Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15748
e-ISSN : 2541-2019
p-ISSN : 2541-044X

An Integrated K-Means and Composite Risk Scoring Framework for Urban Dengue Vulnerability Mapping

Moh. Fachri Alif[1)], Amiq Fahmi[2)*]
[1,2)] Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang, Indonesia
[1)]111202213955@mhs.dinus.ac.id, [2)]amiq.fahmi@dsn.dinus.ac.id

Submitted :Dec 23, 2026 | **Accepted** : Jan 8, 2026 | **Published** : Jan 13, 2026

Abstract: The rising incidence of dengue hemorrhagic fever (DHF) in Indonesian urban areas highlights the urgent need for analytical frameworks capable of capturing spatial heterogeneity in vulnerability while supporting targeted public health interventions. However, most existing dengue vulnerability studies rely on clustering or indicator-based scoring in isolation, limiting interpretability and reducing their operational relevance for policy-driven decision making. This study explicitly addresses this gap by proposing an integrated spatial clustering and epidemiologically weighted composite risk scoring framework for urban dengue vulnerability mapping. Using Semarang Municipality as a case study, K Means based spatial clustering was combined with composite risk scoring to analyze dengue vulnerability across administrative subdistricts. Seven key indicators consisting of population density, area size, total population, morbidity, mortality, incidence rate, and health facility availability were processed through systematic imputation, normalization, and attribute selection to ensure data consistency and analytical robustness. The optimal number of clusters was determined using the Elbow Method and Silhouette Score, after which K-Means clustering was applied to generate spatially coherent vulnerability groupings. A composite risk scoring mechanism was subsequently employed to classify regions into five operational risk categories: Low-Risk, Moderate-Risk, High-Risk, Very High-Risk, and Emergency-Priority. The results reveal clear structural differentiation in dengue vulnerability patterns, where Emergency-Priority and Very High-Risk clusters are not only characterized by elevated epidemiological indicators but also by constrained health service availability, amplifying outbreak susceptibility. Specifically, 13 subdistricts (7.5%) were identified as Emergency-Priority and 22 subdistricts (12.4%) as Very High-Risk, together accounting for approximately 20% of the study area. Beyond numerical classification, the integration of spatial clustering and composite risk scoring enhances interpretability by linking cluster structure with epidemiological severity and service capacity, thereby improving policy relevance compared to conventional clustering-only approaches. Validation through heatmap visualization, risk category distribution, and cluster ranking confirms the stability and interpretive clarity of the proposed framework. By moving beyond descriptive clustering toward an integrated analytical model, this study contributes a scalable and adaptive decision-support framework for dengue risk mapping. The findings provide actionable insights for policymakers, enabling evidence-based prioritization, optimized resource allocation, and the development of responsive intervention strategies to mitigate dengue burden in complex urban environments.

Keywords: Composite risk scoring; dengue hemorrhagic fever; K-Means clustering; vulnerability mapping; Semarang municipality

INTRODUCTION

Dengue hemorrhagic fever (DHF) continues to pose a significant public health challenge in tropical regions, particularly in Indonesia. The spatial distribution of DHF cases in Semarang Municipality during the most recent year (2024) demonstrates substantial heterogeneity across administrative units, both at the district and

*Amiq Fahmi

e-ISSN : 2541-2019
p-ISSN : 2541-044X

neighborhood levels. These variations are strongly influenced by differences in population density, environmental quality, demographic composition, and access to healthcare services. Such complex spatial variability necessitates the adoption of more rigorous analytical frameworks to objectively map regional vulnerability and to strengthen the prioritization of interventions, resource allocation, and evidence-based public health planning.

In contemporary epidemiology, dengue risk mapping has evolved from descriptive approaches toward multidimensional, data-driven methodologies. High-resolution spatial modeling has been shown to identify dengue hotspots more accurately than conventional analytical techniques (Shen et al., 2025). Previous studies emphasize that environmental and demographic determinants, including population density, mobility, and urban spatial structure, play a critical role in shaping dengue risk clusters in densely populated urban settings (Hoque et al., 2025). Beyond spatial determinants, temporal dynamics and climatic conditions also exert considerable influence on risk fluctuations. For instance, a study in Yogyakarta revealed that variations in temperature, humidity, and rainfall significantly affect monthly changes in dengue risk, underscoring the importance of integrated multivariable risk analysis (Salim et al., 2025).

Recent advances in dengue risk mapping increasingly highlight the importance of local-scale spatial analysis to uncover transmission structures across small administrative units. Case variations frequently form distinct clusters at fine spatial resolutions, thereby enabling more precise identification of high-risk areas and supporting targeted public health interventions (Leandro et al., 2024). At the national scale, epidemiological modeling in Brazil demonstrated that climate variability, population density, and land-use change are major determinants of dengue risk clusters, reinforcing the necessity of multidimensional approaches in vulnerability mapping (Araujo et al., 2024).

Despite these advances, several limitations remain in existing dengue vulnerability studies. Although numerous works have applied spatial clustering and GIS-based mapping techniques for dengue risk assessment, most focus primarily on clustering outcomes without integrating epidemiologically weighted composite risk scoring to translate cluster structures into interpretable and operational vulnerability levels. Consequently, many dengue mapping studies remain largely descriptive and lack a standardized analytical framework for converting spatial patterns into actionable public health priorities. This methodological gap limits the effectiveness of clustering-based analyses for decision support, particularly in complex urban environments where clear risk stratification is essential for resource prioritization.

In response to this gap, the present study proposes an integrated analytical framework that combines K-Means–based spatial clustering with epidemiologically weighted risk scoring to assess dengue vulnerability in Semarang Municipality using 2024 data. The analysis incorporates seven key indicators, including dengue incidence rate, number of cases, number of deaths, area size, population density, total population, and availability of healthcare facilities. The analytical workflow includes missing data imputation, variable standardization, optimal cluster determination using the Elbow Method and Silhouette Score, and the construction of composite vulnerability categories consisting of Low Risk, Moderate Risk, High Risk, Very High Risk, and Emergency Priority. By integrating spatial clustering with structured risk scoring, this study moves beyond descriptive mapping and contributes a systematic and operationally meaningful analytical framework that bridges spatial pattern detection and public health prioritization.

This study provides both theoretical and practical contributions. From a theoretical perspective, it enriches the application of spatial epidemiology by integrating K-Means clustering with a composite risk scoring approach to identify dengue vulnerability patterns at the sub-district level. From a practical perspective, the resulting cluster stratification and risk categorization can support local governments and public health authorities in prioritizing surveillance, allocating resources, and implementing early intervention strategies in high-risk areas. The proposed framework offers an interpretable and data-driven approach that can be replicated in other regions with similar epidemiological characteristics.

