid PCA-FCM framework to generate an objective classification of provincial rice distribution characteristics. The research aims to: (1) extract latent dimensions that shape multivariate distribution behavior, (2) group provinces into soft clusters that reflect transitional and overlapping characteristics, and (3) produce a provincial typology that supports targeted policy intervention and enhances national rice-distribution structural understanding [26]. The contribution of this study lies in the national-scale application of integrated multivariate provincial classification to rice-distribution analysis, supporting evidence-based regional typology development and policy-oriented distribution planning in Indonesia.

2. METHODOLOGY

2.1 Research framework

The analytical design of this study is implemented through a structured, step-by-step workflow encompassing data acquisition, indicator grouping, dimensionality reduction, clustering formulation, and model performance evaluation. These stages operate sequentially to ensure that each analytical output becomes the foundation for the next phase. As illustrated in Figure 1, the framework visually integrates the major components of the hybrid PCA-FCM approach, showing how provincial rice-distribution indicators are consolidated, transformed through PCA, and subsequently clustered using FCM to reveal regional performance patterns. The figure clearly depicts the methodological flow, highlighting the interconnections among preprocessing, transformation, and clustering procedures within the overall analytical pipeline.

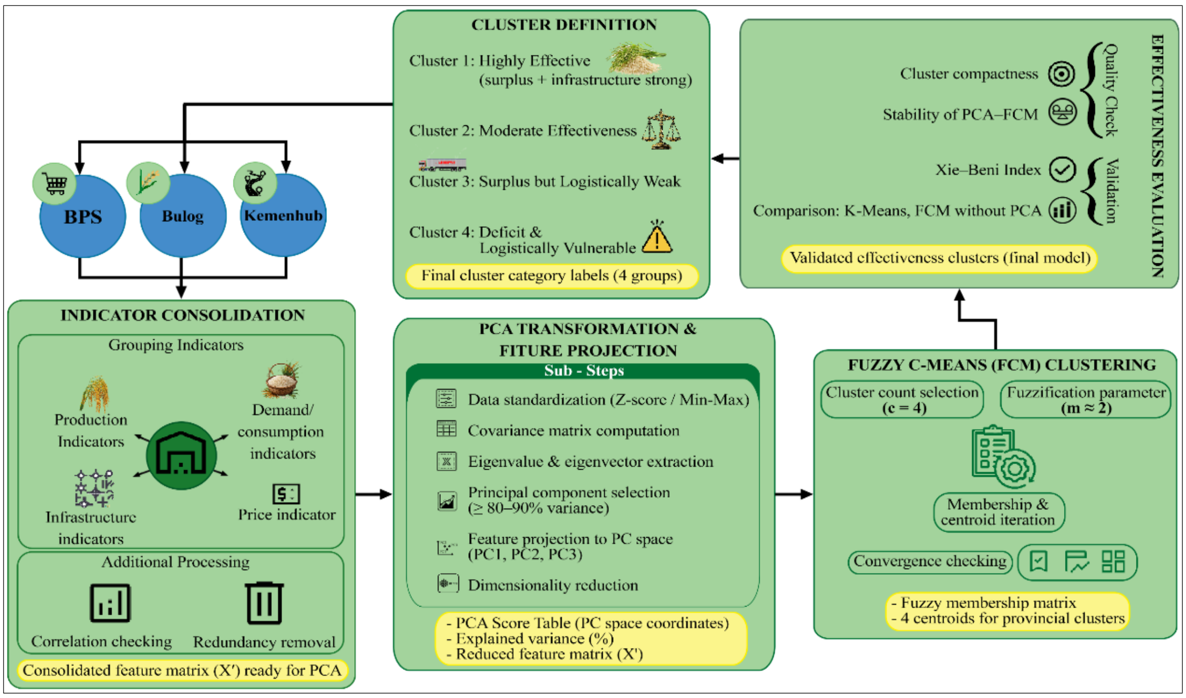

Figure 1. Hybrid Principal Component Analysis-Fuzzy C-Means (PCA-FCM) workflow for evaluating provincial rice-distribution effectiveness

Figure 1 presents the step-by-step analytical pipeline applied in this study, starting from data acquisition and indicator integration, followed by PCA-based transformation, and concluding with FCM clustering and the evaluation of distribution effectiveness. Each component of the workflow is linked to ensure that the resulting cluster interpretation provides a coherent and comprehensive assessment of provincial rice-distribution performance.

2.2 Data collection

The dataset used in this study was compiled from three primary national institutions that collectively represent Indonesia's rice production, consumption patterns, logistics capacity, and supply-flow conditions, and market-price dynamics. Provincial statistics on production volumes, consumption levels, and population were sourced from the Central Bureau of Statistics of Indonesia. Data on logistics performance-specifically road network characteristics and transport infrastructure-were retrieved from the Ministry of Transportation using publicly available infrastructure statistics. Paddy-price information (GKG) representing market dynamics was obtained from Bulog statistical records and integrated with provincial production, demand, and logistics indicators within the analytical framework. All variables collected from these institutions were merged into a unified dataset that captures the multi-dimensional aspects of provincial rice distribution. An excerpt from the compiled dataset is presented in Table 1, illustrating the structure and arrangement of the indicators used in this research.

The analytical dataset consists of 114 province-year observations derived from 38 provinces observed across three annual periods (2022-2024). This structure enables the analytical framework to capture both interprovincial variation and temporal distribution dynamics while maintaining consistent multivariate representation across observation periods.

Table 1. Sample rows of the provincial rice-distribution dataset

Province	Rice Production (Ton)	Rice Consumption per Capita (Kg/Capita/Year)	Population (Thousand People)	Total Rice Consumption (Ton)	Surplus/Deficit (Ton)	Good Road Length (KM)	Paddy Price (IDR/KG)
Aceh (1)	869,572.00	98.80	5,407.90	534,300.52	335,271.48	2,112.00	8,000.00
Sumatera Utara (2)	1,198,045.75	99.40	15,115.20	1,502,450.88	-304,405.13	2,620.00	8,950.00
Sumatera Barat (3)	795,306.36	99.70	5,640.60	562,367.82	232,938.54	1,423.00	9,045.00
Riau (4)	122,561.42	85.90	6,614.40	568,176.96	-445,615.54	1,254.00	9,000.00
Jambi (5)	160,667.47	88.10	3,631.10	319,899.91	-159,232.44	1,319.00	10,075.00
Sumatera Selatan (6)	1,593,597.70	95.30	8,657.00	825,012.10	768,585.60	1,581.00	7,478.00
Bengkulu (7)	162,197.49	98.50	2,060.10	202,919.85	-40,722.36	782.00	7,342.00
Lampung (8)	1,545,296.21	90.10	9,176.60	826,811.66	718,484.55	1,298.00	9,275.00
Kepulauan Riau (9)	290.07	81.70	2,179.80	178,089.66	-177,799.59	589.00	0.00
DKI Jakarta (10)	1,378.00	89.40	10,680.00	954,792.00	-953,414.00	53.00	0.00
Jawa Barat (11)	5,447,806.31	94.70	49,405.80	4,678,729.26	769,077.05	1,783.00	10,300.00
Jawa Tengah (12)	5,380,509.51	90.10	37,032.40	3,336,619.24	2,043,890.27	1,581.00	10,600.00
DI Yogyakarta (13)	319,059.52	80.60	3,761.90	303,209.14	15,850.38	307.00	9,000.00
Jawa Timur (14)	5,500,801.88	88.30	41,150.00	3,633,545.00	1,867,256.88	2,262.00	9,750.00
Banten (15)	1,018,653.08	101.00	12,252.00	1,237,452.00	-218,798.92	568.00	9,600.00
Bali (16)	383,829.16	111.90	4,415.10	494,049.69	-110,220.53	590.00	7,900.00
…	…	…	…	…	…	…	…
Papua (38)	24.20	51.40	1,467.00	75,403.80	-75,379.60	0.00	0.00

The dataset excerpt in Table 1 summarizes the key indicators used in this study, including provincial rice production, consumption intensity, total demand, logistics capacity, and paddy price measures. Collectively, these variables capture the multidimensional characteristics of Indonesia's rice distribution system from both supply and demand perspectives. The variation observed across provinces, reflected in differences in production volumes, surplus-deficit conditions, and road-network quality, highlights the heterogeneity that motivates the use of dimensionality-reduction and clustering techniques. This structured set of multivariate indicators serves as the empirical foundation of the analytical workflow, supporting subsequent stages of indicator integration, PCA-based feature extraction, and FCM clustering. The diversity and completeness of the dataset ensure that the methodological framework can effectively represent interprovincial disparities and enables a comprehensive evaluation of distribution characteristics.

