Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 1, September 2025
e-ISSN 2685-998X
DOI : 10.30865/json.v7i1.9071

Hybrid DAC-GA and K-Means for Spatial Clustering of Stunting Risk in North Sumatra

Andy Satria[1*], Ibnu Rusydi[1], Dian Septiana[2], Fanny Ramadhani[2]

[1,2] Faculty of Engineering and Computer Science, Universitas Dharmawangsa, Medan, Indonesia
[3,4] Faculty of Mathematics and Natural Science, Medan State University, Medan, Indonesia
Email: [1,*]andysatria@dharmawangsa.ac.id, [2]ibnurusydi@dharmawangsa.ac.id, [3]dianseptiana@unimed.ac.id, [4]fannyr@unimed.ac.id
Correspondence Author Email: andysatria@dharmawangsa.ac.id*
Submitted: **27/08/2025**; Accepted: **23/09/2025**; Published: **30/09/2025**

Abstract− Stunting continues to pose a severe global health concern, particularly in Indonesia, where prevalence rates persist above international standards despite recent advances in reduction initiatives. Accurately documenting the regional variation of stunting is critical to facilitate targeted interventions and successful policymaking. This paper offers a hybrid clustering framework that merges the classic K-Means approach with the Dynamic Artificial Chromosomes Genetic approach (DAC-GA) to increase the resilience and reliability of spatial analysis. The dataset used combines demographic and population statistics from the Central Bureau of Statistics (BPS), strategic policy documents from the Regional Medium-Term Development Plan (RPJMD) of North Sumatra, and health indicators including stunting prevalence data from the Ministry of Health of the Republic of Indonesia.
The research approach consists of four primary phases: data preparation, clustering model construction, cluster evaluation, and geographical visualization. Three evaluation metrics Sum of Squared Errors (SSE), Davies–Bouldin Index (DBI), and Silhouette Coefficient were applied to validate clustering performance. Results demonstrate that DAC-GA dynamically determined the ideal number of clusters at k=2 in just 1.171677 seconds, classifying Kota Medan and Deli Serdang into the low-risk cluster, while all other districts were consistently put into the high-risk cluster. Both DAC-GA and standard K-Means yielded similar spatial maps, giving significant methodological validation and strengthening the dependability of the findings. The study reveals not just the technical advantages of DAC-GA in maximizing clustering but also its practical utility in guiding spatially targeted health interventions. Future research is recommended to add dimensionality reduction utilizing Principal Component Analysis (PCA) to improve computing efficiency and enhance the interpretability of clustering results.

Keywords: Stunting; Spatial Clustering; K-Means; Dynamic Artificial Chromosomes Genetic Algorithm (DAC-GA); North Sumatra

1. INTRODUCTION

Stunting, defined as hindered linear growth due to chronic undernutrition and frequent infections in the crucial first 1,000 days of life, is a significant global public health issue [1][2]. Its long-term consequences transcend mere physical development, including protracted cognitive advancement, lower learning ability, decreased economic production in adulthood, and an elevated risk of noncommunicable diseases. These results together strengthen the cycles of poverty and inequality that last for generations [3]. The most recent UNICEF–WHO–World Bank Joint Child Malnutrition Estimates (JME) say that about 150.2 million children under the age of five were stunted in 2024. This is progress, but it is still far from the 2030 global goal [4]. The slow pace of progress shows how important it is to come up with new, accurate, and data-driven solutions.

Indonesia still has one of the highest rates of stunting in Southeast Asia, hence the government has put the National Strategy for Accelerating Stunting Prevention (Stranas Stunting) into effect for 2021–2024 [5][6]. Data from the 2024 Indonesian Nutrition Status Survey (SSGI) suggest optimistic improvement, with the national prevalence of undernutrition dropping from 24.4% in 2021 to 19.8% in 2024 [7]. Despite this improvement, the frequency remains over the World Health Organization's 20% threshold for severe public health problems, with large regional variations remaining across the archipelago [7][8].

North Sumatra is a good example of these differences, as the incidence of these diseases are always higher than the national norm [7]. These discrepancies are significantly shaped by intricate socio-economic factors, such as maternal education, healthcare accessibility, sanitation, and geographical isolation [9]. Addressing this complexity demands analytical frameworks capable of extending beyond aggregate statistics to uncover specific patterns and high-risk areas. Spatial epidemiology has proven particularly beneficial in this regard, regularly proving that stunting exhibits non-random, regionally concentrated patterns. Empirical research conducted in Indonesia, including Wardana [10] in South Lampung and Ramadhani [2] in North Sumatra, substantiates the existence of significant spatial autocorrelation, with clusters closely associated with poverty and restricted access to health services.

The ongoing digital revolution of health information systems has dramatically transformed public health data collecting, management, and analysis, giving new prospects for advanced research and evidence-based

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 1, September 2025
e-ISSN 2685-998X
DOI : 10.30865/json.v7i1.9071

policymaking [11]. Large-scale digital health databases now facilitate the implementation of data-driven methodologies that transcend traditional statistical frameworks, enabling the examination of intricate, multidimensional, and dynamic health phenomena. Within this paradigm, Machine Learning (ML) has evolved as a powerful analytical tool, particularly suited to modeling non-linear relationships, processing diverse data, and generating high-accuracy prediction models [12]. This skill is particularly important to stunting and malnutrition, conditions driven by complex and linked biological, environmental, and socio-economic elements that frequently defy traditional modeling methodologies.

A increasing amount of global research emphasizes the effectiveness of ML in detecting crucial risk factors and predicting nutritional consequences [13]. Techniques such as Random Forests and Support Vector Machines (SVM) consistently outperform classic regression models by allowing complicated interactions among factors, including household income, maternal education, sanitation, and food security [14]. Applications of these methodologies in South Asia and Sub-Saharan Africa, for example, have discovered region-specific factors of stunting, enabling the creation of more context-sensitive therapies. However, despite these gains, the application of ML to stunting research in Indonesia remains limited. In particular, the integration of ML with spatial analysis essential for detecting geographic grouping and disparities has yet to be thoroughly studied or generally used [15].