LITERATURE REVIEW

Spatial Approaches in Dengue Analysis

Spatial analysis has become a critical component in understanding dengue transmission patterns, particularly in densely populated urban environments. Evidence from the Philippines demonstrated that spatial mapping techniques can reveal dengue hotspots not detected through conventional surveillance, thereby assisting governments in defining priority intervention zones (Medina et al., 2023). Similarly, research in Yunnan Province, China, found that spatial clustering approaches successfully detected long-term seasonal shifts in dengue patterns, providing a foundation for early warning systems (Deng et al., 2025). Studies in West Java further indicated that environmental heterogeneity and population density significantly influence dengue distribution patterns (Maramis & Wispriyono, n.d.). In Baubau, Southeast Sulawesi, spatial analysis using Moran's I effectively detected dengue case clustering overlooked by tabular analysis (Agusrawati et al., 2023). Comparable findings in Ho Chi Minh City identified land-use variation and building density as major determinants of dengue risk zones (Thi-Quynh

*Amiq Fahmi

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Nguyen & Thi-Hien Cao, 2023). These studies collectively highlight the methodological importance of spatial approaches in detecting hidden transmission structures and guiding localized interventions.

Collectively, these studies confirm the effectiveness of spatial approaches in uncovering hidden transmission structures and improving localized dengue surveillance. However, most spatial analyses remain focused on pattern detection and hotspot identification, without providing a systematic mechanism to translate spatial clusters into interpretable vulnerability levels that directly support operational public health decision-making.

Environmental and Population Density Factors as Drivers of Dengue Risk

Environmental and climatic factors have consistently been identified as major drivers of dengue transmission dynamics. A study in Semarang Municipality reported that rainfall variability was significantly correlated with DHF incidence between 2017 and 2021, emphasizing the role of water accumulation and mosquito breeding habitats in case surges (Dwi et al., n.d.) Recent spatiotemporal analyses in Indonesia have demonstrated that dengue incidence exhibits strong spatial variability and is significantly associated with environmental and climatic factors, reinforcing the importance of spatial analytical approaches for dengue risk assessment (Mamenun et al., 2024). Recent research in Malaysia demonstrated that temperature, humidity, and vegetation conditions strongly influence dengue hotspot formation, with areas of higher surface temperatures and lower vegetation cover associated with increased mosquito populations and significant case surges (Abdullah et al., 2025). Areas with higher surface temperatures and lower vegetation cover were associated with increased *Aedes aegypti* populations and significant case surges. Climate and urbanization factors have also been shown to shape dengue spatial distribution, where combinations of temperature, rainfall, population density, and urbanization consistently identify high-incidence sub-zones (Gurram et al., 2025). These findings underscore the necessity of integrating environmental and demographic variables into spatial risk models (Magalhães et al., 2023).

While these studies clearly establish the importance of environmental and demographic drivers, most focus predominantly on biophysical variables and population density. Healthcare system capacity and service availability are rarely incorporated as moderating factors, despite their critical role in influencing disease outcomes, surveillance effectiveness, and response capacity in urban settings.

Application of K-Means Clustering in Dengue Risk Scoring

Clustering methods, particularly K-Means, have been increasingly applied in dengue risk analysis to group areas based on shared epidemiological and demographic characteristics. Studies in Jakarta demonstrated that K-Means clustering can effectively classify areas according to DHF risk potential, providing valuable spatial insights for intervention planning (Nurkhairiyyah et al., 2024). Research in East Java further classified districts and cities into dengue risk clusters using K-Means, revealing distinct spatial distributions and facilitating priority area identification (Auditiyah, 2024). Comparative studies have also evaluated hierarchical clustering versus K-Means for grouping provinces based on DHF cases, highlighting the strategic value of clustering methods for disease mitigation planning (Rahmah et al., 2025). These applications confirm the methodological relevance of K-Means clustering in epidemiological vulnerability analysis.

Despite its methodological advantages, most existing applications rely solely on cluster membership as the final analytical output. Quantitative risk scoring and structured vulnerability ranking are generally absent, limiting cluster interpretability and constraining their usefulness for translating analytical findings into prioritized public health actions.

Relevance to the Indonesian Context and Sub-District/Neighborhood-Level Mapping

International studies consistently emphasize the importance of integrating epidemiological, demographic, and environmental variables in dengue clustering models. This approach is particularly relevant for Indonesian regions, including Central Java, which exhibit heterogeneous population density and socio-economic conditions. K-Means clustering is well suited for small administrative units such as sub-districts and urban villages, as it differentiates areas based on incidence, mortality, population density, and healthcare availability. Consequently, cluster-based risk mapping can support local governments in allocating resources more effectively for dengue mitigation and intervention strategies (Lestari, 2025).

Related Work

Several studies share methodological similarities with this research but differ in scope or variables. For example, studies in Semarang applied K-Means clustering without epidemiologically weighted risk ranking. Research in Bandung focused primarily on vegetation indices without incorporating healthcare capacity variables. Studies in Malaysia and Thailand emphasized climate integration but did not employ risk score–based classification. Meanwhile, research in Brazil demonstrated the advantages of big data integration but differed in administrative structure. Addressing these gaps, the present study integrates epidemiological, demographic, and healthcare

*Amiq Fahmi

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15748
e-ISSN : 2541-2019
p-ISSN : 2541-044X

variables within a K-Means clustering framework to produce an operational dengue vulnerability mapping model suitable for sub-district and neighborhood-level decision-making.

However, most Indonesian studies apply clustering techniques in isolation and do not incorporate epidemiologically weighted composite risk scoring. Consequently, dengue vulnerability maps often remain descriptive and lack standardized criteria for translating analytical outputs into operational intervention priorities. Addressing these limitations, the present study integrates epidemiological, demographic, and healthcare capacity indicators within a unified K-Means–based spatial clustering and composite risk scoring framework. More recent studies have also demonstrated the effectiveness of clustering-based approaches and composite indicators for dengue risk mapping, thereby reinforcing the relevance and timeliness of the proposed methodology.

METHOD

A structured methodological framework underpins this study, ensuring that each stage of the research process is systematically organized and logically connected. As depicted in Fig. 1, the sequence commences with data collection, followed by preprocessing to refine and prepare the dataset for subsequent analysis. Modeling is then applied using appropriate computational techniques, with rigorous evaluation conducted to validate performance and reliability. The final stage involves interpreting the model outcomes to generate meaningful insights aligned with the research objectives. This comprehensive workflow reflects a disciplined and iterative approach that strengthens the validity and relevance of the findings.

Fig. 1 Research Sequence

Data Collection

The data employed in this study are secondary data obtained from official institutions, namely the Central Statistics Agency (BPS Kota Semarang, n.d.) and the Semarang Municipality Health Office. These datasets reflect the epidemiological, demographic, and healthcare service capacity conditions of the study area in 2024. The unit of analysis is the administrative area at the urban village level, comprising a total of 177 urban villages with 98 variables initially included in the dataset.