The selected indicators were designed to represent structural provincial distribution conditions related to production capacity, demand pressure, supply balance, infrastructure readiness, and market-price dynamics, supporting comparative classification analysis despite not explicitly incorporating operational logistics variables.

2.3 Indicator consolidation

All indicators sourced from BPS, Bulog, and the Ministry of Transportation are initially integrated into a single multivariate dataset consisting of production, consumption, logistics, infrastructure, and market-price indicators so that each province is represented by a complete set of production, consumption, logistics, and price-related attributes. After the merging process, the dataset is examined through correlation testing and redundancy checks to detect highly collinear variables or overlapping information that could interfere with the accuracy of multivariate analysis. Indicators displaying substantial redundancy are removed or combined to ensure that each feature contributes unique and meaningful analytical value, a process commonly emphasized in high-dimensional preprocessing frameworks [12]. Production-related indicators and surplus-deficit measurements exhibit partial structural dependency because surplus-deficit values are mathematically derived from production-demand balance. However, surplus-deficit indicators were retained because they represent operational supply-demand equilibrium conditions that remain substantively relevant for provincial rice-distribution structural profiling. While rice production captures output capacity and consumption reflect demand pressure, surplus-deficit directly characterizes distribution imbalance conditions that are important for policy-oriented regional classification. PCA transformation is subsequently applied to reduce redundancy effects arising from correlated indicators while preserving analytical interpretability. The remaining variables are subsequently organized into four conceptual categories-production, consumption-demand, logistics, and price-to reflect the underlying structural dimensions of Indonesia's rice-distribution system [27]. Price-related indicators obtained from Bulog are incorporated to represent market dynamics within the provincial rice-distribution structure. This consolidation phase produces a coherent and methodologically consistent dataset that is ready for standardization before being processed through PCA.

2.4 Principal Component Analysis transformation and feature projection

The consolidated dataset is standardized using Z-Score standardization to ensure that all indicators contribute proportionally to the multivariate analysis. This standardization step minimizes scale-induced distortions, which are particularly important in agricultural and supply-chain datasets where production quantities, consumption intensities, road capacity, and price variables differ substantially in magnitude. Prior studies emphasize the importance of such preprocessing when preparing high-dimensional operational data [28, 29]. Missing-value assessment was conducted before preprocessing to identify incomplete observations across provincial indicators. Infrastructure-related indicators contained missing observations affecting 46 province-year observations (40.35% of the analytical dataset), consisting of 4 observations in 2022, 4 observations in 2023, and 38 observations in 2024. The concentration of missing observations in 2024 primarily reflects data-availability limitations rather than random measurement absence. For several province-year observations, paddy-price records were unavailable in the original data source and were therefore recorded as 0.00 in the compiled dataset. These values do not represent actual market prices but indicate data unavailability in the source records. Because the study relied on official statistical datasets and did not perform additional price imputation, the original values were retained to preserve consistency with the reported observations. This limitation should be considered when interpreting market-price patterns for provinces with unavailable paddy-price records. To preserve analytical completeness within the multivariate framework, missing observations in the infrastructure indicator were imputed using mean-value substitution prior to standardization and PCA processing. This approach was adopted to maintain a consistent dataset structure for PCA and FCM analysis while minimizing the loss of province-year observations.

Following Z-Score standardization, PCA is performed by computing the covariance matrix and deriving its eigenvalues and eigenvectors to extract the dominant latent dimensions embedded within the provincial rice-distribution indicators. Principal components were retained based on cumulative explained variance. In this study, PC1–PC3 were selected because they collectively explained 98.5% of the total variance thereby preserving the dominant structural information contained in the provincial rice-distribution indicators. The retention of PC3 was maintained because it captures meaningful logistics- and price-related variation that is not fully represented by PC1 and PC2 [26, 30]. Once the principal components are identified, PCA scores are calculated for each province and projected into a reduced feature space-typically represented using the first two or three components. The transformation can be expressed formally in Eq. (1):

$$Y = Z\, V_k \tag{1}$$

where, Z is the standardized data matrix, V_k is the matrix of selected eigenvectors that retain the majority of the explained variance, and Y denotes the PCA-reduced feature representation. This projection yields a compact and noise-reduced structure that enhances the stability of subsequent clustering. The resulting PCA-reduced feature matrix is then designated as the primary input for the FCM algorithm, supporting improved separability and reducing clustering distortions commonly observed in high-dimensional, correlated datasets.

2.5 Fuzzy C-Means clustering

The PCA-reduced feature matrix is subsequently processed using the FCM algorithm to classify provinces based on similarities in their rice-distribution characteristics [31]. The number of clusters is pre-defined as c = 4, reflecting four conceptual levels of distribution effectiveness established in the research design. Initial cluster centroids are positioned within the PCA feature space, after which the algorithm iteratively updates both membership degrees and centroid locations [32]. At each iteration, membership values are adjusted proportionally to the relative distance between each province and the centroids, while updated centroid positions are computed using weighted membership contributions. The optimization FCM objective function is expressed in Eq. (2):

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^{m} \, \| x_i - c_j \|^2 \tag{2}$$

where, u_{ij} denotes the membership degree of province i in cluster j, m is the fuzzification coefficient, x_i represents the PCA-reduced feature vector, and c_j is the centroid of cluster j. The algorithm continues updating membership values and centroid positions until convergence.

2.6 Effectiveness evaluation

The quality of the clustering results is assessed using the Xie-Beni (XB) index, which evaluates cluster compactness and separability based on distances between membership-weighted provincial features and their respective centroids [33]. This measure provides an objective indication of whether the four-cluster structure generated by the PCA-FCM model represents meaningful distinctions in provincial distribution characteristics. The XB index is shown in Eq. (3):

$$XB = \frac{\sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^2 \parallel x_i - c_j \parallel^2}{N \cdot \min_{j \neq k} \parallel c_j - c_k \parallel^2} \tag{3}$$

where, u_{ij} represents the membership degree of province i in cluster j, x_i are the PCA-reduced feature vector, and c_j and c_k denote centroid positions. To benchmark performance, the PCA-FCM model is compared against two baseline clustering approaches: conventional K-Means and FCM applied directly to the original high-dimensional indicators without PCA. This comparison reveals the extent to which dimensionality reduction improves cluster compactness, stability, and interpretability. Model robustness is further evaluated through sensitivity tests that vary the fuzzification coefficient m, initialization parameters, and the number of clusters. Together, these procedures validate the reliability of the PCA-FCM hybrid model and confirm that the resulting clusters represent stable and structurally coherent groupings of provincial rice-distribution effectiveness.

3. RESULTS

3.1 Principal Component Analysis result

PCA analysis was applied to five key indicators representing the dimensions of production, demand, surplus-deficit, road-infrastructure capacity, and price stability, namely Rice-Production, Total-Consumption, Surplus-Deficit, Good-Road-Length, and Paddy-Price. This multidimensional indicator set reflects the structural attributes of Indonesia's rice-distribution system and aligns with analytical practices in supply-chain and agrifood studies [12, 15]. A total of 114 provincial observations (province-year) were included in the analysis. These indicators collectively capture the core mechanics of distribution effectiveness-production-demand balance, distribution capacity, infrastructure accessibility, and market dynamics. PCA is therefore employed to reduce dimensionality and extract the dominant latent components that explain the structural variance across provinces. This enables a clearer understanding of the underlying patterns shaping distribution performance before integrating them into the FCM clustering phase. The subsequent subsections examine the standardization process, covariance structure, variance contributions, loading interpretations, and spatial distribution of PCA scores. Together, these analytical stages form the foundation for the clustering and effectiveness evaluation presented in later sections.

3.2 Construction of the analysis matrix

Before PCA, each indicator X_j was standardized using the Z-Score to ensure comparability across variables with different scales, following common procedures in multivariate agricultural and logistics research [13]. Standardization transforms all indicators into a unified metric so that large-magnitude variables do not dominate the principal components. Following common practice in multivariate agricultural and logistics studies using Eq. (4):

$$Z_{ij} = \frac{X_{ij} - \mu_j}{\sigma_j} \tag{4}$$

where, X_{ij}: The value of the indicator to - j on observation (province-year) to - i, μ_j: Average Indicator j, σ_j: Standard deviation indicator j. For Good_Road_Length_km, missing values were imputed using the mean so that the analysis matrix is complete and can be processed in multivariate form, a practice that is also used in PCA research for environmental and supply chain data. Table 2 presents the mean and standard deviation values for each variable used in the analysis.

