Journal of Computing and Smart Ecosystem

Universitas Muhammadiyah Semarang

e-ISSN 3110-5777

Vol. 2 No.1, 2026: 83-93

Provincial Classification Of Child Malnutrition In Indonesia: A Hierarchical Clustering Approach For Evidence-Based Stunting Intervention Prioritization

Inas Nilasari[1*], Nia Eka Safitri[2]

[1] Department Statistic, Faculty of Science and Technology, Universitas Airlangga, Indonesia

[2] Department Environmental Health, Faculty of Public Health, Universitas Airlangga, Indonesia

***Corresponding author**: *inas.nilasari-2023@fst.unair.ac.id*

Article Info:
Received: May 20, 2026
Accepted: June 12, 2026
Available Online: June 30, 2026

Keywords:
Stunting; Hierarchical clustering; K-Means; Indonesia.
.

Abstract: Indonesia's stunting prevalence of 21.5% recorded in SSGI 2024 exceeds the WHO emergency threshold, yet provincial data are commonly presented in tabular form that obscures spatial risk patterns and limits evidence-based resource allocation. This study applies hierarchical clustering to 35 Indonesian provinces using five child nutritional indicators stunting, underweight, severe wasting, wasting, and overweight drawn from SSGI 2024. Prior to clustering, all variables were standardized to place the indicators on a comparable scale and reduce dominance by higher-variance measures. Four linkage methods (single, complete, average, and Ward.D2) were evaluated across k = 2 to 7 using Silhouette Score, Dunn Index, Connectivity, and Cophenetic Correlation Coefficient. Ward.D2 at k = 2 produced the best overall solution, yielding two substantively distinct clusters: Cluster 1 (18 provinces) with a relatively lower undernutrition burden (mean stunting 17.98%; mean wasting 5.72%) and Cluster 2 (17 provinces) with consistently higher risk across undernutrition indicators (mean stunting 26.60%; mean wasting 9.11%). Papua remained structurally outlying across linkage methods, indicating a profile that requires special policy attention. The findings provide an empirical basis for geographically differentiated nutrition policy, enabling priority deployment of intensive interventions in high-burden provinces while sustaining preventive programming in lower-risk provinces.

1. INTRODUCTION

Stunting in children under five constitutes one of the gravest structural threats to Indonesia's human capital trajectory. The condition, defined as height-for-age below two standard deviations of the WHO child growth standard, is not a transient growth delay but a cumulative manifestation of chronic undernutrition during the first 1,000 days of life [1]. Children who experience stunting face measurable deficits in cognitive development, school achievement, and adult productivity, generating macroeconomic losses estimated at 2–3% of GDP annually in high-burden settings [2], [3]. Indonesia has therefore positioned stunting reduction as a national priority through the RPJMN 2020–2024 and the broader Generasi Emas 2045 agenda [4], [5]. Yet national targets cannot be met through uniform interventions alone,

OPEN ACCESS CC BY SA

because the intensity and composition of nutritional risk vary substantially across Indonesia's geographically diverse provinces.

The principal source for monitoring this challenge is the Survei Status Gizi Indonesia (SSGI), which reports provincial prevalence of multiple child nutrition outcomes including stunting, underweight, wasting, severe wasting, and overweight [6]. SSGI 2024 records a national stunting prevalence of 21.5%, still above the WHO public-health threshold of 20%, while the burden is unevenly distributed across provinces [6], [7]. Although these data are routinely published, they are usually communicated as isolated provincial percentages. Such presentation is descriptively informative but analytically limited because it does not reveal which provinces share similar nutritional profiles, which provinces should be prioritized together, or how interventions should be differentiated according to the pattern of deprivation [8]. A grouping approach that organizes provinces by multivariate similarity is therefore needed to translate surveillance data into actionable policy categories.