The raw dataset encompasses a broad range of variables describing regional characteristics, including population indicators, environmental conditions, and public health attributes. However, not all available variables were incorporated into the modeling process. Variable selection was conducted based on theoretical justification and empirical relevance to dengue hemorrhagic fever (DHF) risk analysis, ensuring methodological rigor and epidemiological validity.

Following this selection process, seven key variables were retained as input features in the clustering model: dengue incidence rate (IR_DBD-2024), total number of dengue cases, number of dengue-related deaths, area size, total population, population density, and number of healthcare facilities. The dependent variable in this study was not directly observed but was derived through an unsupervised clustering process, which grouped regions according to similarities in epidemiological and socio-demographic characteristics.

In addition to the collection of core variables, descriptive statistical analyses were performed to summarize the distribution and variability of values across regions. These statistics provide an initial overview of the heterogeneity in epidemiological burden and socio-demographic conditions, thereby supporting the robustness of the clustering framework. A summary of the dataset is presented in Table 1, while the selected input variables are listed in Table 2. Together, these tables illustrate the multidimensional structure of the data and establish the foundation for the cluster-based dengue vulnerability model.

Table 1. Dataset Description

Variable	Description	Type
DBD Incidence Rate 2024	Dengue cases per 100,000 population in 2024	Numeric
Number of DBD Patients 2024	Total dengue cases in 2024	Numeric
Area Size	Administrative area in km²	Numeric
Number of DBD deaths 2024	Total dengue related deaths in 2024	Numeric
Population Density	Number of residents per km²	Numeric
Total Population	Total population of the area	Numeric
Number of healthcare facilities	Number of hospitals, health centers, clinics, pharmacies, and other healthcare facilities	Numeric

*Amiq Fahmi

Per Capita Income	Average annual income per capita	Numeric
CFR DBD 2024	Dengue case fatality rate in 2024	Numeric
Dengue Fever Incidence 2024	Dengue fever cases per 100,000 population	Numeric
CFR Demam Dengue 2024	Dengue fever case fatality rate in 2024	Numeric
Total Incidence 2024	Combined dengue and dengue fever incidence rates	Numeric
Total CFR 2024	Combined dengue and dengue fever CFR	Numeric
District	District identifier in Semarang City	Categorical
Urban Village	Urban village identifier in Semarang City	Categorical

Table 2. Selected Variables

Variable	Description	Type
DBD Incidence Rate 2024	Dengue cases per 100,000 population in 2024	Numeric
Number of DBD Patients 2024	Total dengue cases in 2024	Numeric
Area Size	Administrative area in km²	Numeric
Number of DBD deaths 2024	Total dengue related deaths in 2024	Numeric
Population Density	Number of residents per km²	Numeric
Total Population	Total population of the area	Numeric
Number of healthcare facilities	Number of hospitals, health centers, clinics, pharmacies, and other healthcare facilities	Numeric

Data Preprocessing

Several preprocessing techniques were applied to enhance data quality and ensure compatibility with the clustering procedure. First, identity variables such as district and urban village names were excluded, as they do not provide numerical information relevant to distance-based algorithms such as K-Means. Missing values were addressed using median imputation, a method chosen for its robustness against outliers and non-normal distributions, thereby preserving the stability of epidemiological variables.

Subsequently, all numerical variables were normalized using the StandardScaler, which transforms each feature to have a mean of zero and a standard deviation of one (Wongoutong, 2024). The standardization process is mathematically defined as (1).

$$z = \frac{x-\mu}{\sigma} \quad (1)$$

Where x represents the original value of a variable, μ denotes the mean of the variable, and σrepresents the standard deviation. This standardization step was essential because the K-Means algorithm is highly sensitive to differences in variable scales. Previous comparative studies on clustering high-dimensional data have demonstrated that normalization improves cluster stability and interpretability by preventing variables with larger magnitudes from disproportionately influencing the clustering results (Baligodugula & Amsaad, 2025).

To further ensure methodological rigor, Pearson correlation analysis was conducted to assess potential multicollinearity, defined as high correlations among features that may destabilize cluster formation. The Pearson correlation coefficient is defined as (2).

$$r = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum(X-\bar{X})^2(Y-\bar{Y})^2}} \quad (2)$$

Where X and Y represent two variables being compared, and $\bar{X}$ and $\bar{Y}$ denote their respective means. A correlation coefficient value greater than 0.8 indicates a strong correlation and potential multicollinearity. The correlation structure was visualized using a heatmap to facilitate interpretation and guide feature evaluation. Variables exhibiting very high pairwise correlations were carefully evaluated to ensure conceptual distinctiveness, and retained only when they represented different epidemiological dimensions, thereby preventing redundancy while preserving interpretability of the clustering structure. This step ensured that the dataset used for clustering was not only clean and consistent but also statistically interpretable in a transparent manner, thereby strengthening the validity of the subsequent clustering process (Li et al., 2024).

*Amiq Fahmi

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15748
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Modeling

The preprocessed dataset was subsequently utilized as input for the K-Means clustering algorithm to classify administrative areas according to their level of vulnerability to dengue hemorrhagic fever (DHF). The input variables comprised three epidemiological indicators—dengue incidence rate in 2024, total number of dengue cases, and number of dengue-related deaths; two demographic indicators—total population and population density; and two regional and healthcare capacity indicators—area size and number of healthcare facilities. Together, these seven variables capture multidimensional aspects of dengue vulnerability, integrating epidemiological burden, demographic pressure, and healthcare availability.

The K-Means algorithm partitions the dataset into k clusters by minimizing the total within-cluster sum of squared distances between observations and their nearest cluster centroids. Mathematically, the objective function is expressed as (3).

$$Inertia = \sum_{i=1}^{n} \min_{k} \|xi - \mu k\|^2 \quad (3)$$

Where x_i denotes the feature vector of the $i-th$ observation and $\mu_k$represents the centroid of the $k-th$ cluster. This approach has been widely applied in spatial epidemiological studies to identify distribution patterns and high-risk dengue areas, particularly in Southeast Asia.

Table 3. K-Means Parameters

Parameter	Value	Source
Algorithm	K-Means	Scikit-learn
Distance metric	Euclidean	Default
Number of clusters (k)	best_k	Silhouette score
Initialization method	k-means++	Default
Number of initializations	10	Model setting
Random seed	42	Reproducibility
Missing value handling	Median imputation	Preprocessing
Feature scaling	Z-score normalization	StandardScaler
Risk normalization	Min–Max scaling	MinMaxScaler
Evaluation metric	Inertia & Silhouette score	Internal validation

The K-Means algorithm was implemented using the parameters listed in Table 3 to ensure convergence stability and reproducibility of the clustering results. The Euclidean distance metric was employed to measure similarity between observations, while multiple centroid initializations were applied to reduce sensitivity to initial cluster placement. This optimization ensures that each cluster groups administrative areas with similar epidemiological, demographic, and healthcare characteristics, thereby producing spatially coherent vulnerability categories. The clustering results were subsequently integrated with composite risk scoring to generate five vulnerability levels: Low Risk, Moderate Risk, High Risk, Very High Risk, and Emergency Priority.