Table 2. Mean and standard deviation of Principal Component Analysis (PCA)

Indicator	Mean	Std Dev
Rice-Production (Ton)	818,096.98	1,448,437.39
Total-Consumption (Ton)	683,637.91	1,016,087.52
Surplus-Deficit (Ton)	817,413.34	1,447,479.14
Good-Road-Length (KM)	1,258.34	783.49
Paddy-Price (KG)	5,807.22	3,942.08

These descriptive values demonstrate the variability across indicators and justify the need for standardization prior to PCA, as shown in Table 2. As an illustration, for Aceh in 2022, the original indicator values are: Rice-Production = 869,572.00, Total-Consumption = 534,300.52, Surplus-Deficit = 869.037.70, Good-Road-Length = 2,112.00, and Paddy_Price = 8.000. After standardization, the Z-Score values in Table 3 were obtained.

Table 3. Example of Z-Score values for Aceh 2022

Indicator	Raw Value	Z-Score
Rice-Production (Ton)	869,572.00	0.036
Total-Consumption (Ton)	534,300.52	-0.147
Surplus-Deficit (Ton)	869,03770	0.036
Good-Road-Length (KM)	2,112.00	1.090
Paddy-Price (IDR/KG)	8,000.00	0.556

The Z-Score pattern for Aceh demonstrates slightly above-average performance in both rice production and surplus (positive Z-scores), indicating strong supply potential. The province also shows higher road-network capacity relative to the national distribution, suggesting logistical readiness. Conversely, Aceh's grain price level is higher than average, reflecting market pressure despite strong logistics.

3.3 Covariance structure and component extraction

After standardisation, the covariance matrix Σs formed from the standardized Z data matrix using Eq. (5):

$$\Sigma s = \frac{1}{n-1} Z^T Z \quad (5)$$

The way that section titles and other headings are displayed in these instructions is meant to be followed in your paper. Because PCA was performed on Z-score standardized variables, the resulting covariance matrix is numerically equivalent to a correlation matrix. Minor deviations of reported diagonal values from exactly 1.000 (e.g., 1.009) arise from numerical rounding and sample-based covariance estimation and do not affect the PCA extraction process.

Table 4. Covariance matrix of standardized indicators

Indicator	Rice-Production	Total-Consumption	Surplus-Deficit	Good-Road-Length	Paddy-Price
Rice-Production	1.009	0.951	1.009	0.281	0.302
Total-Consumption	0.951	1.009	0.951	0.270	0.282
Surplus-Deficit	1.009	0.951	1.009	0.281	0.302
Good-Road-Length	0.281	0.270	0.281	1.009	0.335
Paddy-Price	0.302	0.282	0.302	0.335	1.009

The covariance values indicate very strong relationships among three indicators, such as Rice-Production, Total-Consumption, and Surplus-Deficit (≈0.95-1.00) suggesting high multicollinearity in the production-demand dimension of the rice system. This aligns with literature noting that surplus is structurally tied to both production capacity and consumption intensity in agrifood supply chains. In contrast, Good-Road-Length_km and Paddy-Price show moderate but positive covariance with the volume-related indicators (≈0.27-0.33). This pattern indicates that logistics conditions and market prices introduce distinct variability into the dataset, contributing additional dimensions beyond pure production-demand metrics. Such differentiation is important for capturing spatial disparities in infrastructure and price stability that influence distribution resilience. Principal components were derived from the covariance matrix by solving the eigenvalue problem as defined in Eq. (6):

$$\Sigma v_k = \lambda_k v_k \quad (6)$$

where, λ_k denotes the eigenvalue and v_k is the eigenvector corresponding to component k. This spectral decomposition approach is widely used in dimensionality-reduction frameworks for high-dimensional datasets and complex supply chain systems [13]. The clear multicollinearity observed in Table 4 therefore justifies the application of PCA to condense redundant variables while preserving the structural contrasts arising from logistics capacity and price volatility. These contrasts later emerge as major axes of variation in the PCA-FCM clustering process, shaping the separation between surplus provinces, logistics-weak regions, and price-volatile areas.

3.4 Eigenvalues and variance explained

Eigenvalues were extracted to determine the relative contribution of each principal component to the total variance of the dataset. Table 5 presents the five eigenvalues obtained from the spectral decomposition of the covariance matrix Σ, such as $\lambda_1 = 3.213$, $\lambda_2 = 1.080$, $\lambda_3 = 0.674$, $\lambda_4 = 0.077$, $\lambda_5 \approx 0.000$.

The first two components have eigenvalues greater than 1, indicating that they each capture a substantial portion of the multivariate structure, consistent with the Kaiser retention rule commonly applied in PCA-based agricultural and supply-chain studies.

The proportion of variance explained by each component is summarized in Table 5 and visualized in the scree plot in Figure 2.

Table 5. Variance explained by principal components

Component	Eigenvalue	Variance Explained
PC1	3.213	0.637
PC2	1.080	0.214
PC3	0.674	0.134
PC4	0.077	0.015
PC5	~0.000	~0.000

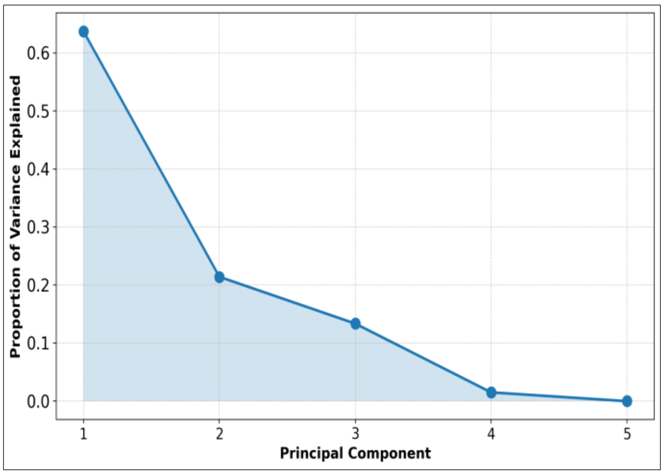

Figure 2. Scree plot of principal components for the rice-distribution indicators

The scree plot in Figure 2 clearly shows a sharp decline after PC2, confirming that PC1 and PC2 together account for more than 85% of total variance, while subsequent components contribute marginally. This pattern is typical in systems where production-demand indicators are highly correlated, while logistics and market-price factors introduce secondary but meaningful variation.

The dominance of PC1 reflects the strong multicollinearity among production, consumption, and surplus indicators, representing the core rice volume axis. PC2 contributes additional variance related to logistics capacity and price structure, capturing the heterogeneity observed across Indonesian provinces, especially between regions with strong versus limited infrastructure. PC3 explains small but relevant fluctuations associated with price-demand volatility. Together, these components form the structural basis for the PCA-FCM clustering phase, determining how provinces are separated into high-surplus regions, logistics-strong regions, transitional areas, and vulnerable clusters.

3.5 Loading matrix and interpretation of principal components

The loading matrix summarizes the contribution of each indicator to the principal components extracted in the PCA. Table 6 presents the loading coefficients for the first three components, which together account for more than 98% of the total variance.

Table 6. Principal Component Analysis (PCA) loading matrix for the first three components

Indicator	PC1	PC2	PC3
Rice-Production	0.545	−0.200	0.005
Total-Consumption	0.532	−0.208	0.016
Surplus-Deficit	0.545	−0.200	0.005
Good-Road-Length	0.242	0.676	0.696
Paddy-Price	0.254	0.648	−0.718

The loading matrix indicates that PC1 is strongly defined by Rice-Production, Total-Consumption, and Surplus-Deficit (loadings ≈ 0.53-0.55), forming a clear production-surplus axis that reflects variations in provincial production capacity and overall supply-demand balance. Provinces with high PC1 scores therefore tend to exhibit larger outputs, higher surpluses, and stronger volume-driven distribution potential, consistent with literature identifying production-surplus strength as a primary determinant of food-system performance. PC2, dominated by Good_Road_Length_Km (0.676) and Paddy_Price_IDRkg (0.648), represents a logistics-price dimension, where better road networks correspond to improved accessibility while higher prices signal market pressure or transport inefficiencies.

This component differentiates logistics-advantaged provinces from those with weak infrastructure or remote positioning, aligning with spatial inequality patterns commonly observed in Indonesian distribution systems. Meanwhile, PC3 exhibits a strong positive loading on Good-Road-Length (0.696) and a strong negative loading on Paddy-Price (-0.718), indicating an infrastructure-price stability trade-off, where stronger logistical conditions tend to suppress price volatility, whereas limited accessibility amplifies market fluctuations. Although PC3 contributes a smaller portion of variance, it provides important discrimination among provinces with similar logistics capacity but differing price dynamics. Collectively, these three components establish the structural foundation for subsequent clustering, with PC1 separating surplus versus deficit regions, PC2 highlighting logistics readiness, and PC3 revealing hidden disparities in price stability.