In spatial epidemiology, clustering algorithms serve as crucial tools for finding geographic concentrations of disease or malnutrition. Among these strategies, the K-Means algorithm has been widely adopted due to its computational simplicity and scalability for big, multivariate datasets [16]. Nonetheless, K-Means is restricted by inherent restrictions. Its sensitivity to initial centroid selection can lead to unstable results and convergence to local optima, while its performance declines in the presence of high-dimensional, noisy, or diverse data characteristics typical of nutritional and socio-demographic datasets. Such flaws affect the reliability of clustering outcomes, potentially distorting spatial risk evaluations. These mistakes entail real-world consequences, including the misallocation of scarce public health resources, the entrance of biases into policymaking, and the overall reduction in the effectiveness of stunting reduction efforts.

To minimize these constraints, hybrid models merging K-Means with metaheuristic and evolutionary optimization strategies have been developed. Genetic Algorithms (GA), inspired by concepts of natural selection and evolutionary adaptation, are particularly promising in this setting [17]. A new development, the Dynamic Artificial Chromosomes Genetic Algorithm (DAC-GA), presents adaptive solutions for centroid initialization through dynamic chromosomal evolution. This method accelerates the global search process, minimizes the risk of premature convergence, and considerably improves clustering robustness. Prior applications of DAC-GA in health-related fields, like tuberculosis clustering and outbreak prediction, demonstrate its capacity to effectively manage complex, multidimensional datasets characterized by high variability.

Nevertheless, research applying DAC-GA to stunting remains scarce, particularly in Indonesia, where geographical heterogeneity and regional inequities are especially apparent[18]. Against this setting, the present work tackles two key research gaps. First, it tries to extend the underexplored application of optimal spatial ML approaches for stunting study in Indonesia by building and testing a hybrid K-Means–DAC-GA model. Second, it intends to build a robust and fine-grained analytical framework capable of providing high-resolution stunting risk maps specific to the province of North Sumatra. By adopting this innovative analytical approach, the study not only helps to increasing scientific debate on spatial ML applications but also gives a practical decision-support tool for local governments and communities [19]. Importantly, this tool is designed to enhance community health autonomy by enabling stakeholders in North Sumatra to independently identify, evaluate, and resolve important health concerns. Through interventions that are evidence-based, participatory, and context-sensitive, this research emphasizes the transformative potential of digital health technologies in promoting sustainable and resilient public health systems [20].

2. RESEARCH METHODOLOGY

The study follows a defined technique to analyze the performance of a hybrid clustering method that combines the K-Means algorithm with the Dynamic Artificial Chromosomes Genetic Algorithm (DAC-GA). The hybrid model is compared to the classic K-Means technique to test its capacity to increase clustering accuracy, stability, and spatial representation. To ensure clarity and rigor, the entire research design is separated into four phases, as indicated in Figure 1.

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 1, September 2025
e-ISSN 2685-998X
DOI : 10.30865/json.v7i1.9071

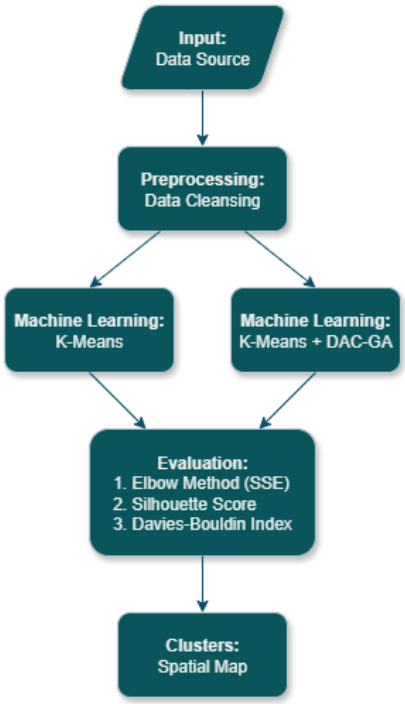

Figure 1. Hybrid DAC-GA K-Means Research Flow

The first phase comprises data preprocessing, which includes data cleaning, transformation, and standardization to assure quality and consistency. The second step focuses on creating clustering models, utilizing both the regular K-Means and the proposed K-Means-DAC-GA hybrid technique. Next, a cluster validation phase assesses the quality and strength of the clustering results using relevant assessment measures. Finally, the data are translated into spatial representations of stunting risk, allowing for the discovery of geographic patterns and clusters with major public health implications. Overall, this method offers a strong workflow that promotes analytical reliability and deepens the relationship between computational methods and practical insights for spatial epidemiology.

2.1 Datasets and Preprocessing Phase

The initial part of the research focuses on preparing the input data through a series of refinement methods aimed at ensuring its quality and dependability for further analysis. This method involves resolving missing values, removing noise, and applying normalization techniques to normalize feature scales. Such preparation is necessary because clustering algorithms particularly distance-based techniques like K-Means are highly sensitive to scale fluctuations and the presence of outliers, which can distort clustering conclusions [21]. For this study, the dataset was compiled from three primary and authoritative sources: demographic and population statistics provided by the Central Bureau of Statistics (BPS), policy information derived from the 2023 Regional Medium-Term Development Plan (RPJMD) of North Sumatra, and health-related indicators, including stunting prevalence, obtained from the Ministry of Health of the Republic of Indonesia. Together, these sources gave comprehensive and dependable data to support the analysis.

2.2 Model Creation

In this stage, a dual-path experimental design conducts the clustering process. The first pathway uses the traditional K-Means algorithm as the baseline model for comparison. The second pathway combines K-Means with the Dynamic Artificial Chromosomes Genetic Algorithm (DAC-GA) to improve the clustering process. This design helps to systematically evaluate the performance of the hybrid model against the conventional K-Means method.