Determination of Optimal Number of Clusters

The optimal number of clusters was determined by combining the Elbow Method and Silhouette Score analysis. The Elbow Method evaluates the within-cluster sum of squares to identify a point at which further increases in the number of clusters result in diminishing improvements in cluster compactness. Meanwhile, the Silhouette Score measures both cluster cohesion and separation, with higher values indicating better-defined and more distinct clusters. The final number of clusters was selected based on the value of k that achieved the highest silhouette score while remaining consistent with the elbow point observed in the inertia curve.

Model Evaluation

To evaluate the quality of the resulting clusters, internal validation was conducted using the Silhouette Score and the Elbow Method. The Silhouette Score measures how well each observation fits within its assigned cluster compared to other clusters, with values closer to +1 indicating well-defined clusters, values near 0 suggesting overlapping clusters, and negative values indicating potential misclassification. The Elbow Method evaluates total inertia (within-cluster sum of squared distances) across different numbers of clusters, where the “elbow point” represents the optimal number of clusters (k) that balances model complexity and within-cluster variation.

*Amiq Fahmi

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15748
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Although the optimal number of clusters was statistically identified using the Elbow Method and Silhouette Score, the final selection of five clusters was also guided by epidemiological interpretability. This stratification aligns with commonly used public health risk categorization frameworks, which distinguish multiple levels of vulnerability ranging from low transmission risk to emergency priority conditions. The five cluster structure enables direct translation of analytical results into operational categories that support tiered intervention planning and resource prioritization in urban dengue control programs.

After determining the optimal number of clusters, each cluster was further analyzed through the construction of a Composite Risk Score, which integrates multiple indicators: dengue epidemiological variables (incidence rate, total cases, mortality), demographic variables (population density, total population), regional characteristics (area size), and healthcare service capacity (number of facilities). The Risk Score was defined as (4).

$$Risk\ Score = \varpi_1 \cdot IR_{DBD} + \varpi_2 \cdot J\_P + \varpi_3 \cdot J\_M + \varpi_4 \cdot Kepadatan + \varpi_5 \cdot Jumlah\ Penduduk - w_6 \cdot Fasilitas\ Kesehatan - \varpi_7 \cdot Luas\ Wilayah \quad (4)$$

The composite risk score was calculated using a weighted linear combination of normalized indicators. The weighting scheme was determined based on epidemiological relevance and findings from previous dengue surveillance studies. Indicators directly related to dengue incidence and mortality were assigned higher weights, as they reflect immediate disease burden and outbreak intensity, while demographic and infrastructure-related variables were assigned lower weights to represent their moderating effects on transmission risk.

The weights w_i were assigned based on epidemiological relevance. For example, the dengue incidence rate (IR_DBD-2024) received the highest weight (0.35), as it directly reflects transmission intensity. In contrast, area size and healthcare facilities were assigned negative weights, since greater territorial capacity and stronger healthcare infrastructure are associated with reduced vulnerability.

The resulting risk scores were used to rank clusters from lowest to highest risk and to classify regions into five categories: Low Risk, Moderate Risk, High Risk, Very High Risk, and Emergency Priority. This classification framework provides a structured basis for prioritizing public health interventions and resource allocation.

This evaluation procedure ensures that the resulting clusters exhibit both structural validity (through statistical coherence of the clustering process) and epidemiological relevance (through weighted risk interpretation). Consequently, the reliability of regional risk assessment is strengthened, and the recommendations for public health intervention become more defensible.

Furthermore, this approach aligns with international practices in spatial epidemiology and infectious disease hotspot modeling, which emphasize multidimensional risk assessment for evidence-based regional classification and intervention planning (Listyono et al., 2025); (Sena et al., 2025).

Justification of Risk Weighting Scheme

The composite risk score applied in this study is designed as a conceptual epidemiological vulnerability model rather than a purely mathematical aggregation. The weighting scheme reflects established principles in dengue risk assessment frameworks proposed by international public health institutions such as the World Health Organization and the Centers for Disease Control and Prevention, which emphasize transmission intensity, population exposure, and health system capacity as core dimensions of dengue vulnerability.

Epidemiological indicators, particularly dengue incidence rate, were assigned the highest weight because they directly represent active transmission intensity and outbreak potential, consistent with the WHO dengue risk stratification guidelines and previously developed Dengue Vulnerability Indices. Mortality and case counts were weighted as secondary indicators to reflect disease severity and burden. Demographic indicators, including population density and total population, were incorporated to represent exposure pressure and amplification risk in densely populated urban environments.

Conversely, healthcare service availability and area size were assigned negative weights, reflecting their mitigating role in reducing vulnerability through improved access to diagnosis, treatment, and outbreak response capacity. Similar inverse weighting of healthcare infrastructure has been applied in spatial epidemiological risk models to capture system resilience and adaptive capacity. Thus, the resulting risk score functions as an epidemiologically informed vulnerability index that integrates hazard, exposure, and capacity dimensions within a unified analytical framework.

Model Interpretation

After cluster formation and Risk Score calculation, model interpretation was conducted to examine the characteristics of each cluster and their implications for public health interventions. Each cluster was analyzed based on the mean values of key indicators, including dengue incidence, cases per population, population density, total population, area size, and healthcare service capacity. The results revealed clear differentiation of risk levels

*Amiq Fahmi

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15748
e-ISSN : 2541-2019
p-ISSN : 2541-044X

across regions. Clusters with High-Risk Scores were typically characterized by elevated dengue incidence and high population density, combined with limited healthcare facilities, thereby reflecting conditions of heightened vulnerability.

Based on the Risk Score ranking, clusters were classified into five categories: Low Risk, Moderate Risk, High Risk, Very High Risk, and Emergency Priority. This classification framework facilitates spatial mapping and prioritization of interventions such as vector control measures, community education programs, and expansion of healthcare facilities. Correlation analysis among features indicated that population density was positively associated with dengue incidence, whereas the number of healthcare facilities exhibited a negative correlation with risk, underscoring the protective role of healthcare infrastructure.

Visualization of risk category distributions using pie charts and spatial maps enables rapid identification of priority areas and enhances interpretability for policymakers. Overall, this approach supports evidence-based decision-making and promotes efficient, targeted, and data-driven public health strategies. The methodological integration of clustering and composite risk scoring is consistent with international best practices in spatial epidemiology and infectious disease hotspot modeling, which emphasize multidimensional risk assessment for regional classification and intervention planning (Lin et al., 2022; Rahman, 2022).