3.6 Principal Component Analysis scores and spatial separation of provinces

The PCA scores represent the transformed positions of each province in the reduced feature space defined by the principal components. Table 6 illustrates an example for Aceh (2022), showing positive values on PC1 and PC3 and a relatively high score on PC2. These values indicate that Aceh possesses moderate production-surplus strength (PC1), strong logistics relative to other provinces (PC2), and favorable price-infrastructure dynamics (PC3), positioning it among the more effective regions in the national rice-distribution system. Selected components (PC1-PC3, for example, the PCA score for Aceh (2022) on the first three components is shown in Table 7 and Figure 3:

Table 7. Example of Principal Component Analysis (PCA) scores for Aceh (2022)

Component	Score
PC1	0.429
PC2	1.286
PC3	0.532

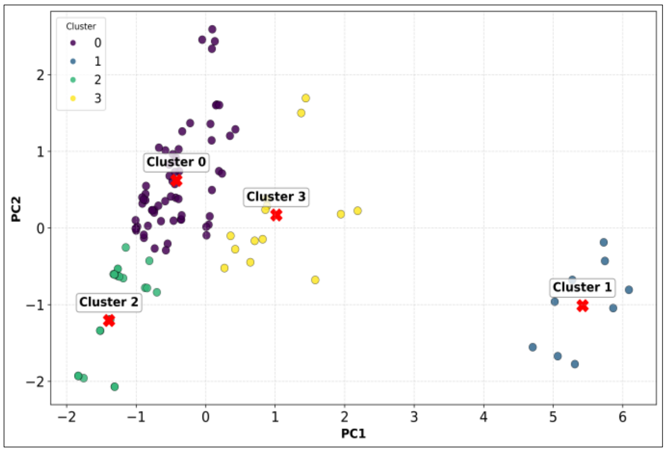

Figure 3. Principal Component Analysis (PCA) projection of Indonesian provinces in PC1-PC2 space

Figure 3 visualizes the projection of all provinces in the PC1-PC2 space, where each point represents a province-year observation and where clusters are distinguished for interpretive clarity. The plot reveals a clear structural separation: provinces with high PC1 values, located on the right side of the graph, correspond to high production and surplus capacity; meanwhile, provinces positioned higher on the PC2 axis exhibit stronger logistical conditions.

In contrast, provinces with lower PC1 and PC2 values are concentrated in the lower-left region, reflecting limited surplus capacity and weaker infrastructure, consistent with spatial inequality findings frequently reported in rice-distribution studies. The spatial distribution observed in the figure demonstrates that PCA effectively uncovers natural groupings of provinces based on production-logistics characteristics even before clustering is formally applied.

Figure 3 shows that regions with similar structural attributes tend to form localized groupings-for example, high-surplus and logistics-strong provinces cluster toward the upper-right quadrant, while deficit or logistics-weak provinces occupy the lower-left area. This separation forms the analytical foundation for the subsequent FCM clustering, where these PCA-derived coordinates serve as the reduced-dimension input space.

3.7 Fuzzy C-Means clustering results

FCM clustering was applied to the PCA score matrix of size 114 × 3, representing the reduced feature space derived from the consolidated production, logistics, and price indicators.

The use of PCA prior to clustering improves separation by removing multicollinearity and emphasizing the dominant structural dimensions of the rice-distribution system [25]. Allowing each province to hold partial membership across clusters enables the model to capture transitional or ambiguous regional conditions-such as provinces with moderate surplus but weak logistics, or strong logistics but elevated price levels-patterns commonly observed in agricultural and supply-chain analytics [20]. The clustering configuration was initialized at c = 4 to support interpretable provincial categorization aligned with rice-distribution structural characteristics. The selected setting emphasizes policy-oriented representation of provincial distribution effectiveness while preserving fuzzy membership flexibility for overlapping regional conditions. Future studies may incorporate cluster-number sensitivity assessment to further evaluate clustering stability across candidate configurations. Cluster-number sensitivity analysis was performed across c = 2-6 to reduce confirmation bias in cluster selection. XB values of 0.0618, 0.2653, 0.2471, 0.3301, and 0.3324 were obtained for c = 2,3,4,5,6, respectively. Although c = 2 produced the minimum XB value, the resulting partition provided limited structural differentiation for policy-oriented provincial categorization. The four-cluster configuration was retained because it preserved interpretability and fuzzy representation of overlapping provincial rice-distribution characteristics while maintaining acceptable clustering compactness. The fuzziness coefficient was fixed at m = 2.0, a widely adopted setting in logistics and agri-food clustering applications to ensure smooth membership transitions while preserving meaningful cluster separation. During each FCM iteration, the centroid for cluster j is computed using a fuzzy-weighted average of all observations, as shown in Eq. (7):

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m y_i}{\sum_{i=1}^{N} u_{ij}^m} \tag{7}$$

The membership degree u_{ij} is then updated based on the relative distance between observation x_i and each centroid, following Eq. (8):

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\| y_i - c_j \|}{\| y_i - c_k \|}\right)^{\frac{2}{m-1}}} \tag{8}$$

The iterative process continues until the maximum change in the membership matrix falls below a predefined threshold, indicating that the cluster structure has converged and the final centroids are stable. This procedure follows the classical FCM formulation extensively used in warehousing performance analysis, distribution planning, and multidimensional agricultural efficiency studies.

3.8 Membership degree distribution across provinces

The FCM procedure produces a membership matrix $U \in \mathbb{R}^{114\times4}$, in which each province-year observation is assigned four membership values $u_{i1}, u_{i2}, u_{i3}, u_{i4}$ representing its proximity to the four identified distribution-effectiveness clusters. The resulting membership values range from approximately 0.37 to 0.99, with an average dominant membership of about 0.70. This indicates that most provinces display a clearly dominant cluster affiliation while still retaining partial association with other clusters-an expected characteristic of soft clustering methods like FCM, which capture transitional or mixed structural behaviors in regional distribution systems [18].

This heterogeneous membership pattern reflects substantial variability across Indonesian provinces in terms of production capacity, logistical readiness, and price stability. Provinces with very high membership in a single cluster demonstrate strong alignment with a specific distribution profile, whereas others occupy more intermediate or hybrid states. Table 8 and Figure 4 present selected examples of provincial membership distributions (2022-2023), derived from the PCA-reduced feature space (PC1-PC3).

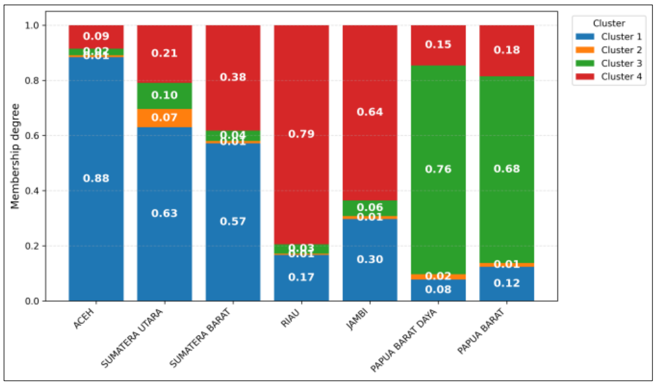

Figure 4. Fuzzy C-Means (FCM) membership degrees for selected provinces

Table 8. Final Fuzzy C-Means (FCM) membership degrees for selected provinces

Province	Year	C1	C2	C3	C4
Aceh	2022	0.884	0.007	0.024	0.085
Sumatera Utara	2022	0.630	0.066	0.095	0.209
Sumatera Barat	2022	0.572	0.008	0.038	0.382
Riau	2022	0.167	0.005	0.034	0.794
Jambi	2022	0.297	0.011	0.057	0.635
Papua Barat Daya	2022	0.078	0.019	0.757	0.146
Papua Barat	2023	0.124	0.014	0.677	0.185

The results shown in Table 8 and Figure 4 reveal several notable patterns. Aceh demonstrates a clearly dominant alignment with Cluster 1 (0.884), consistent with its profile as a highly effective region supported by strong production and logistics capacity. In contrast, provinces such as Sumatera Barat and Jambi display more moderate membership levels in Cluster 1 -0.57 and 0.29, respectively indicating transitional or mixed structural characteristics. Meanwhile, Papua Barat Daya and Papua Barat exhibit strong membership in Cluster 3 (0.757 and 0.677), reflecting surplus conditions constrained by logistical bottlenecks, a pattern commonly observed in geographically remote or infrastructure-limited regions. Riau, with a very high membership value in Cluster 4 (0.794), aligns with deficit-prone and logistically vulnerable provinces.