2.2.1 K-Means (Baseline Model)

The baseline model uses the K-Means algorithm, which divides the dataset into k clusters while minimizing variance within each cluster. The process starts with randomly choosing k centroids. It then moves through two main steps iteratively [20]. Given a dataset,

$$D = \{x_1, x_2, \dots, x_n\}, \quad x_i \in \mathbb{R}^m \tag{1}$$

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 1, September 2025
e-ISSN 2685-998X
DOI : 10.30865/json.v7i1.9071

Each observation is allocated to the cluster with the closest centroid according to the Euclidean metric. The average of each cluster's currently assigned points is then used to recalculate its centroid. Until a convergence criterion is satisfied either when the centroid locations stabilize within a predetermined tolerance or when the objective function typically the within-cluster sum of squared distances can no longer be reduced these two steps are repeated. Stated differently, until stability is reached, the approach alternates between an assignment step and a centroid reestimation step. Formally, the goal of K-Means optimization is stated as [22]:

$$min_C \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2, \quad (2)$$

Where $C = \{C_1, C_2, \ldots, C_k\}$ represents the set of clusters, and μ_i denotes the centroid of cluster C_i. calculated as the mean of all data points assigned to that cluster.

The K-Means optimization process works in two main steps. First, in the assignment step, we allocate each data point xp to the cluster with the nearest centroid based on Euclidean distance:

$$c_p = \arg \min_{i \in \{1,\ldots,k\}} \|x_p - \mu_i\|^2, \quad (3)$$

where c_p represents the cluster label of observation x_p. Second, in the update step, the centroids are recalculated to reflect the new cluster memberships:

$$\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x_p, \quad (4)$$

with $|C_i|$ denoting the number of data points in cluster C_i.

The method iteratively repeats these two steps until convergence is achieved. Convergence is generally defined by the stabilizing of centroid positions, wherein subsequent updates yield negligible alterations in the cluster centers, or by the reduction of the objective function, particularly the within-cluster sum of squares (WCSS). In practical applications, convergence may also be assessed using specified stopping conditions, for a maximum iteration count or a tolerance level for centroid displacement. This recurrent refinement guarantees that the final clustering configuration reflects a locally optimal partition of the data, although it may not align with the global optimum due to the algorithm's sensitivity to initial centroid selection. The procedural flow of the typical K-Means algorithm is delineated in the pseudocode below, demonstrating the initialization, assignment, and update phases until the convergence requirements are met:

Input:

- Dataset $D = \{x\_1, x\_2, \ldots, x\_n\}, \& x\_i \in \mathbb{R}\text{^}m$
- Number of clusters k

Output:

- Partition of data into clusters $C = \{C_1, C_2, \ldots, C_k\}$
- Final centroids $\mu = \{\mu_1, \mu_2, \ldots, \mu_k\}$

```
Initialize centroids μ1, μ2, …, μk by randomly selecting k points from X
repeat
    # Assignment Step
    for each data point xp ∈ X do
        Assign xp to the nearest centroid:
            cp ← argmin_i || xp - μi ||²
    end for

    # Update Step
    for each cluster Ci, i = 1, …, k do
```

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 1, September 2025
e-ISSN 2685-998X
DOI : 10.30865/json.v7i1.9071


```
11:         Recalculate centroid μi:
12:             μi ← (1 / |Ci|) * Σ (xp ∈ Ci) xp
13:     end for
14:
15:     # Convergence Check
16: until centroids μi stabilize or maximum iterations reached
17: return final clusters C and centroids μ
```

Although the K-Means technique is widely known and commonly used and is computationally efficient, it comes with a number of acknowledged limitations. One of the key concerns is that K-Means is sensitive to the beginning position of the centroids, which might yield inconsistent results and unsatisfactory partitions. An additional difficulty with the K-Means algorithm is that it has no means of assisting a user determine the ideal number of clusters. Thus, the selection of clusters (k) will still be a challenge. Moreover, K-Means usually converge to the local minimum of the goal function and does not identify the global maximum. This is particularly problematic when doing clustering in the presence of noise or in high-dimensionality datasets [23], [24]. These shortcomings indicate to the requirement of applying an extra optimization strategy in the second experimental branch [25]. The purpose of the new optimization is to improve the centroid initialization, better support a global search, and overall provide more stable and consistent clustering conclusions.

2.2.2 K-Means Optimization with the DAC-GA

The second part of the experimental design uses the Dynamic Artificial Chromosomes Genetic Algorithm (DAC-GA) as an optimization tool to improve the clustering process[26]. It specifically enhances how centroids are initialized in K-means. DAC-GA is an evolutionary computation method inspired by natural selection and adaptability. It builds on the traditional Genetic Algorithm (GA) by introducing chromosome structures that change during the search process. This adaptive feature helps the algorithm keep population diversity, prevent premature convergence, and balance exploration and exploitation of the solution space [27]. As a result, it addresses the limitations found in standard K-means and traditional GA approaches [28].

In this framework, each candidate solution is represented as an artificial chromosome $X = (M, \Sigma)$, where $M \in \mathbb{R}^{k \times d}$ represents the matrix of centroids, and $\Sigma \in \mathbb{R}_{>0}^{k \times d}$ contains the mutation strengths associated with each gene. The objective function for assessing the quality of a chromosome is based on the within-cluster sum of squares (WCSS), also known as the sum of squared errors (SSE). This is defined as:

$$J(M) = \sum_{i=1}^{k} \sum_{x_p \in C_i(M)} \|x_p - \mu_i\|^2, \tag{5}$$

Where each data point x_p is assigned to the nearest centroid μ_i. For selection, DAC-GA uses fitness-proportionate sampling. In this method, probabilities are proportional to $F(X) = 1 / (J(X) + \varepsilon)$, or tournament-based selection. This approach ensures that higher-quality solutions are more likely to survive.

Reproduction in DAC-GA occurs through recombination and mutation. In crossover, offspring centroids are created by linearly combining the centroids of two parents. Meanwhile, strategy parameters are updated using geometric averaging. Mutation follows a self-adaptive Gaussian process. In this process, both centroids and mutation strengths evolve based on:

$$\sigma'_{il} = \sigma_{il} exp(\tau' z + \tau z_{il}), \quad \mu'_{il} = \mu_{il} + \sigma'_{il} \cdot \epsilon_{il}, \tag{6}$$

where $z, z_{il}, \epsilon_{il} \sim \mathcal{N}(0,1)$, and τ, τ' are global learning rates. A unique feature of DAC-GA is its dynamic adaptation rule. In this rule, mutation strengths Σ change based on search performance. This allows the method to balance exploration and exploitation of the solution space.