Cluster analysis offers a suitable framework for this purpose because it identifies internally similar and externally distinct groups directly from the data [9]. In nutrition and public-health research, both partitioning methods such as K-Means and agglomerative hierarchical clustering are commonly used [10], [15]. K-Means is computationally efficient, but it requires the number of clusters to be specified a priori, is sensitive to initial centroid selection, and can be influenced by outliers. Hierarchical clustering is methodologically preferable for the present dataset because the number of provinces is small, exploratory structure discovery is required, and dendrograms provide an interpretable visual summary of inter-provincial similarity. In addition, hierarchical clustering enables comparison across linkage rules single, complete, average, and Ward.D2 so that the final solution is not tied to a single optimization routine. Because the five indicators are measured on the same percentage scale but exhibit different variances, standardization remains essential before distance-based clustering to ensure that no single indicator dominates the grouping process. Cluster validity must also be assessed systematically; for that reason this study uses Silhouette Score, Dunn Index, Connectivity, and Cophenetic Correlation Coefficient to evaluate cohesion, separation, local neighborhood consistency, and fidelity of the dendrogram structure [14].

The policy value of such clustering is direct. Provinces belonging to the same nutritional-risk cluster can be treated as analytically comparable planning units, allowing high-burden clusters to receive intensified supplementary feeding, maternal-child health outreach, and sanitation support, while lower-risk clusters can emphasize prevention and surveillance [13]. Despite the importance of this perspective, province-level studies using recent SSGI data and a transparent cluster-validation framework remain limited in Indonesia. This study therefore aims to classify 35 Indonesian provinces using hierarchical clustering on five nutritional indicators from SSGI 2024, identify the most defensible linkage and cluster configuration, and interpret the resulting clusters for public-health prioritization.

[DOI: 10.14710/JCASE.XX.X.XX-XX]

2. METHODOLOGY

Data Source

This study utilized secondary data from the Indonesian Nutritional Status Survey (*Survei Status Gizi Indonesia*/SSGI), which serves as the primary national reference for child nutritional indicators across Indonesian provinces. The dataset encompassed 38 provinces in total; however, 3 provinces such as Papua Selatan, Papua Tengah, and Papua Pegunungan were excluded from analysis due to missing values on key indicators (Severe Wasting and Wasting), resulting in a final analytical sample of 35 provinces.

Research Variables

Five nutritional indicators for children under five (balita) were employed as clustering variables: (1) Stunting (%), (2) Underweight (%), (3) Severe Wasting (%), (4) Wasting (%), and (5) Overweight (%). These variables collectively capture both undernutrition and overnutrition dimensions, enabling a comprehensive assessment of the double burden of malnutrition at the provincial level.

Sampling Technique

This study adopts a total population sampling approach at the provincial level, incorporating all 35 provinces for which complete data were available. No probabilistic sampling was applied, as the unit of analysis is the administrative province rather than individual children.

Data Collection Method

Data were collected through desk research and secondary data retrieval from the SSGI dataset. All five nutritional indicators were extracted, compiled, and structured into a province-level matrix prior to analysis. Data cleaning procedures were applied to handle missing values, resulting in listwise deletion of the three aforementioned provinces.

Data Analysis Method

The analysis employed agglomerative hierarchical clustering to classify Indonesian provinces into groups with similar nutritional profiles. Euclidean distance was computed from z-score standardized indicators so that each variable contributed comparably to the distance matrix. Four linkage methods single, complete, average, and Ward.D2 were compared across candidate solutions from k = 2 to k = 7. The analytical procedure comprised:

1. Data cleaning and standardization;
2. Construction of the inter-provincial distance matrix;
3. Estimation of hierarchical cluster solutions under the four linkage methods;
4. Selection of the optimal cluster solution using Silhouette Score as the primary criterion, supported by Dunn Index, Connectivity, and Cophenetic Correlation Coefficient;
5. Substantive interpretation of cluster membership and cluster mean profiles for public-health policy;
6. Sensitivity to extreme observations was considered during the interpretation stage because single linkage may produce chaining effects or isolate outliers, whereas Ward.D2 generally generates more compact and internally homogeneous clusters.

3. RESULTS AND DISCUSSION

Hierarchical clustering of 35 provinces based on seven normalized numerical variables yielded a coherent cluster structure, with some differences across linkage methods. Clusters were constructed for k = 2 to k = 7 using single, complete, average, and Ward.D2 linkages, and their quality was assessed using Connectivity, Dunn Index, Silhouette Score, DB Index, and

[DOI: 10.14710/JCASE.XX.X.XX-XX]

Cophenetic Correlation Coefficient as summarized in Table 1 [14]. The upper limit of k = 7 was chosen to avoid over-segmentation (where clusters would contain too few provinces for meaningful interpretation) while still allowing exploration of moderately detailed groupings [15].