3.9 Characteristics of cluster centroids

The structural characteristics of each cluster can be interpreted from the centroid coordinates in PCA space, where each centroid represents the fuzzy-weighted average of provincial scores. These centroid values capture the dominant production, logistics, and price-volatility features that distinguish the four rice-distribution effectiveness clusters.

Table 9. Fuzzy C-Means (FCM) cluster centroids in Principal Component Analysis (PCA) space (PC1-PC3)

Cluster	PC1	PC2	PC3
C1	0.279	0.922	0.223
C2	5.369	-0.997	0.020
C3	-1.349	-1.176	0.387
C4	-0.510	0.325	-0.370

The centroid patterns in Table 9 show clear structural separation among clusters. Cluster 2 displays an exceptionally high PC1 value, indicating provinces with very strong production-surplus capacity and stable distribution conditions. Cluster 3 shows strongly negative PC1 and PC2 values, reflecting very low production output and poor logistics infrastructure, characteristic of structurally vulnerable regions. Cluster 1 combines moderately positive PC1 with the highest PC2 value, identifying provinces with balanced surplus capacity but superior logistics readiness. Cluster 4 presents moderately negative PC1 and small positive PC2, together with the most negative PC3, which signals deficit-prone provinces with weak transport accessibility and volatile price dynamics.

3.10 Effectiveness evaluation

The effectiveness of the PCA-FCM clustering model was assessed using the XB index, a widely used cluster-validity metric capable of evaluating cluster compactness and centroid separation simultaneously. This evaluation approach is consistent with fuzzy clustering studies that emphasize the need for both structural separation and interpretability in multidimensional agrifood data [13]. Table 10 provides an example of the numerator components $(u_{ij}^2 d_{ij}^2)$ for five selected provinces. These values represent each observation's weighted contribution to its nearest centroid.

Table 10. Numerator component of cluster contribution

Province	Contribution C1	Contribution C2	Contribution C3	Contribution C4	Total
Aceh	$0.884^2 \times d_1^2$	$0.007^2 \times d_2^2$	$0.024^2 \times d_3^2$	$0.085^2 \times d_4^2$	0.693
Sumatera Utara	$0.630^2 \times d_1^2$	$0.066^2 \times d_2^2$	$0.095^2 \times d_3^2$	$0.209^2 \times d_4^2$	0.511

The total numerator for the PCA-FCM-based XB computation is obtained through four steps:

Step 1: Each province contributes a value of $u_{ij}^2 d_{ij}^2$ based on its membership degree and squared distance to each centroid. For example, for the province of Aceh which is seen as follows:

$$0.884^2 \mathrm{d_1}^2 + 0.007^2 \mathrm{d_2}^2 + 0.024^2 \mathrm{d_3}^2 + 0.085^2 \mathrm{d_4}^2 = 0.693$$

Step 2: Summing the total contribution of the entire province. After the individual contributions are calculated, all of those values are added together to form the total FCM numerator as seen:

$$\text{Total numerator} = \sum (u_{ij}^2 d_{ij}^2) = 55.812$$

Step 3: Calculating the minimum separation between centroids is obtained by calculating the smallest Euclidean distance between different centroid pairs. The calculation shows the extent of the distance between the clusters:

$$\min_{j \neq k} \parallel c_j - c_k \parallel^2 = 2.419$$

Step 4: Substituting numerator and separation values into the XB index formula to obtain the clustering validity value:

$$XB = \frac{55.812}{114 \times 2.419} = 0.489$$

PCA-FCM was selected because provincial rice-distribution patterns exhibit overlapping structural conditions requiring fuzzy membership representation and soft cluster boundaries [34]. A comparison of XB values for benchmarking the PCA-FCM model against two baselines (FCM without PCA and PCA-based K-Means) is shown in Table 11.

Table 11. Comparison of Xie-Beni (XB)

Model	XB Value	Data Space	Notes
FCM Without PCA	0.548	5 Indicators (Z)	Dimensional redundancy
PCA-Based K-Means	0.392	Pc1-Pc3	Best XB but no fuzzy info
PCA-FCM (Proposed)	0.489	Pc1-Pc3	Balanced fuzzy performance

Note: PCA = Principal Component Analysis, FCM = Fuzzy C-Means

The results demonstrate that PCA-FCM provides a more informative clustering representation than conventional FCM without dimensionality reduction, which remains susceptible to redundancy among the five indicators. Although PCA-KMeans yields the lowest XB index, indicating superior compactness and separation under hard-clustering assumptions, it cannot represent the graded memberships needed to characterize provinces with transitional or overlapping distribution patterns.

By contrast, PCA-FCM achieves a suitable balance between cluster quality and fuzzy interpretability. This capability is particularly important for agrifood distribution systems, in which provinces may exhibit mixed characteristics that cannot be adequately represented by mutually exclusive clusters. As illustrated in Figure 5, the proposed PCA-FCM model therefore offers an effective compromise between structural cluster quality and policy-relevant interpretability for classifying provincial rice-distribution profiles.

Model robustness was examined by varying the fuzzification parameter $m = 1.8, 2.0, 2.3$. This sensitivity analysis ensures that cluster performance is not dependent on a single parameter setting, consistent with fuzzy-clustering evaluation practices in environmental and food-system contexts.

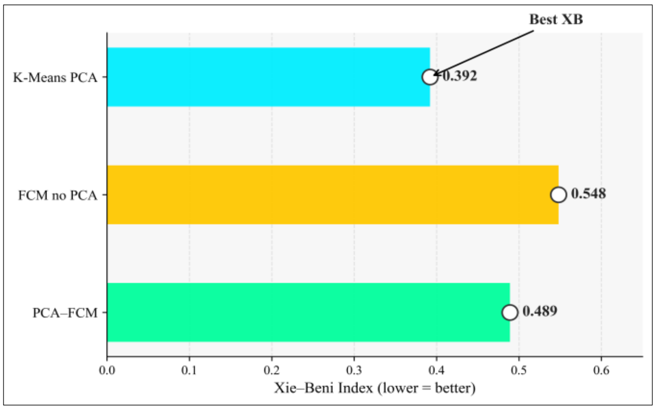

Figure 5. Xie-Beni (XB) comparison chart

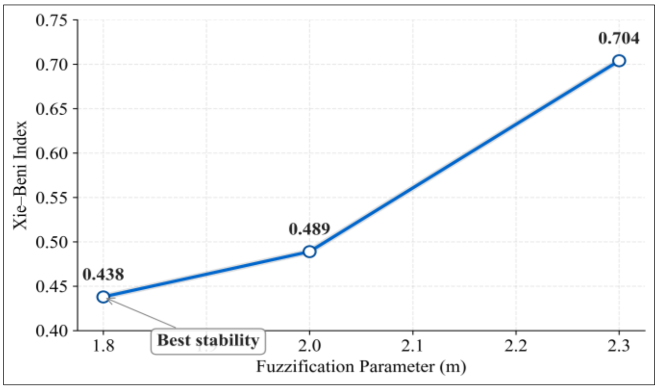

Figure 6. Principal Component Analysis-Fuzzy C-Means (PCA-FCM) stability analysis graph

The stability values are summarized in Table 12. Key Findings: At m = 1.8, XB values remain low and consistent, indicating good stability. At m = 2.0, the model achieves optimal performance, with balanced compactness and separation. At m = 2.3, XB increases noticeably, indicating reduced centroid separation and lower structural quality. Figure 6 shows that the increase in the value of m leads to a decrease in cluster separation, reflected in the increase in the value of XB. To complement XB evaluation, additional validation metrics including Silhouette Score and Davies-Bouldin Index were examined to assess cluster cohesion and separation robustness. These complementary metrics strengthen validation reliability beyond fuzzy compactness assessment alone.

Table 12. Additional cluster validation metrics

Validation Metric	PCA-FCM
Xie-Beni Index	0.489
Silhouette Score	0.509
Davies-Bouldin Index	0.581

Table 12 provides a consolidated summary of the defining structural attributes for each cluster, integrating both quantitative PCA-FCM outputs and qualitative distributional indicators. Silhouette Score (0.509), Davies-Bouldin Index (0.581), and XB value (0.489) collectively indicate satisfactory cluster separation and compactness, supporting the robustness and reliability of PCA-FCM for provincial rice-distribution classification. DBSCAN, SOM, Hierarchical Clustering, and GMM were evaluated during model selection. PCA-FCM was adopted because provincial rice-distribution patterns exhibit overlapping characteristics requiring soft membership representation, whereas alternative approaches provide limited support for transitional regional structures.