RESULT

Based on the clustering process using the K-Means algorithm with seven key features, the optimal number of clusters was determined to be five clusters (k = 5). The selection of the optimal number of clusters was guided by a combination of the Elbow Method and the Silhouette Score, ensuring both the stability and quality of cluster separation.

The Elbow Method was applied to evaluate changes in total inertia across different numbers of clusters. As illustrated in Fig. 2, the Elbow Method graph indicates a substantial decrease in inertia up to k = 5, followed by a gradual flattening. This pattern suggests an optimal balance between model complexity and intra-cluster variation, thereby supporting the selection of five clusters as the most appropriate solution.

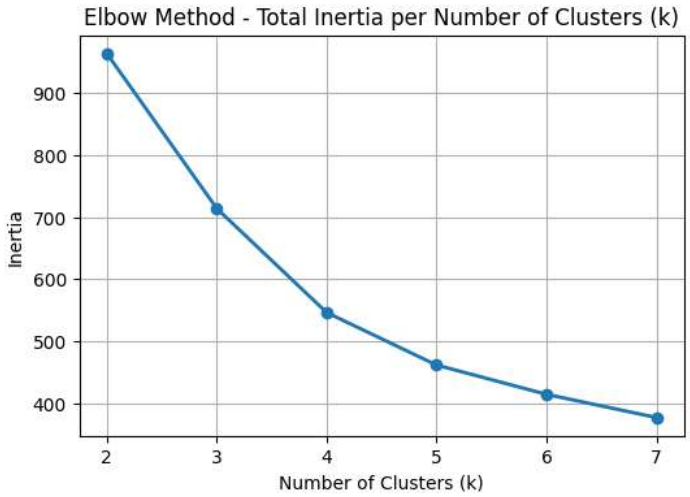

Fig. 2 Elbow Method Graph

To further validate the clustering quality, the Silhouette Score was employed to measure intra-cluster cohesion and inter-cluster separation. As shown in Fig. 3, the Silhouette Score graph demonstrates the highest value at k = 5, indicating that the resulting cluster structure provides the most optimal separation and compactness compared to other cluster configurations.

After determining the optimal number of clusters, a Composite Risk Score was calculated for each cluster based on the weighted contributions of epidemiological indicators, population density, total population, area size, and the availability of healthcare facilities. The summary of average feature values and corresponding Risk Scores for each cluster is presented in Fig. 4, clearly illustrating distinct risk characteristics across clusters.

*Amiq Fahmi

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15748
e-ISSN : 2541-2019
p-ISSN : 2541-044X

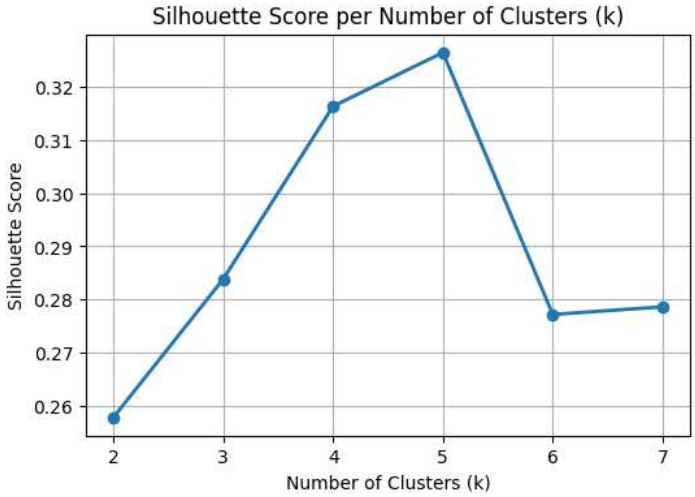

Fig. 3 Silhouette Score Graph

Cluster Characteristics and Risk Score (Normalized Features)

Cluster	DBD Incidence (Norm)	DBD Cases (Norm)	DBD Deaths (Norm)	Area (Norm)	Population Density (Norm)	Total Population (Norm)	Health Facilities (Norm)	Risk Score
0.0	0.13	0.089	0.0	0.096	0.776	0.189	0.178	0.146
1.0	0.214	0.345	0.0	0.431	0.448	0.458	0.229	0.202
2.0	0.268	0.576	1.0	0.435	0.563	0.574	0.533	0.424
3.0	0.076	0.043	0.0	0.779	0.13	0.137	0.074	0.019
4.0	0.615	0.357	0.0	0.641	0.208	0.158	0.032	0.29

Fig. 4 Cluster Summary and Risk Score

To provide a more detailed quantitative comparison between clusters, a numerical summary of key variables was computed in the form of mean and standard deviation (mean ± SD) for each cluster. This summary captures the central tendency and variability of epidemiological, demographic, and healthcare related indicators within clusters, thereby enabling a clearer interpretation of inter cluster differences. The numerical summary confirms that clusters with higher Risk Scores consistently exhibit higher mean values of dengue incidence and case counts, while clusters with lower Risk Scores tend to show lower epidemiological burdens and relatively better healthcare availability.

Table 4. Cluster Characteristics and Risk Score Based on Normalized Features

Cluster	IR_DBD-2024	J_P_DBD-2024	J_M_DBD-2024	Area	Population Density	Total Population	Health Facilities	Risk Score
0	0.130 ± 0.142	0.089 ± 0.089	0.000 ± 0.000	0.096 ± 0.125	0.776 ± 0.188	0.189 ± 0.098	0.178 ± 0.129	0.146
1	0.214 ± 0.085	0.345 ± 0.159	0.000 ± 0.000	0.431 ± 0.193	0.448 ± 0.208	0.458 ± 0.142	0.229 ± 0.121	0.202
2	0.268 ± 0.151	0.576 ± 0.389	1.000 ± 0.000	0.435 ± 0.278	0.563 ± 0.307	0.574 ± 0.304	0.533 ± 0.313	0.424
3	0.076 ± 0.112	0.043 ± 0.063	0.000 ± 0.000	0.779 ± 0.254	0.130 ± 0.141	0.137 ± 0.074	0.074 ± 0.085	0.019
4	0.615 ± 0.205	0.357 ± 0.224	0.000 ± 0.000	0.641 ± 0.310	0.208 ± 0.239	0.158 ± 0.104	0.032 ± 0.054	0.290

A closer examination of the cluster characteristics reveals distinct epidemiological profiles across the five clusters.

Cluster 0 represents the Low-Risk group, characterized by low average dengue incidence and case counts, moderate population density, and relatively adequate healthcare facility availability. These conditions indicate

*Amiq Fahmi

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15748
e-ISSN : 2541-2019
p-ISSN : 2541-044X

minimal dengue vulnerability, suggesting that routine surveillance and preventive measures are sufficient for this group.

Cluster 1 corresponds to the Moderate Risk category, exhibiting higher mean dengue cases and population density compared to the low-risk cluster, while mortality remains low. The demographic pressure in this cluster may facilitate dengue transmission, highlighting the need for strengthened vector control and community awareness programs.