Table 1. Cluster Validation Matrix

Linkage	k	Cophenetic Corr.	Connectivity	Dunn Index	Silhouette Avg
single	2	0.684497	2.0	0.343865	0.377316
single	3	0.684497	4.5	0.311457	0.227636
single	4	0.684497	7.5	0.287094	0.206293
single	5	0.684497	12.5	0.310265	0.274430
single	6	0.684497	15.5	0.301498	0.235226
single	7	0.684497	17.0	0.301047	0.211872
complete	2	0.728579	3.5	0.196048	0.394938
complete	3	0.728579	5.5	0.287094	0.277908
complete	4	0.728579	10.5	0.263578	0.330528
complete	5	0.728579	14.0	0.267948	0.324242
complete	6	0.728579	17.0	0.238263	0.304075
complete	7	0.728579	19.0	0.250644	0.290133
average	2	0.758749	2.0	0.343865	0.377316
average	3	0.758749	5.5	0.287094	0.277908
average	4	0.758749	11.0	0.263578	0.330528
average	5	0.758749	14.5	0.267948	0.324242
average	6	0.758749	16.5	0.267948	0.318843
average	7	0.758749	19.5	0.257446	0.302863
ward	2	0.608907	5.5	0.182014	0.413715
ward	3	0.608907	9.0	0.184183	0.354873
ward	4	0.608907	13.5	0.178133	0.305926
ward	5	0.608907	17.0	0.178133	0.316055
ward	6	0.608907	17.5	0.238263	0.317804
ward	7	0.608907	20.0	0.238263	0.325368

[DOI: 10.14710/JCASE.XX.X.XX-XX]

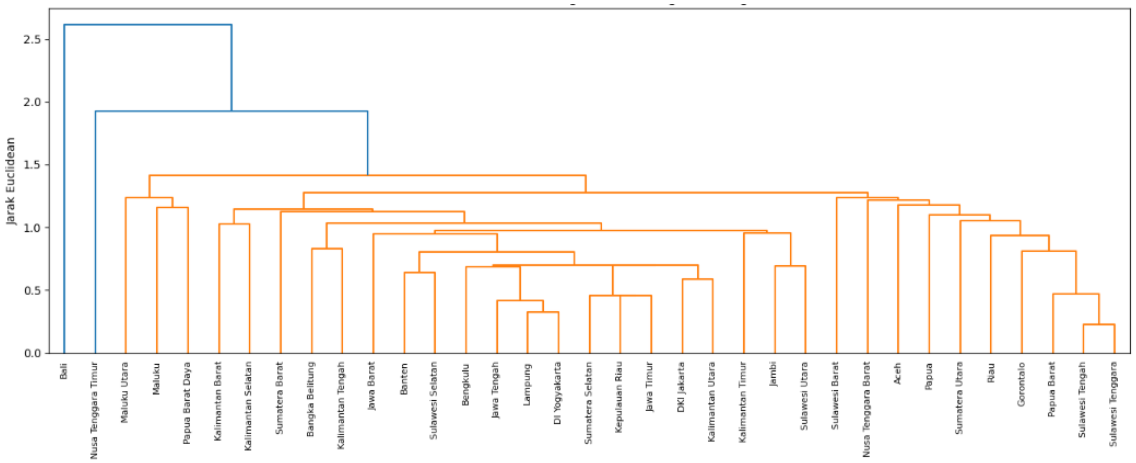


Fig 1. Dendrogram for Single Linkage

The single-linkage dendrogram (Figure 1) exhibits a pronounced chaining pattern, with most provinces merging gradually into one large cluster while Papua appears as a clearly separated branch. For k = 2, single linkage produces an imbalanced solution in which one cluster contains 34 provinces and the other contains only Papua. Its Cophenetic Correlation Coefficient is 0.684497, indicating moderate preservation of the original pairwise distance structure, but its overall interpretability is weaker than that of the Ward.D2 solution because chaining reduces cluster compactness.