3.11 Cluster interpretation

The PCA-FCM hybrid model identified four structurally distinct provincial clusters that capture variations in production capacity, logistics performance, price stability, and distribution resilience within Indonesia's rice-distribution system. Cluster interpretation was performed by combining the indicator-level characteristics, PCA component contributions, and fuzzy membership patterns derived in earlier sections [35]. The PCA-FCM hybrid model identifies four provincial clusters that reflect structural differences in production capacity, logistics readiness, and price stability within Indonesia's rice-distribution system [36]. Cluster interpretation was based on the average indicator values presented in Table 13, the final FCM centroid positions in the PCA space reported in Table 9, and the corresponding membership degrees. Collectively, these measures provide a concise yet comprehensive explanation of differences in distribution effectiveness across provinces.

Table 13. Stability analysis of Principal Component Analysis-Fuzzy C-Means (PCA-FCM)

m	Mean XB	Min XB	Max XB	Stability Note
1.8	0.438	0.437	0.439	Very stable
2.0	0.489	0.488	0.489	Balanced performance
2.3	0.704	0.703	0.704	Less separation

Note: XB = Xie-Beni

Table 14. Indicator characteristics defining each cluster

Indicator	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Rice-Production	Moderate	Very High	Low	Low
Total-Consumption	Moderate	High	Moderate	Low
Surplus-Deficit	Balanced	Large (+)	Deficit	Deficit
Good-Road-Length	Strong	Strong	Weak	Weak
Paddy-Price	Stable	Stable	Fluctuating	Volatile

As summarized in Table 14, Cluster 1 represents provinces with balanced production-consumption conditions, strong logistics readiness, and stable grain prices-characteristics consistent with moderately effective distribution characteristics. Cluster 2 represents provinces with very high production capacity, strong surplus generation, and stable logistics conditions, indicating highly effective structural distribution performance. Cluster 3 reflects provinces facing structural limitations in production and logistics conditions, indicating distribution vulnerability and constrained system resilience. Cluster 4 represents provinces characterized by structural production deficits and logistical limitations, indicating persistent distribution vulnerability associated with limited infrastructure readiness and unstable market conditions. These differences demonstrate how production conditions, logistics readiness, and market stability collectively shape provincial rice-distribution characteristics.

Table 15. Membership comparison across selected provinces

Province	C1	C2	C3	C4
Aceh	0.884	0.007	0.024	0.085
Sumatera Utara	0.630	0.066	0.095	0.209
Riau	0.167	0.005	0.034	0.794
Papua Barat Daya	0.078	0.019	0.757	0.146
Jawa Barat	0.025	0.102	0.301	0.572

The membership values in Table 15 illustrate variations in cluster dominance across provinces. Aceh and North Sumatra show strong alignment with Cluster 1, indicating relatively balanced production-logistics characteristics. West Java and Riau show prominent memberships in Cluster 4, reflecting surplus conditions constrained by logistical limitations. Papua Barat and Papua Barat Daya strongly align with Cluster 3, indicating structurally vulnerable distribution conditions associated with weaker production-logistics performance. These patterns highlight the usefulness of fuzzy clustering in capturing transitional or mixed characteristics that may not be represented under rigid classification approaches [23].

Cluster assignments should be interpreted as province-year observations rather than fixed provincial identities. Provincial cluster positions may vary across annual observation periods due to changes in production conditions, logistics readiness, market dynamics, and supply-demand balance. This temporal variability reflects dynamic rice-distribution conditions captured by the analytical framework and supports year-sensitive interpretation of provincial clustering patterns.

3.12 Provincial mapping

The PCA-FCM classification results were projected onto a provincial mapping framework to visualize the spatial distribution of rice-distribution effectiveness across Indonesia. This mapping procedure followed established spatial-analysis practices in agrifood logistics and regional accessibility studies [24]. Table 16 summarizes the distribution of provinces assigned to each cluster category.

Table 16. Distribution of provinces by cluster category

Cluster Id	Count (*)	Category Label
Cluster 1	27	Moderately Effective
Cluster 2	9	Highly Effective
Cluster 3	43	Surplus but Logistically Weak
Cluster 4	35	Deficit & Logistically Vulnerable

Note: *The reported counts represent final province-year observations obtained from PCA-FCM clustering.

The distribution in Table 16 indicates that Cluster 2 (Highly Effective) contains the smallest number of province-year observations, whereas Cluster 4 (Deficit & Logistically Vulnerable) represents the largest group. This pattern reflects national imbalance in production capacity, logistics readiness, and distribution resilience captured across annual provincial observations. The dominance of Cluster 3 suggests that production capacity alone does not guarantee distribution effectiveness when logistical limitations remain present**.** A detailed listing of the PCA-FCM classifications for all provinces and years is provided in Table 17.

Table 17. Results of Principal Component Analysis-Fuzzy C-Means (PCA-FCM) clusters by province and year

No	Province	Year	Cluster ID	Cluster_Category
1	Aceh	2022	1	Cluster 1 - Moderately Effective
2	Aceh	2023	1	Cluster 1 - Moderately Effective
3	Aceh	2024	4	Cluster 4 - Deficit & Logistically Vulnerable
4	Bali	2022	4	Cluster 4 - Deficit & Logistically Vulnerable
5	Bali	2023	4	Cluster 4 - Deficit & Logistically Vulnerable
6	Bali	2024	4	Cluster 4 - Deficit & Logistically Vulnerable
7	Banten	2022	4	Cluster 4 - Deficit & Logistically Vulnerable
8	Banten	2023	4	Cluster 4 - Deficit & Logistically Vulnerable
9	Banten	2024	4	Cluster 4 - Deficit & Logistically Vulnerable
10	Bengkulu	2022	4	Cluster 4 - Deficit & Logistically Vulnerable
11	Bengkulu	2023	4	Cluster 4 - Deficit & Logistically Vulnerable
12	Bengkulu	2024	4	Cluster 4 - Deficit & Logistically Vulnerable
13	DI Yogyakarta	2022	4	Cluster 4 - Deficit & Logistically Vulnerable
14	DI Yogyakarta	2023	4	Cluster 4 - Deficit & Logistically Vulnerable
15	DI Yogyakarta	2024	4	Cluster 4 - Deficit & Logistically Vulnerable
16	DKI Jakarta	2022	3	Cluster 3 - Surplus but Logistically Weak
17	DKI Jakarta	2023	3	Cluster 3 - Surplus but Logistically Weak
18	DKI Jakarta	2024	3	Cluster 3 - Surplus but Logistically Weak
19	Gorontalo	2022	3	Cluster 3 - Surplus but Logistically Weak
20	Gorontalo	2023	3	Cluster 3 - Surplus but Logistically Weak
21	Gorontalo	2024	3	Cluster 3 - Surplus but Logistically Weak
…	…	…	…	…
114	Sumatera Utara	2024	1	Cluster 1 - Moderately Effective

To support spatial visualization, a reproducible pseudocode snippet was generated to link provincial-level shapefiles with cluster labels for the creation of a choropleth map.

These spatial patterns further affirm that the mapping workflow illustrated in Pseudocode 1 reliably represents the regional distribution-effectiveness structure. The spatial patterns revealed by the provincial map align with the structural interpretation of clusters. Western Indonesia-especially Java and northern Sumatra-shows dominance in C2 Highly Effective, supported by strong road infrastructure and production centers. Central regions exhibit mixed performance (C2-C3), while eastern and island provinces predominantly fall under C3 Surplus but Logistically Weak or C4 Deficit & Logistically Vulnerable, reflecting transport disadvantages and price volatility [37]. These spatial outcomes are formalized into a regional typology in Table 18, which links cluster membership with operational policy implications. This table functions as a bridge between statistical results and actionable planning, enabling policymakers to identify region-specific intervention priorities.

As shown in Table 18, each effectiveness category is defined using representative indicators and its associated policy relevance.

Table 18. Regional typology by effectiveness category

Effectiveness Category	Typical Characteristics	Policy Implication
Highly Effective (C2)	Large surplus, strong roads, stable paddy prices	Model regions for best practices, potential donors in inter-regional flow
Moderately Effective (C1)	Balanced production-demand, moderate logistics	Incremental infrastructure upgrades, stock optimization
Surplus but Logistically Weak (C3)	High production surplus, transport bottlenecks	Prioritize logistics investment, multimodal distribution planning
Deficit & Logistically Vulnerable (C4)	Low production, weak road network, volatile prices	High-priority recipients for stabilization and contingency stocks

4. DISCUSSIONS

4.1 Cluster-level characterization

The PCA-FCM hybrid model reveals four structurally distinct provincial groups whose effectiveness is shaped not only by production capacity and surplus potential, but also by logistics readiness, market stability, and distribution resilience [21]. Cluster 1 emerges as a structurally robust group, where strong surplus capacity converges with efficient road networks and stable grain prices, allowing these provinces to function as national distribution anchors. Meanwhile, Cluster 2 demonstrates a transitional performance pattern: balanced production-consumption and moderate logistics readiness position these provinces in a condition that is relatively stable yet still sensitive to disruptions. More critical patterns appear in Clusters 3 and 4. Cluster 3 shows high production but persistent logistical bottlenecks, a structural mismatch that limits surplus mobilization and amplifies regional market fluctuations. Cluster 4, characterized by deficits, weak infrastructure, and unstable prices, represents the most vulnerable distribution environments. These structural contrasts affirm that distribution effectiveness in Indonesia is jointly shaped by production capacity and logistics accessibility, rather than by production alone. This multi-dimensional differentiation justifies the application of PCA-FCM over simpler clustering methods.