Following the creation of offspring, the population is updated using an elitist replacement method that guarantees the survival of the best-performing solutions. The evolutionary cycle progresses until the maximum number of generations T_{max} is reached or convergence is attained, characterized by minor improvement in the objective function across consecutive generations.The resulting solution $X^* = (M^*, \Sigma^*)$, spesifically the optimized centroid matrix M^*, is then passed as the initialization stage for the K-Means clustering process.

Through this integration, the hybrid DAC-GA and K-Means model is aimed to solve two major flaws of standard K-Means: sensitivity to initial centroid placement and vulnerability to local minima. This

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 1, September 2025
e-ISSN 2685-998X
DOI : 10.30865/json.v7i1.9071

methodological innovation offers more consistent, reliable, and interpretable clustering results, particularly in the context of geographic risk mapping for stunting prevalence.

To operationalize the suggested methodology, the hybrid model utilizes DAC-GA as an optimization method for centroid creation, followed by the usual K-means clustering procedure. This combination ensures that the centroids utilized in the clustering step are optimized by evolutionary adaptation, thereby limiting the limits of standard K-means. For clarity and reproducibility, the procedural workflow of the hybrid DAC-GA and K-means method is detailed in the pseudocode below:

```
Input:
```

- Dataset X = {x_1, x_2,…, x_n}, with n data points in $\mathbb{R}^d$
- Number of clusters k
- Population size P
- Maximum generations T_{max}

```
Output:
```

- Optimized centroid set M* = {$\mu_1,\mu_2,\ldots,\mu_k$}
- Final clustering assignment of dataset X

```
Initialize a population of P chromosomes χ = (M, Σ),
    where M is centroid matrix and Σ is mutation strength.
Evaluate fitness of each chromosome:
    F(χ) = 1 / ( Σ Σ ||xp − μi||² + ε )
repeat
    # Selection
    Select parents from population based on fitness.

    # Crossover
    For each pair of parents:
        Generate offspring centroids:
            M_off ← α·M_parent1 + (1−α)·M_parent2
        Update mutation strength Σ_off accordingly.

    # Mutation (self-adaptive)
    For each gene μiℓ in chromosome:
        σiℓ' ← σiℓ * exp(τ'·N(0,1) + τ·Niℓ(0,1))
        μiℓ' ← μiℓ + σiℓ'·N(0,1)

    # Evaluation
    Compute fitness of offspring F(χ_off).

    # Replacement
    Form new population using elitism.

    # Dynamic Adaptation
    Adjust Σ according to diversity and convergence level.

until maximum generations Tmax reached
```

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 1, September 2025
e-ISSN 2685-998X
DOI : 10.30865/json.v7i1.9071


```
27:
28: Select best chromosome χ* = (M*, Σ*).
29:
30: # K-Means Refinement
31: Initialize centroids with M*.
32: repeat
33:     # Assignment Step
34:     For each data point xp ∈ X:
35:         Assign xp to nearest centroid μi:
                cp ← argmin_i ||xp − μi||²
36:
37:     # Update Step
38:     For each cluster Ci, i = 1,…,k:
39:         μi ← (1 / |Ci|) * Σ (xp ∈ Ci) xp
40:
41: until centroids μi stabilize or maximum iterations reached
42:
43: Return final clusters C and optimized centroids M*
```

The pseudocode demonstrates the initial setup of the population, evaluation of potential solutions, implementation of evolutionary operators, and the final refining of clusters using the optimized centroids.

2.3 Model Evaluation

To thoroughly assess the validity of the clustering findings, three internal validation metrics were applied: the Sum of Squared Errors (SSE), the Davies–Bouldin Index (DBI), and the Silhouette Coefficient. These indicators provide distinct viewpoints on cluster compactness, isolation, and assignment appropriateness, which are crucial for confirming the robustness of spatial stunting cluster analysis.

2.3.1 Sum of Squared Errors (SSE)

The Sum of Squared Errors (SSE), commonly referred to as the Within-Cluster Sum of Squares (WCSS), analyzes cluster homogeneity by measuring the squared distance of each data point to its assigned centroid [22]. The formal expression is:

$$SSE = \sum_{i=1}^{k} \sum_{x_p \in C_i} \|\mu_p - \mu_i\|^2 \quad (7)$$

Where k is the number of clusters, x_p denotes a data point in cluster C_i and μ_i represents the centroid of cluster C_i. A lower SSE value suggests tighter clusters and greater homogeneity within each group. In the context of spatial stunting study, lower SSE shows that districts grouped together share very consistent prevalence rates and socio-economic features. However, SSE reduces monotonically as the number of clusters grows, making it necessary to match this measure with additional evaluation criteria or heuristic methods to prevent overestimating the ideal cluster count.

2.3.2 Davies - Bouldin Index (DBI)

The DBI analyzes clustering performance by considering both intra-cluster scatter and inter-cluster separation [29]. It is defined as:

$$DBI = \frac{1}{k}\sum_{i=1}^{k} \max_{j \neq 1} \left(\frac{S_i + S_j}{d(\mu_i, \mu_j)}\right) \quad (8)$$

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 1, September 2025
e-ISSN 2685-998X
DOI : 10.30865/json.v7i1.9071

where $S_i = \frac{1}{|C_i|}\sum_{x_p \in C_i}\|x_p - \mu_i\|$ represents the average dispersion of cluster i, and $d(\mu_i, \mu_j)$ is the Euclidean distance between centroids μ_i and μ_j.

Lower DBI values indicate more compact and well-separated clusters. In spatial stunting analysis, a lower DBI suggests that regions in distinct clusters show clearly different epidemiological characteristics. This difference provides stronger evidence for region-specific intervention policies. However, DBI may be influenced by uneven cluster sizes, so it requires careful interpretation with contextual knowledge [30].