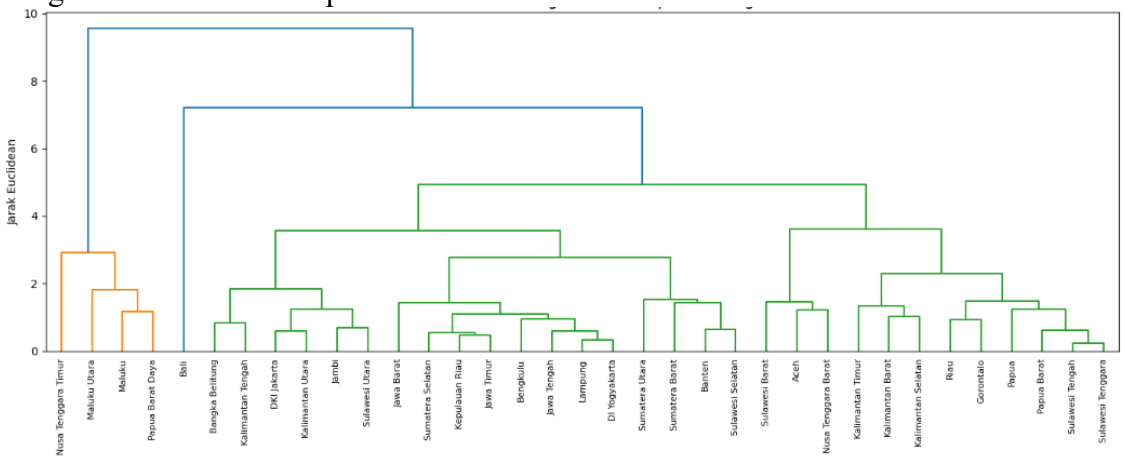
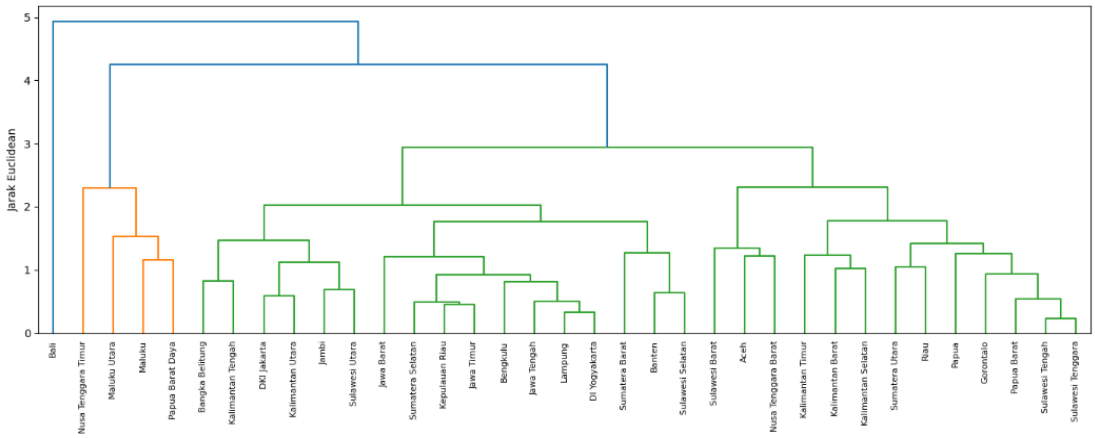


Fig 2. Dendrogram for Complete Linkage



[DOI: 10.14710/JCASE.XX.X.XX-XX]