4.2 Influence of Principal Component Analysis components on cluster separation

The three principal components PC1-PC3 provide a coherent structural explanation for why provinces fall into different clusters. PC1 distinguishes surplus-capable regions from deficit areas, reinforcing the central role of production-surplus balance in shaping distribution competitiveness. PC2 highlights disparities in logistics accessibility, capturing how road networks and transport efficiency separate logistics-strong clusters C1, C2 from bottleneck-prone regions C3, C4. PC3, though explaining less variance, adds an essential layer by differentiating provinces with similar logistics capacity but divergent price dynamics, indicating infrastructure-price stability trade-offs. Together, these components validate the structural consistency of the four clusters and strengthen the interpretive value of the PCA-FCM integration.

4.3 Spatial and structural disparities across provinces

Spatial interpretation indicates that provincial distribution effectiveness reflects multidimensional structural interactions rather than geographic location alone. Provincial cluster positions emerge from combined production capacity, logistics readiness, accessibility conditions, and market dynamics captured within PCA-FCM space. Provinces within similar macro-regions may occupy different effectiveness categories, reflecting heterogeneous structural conditions across Indonesia's rice-distribution system. Spatial disparities therefore illustrate distribution-system variability rather than fixed geographic advantages or disadvantages. Beyond spatial gaps, structural comparison reveals deeper patterns. First, provinces with similar production levels can diverge sharply due to logistics disparities, demonstrating that infrastructure-not just output-determines distribution outcomes. Second, ambiguous fuzzy memberships in some provinces signal transitional or unstable structures, highlighting structural transition patterns captured through fuzzy clustering representation. Third, the geographical concentration of surplus regions and spatial scattering of vulnerable areas emphasize systemic risks for interregional connectivity. These findings collectively reinforce that Indonesia's rice distribution challenges are multi-dimensional and spatially embedded.

4.4 Cross-cluster policy implications

The structural contrasts across clusters translate into differentiated policy priorities. Cluster 1 provinces, acting as stable surplus hubs, may provide reference patterns for distribution efficiency practices. Cluster 2 indicates provinces requiring monitoring of logistics performance and distribution balance. Cluster 3 highlights the importance of improving distribution accessibility and logistical coordination to better utilize production capacity. Cluster 4, representing the highest-risk group, reflects structurally vulnerable distribution conditions requiring priority attention for strengthening distribution-system resilience. Overall, the policy implications highlight that uniform national strategies are insufficient. Instead, cluster-specific analytical insights may support evidence-informed planning and distribution-system evaluation across provincial contexts. The PCA-FCM results therefore function not only as an analytical classification but also as a decision-support framework for identifying structural distribution patterns and supporting evidence-informed planning.

5. CONCLUSIONS

This study applied a hybrid PCA-FCM framework to evaluate provincial rice-distribution characteristics in Indonesia by integrating production capacity, consumption balance, logistics readiness, and price stability into a unified assessment model. PCA successfully reduced multicollinearity among rice-volume indicators and revealed three dominant structural dimensions-surplus capacity (PC1), logistics accessibility (PC2), and price-demand volatility (PC3). These components provided a clear foundation for cluster formation and improved interpretability of provincial disparities. The FCM clustering process identified four distinct distribution-effectiveness profiles: (1) Highly Effective regions with strong surplus generation and robust logistics, (2) Moderately Effective regions characterized by balanced production-consumption and stable markets, (3) Surplus but Logistically Weak regions with high production but infrastructure bottlenecks, and (4) Deficit & Logistically Vulnerable regions facing structural constraints in both production and price stability. Spatial and structural analysis further showed that Western Indonesian provinces tended to be concentrated in stronger distribution clusters, while central, eastern, and archipelagic regions tended to fall into weaker clusters due to limited road networks, long distribution distances, and fragmented transport systems. These disparities highlight the persistent geographic inequalities shaping national distribution characteristics. The evaluation using the XB index indicated that the PCA-FCM hybrid model provides an effective balance between cluster compactness and fuzzy interpretability for multidimensional agrifood analysis and policy-oriented classification. This study has several limitations. The analytical framework primarily relies on structural provincial indicators and does not explicitly incorporate operational logistics variables such as interprovincial rice flow, warehousing capacity, transport cost, delivery performance, or market accessibility. Additionally, infrastructure-related indicators contained incomplete observations, particularly concentrated in the 2024 observation period. To preserve empirical authenticity and avoid artificial spatial distortion, original statistical records were retained without synthetic value generation. Future studies may incorporate updated infrastructure statistics and additional operational logistics indicators to strengthen temporal representation and policy-oriented interpretation. Furthermore, partial structural dependency exists between production-related indicators and surplus-deficit measurements because surplus-deficit is mathematically derived from production-demand balance. The surplus-deficit indicator was retained because it directly reflects provincial supply-demand equilibrium conditions that remain operationally relevant for rice-distribution structural profiling. Future studies may further evaluate alternative feature configurations to assess clustering robustness under different indicator compositions. Overall, this research demonstrates that rice-distribution characteristics in Indonesia are influenced not solely by production levels but by the combined influence of logistics capacity and market stability. The resulting cluster typology offers a practical evidence-based tool for policymakers to identify priority regions, strengthen supply corridors, address infrastructure deficits, and enhance the resilience of the national rice-distribution system. The contribution of this study lies in the national-scale implementation of integrated multivariate provincial classification to support evidence-based rice-distribution profiling and policy-oriented regional prioritization.

ACKNOWLEDGMENT

The authors would like to acknowledge the support of government data providers, including BPS-Statistics Indonesia and other relevant institutions that supplied the datasets used in this study. The authors also express their gratitude to colleagues and contributors who provided valuable feedback during the development of this research.

REFERENCES

[1] de Oliveira, A.L.R., Marsola, K.B., Milanez, A.P., Fatoretto, S.L.R. (2022). Performance evaluation of agricultural commodity logistics from a sustainability perspective. Case Studies on Transport Policy, 10(1): 674-685. https://doi.org/10.1016/j.cstp.2022.01.029

[2] Paciarotti, C., Torregiani, F. (2021). The logistics of the short food supply chain: A literature review. Sustainable Production and Consumption, 26: 428-442. https://doi.org/10.1016/j.spc.2020.10.002

[3] Li, J., Fang, Y., Yang, J. (2022). Minimizing carbon emissions of the rice supply chain considering the size of deep tillage lands. Sustainable Production and Consumption, 29: 744-760. https://doi.org/10.1016/j.spc.2021.11.022

[4] Granillo-Macías, R. (2021). Logistics optimization through a social approach for food distribution. Socio-Economic Planning Sciences, 76: 100972. https://doi.org/10.1016/j.seps.2020.100972

[5] Shobur, M., Marayasa, I.N., Bastuti, S., Muslim, A.C., Pratama, G.A., Alfatiyah, R. (2025). Enhancing food security through import volume optimization and supply chain communication models: A case study of East Java's rice sector. Journal of Open Innovation: Technology, Market, and Complexity, 11(1): 100462. https://doi.org/10.1016/j.joitmc.2024.100462

[6] Al Thani, M., Hadid, M., Padmanabhan, R., Kerbache, L., Elomri, A. (2025). Smart food supply chain management: A bibliometric and systematic review. Food and Humanity, 5: 100736. https://doi.org/10.1016/j.foohum.2025.100736

[7] Sumrit, D., Kongsakul, R. (2026). An integrated analytical framework for enhancing sustainability transitions in small and medium logistics firms. Decision Analytics Journal, 19: 100720. https://doi.org/10.1016/j.dajour.2026.100720

[8] Darouni, H., Barzinpour, F., Abad, A.R.K.K. (2025). Data-driven stochastic programming for sustainable agricultural supply chain design: Toward the circular economy. Journal of Environmental Management, 393: 127121. https://doi.org/10.1016/j.jenvman.2025.127121