2.3.3 Silhouette Coefficient

The Silhouette Coefficient assesses how appropriately each data point is assigned to its cluster relative to other alternative clusters. It is calculated as [19], [31]:

$$s(x_p) = \frac{b(x_p) - a(x_p)}{max\{a(x_p), b(x_p)\}} \quad (9)$$

where $a(x_p)$ is the average distance of data point x_p to all other points in the same cluster, and $b(x_p)$ is the minimum average distance from x_p to points in a different cluster. The overall silhouette score is obtained by averaging $s(x_p)$ across all data points.

Values vary from –1 to +1, with higher values suggest better-defined clusters. In spatial epidemiology, a high silhouette score suggests that districts are firmly connected with their allocated clusters, while lower scores highlight boundary zones where stunting prevalence may overlap between categories. This diagnostic capacity is particularly crucial for local governments, as transitional or uncertain areas may demand personalized policy responses rather than uniform measures.

3. RESULT AND DISCUSSION

This section delineates the experimental results derived from the implementation of both the conventional K-Means algorithm and the proposed hybrid method, which integrates K-Means with a Dynamic Artificial Chromosome Genetic Algorithm (DAC-GA). The analysis evaluates the efficacy of the algorithms in determining the appropriate number of clusters (k), compares the quality of the resultant clusters, and interprets their geographical distribution for stunting cases in North Sumatra. The assessment is based on three primary metrics: the Elbow Method (SSE), Silhouette Score, and Davies-Bouldin Index (DBI), in addition to Execution Time and Spatial Visualization.

3.1 Performance Analysis of the Baseline K-Means Algorithm

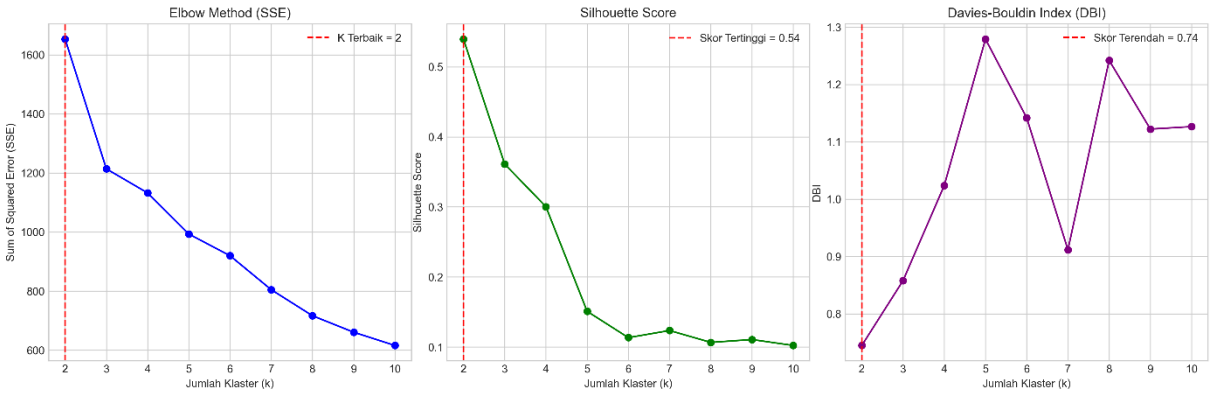

Figure 2. Evaluation Metrics of K-Means for Different K Values

Figure 1 depicts the efficacy of the K-Means clustering model across various k values by showcasing three evaluative dimensions: compactness, separation, and overall clustering quality. The initial graphic illustrates the progression of cluster compactness as the quantity of clusters rises, demonstrating a consistent enhancement with the introduction of additional partitions. The second graphic illustrates how well the clusters are separated, exhibiting sharper distinctions at lower cluster counts and a progressive reduction as the number climbs. Simultaneously, the third graphic illustrates the stability of cluster boundaries, where diminished values signify more dependable structures. Collectively, these visuals offer a summary of the clustering behavior across different

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 1, September 2025
e-ISSN 2685-998X
DOI : 10.30865/json.v7i1.9071

settings. They illustrate how augmenting the number of clusters affects the equilibrium between compactness and separation, while also indicating the range in which cluster quality is most constant. This visual evidence serves as the foundation for evaluating the quantitative facts offered in the ensuing table.

Table 1. *Summary of K-Means Result (k = 2 to k = 10)*

Number of Clusters (k)	Time (s)	SSE	DBI	Silhouette Score
2	1.854508	1653.149229	0.744897	0.539733
3	0.006997	1213.793528	0.857971	0.361179
4	0.009023	1132.367455	1.023950	0.300367
5	0.008425	993.111165	1.279199	0.150877
6	0.010049	919.676259	1.141842	0.113342
7	0.011009	804.601005	0.911717	0.123469
8	0.007514	716.445699	1.242122	0.106540
9	0.009998	660.584178	1.122229	0.110606
10	0.010749	615.754706	1.126576	0.102144

Table 1 contains the detailed numerical values of the evaluation measures across multiple values of *k*. Consistent with the graphical results, the Sum of Squared Errors (SSE) reduces monotonically as the number of clusters grows, showing tighter within-cluster compactness but without necessarily enhancing interpretability. The Silhouette Score gets its highest value of 0.539 with $k = 2$, reinforcing that this configuration generates the most distinct and well-separated clusters. Similarly, the Davies–Bouldin Index (DBI) records its lowest value of 0.744 at $k = 2$, further demonstrating the superiority of the two-cluster approach. While execution times vary significantly between different *k*, they remain within a fairly small range (milliseconds), demonstrating that computational performance is not a limiting constraint. Overall, the tabular evidence substantiates the conclusion drawn from the visual analysis, namely that $k = 2$ provides the ideal balance of compactness, separation, and efficiency for the given dataset.