Fig 3. Dendrogram for Average Linkage

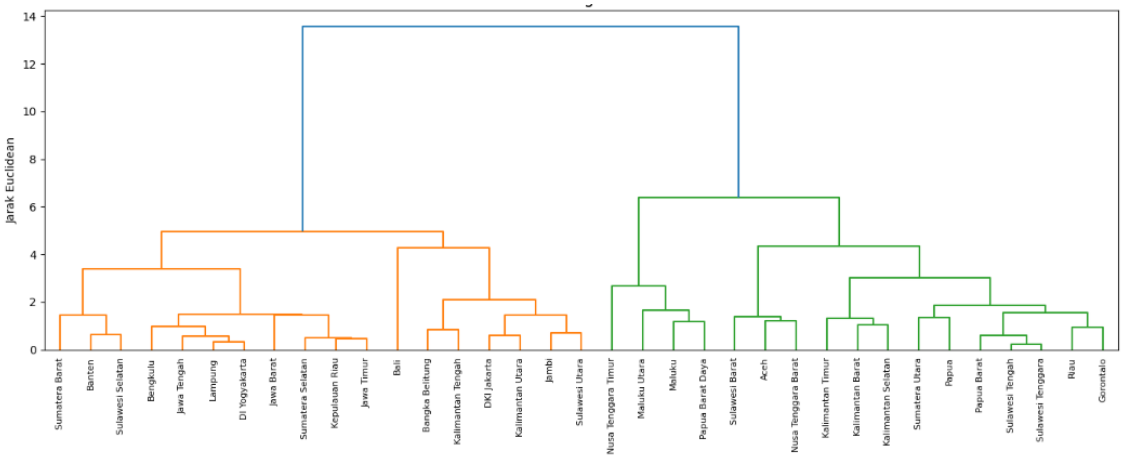


Fig 4. Dendrogram for Ward.D2

The complete linkage dendrogram (Figure 2) reveals more compact clusters, with separations that tend to produce more balanced groupings than those of single linkage. The average linkage dendrogram (Figure 3) yields a pattern intermediate between single and complete linkage, with more regular inter-cluster distances and smoother agglomeration steps. In contrast, the Ward.D2 dendrogram (Figure 4) shows relatively clear cut heights, consistent with Ward's criterion of minimizing the increase in within-cluster variance, which leads to compact and well-separated clusters.

Table 1 reports Connectivity, Dunn Index, Silhouette Score, and Cophenetic Correlation Coefficient for each linkage method and candidate number of clusters. The Silhouette Score was used as the primary selection criterion because it simultaneously reflects within-cluster cohesion and between-cluster separation. In this study, Ward.D2 at k = 2 achieved the highest Silhouette value (0.413715) and was therefore selected as the principal solution for substantive interpretation, while the other linkage methods served as robustness comparisons.

Beyond the Silhouette Score, the Dunn Index is expected to be larger when clusters are dense and well separated, whereas lower Connectivity indicates that neighboring observations in the feature space tend to be assigned to the same cluster. Cophenetic Correlation Coefficients closer to 1 suggest better preservation of the original distance relationships by the dendrogram. Taken together, these metrics support the use of Ward.D2 as the most balanced solution, even though single linkage yields a slightly different representation of distance fidelity.

Table 2. Cluster membership of 35 provinces for the best linkage (Ward.D2)

Cluster 1	Cluster 2
Sumatera Barat, Jambi, Sumatera Selatan, Bengkulu, Lampung, Bangka Belitung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Kalimantan Tengah, Kalimantan Utara, Sulawesi Utara, Sulawesi Selatan	Aceh, Sumatera Utara, Riau, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Selatan, Kalimantan Timur, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua Barat Daya, Papua

[DOI: 10.14710/JCASE.XX.X.XX-XX]

Table 2 presents the distribution of provinces across the two clusters generated by the selected Ward.D2 solution. Cluster 1 contains 18 provinces and is dominated by western and more urbanized areas with comparatively lower burdens of undernutrition, whereas Cluster 2 contains 17 provinces with systematically less favorable nutritional outcomes. Papua remains part of the higher-risk cluster and also appears structurally distant from the main provincial grouping in alternative linkage methods, reinforcing its status as a policy-sensitive outlier.

Table 3. Cross-linkage comparison of cluster assignments for each province

Provinsi	*Cluster_single*	*Cluster_complete*	*Cluster_average*	*Cluster_ward*
Aceh	1	2	1	2
Sumatera Utara	1	2	1	2
Sumatera Barat	1	2	1	1
Riau	1	2	1	2
Jambi	1	2	1	1
Sumatera Selatan	1	2	1	1
Bengkulu	1	2	1	1
Lampung	1	2	1	1
Bangka Belitung	1	2	1	1
Kepulauan Riau	1	2	1	1
DKI Jakarta	1	2	1	1
Jawa Barat	1	2	1	1
Jawa Tengah	1	2	1	1
DI Yogyakarta	1	2	1	1
Jawa Timur	1	2	1	1
Banten	1	2	1	1
Bali	2	2	2	1
Nusa Tenggara Barat	1	2	1	2
Nusa Tenggara Timur	1	1	1	2
Kalimantan Barat	1	2	1	2
Kalimantan Tengah	1	2	1	1
Kalimantan Selatan	1	2	1	2
Kalimantan Timur	1	2	1	2
Kalimantan Utara	1	2	1	1
Sulawesi Utara	1	2	1	1
Sulawesi Tengah	1	2	1	1
Sulawesi Selatan	1	2	1	1
Sulawesi Tenggara	1	2	1	2
Gorontalo	1	2	1	2
Sulawesi Barat	1	2	1	2
Maluku	1	1	1	2
Maluku Utara	1	1	1	2
Papua Barat	1	2	1	2
Papua Barat Daya	1	1	1	2
Papua	1	2	1	2