Cluster 2 is classified as High Risk, with substantially elevated mean values of dengue incidence, case counts, and population size. This cluster also shows limited healthcare capacity relative to its epidemiological burden, resulting in a high-Risk Score. Areas in this cluster should be prioritized for intensive public health interventions and healthcare resource allocation.

Cluster 3 falls into the Very High-Risk category, marked by extreme dengue incidence levels despite lower population density in some areas. This pattern suggests the influence of environmental or behavioral factors beyond population concentration, indicating the importance of targeted environmental management and focused epidemiological monitoring.

Cluster 4 represents the Emergency Priority group, characterized by the highest Risk Score among all clusters. This cluster combines high epidemiological burden with limited healthcare facility availability, placing it in a critical condition. Immediate and coordinated intervention strategies are required for this cluster to prevent further escalation of dengue transmission and adverse health outcomes.

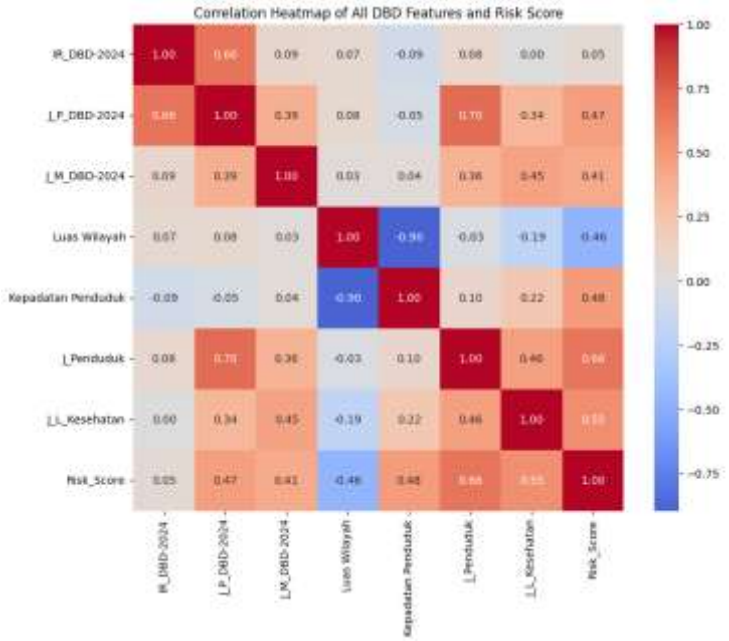

Fig. 5 Correlation Heatmap of All DBD Features and Risk Score

To examine the relationships among the variables used in the clustering process and their association with the Risk Score, a Pearson correlation analysis was conducted and visualized using a heatmap in Fig. 5. This visualization reveals the correlation patterns between epidemiological indicators, demographic factors, healthcare capacity, and the Risk Score, providing additional insight into the relative influence of each variable on dengue risk levels.

Based on the Risk Score, each area was subsequently classified into five dengue risk categories: Low Risk, Moderate Risk, High Risk, Very High Risk, and Emergency Priority. The proportional distribution of these risk categories is illustrated using a pie chart in Fig. 6, highlighting the relative share of each category across the study area.

In addition, the distribution of the number of areas across clusters is presented using a count plot in Fig. 7 to further illustrate the spatial distribution of regions according to their risk levels. This visualization demonstrates that the distribution of areas across clusters is not homogeneous, reinforcing the relevance of a cluster-based approach for dengue risk assessment and public health intervention planning.

*Amiq Fahmi

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15748
e-ISSN : 2541-2019
p-ISSN : 2541-044X

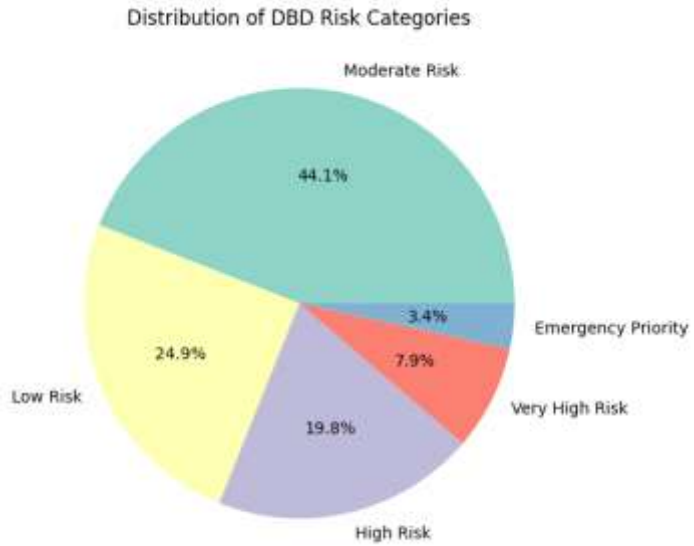

Fig. 6 Distribution of DBD Risk Categories Chart

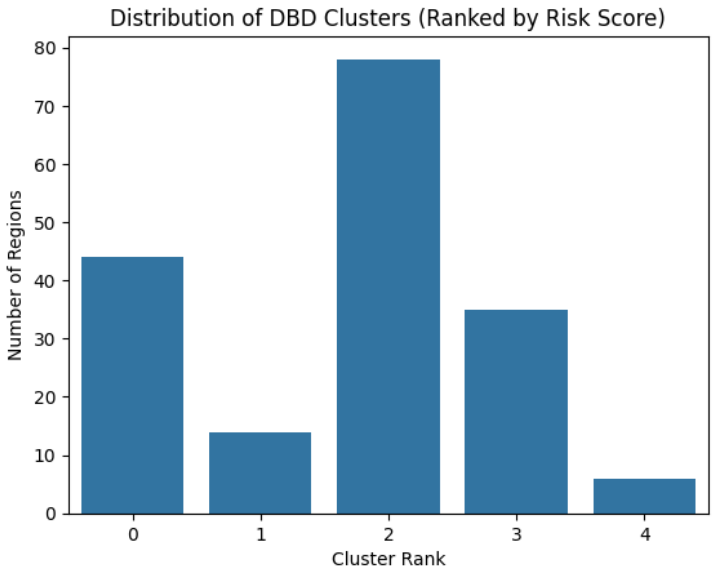

Fig. 7 Distribution of DBD Cluster plot

DISCUSSIONS

The results of this study demonstrate that the application of the K-Means clustering algorithm using seven key features, including dengue epidemiological indicators, population density, total population, area size, and healthcare service capacity, is effective in grouping urban regions according to dengue vulnerability in a structured and objective manner. The optimal number of clusters was determined through the combined use of the Elbow Method and the Silhouette Score, ensuring both statistical robustness and meaningful cluster separation.