3.2 Performance Analysis of the Optimized K-Means Algorithm with DAC-GA

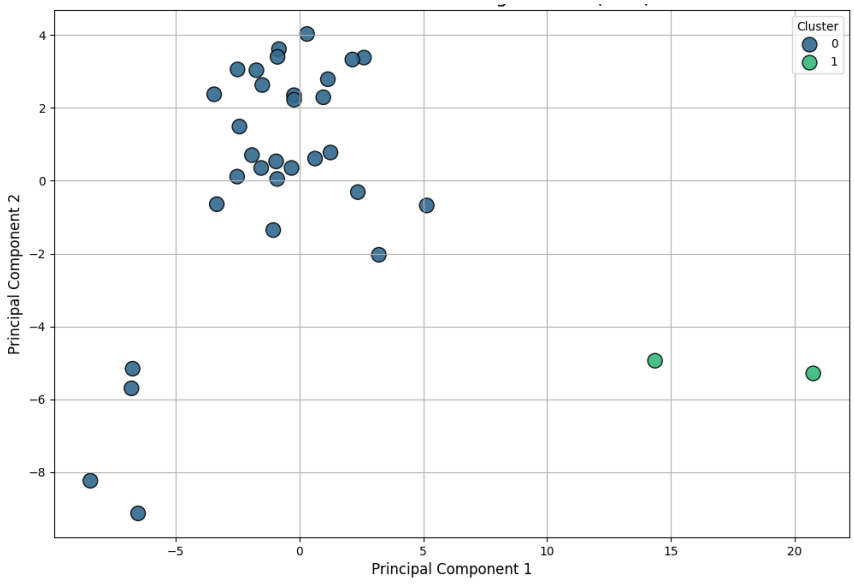

Figure 3. Visualization of Clustering Results

The resulting cluster visualization, displayed in Figure 3, separates the dataset into two distinct groups depending on the best k. The scatterplot, constructed using the first two major components, displays how the algorithm isolates the data points into compact and clearly defined clusters. Cluster 0, displayed in blue, contains most of the observations with a fairly dense spatial distribution. In contrast, Cluster 1, shown in green, isolates a smaller group of data that significantly differs from the main cluster. This division reveals important structural variations within the data, which may relate to areas with different rates of stunting and socio-economic conditions.

The Dynamic Artificial Chromosomes Genetic Algorithm (DAC-GA) showed its effectiveness in determining the best number of clusters without needing extensive calculations for different values of *k*. Unlike

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 1, September 2025
e-ISSN 2685-998X
DOI : 10.30865/json.v7i1.9071

the standard K-Means method, which requires repeated assessments for each possible k, DAC-GA integrates the optimization process within the clustering framework. This leads to a faster convergence toward the best solution. As shown in Table 2, the algorithm found $k = 2$ to be the best cluster configuration in just 1.171677 seconds. This eliminates the need to test other cluster counts. This capability highlights DAC-GA's potential for working with high-dimensional and spatially varied datasets, where speed and reliability are crucial.

Table 2. Optimized K-Means with DAC-GA Evaluation Result

Number of Clusters (k)	Time (s)	SSE	DBI	Silhouette Score
2	1.171677	1653.1492	0.7449	0.5397

Furthermore, the cluster quality evaluation verifies the suitability of the chosen configuration, with a Sum of Squared Error (SSE) of 1653.1492, a Davies–Bouldin Index (DBI) of 0.7449, and a Silhouette Score of 0.5397. These numbers collectively confirm that the clustering outcome generated by DAC-GA achieves a compromise between compactness and separation, demonstrating the robustness of the technique in capturing spatial inequalities in stunting.

3.3 The Spatial Maps

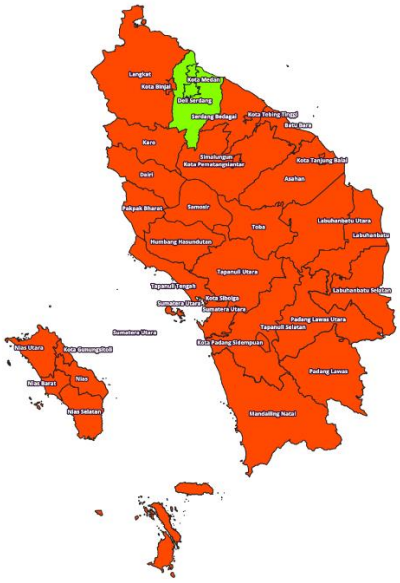

Figure 4. K-Means Spatial Cluster

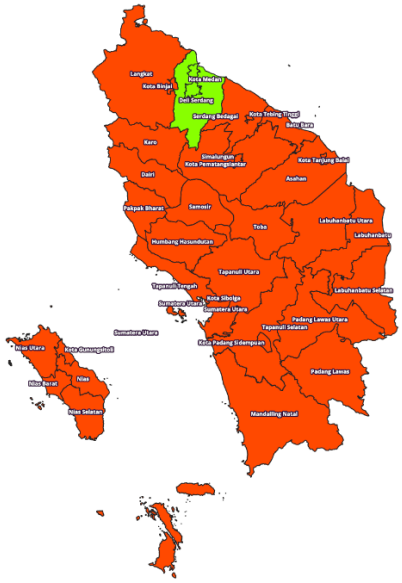

Figure 5. Optimized K-Means + DAC-GA Spatial Cluster

Table 3. Distribution of Stunting by Both Study

Cluster (Color)	Category	Districts/Cities
Green	Low	Kota Medan, Deli Serdang
Red	High	Langkat, Simalungun, Asahan, Kota Binjai, Kota Tanjung Balai, Kota Gunungsitoli, Tapanuli Utara, Tapanuli, Tengah, Tapanuli Selatan, Padang Lawas, Padang Lawas Utara, Mandailing Natal, Humbang Hasundutan, Pakpak Bharat, Labuhan Batu, Labuhan Batu Utara, Labuhan Batu Selatan, Toba, Samosir, Dairi, Karo, Serdang Bedagai, Tebing Tinggi, Batu Bara, Padang Sidempuan, Nias, Nias Utara, Nias Selatan, Nias Barat

Both the K-Means and Optimized K-Means with DAC-GA methods produced the same spatial clustering results, as shown in the maps at Figure 4 and 5. Table 3 lists the cities and districts divided into low-risk stunting and high-risk stunting clusters. It confirms that in both methods, Kota Medan and Deli Serdang consistently fell into the low-risk cluster (green). In contrast, all other districts in North Sumatra were placed in the high-risk cluster (red). This highlights the uniqueness of Kota Medan and Deli Serdang, which have much lower stunting rates compared to the rest of the province.