[DOI: 10.14710/JCASE.XX.X.XX-XX]

Table 3 compares cluster assignments across linkage methods for a common k, showing that most provinces remain in analogous clusters under complete, average, and Ward.D2 linkages, whereas single linkage introduces greater variability due to chaining effects. A few provinces switch clusters when the linkage method changes, highlighting the sensitivity of the cluster structure to the definition of inter-cluster distance.

Table 4. Mean values of variables for each cluster under the best linkage (ward.D2)

Cluster_Hier	Stunting	Underweight	Severe Wasting	Wasting	Overweight
1	17.98	15.01	1.11	5.72	4.07
2	26.60	24.09	2.42	9.11	2

Table 4 shows that Cluster 1 (18 provinces) has lower mean stunting (17.98%), underweight (15.01%), severe wasting (1.11%), and wasting (5.72%), alongside higher overweight (4.07%), indicating a relatively better overall nutritional profile. Cluster 2 (17 provinces) shows markedly higher stunting (26.60%), underweight (24.09%), severe wasting (2.42%), and wasting (9.11%), with lower overweight (2.00%), reflecting a substantially heavier undernutrition burden. These contrasts confirm that the selected clustering solution captures meaningful heterogeneity across provinces.

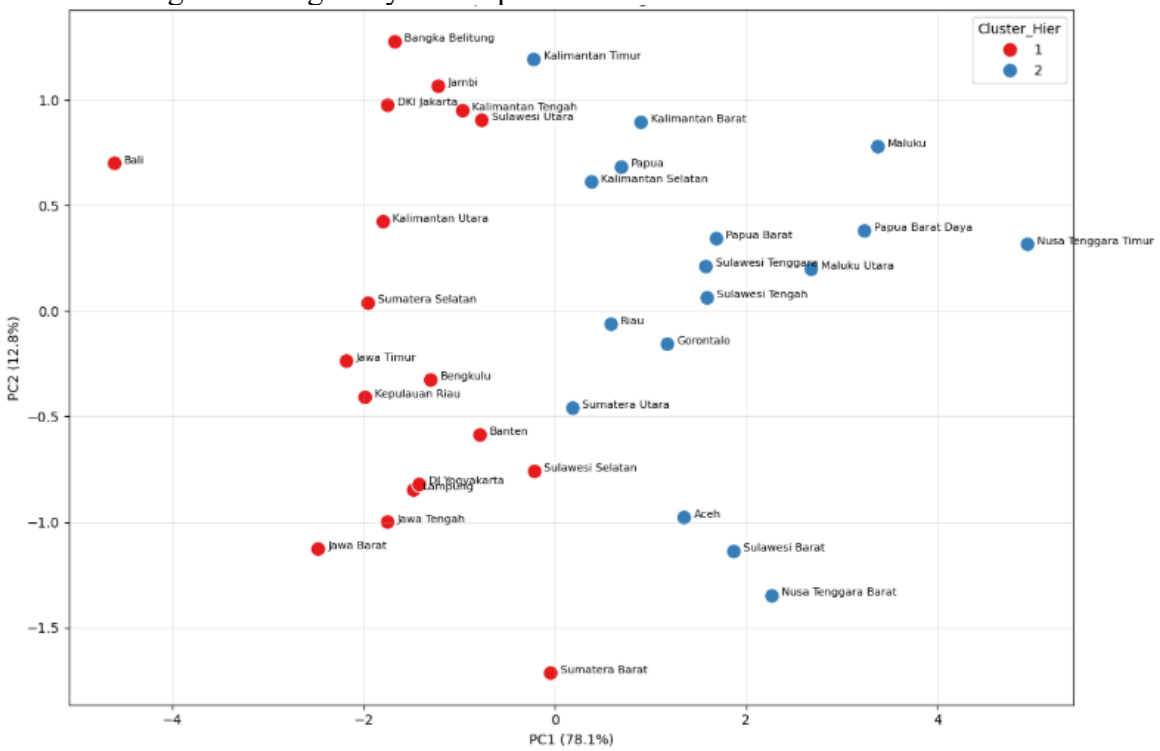


Fig 5. PCA biplot of 34 provinces coloured by Ward.D2 clusters

Figure 5 displays the PCA-based cluster plot using the first two principal components for the selected solution, projecting the 35 provinces into a two-dimensional space coloured by Ward.D2 cluster membership. The plot indicates relatively clear separation between the two clusters, with limited overlap among provinces near the cluster boundary. Papua lies toward the periphery of the point cloud, reinforcing the conclusion that it behaves as a structurally distinct province.