As illustrated in Fig. 2, the Elbow Method reveals a pronounced reduction in total inertia up to k = 5, followed by a gradual flattening of the curve. This pattern indicates an optimal balance between model complexity and the ability of clusters to capture data variability. Increasing the number of clusters beyond this point yields diminishing improvements in clustering performance. This finding is further supported by the Silhouette Score analysis shown in Fig. 3, where the highest silhouette value, approximately 0.33, is achieved at k = 5. Together, these validation results confirm that the five-cluster configuration provides relatively strong inter-cluster separation and satisfactory intra-cluster cohesion.

From an epidemiological perspective, the resulting clusters reflect heterogeneous dengue risk profiles shaped by the interaction of disease burden, population characteristics, and healthcare capacity. Clusters with higher Risk Scores are consistently associated with elevated dengue incidence rates, higher case counts, and larger populations,

*Amiq Fahmi

indicating intensified transmission potential and increased exposure. Conversely, clusters characterized by lower Risk Scores tend to exhibit lower epidemiological burdens and relatively stronger healthcare availability, which may contribute to enhanced disease management and mitigation capacity. These findings suggest that dengue vulnerability is not solely determined by population density, but rather emerges from the combined influence of epidemiological intensity, demographic pressure, and health service accessibility.

Methodologically, this study contributes to the field of spatial epidemiology by demonstrating that integrating unsupervised spatial clustering with an epidemiologically weighted composite risk scoring framework enhances both interpretability and operational usability of dengue vulnerability assessments. Unlike conventional hotspot mapping techniques or purely regression-based approaches that focus on individual predictors, the proposed framework captures multidimensional risk patterns and translates them into clearly defined vulnerability categories. This integration enables a more holistic representation of dengue risk at the regional level while remaining computationally efficient and adaptable to data-limited settings.

The Pearson correlation analysis presented in Fig. 5 further supports the validity of the composite Risk Score. Strong to moderate positive correlations are observed between the Risk Score and key epidemiological indicators, including dengue incidence rate, number of cases, and mortality, indicating that the weighting scheme appropriately emphasizes variables directly related to disease transmission and severity. In contrast, area size exhibits a negative correlation with the Risk Score, suggesting that larger administrative areas may experience lower vulnerability due to reduced population concentration or more distributed exposure patterns. These relationships reinforce the epidemiological plausibility of the proposed risk scoring approach.

From a policy and public health planning perspective, the classification of regions into five risk categories—Low Risk, Moderate Risk, High Risk, Very High Risk, and Emergency Priority—provides a practical decision-support tool for dengue control programs. As shown in Fig. 6, although most regions fall within the low to moderate risk categories, the presence of Very High Risk and Emergency Priority clusters highlights localized areas requiring immediate and intensified intervention. Low-risk clusters may be adequately managed through routine surveillance and preventive education, whereas high-risk and emergency-priority clusters warrant prioritized vector control measures, enhanced healthcare resource allocation, and continuous epidemiological monitoring.

Finally, the uneven distribution of regions across clusters illustrated in Fig. 7 underscores the spatial heterogeneity of dengue vulnerability within the study area. This heterogeneity reinforces the importance of cluster-based analytical approaches for designing targeted and efficient public health interventions. Overall, the findings confirm that the proposed K-Means-based clustering combined with composite risk scoring offers a robust, interpretable, and policy-relevant framework for dengue vulnerability assessment. This approach aligns with international best practices in spatial epidemiology and infectious disease risk mapping, and it holds strong potential for supporting data-driven decision-making in urban dengue control strategies.

CONCLUSION

Based on the findings of this study, the application of K-Means clustering using seven main features including epidemiological indicators, demographic characteristics, and healthcare facility availability successfully produced a data driven mapping of dengue vulnerability that captures the heterogeneity of conditions across regions. The determination of the optimal number of clusters using the Elbow Method and Silhouette Score resulted in the best configuration at k = 5, indicating an appropriate balance between model complexity and the quality of cluster separation. These results confirm that an unsupervised learning approach can be effectively applied to classify regions according to dengue vulnerability by integrating multidimensional variables.

The clustering results reveal clear differences among clusters in terms of Risk Score values and the average characteristics of each feature. Correlation analysis confirms that key epidemiological indicators of dengue hemorrhagic fever particularly incidence rate, number of cases, and number of deaths make the most dominant contribution to increased Risk Scores. Total population shows a moderate positive relationship with dengue risk, while population density exhibits a relatively weak positive relationship, suggesting that it functions as a supporting rather than a primary determining factor. Conversely, area size demonstrates a negative relationship with the Risk Score, indicating that regions with larger geographic areas tend to have lower levels of dengue vulnerability.

From a scientific perspective, this study contributes to the field of spatial epidemiology by providing empirical evidence that dengue vulnerability in urban settings is shaped by the interaction of epidemiological burden, demographic pressure, and healthcare service capacity rather than by single risk factors in isolation. The findings reinforce the importance of multidimensional analytical frameworks for understanding the spatial structure of infectious disease risk.

From a methodological perspective, this study demonstrates that integrating unsupervised clustering with an epidemiologically weighted composite Risk Score enhances both interpretability and operational relevance of dengue vulnerability mapping. This approach extends beyond descriptive clustering by translating cluster

*Amiq Fahmi

structures into standardized risk categories that can be directly applied in public health decision making. The proposed framework is computationally efficient, transparent, and does not require labeled data, making it suitable for regions with limited surveillance resources.

From a policy and implementation perspective, the classification of regions into five risk categories namely Low Risk, Moderate Risk, High Risk, Very High Risk, and Emergency Priority provides a clear and actionable basis for prioritizing dengue control interventions. This information can assist policymakers in allocating healthcare resources, strengthening prevention programs, and designing targeted vector control strategies based on local risk profiles.

Despite these contributions, this study has several limitations, including its reliance on data from a single time period and the exclusion of environmental and socioeconomic variables such as rainfall, environmental sanitation, and socioeconomic conditions that may influence dengue transmission dynamics. Future studies are therefore encouraged to incorporate additional variables, temporal data, and alternative clustering approaches to further improve the robustness of dengue risk assessment.

Importantly, the analytical framework proposed in this study is replicable and adaptable. By adjusting indicator selection, weighting schemes, and spatial units, the model can be applied to other regions or extended to different vector borne diseases. Accordingly, the findings of this study offer a methodological foundation that supports evidence based and sustainable public health planning for dengue prevention and control.

ACKNOWLEDGMENT

The author expresses sincere gratitude to the Faculty of Computer Science, Dian Nuswantoro University, for the support, guidance, and conducive academic environment provided throughout the completion of this research. The author also extends appreciation to the Central Java Provincial Office of Statistics for providing essential datasets that made this study possible. The contributions of these institutions and all parties involved were highly valuable to the successful completion of this research.