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 1, September 2025
e-ISSN 2685-998X
DOI : 10.30865/json.v7i1.9071

The consistency between the Optimized K-Means with DAC-GA method, which finds the best cluster arrangement, and the standard K-Means approach provides solid support for the methods. Even though their computational processes differ, both approaches reached the same cluster. This shows that the division between low-risk and high-risk areas is not just a result of method choice, it reflects real differences in the spatial data.

Academically, this result strengthens the trustworthiness and accuracy of the clustering findings. It shows that they are useful for planning spatial policies and directing focused stunting interventions within the province. From a policy view, identifying Kota Medan and Deli Serdang as low-risk areas, compared to most districts in the high-risk group, offers useful insights for the provincial government. The government can focus on targeted interventions and allocate resources to high-risk districts. Successful practices in low-risk areas may serve as models to replicate. This way, the spatial clustering results not only confirm the methodological framework but also provide clear guidance for speeding up stunting reduction efforts in North Sumatra.

4. CONCLUSION

This study shows that combining the Dynamic Artificial Chromosomes Genetic Algorithm (DAC-GA) with the K-Means method effectively analyzes the spatial patterns of stunting prevalence in North Sumatra. Both methods consistently identified Medan City and Deli Serdang as low-risk clusters, while other districts fell into the high-risk cluster. This confirms real spatial differences. Methodologically, DAC-GA can dynamically determine the number of clusters efficiently, converging at k=2 in just 1.171677 seconds without needing multiple testing scenarios. It also handles complex and varied health datasets effectively. Evaluation metrics further confirmed the strength of this approach, ensuring reliable clusters for practical interpretation. These findings offer valuable insights for stunting prevention programs. They allow for more targeted interventions in high-risk districts and provide lessons from low-risk areas. For future research, it is advised to include Principal Component Analysis (PCA) in the preprocessing stage. This can help reduce redundancy, minimize noise, improve efficiency, and clarify cluster formation as well as the understanding of spatial stunting risk patterns.

REFERENCES

[1] P. Ssentongo *et al.*, “Global, regional and national epidemiology and prevalence of child stunting, wasting and underweight in low- and middle-income countries, 2006–2018,” *Sci Rep*, vol. 11, no. 1, pp. 1–12, Dec. 2021, doi: 10.1038/S41598-021-84302-W;SUBJMETA=1702,255,692,699;KWRD=INFECTIOUS+DISEASES,NUTRITION+DISORDERS.

[2] F. Ramadhani, D. Septiana, S. N. Amalia, P. M. Fadilah, and A. Satria, “Spatial Clustering Analysis of Stunting in North Sumatra Based on Environmental Factors Using K-Means Algorithm,” *Data Science: Journal of Computing and Applied Informatics*, vol. 9, no. 2, pp. 18–25, Jul. 2025, doi: 10.32734/JOCAI.V9.I2-17179.

[3] S. Grantham-McGregor, Y. B. Cheung, S. Cueto, P. Glewwe, L. Richter, and B. Strupp, “Developmental potential in the first 5 years for children in developing countries,” *Lancet*, vol. 369, no. 9555, pp. 60–70, Jan. 2007, doi: 10.1016/S0140-6736(07)60032-4/ATTACHMENT/2F933B60-1E68-4B3E-A529-97892C41953B/MMC2.PDF.

[4] C. Hayashi *et al.*, “Levels and trends in child malnutrition - Key findings of the 2025 edition,” Geneva, May 2025. Accessed: Aug. 28, 2025. [Online]. Available: https://iris.who.int/bitstream/handle/10665/381846/9789240112308-eng.pdf

[5] “STRATEGI NASIONAL PERCEPATAN PENCEGAHAN ANAK KERDIL (STUNTING),” Jakarta, Jul. 2019. Accessed: Aug. 28, 2025. [Online]. Available: www.wapresri.go.id

[6] T. Sipahutar, T. Eryando, and M. P. Budhiharsana, “Spatial Analysis of Seven Islands in Indonesia to Determine Stunting Hotspots,” *Health Services Research Commons*, vol. 17, no. 3, Aug. 2022, doi: 10.21109/kesmas.v17i3.6201.

[7] Kementerian Kesehatan, “SSGI 2024 SURVEI STATUS GIZI INDONESIA DALAM ANGKA,” Jakarta, 2025. Accessed: Aug. 28, 2025. [Online]. Available: kemkes.go.id

[8] Development Initiatives, “2022 Global Nutrition Report: Stronger commitments for greater action,” Bristol, 2022.

[9] N. Juniarti, E. Alsharaydeh, C. W. M. Sari, D. I. Yani, and A. Hutton, “Determinant factors influencing stunting prevention behaviors among working mothers in West Java Province, Indonesia: a cross-sectional study,” *BMC Public Health*, vol. 25, no. 1, p. 2719, Dec. 2025, doi: 10.1186/S12889-025-24078-0.

[10] W. Wardana, K. Munibah, and Y. F. Baliwati, “Pola Sebaran Spasial Stunting di Kabupaten Lampung Selatan dengan Pendekatan Autokorelasi Spasial,” *Journal of Regional and Rural Development Planning (Jurnal Perencanaan Pembangunan Wilayah dan Perdesaan)*, vol. 7, no. 1, pp. 68–78, Feb. 2023, doi: 10.29244/JP2WD.2023.7.1.68-78.

[11] M. Bertl, P. Ross, and D. Draheim, "Systematic AI Support for Decision-Making in the Healthcare Sector: Obstacles and Success Factors," *Health Policy Technol*, vol. 12, no. 3, p. 100748, Sep. 2023, doi: 10.1016/J.HLPT.2023.100748.

[12] Y. Kumar, A. Koul, R. Singla, and M. F. Ijaz, "Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda," *Journal of Ambient Intelligence and Humanized Computing 2021 14:7*, vol. 14, no. 7, pp. 8459–8486, Jan. 2022, doi: 10.1007/S12652-021-03612-Z.

[13] S. Kar, S. Pratihar, S. Nayak, S. Bal, H. L. Gururaj, and V. Ravikumar, "Prediction of Child Malnutrition using Machine Learning," *IEMECON 2021 - 10th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks*, 2021, doi: 10.1109/IEMECON53809.2021.9689083.