The analysis demonstrates that hierarchical clustering can successfully identify meaningful groups of Indonesian provinces based on five nutritional indicators, with Ward.D2 emerging as the most appropriate linkage in terms of overall cluster validity. Its selection is

[DOI: 10.14710/JCASE.XX.X.XX-XX]

supported by the highest Silhouette Score and by a balanced pattern across the supporting validation indices, which together indicate compact and reasonably well-separated clusters.

The distinct position of Papua in the dendrograms, particularly its isolation under single linkage and its membership in the higher-risk Ward.D2 cluster, underscores substantial differences in nutritional burden relative to many other provinces. Methodologically, this pattern illustrates outlier sensitivity in hierarchical clustering: single linkage is prone to forming chains and isolating extreme observations, whereas Ward.D2 tends to retain such observations within a more policy-interpretable compact structure.

Differences in cluster structure across linkage methods highlight the importance of methodological choice in spatial nutrition mapping. Single linkage is prone to chaining, complete and average linkages yield more compact solutions but may still produce unstable boundaries for borderline provinces, and Ward.D2 minimizes within-cluster variance, making it especially suitable for continuous standardized indicators of this kind [10], [15].

The moderate Silhouette values obtained for Ward.D2 indicate that most provinces are reasonably well assigned, exhibiting shorter distances to members of their own cluster than to members of the alternative cluster. The supporting Dunn, Connectivity, and Cophenetic results reinforce the interpretation that the two-cluster solution is empirically grounded rather than arbitrary.

Cluster-wise mean profiles in Table 4 show two substantively distinct policy groupings: a lower-risk cluster with comparatively better nutritional outcomes and a higher-risk cluster with elevated burdens across all undernutrition indicators. This pattern is directly meaningful for public-health decision making because it supports differentiated intervention intensity rather than one-size-fits-all programming [13].

The comparison of cluster assignments in Table 3 indicates that, despite some method-dependent reassignments, the broad distinction between lower-risk and higher-risk provincial groupings is stable across linkage approaches. Provinces that change clusters across methods should be interpreted as borderline cases requiring contextual policy judgment rather than rigid categorization.

Overall, these findings suggest that hierarchical clustering, particularly Ward.D2 linkage, is a useful tool for mapping heterogeneity in provincial child malnutrition. The identification of two policy-relevant clusters, together with the recurring structural distinctiveness of Papua, provides an empirical basis for geographically targeted interventions while also showing that cluster interpretation should remain attentive to methodological sensitivity.

4. CONCLUSION

This study classified 35 Indonesian provinces into two distinct nutritional-risk clusters using hierarchical clustering with Ward.D2 linkage applied to SSGI 2024 data. Cluster 2, consisting of 17 provinces, demonstrated substantially higher burdens across all undernutrition indicators than Cluster 1, which contained 18 provinces. Papua also appeared as a structurally distinctive unit across linkage methods, indicating the need for tailored policy attention. These findings confirm that Indonesia's child malnutrition burden is geographically heterogeneous in a way that is not fully visible in standard descriptive reporting and support spatially targeted allocation of nutrition-programme resources.

[DOI: 10.14710/JCASE.XX.X.XX-XX]

This study is limited by the exclusion of three provinces because of incomplete data and by its cross-sectional design, which precludes assessment of temporal change. Future research should integrate multi-year SSGI data, incorporate upstream determinants such as poverty, sanitation, and health-service access, and compare the present hierarchical solution with complementary algorithms such as K-Means or DBSCAN to test clustering stability.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Universitas Airlangga for the institutional support and academic environment that facilitated the completion of this research. The authors also extend their appreciation to all parties who have contributed, both directly and indirectly, to the development of this study, including the providers of the Survei Status Gizi Indonesia (SSGI) data, which served as the empirical foundation of the analysis. This study was carried out independently by the authors, **Inas Nilasari** and **Nia Eka Safitri**, as part of their academic contribution toward evidence-based discourse on child nutritional issues in Indonesia.