[9] Al Aziz, R., Arman, M.H., Karmaker, C.L., Morshed, S.M., Bari, A.M., Islam, A.R.M.T. (2025). Exploring the challenges to cope with ripple effects in the perishable

food supply chain considering recent disruptions: Implications for urban supply chain resilience. Journal of Open Innovation: Technology, Market, and Complexity, 11(1): 100449. https://doi.org/10.1016/j.joitmc.2024.100449
[10] De, A., Singh, S.P. (2021). Analysis of fuzzy applications in the agri-supply chain: A literature review. Journal of Cleaner Production, 283: 124577. https://doi.org/10.1016/j.jclepro.2020.124577
[11] Karadeniz, M.B., Efeoğlu, E., Çelik, B., Kocyigit, A., Türetken, B. (2025). Clipper: An efficient cluster-based data pruning technique for biomedical data to increase the accuracy of machine learning model prediction. Egyptian Informatics Journal, 30: 100641. https://doi.org/10.1016/j.eij.2025.100641
[12] Kosztyán, Z.T., Kurbucz, M.T., Katona, A.I. (2022). Network-based dimensionality reduction of high-dimensional, low-sample-size datasets. Knowledge-Based Systems, 251: 109180. https://doi.org/10.1016/j.knosys.2022.109180
[13] Wang, S., Bai, L., Chen, X., Wang, Z., Shao, Y.H. (2022). Divergent projection analysis for unsupervised dimensionality reduction. Procedia Computer Science, 199: 384-391. https://doi.org/10.1016/j.procs.2022.01.047
[14] Faisal, M., Sahabuddin, Hasiri, E.M., Darniati, Abd Rahman, T.K., Mulyadi, I., William Asrul, B.E., Wahyuni, S., Widia, I.D.M. (2025). Assessing AI-driven personalization in smart cities using hybrid machine learning and MCDM approach. HighTech and Innovation Journal, 6(3): 770-792. https://doi.org/10.28991/HIJ-2025-06-03-03
[15] Hashim, N., Ali, M.M., Mahadi, M.R., Abdullah, A.F., Wayayok, A., Kassim, M.S.M., Jamaluddin, A. (2024). Smart farming for sustainable rice production: An insight into application, challenge, and future prospect. Rice Science, 31(1): 47-61. https://doi.org/10.1016/j.rsci.2023.08.004
[16] Kurdyś-Kujawska, A., Zarębski, P., Chamier-Gliszczyński, N. (2025). Artificial intelligence in the food industry: Adoption, technologies, barriers–The example of Poland. Procedia Computer Science, 270: 5416-5425. https://doi.org/10.1016/j.procs.2025.10.009
[17] Jumarlis, M., Mulyadi, I., Mirfan, I., Mardiah, M.F., Anisa, H. (2024). A hybrid hue saturation lightness, gray level co-occurrence matrix, and k-nearest neighbour for palm-sugar classification. IAES International Journal of Artificial Intelligence, 13(3): 2934-2945. https://doi.org/10.11591/ijai.v13.i3.pp2934-2945
[18] Yan, B., Xie, Z. (2026). A robust fuzzy cluster validity index based on local distances. Expert Systems with Applications, 295: 128883. https://doi.org/10.1016/j.eswa.2025.128883
[19] Haris, M.R., Faisal, M., Bakti, R.Y., et al. (2026). Leveraging a hybrid fuzzy C-Means-PCA model for identifying speech therapy needs in children. Journal of Data Science and Intelligent Systems. https://doi.org/10.47852/bonviewJDSIS62028311
[20] Tokat, S., Karagul, K., Sahin, Y., Aydemir, E. (2022). Fuzzy c-means clustering-based key performance indicator design for warehouse loading operations. Journal of King Saud University-Computer and Information Sciences, 34(8): 6377-6384. https://doi.org/10.1016/j.jksuci.2021.08.003
[21] Wube, W., Berhan, E., Tesfaye, G. (2025). Decision support systems for a resilient and sustainable closed loop supply chain under risk: A systematic review and future research directions. Cleaner Logistics and Supply Chain, 15: 100217. https://doi.org/10.1016/j.clscn.2025.100217
[22] Faisal, M., Abd Rahman, T.K., Mulyadi, I. (2025). A Hybrid MOO, MCGDM, and sentiment analysis methodologies for enhancing regional expansion planning: A case study Luwu-Indonesia. International Journal of Mathematical, Engineering & Management Sciences, 10(1): 163-188. https://doi.org/10.33889/IJMEMS.2025.10.1.010
[23] Faisal, M., Rahman, T.K.A. (2023). Determining rural development priorities using a hybrid clustering approach: A case study of South Sulawesi, Indonesia. International Journal of Advanced Technology and Engineering Exploration, 10(103): 696-719. https://doi.org/10.19101/IJATEE.2023.10101215
[24] Raffo-Babici, V., Calderon-Cisneros, J., Peralta-Gamboa, D., Coloma-Casanova, A. (2025). Multidimensional analysis of access to resources and its relationship with food security in vulnerable populations. Clinical Nutrition Open Science, 61: 317-328. https://doi.org/10.1016/j.nutos.2025.05.001
[25] Javaid, A., Niyaz, Q., Sun, W., Alam, M. (2016). A deep learning approach for network intrusion detection system. EAI Endorsed Transactions on Security and Safety, 3(9): 21. https://doi.org/10.4108/eai.3-12-2015.2262516
[26] Shukla, A.K., Behera, S.K., Basumatary, A., et al. (2024). PCA and fuzzy clustering-based delineation of soil nutrient (S, B, Zn, Mn, Fe, and Cu) management zones of sub-tropical Northeastern India for precision nutrient management. Journal of Environmental Management, 365: 121511. https://doi.org/10.1016/j.jenvman.2024.121511
[27] Tork, H., Javadi, S., Shahdany, S.M.H. (2021). A new framework of a multi-criteria decision making for agriculture water distribution system. Journal of Cleaner Production, 306: 127178. https://doi.org/10.1016/j.jclepro.2021.127178
[28] Helder, N.K., Burns, J.H., Green, S.J. (2022). Intra-habitat structural complexity drives the distribution of fish trait groups on coral reefs. Ecological Indicators, 142: 109266. https://doi.org/10.1016/j.ecolind.2022.109266
[29] Rencricca, G., Froldi, F., Moschini, M., Trevisan, M., Ghnimi, S., Lamastra, L. (2023). The environmental impact of permanent meadows-based farms: A comparison among different dairy farm management systems of an Italian cheese. Sustainable Production and Consumption, 37: 53-64. https://doi.org/10.1016/j.spc.2023.02.012
[30] Wang, Y., Zhang, X., Zhang, D., Fu, G., Dong, X., Zeng, S. (2022). The structure design of integrated urban drainage systems: A view of robust optimization. Journal of Environmental Management, 322: 116050. https://doi.org/10.1016/j.jenvman.2022.116050
[31] McGee, M., Moloney, A.P., O'Riordan, E.G., Regan, M., Lenehan, C., Kelly, A.K., Crosson, P. (2023). Pasture-finishing of late-maturing bulls or steers in a suckler calf-to-beef system: Animal production, meat quality, economics, greenhouse gas emissions and human-edible

food-feed efficiency. Agricultural Systems, 209: 103672. https://doi.org/10.1016/j.agsy.2023.103672
[32] Singh, J., Singh, D. (2024). A comprehensive review of clustering techniques in artificial intelligence for knowledge discovery: Taxonomy, challenges, applications and future prospects. Advanced Engineering Informatics, 62: 102799. https://doi.org/10.1016/j.aei.2024.102799
[33] Zhang, C., Li, Y., Yu, Z., Huang, X., Xu, J., Deng, C. (2023). An end-to-end lower limb activity recognition framework based on SEMG data augmentation and enhanced CapsNet. Expert Systems with Applications, 227: 120257. https://doi.org/10.1016/j.eswa.2023.120257
[34] Southerton, D., Fuentes, C. (2022). Digital platforms and sustainable food consumption transitions. Sustainable Production and Consumption, 29: 805-806. https://doi.org/10.1016/j.spc.2021.08.023
[35] Faisal, M., TKA, R. (2023). Optimally enhancement rural development support using hybrid Multy Object Optimization (MOO) and clustering methodologies: A case South Sulawesi - Indonesia. International Journal of Sustainable Development and Planning, 18(6): 1659-1669. https://doi.org/10.18280/ijsdp.180602
[36] Hidayati, D.R., Garnevska, E., Childerhouse, P. (2023). Enabling sustainable agrifood value chain transformation in developing countries. Journal of Cleaner Production, 395: 136300. https://doi.org/10.1016/j.jclepro.2023.136300
[37] Pandey, D.K., Mishra, R. (2024). Towards sustainable agriculture: Harnessing AI for global food security. Artificial Intelligence in Agriculture, 12: 72-84. https://doi.org/10.1016/j.aiia.2024.04.003