[14] A. D. Purwanto, K. Wikantika, A. Deliar, and S. Darmawan, "Decision Tree and Random Forest Classification Algorithms for Mangrove Forest Mapping in Sembilang National Park, Indonesia," *Remote Sensing 2023, Vol. 15, Page 16*, vol. 15, no. 1, p. 16, Dec. 2022, doi: 10.3390/RS15010016.

[15] R. Bag *et al.*, "Modelling and mapping of soil erosion susceptibility using machine learning in a tropical hot sub-humid environment," *J Clean Prod*, vol. 364, p. 132428, Sep. 2022, doi: 10.1016/J.JCLEPRO.2022.132428.

[16] W. Hadikurniawati, K. D. Hartomo, and I. Sembiring, "Spatial Clustering of Child Malnutrition in Central Java: A Comparative Analysis Using K-Means and DBSCAN," *Proceedings: ICMERALDA 2023 - International Conference on Modeling and E-Information Research, Artificial Learning and Digital Applications*, pp. 242–247, 2023, doi: 10.1109/ICMERALDA60125.2023.10458202.

[17] S. Katoch, S. S. Chauhan, and V. Kumar, "A review on genetic algorithm: past, present, and future," *Multimed Tools Appl*, vol. 80, no. 5, pp. 8091–8126, Feb. 2021, doi: 10.1007/S11042-020-10139-6/METRICS.

[18] T. Sipahutar, T. Eryando, and M. P. Budhiharsana, "Spatial Analysis of Seven Islands in Indonesia to Determine Stunting Hotspots," *Kesmas*, vol. 17, no. 3, pp. 228–234, Aug. 2022, doi: 10.21109/kesmas.v17i3.6201.

[19] F. Ramadhani, P. M. Fadillah, D. Septiana, A. Satria, and S. N. Amalia, "A Spatial Approach to Analyze the Distribution and Risk Factors of Stunting in North Sumatra With the K-Means Algorithm," *Proceedings of the International Conference on Electrical Engineering and Informatics*, pp. 1–6, 2024, doi: 10.1109/ICELTICS62730.2024.10776521.

[20] D. N. Aisyah *et al.*, "The Information and Communication Technology Maturity Assessment at Primary Health Care Services Across 9 Provinces in Indonesia: Evaluation Study.," *JMIR Med Inform*, vol. 12, no. 1, p. e55959, Jul. 2024, doi: 10.2196/55959.

[21] A. Satria, O. S. Sitompul, and H. Mawengkang, "5-Fold Cross Validation on Supporting K-Nearest Neighbour Accuration of Making Consimilar Symptoms Disease Classification," in *Proceedings - 2nd International Conference on Computer Science and Engineering: The Effects of the Digital World After Pandemic (EDWAP), IC2SE 2021*, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/IC2SE52832.2021.9792094.

[22] X. Ran, X. Zhou, M. Lei, W. Tepsan, and W. Deng, "A Novel K-Means Clustering Algorithm with a Noise Algorithm for Capturing Urban Hotspots," *Applied Sciences 2021, Vol. 11, Page 11202*, vol. 11, no. 23, p. 11202, Nov. 2021, doi: 10.3390/APP112311202.

[23] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Inf Sci (N Y)*, vol. 622, pp. 178–210, Apr. 2023, doi: 10.1016/J.INS.2022.11.139.

[24] I. D. Borlea, R. E. Precup, A. B. Borlea, and D. Iercan, "A Unified Form of Fuzzy C-Means and K-Means algorithms and its Partitional Implementation," *Knowl Based Syst*, vol. 214, p. 106731, Feb. 2021, doi: 10.1016/J.KNOSYS.2020.106731.

[25] A. Biswas and M. S. Islam, "Brain Tumor Types Classification using K-means Clustering and ANN Approach," *International Conference on Robotics, Electrical and Signal Processing Techniques*, pp. 654–658, 2021, doi: 10.1109/ICREST51555.2021.9331115.

[26] M. Mursalim, P. Purwanto, and M. A. Soeleman, "Penentuan Centroid Awal Pada Algoritma K-Means Dengan Dynamic Artificial Chromosomes Genetic Algorithm Untuk Tuberculosis Dataset," *Techno.Com*, vol. 20, no. 1, pp. 97–108, Feb. 2021, doi: 10.33633/TC.V20I1.4230.

[27] W. Zhao, L. Wang, and S. Mirjalili, "Artificial hummingbird algorithm: A new bio-inspired optimizer with its engineering applications," *Comput Methods Appl Mech Eng*, vol. 388, p. 114194, Jan. 2022, doi: 10.1016/J.CMA.2021.114194.

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 7, Nomor 1, September 2025
e-ISSN 2685-998X
DOI : 10.30865/json.v7i1.9071

[28] K. Krishna and M. N. Murty, "Genetic K-means algorithm," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 29, no. 3, pp. 433–439, 1999, doi: 10.1109/3477.764879,.

[29] Y. A. Wijaya, D. A. Kurniady, E. Setyanto, W. S. Tarihoran, D. Rusmana, and R. Rahim, "Davies Bouldin Index Algorithm for Optimizing Clustering Case Studies Mapping School Facilities," *TEM Journal*, vol. 10, no. 3, pp. 1099–1103, Aug. 2021, doi: 10.18421/TEM103-13,.

[30] X. Ran, N. Suyaroj, W. Tepsan, J. Ma, X. Zhou, and W. Deng, "A hybrid genetic-fuzzy ant colony optimization algorithm for automatic K-means clustering in urban global positioning system," *Eng Appl Artif Intell*, vol. 137, p. 109237, Nov. 2024, doi: 10.1016/J.ENGAPPAI.2024.109237.

[31] F. Ramadhani, A. Satria, and S. Salamah, "Implementasi Algoritma Convolutional Neural Network dalam Mengidentifikasi Dini Penyakit pada Mata Katarak," *sudo Jurnal Teknik Informatika*, vol. 2, no. 4, pp. 167–175, Dec. 2023, doi: 10.56211/SUDO.V2I4.408.