REFERENCES

[1] WHO, "Stunting in a nutshell," World Health Organization, 2015. [Online]. Available: https://www.who.int/news/item/19-11-2015-stunting-in-a-nutshell
[2] R. E. Black, C. G. Victora, S. P. Walker, Z. A. Bhutta, P. Christian, M. de Onis, M. Ezzati, S. Grantham-McGregor, J. Katz, R. Martorell, and R. Uauy, "Maternal and child undernutrition and overweight in low-income and middle-income countries," *The Lancet*, vol. 382, no. 9890, pp. 427-451, Aug. 2013. Available: https://doi.org/10.1016/S0140-6736(13)60937-X
[3] J. Hoddinott, J. R. Behrman, J. A. Maluccio, P. Melgar, A. R. Quisumbing, M. Ramirez-Zea, A. D. Stein, K. M. Yount, and R. Martorell, "Adult consequences of growth failure in early childhood," *Am. J. Clin. Nutr.*, vol. 98, no. 5, pp. 1170-1178, Nov. 2013. Available: https://doi.org/10.3945/ajcn.113.064584
[4] Bappenas, "Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2020–2024," Bappenas RI, 2019.
[5] Bappenas, "Indonesia 2045: Berdaulat, Maju, Adil dan Makmur," Bappenas RI, 2019.
[6] Kementerian Kesehatan RI, "Survei Status Gizi Indonesia (SSGI) 2024," Badan Kebijakan Pembangunan Kesehatan Kemenkes RI, 2024.
[7] P. Webb, G. A. Stordalen, S. Singh, R. Wijesinha-Bettoni, P. Shetty, and A. Lartey, "Hunger and malnutrition in the 21st century," *BMJ*, vol. 361, p. k2238, Jun. 2018. Available: https://doi.org/10.1136/bmj.k2238
[8] Z. A. Bhutta, J. K. Das, A. Rizvi, M. F. Gaffey, N. Walker, S. Horton, P. Webb, A. Lartey, and R. E. Black, "Evidence-based interventions for improvement of maternal and child nutrition: what can be done and at what cost?" *The Lancet*, vol. 382, no. 9890, pp. 452-477, Aug. 2013. Available: https://doi.org/10.1016/S0140-6736(13)60996-4
[9] J. MacQueen, "Some methods for classification and analysis of multivariate observations," Proc. 5th Berkeley Symp. Math. Stat. Prob., vol. 1, pp. 281-297, 1967.
[10] A. K. Jain, "Data clustering: 50 years beyond K-Means," Pattern Recognition Letters, vol. 31, no. 8, pp. 651-666, 2010. Available: https://doi.org/10.1016/j.patrec.2009.09.011
[11] T. Liao, Y. Wang, J. Zhang, and J. Wang, "Identifying nutritional risk clusters among under-five children in China using K-Means algorithm," *Nutrients*, vol. 14, no. 3, p. 612, Jan. 2022. Available: https://doi.org/10.3390/nu14030612

[12] P. Menon, C. M. McDonald, and S. Chakrabarti, "Cluster-based subnational nutrition targeting in South Asia: implications for programme design," *Food Policy*, vol. 52, pp. S1-S5, Apr. 2015. Available: https://doi.org/10.1016/j.foodpol.2015.03.003
[13] M. T. Ruel and H. Alderman, "Nutrition-sensitive interventions and programmes: how can they help to accelerate progress in improving maternal and child nutrition?" *The Lancet*, vol. 382, no. 9891, pp. 536-551, Aug. 2013. Available: https://doi.org/10.1016/S0140-6736(13)60843-0
[14] Brock, G., Pihur, V., Datta, S., & Datta, S. (2008). *clValid: An R package for cluster validation*. Journal of Statistical Software, 25(4), 1–22. https://doi.org/10.18637/jss.v025.i04
[15] Wulandari, F. S., et al. (2022). Comparison of hierarchical clustering, K-means, K-medoids, and fuzzy C-means. *Indonesian Journal of Statistics and Its Applications*, 6(2). https://doi.org/10.29244/ijsa.v6i2p281-294