REFERENCES

Abdullah, N. A. M. H., Dom, N. C., Salleh, S. A., Dapari, R., & Precha, N. (2025). Dengue's climate conundrum: how vegetation and temperature shape mosquito populations and disease outbreaks. *BMC Public Health*, *25*(1). https://doi.org/10.1186/s12889-024-21105-4

Agusrawati, A., Fithria, F., Wibawa, G. N. A., Ruslan, R., Hadini, H., Baharuddin, B., Yahya, I., & Abapihi, B. (2023). Spatial Analysis on the Spread of Dengue Hemorrhagic Fever in Baubau, Southeast Sulawesi, Indonesia. *International Journal of Science, Technology, Engineering and Mathematics*, *3*(4), 51–72. https://doi.org/10.53378/353033

Araujo, E. C., Codeço, C. T., Loch, S., Vacaro, L. B., Freitas, L. P., Lana, R. M., Bastos, L. S., de Almeida, I. F., Valente, F., Carvalho, L. M., & Coelho, F. C. (2024). *Large-scale Epidemiological modeling: Scanning for Mosquito-Borne Diseases Spatio-temporal Patterns in Brazil*. http://arxiv.org/abs/2407.21286

Auditiyah, C. (2024). Pengelompokkan Daerah Rawan Demam Berdarah (DBD) di Jawa Timur Menggunakan Metode K-Means. *ESTIMASI: Journal of Statistics and Its Application*, 205–215. https://doi.org/10.20956/ejsa.v5i2.27091

Baligodugula, V. V., & Amsaad, F. (2025). *Unsupervised Learning: Comparative Analysis of Clustering Techniques on High-Dimensional Data*. http://arxiv.org/abs/2503.23215

BPS Kota Semarang, B. P. S. (n.d.). *Badan Pusat Statistik Kota Semarang*. Badan Pusat Statistik Kota Semarang. Retrieved December 23, 2025, from https://semarangkota.bps.go.id

Deng, S. Z., Liu, X. Y., Su, J. J., Xiang, L. H., Chang, L. T., Zhu, J. J., & Zhang, H. L. (2025). Epidemiological and cluster characteristics of dengue fever in Yunnan Province, Southwestern China, 2013–2023. *BMC Infectious Diseases*, *25*(1). https://doi.org/10.1186/s12879-024-10403-2

Dwi, S., Ningsih, O., Fibriana, A. I., Masyarakat, I. K., Keolahragaan, I., & Semarang, U. N. (n.d.). *Ecological Study: Correlation Of Rainfall with Dengue Hemorrhagic Fever Using Time Series Analysis nd Gis in Semarang City (2017-2021)*.

Gurram, M. K., Gwee, S., Wang, Y. C., & Pang, J. (2025). Spatiotemporal distribution of sustained dengue hotspots associated with climate and urbanisation in Singapore. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-07296-9

Hoque, M. A. A., Sardar, M. L., Mukul, S. A., & Pradhan, B. (2025). Mapping dengue susceptibility in Dhaka city: a geospatial multi-criteria approach integrating environmental and demographic factors. *Spatial Information Research*, *33*(4). https://doi.org/10.1007/s41324-025-00635-y

*Amiq Fahmi

Indonesia, T. K. (2025). *Indonesia dan WHO Memperkuat Upaya Melawan Dengue dengan Inovasi Surveilans*. https://www.who.int/indonesia/id/news/detail/16-05-2025-indonesia-and-who-ramp-up-dengue-fight-with-smarter-surveillance

Mamenun, Koesmaryono, Y., Sopaheluwakan, A., Hidayati, R., Dasanto, B. D., & Aryati, R. (2024). Spatiotemporal Characterization of Dengue Incidence and Its Correlation to Climate Parameters in Indonesia. *Insects*, *15*(5), 366. https://doi.org/10.3390/insects15050366

Leandro, A. S., Chiba de Castro, W. A., Garey, M. V., & Maciel-de-Freitas, R. (2024). Spatial analysis of dengue transmission in an endemic city in Brazil reveals high spatial structuring on local dengue transmission dynamics. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-59537-y

Lestari, T. R. P. (2025). Maintaining National Commitment to Dengue Control Towards Zero Deaths by 2030. *Commision IX Health, Manpower, and Social Security*, *XVII* (11), 1–5.

Li, J. J., Zhou, H. J., Bickel, P. J., & Tong, X. (2024). Dissecting Gene Expression Heterogeneity: Generalized Pearson Correlation Squares and the *K* -Lines Clustering Algorithm. *Journal of the American Statistical Association*, *119*(548), 2450–2463. https://doi.org/10.1080/01621459.2024.2342639

Lin, C.-H., Wen, T.-H., Lin, C.-H., & Wen, T.-H. (2022). How Spatial Epidemiology Helps Understand Infectious Human Disease Transmission. *Tropical Medicine and Infectious Disease*, *7*(8). https://doi.org/10.3390/tropicalmed7080164

Listyono, G. M., Oinike, A., & Hambali, D. (2025). Uncovering the spatial link between environmental risks, diarrhea incidence, and health service accessibility. *Environmental and Materials*, *3*(1). https://doi.org/10.61511/eam.v3i1.2025.1946

Magalhães, A. R., Codeço, C. T., Svenning, J.-C., Escobar, L. E., Van De Vuurst, P., & Gonçalves-Souza, T. (2023). Neglected tropical diseases risk correlates with poverty and early ecosystem destruction. *Infectious Diseases of Poverty*, *12*(1), 32. https://doi.org/10.1186/s40249-023-01084-1

Maramis, A., & Wispriyono, B. (n.d.). *Vulnerability Mapping of Dengue Hemorrhagic Fever (DHF) Cases in West Java Province in 2023*. https://doi.org/10.37287/ijghr.v7i5.6720

Medina, J. R. C., Takeuchi, R., Mercado, C. E. G., de los Reyes, C. S., Cruz, R. V., Abrigo, M. D. R., Hernandez, P. M. R., Garcia, F. B., Salanguit, M., Gregorio, E. R., Kawamura, S., Hung, K. E., Kaneko, M., Nonaka, D., Maude, R. J., & Kobayashi, J. (2023). Spatial and temporal distribution of reported dengue cases and hot spot identification in Quezon City, Philippines, 2010–2017. *Tropical Medicine and Health*, *51*(1). https://doi.org/10.1186/s41182-023-00523-x

Nurkhairiyyah, H. D., Zaidiah, A., & Irmanda, H. N. (2024). Analisis Daerah Potensi Persebaran Demam Berdarah Dengue di DKI Jakarta dengan Menggunakan Metode Clustering K-Means. *JOINS (Journal of Information System)*, *9*(1), 34–42. https://doi.org/10.33633/joins.v9i1.7050

NTD, WHE, W. T. (2024). *Global strategic preparedness, readiness and response plan for dengue and other Aedes-borne arboviruses*. https://www.who.int/publications/m/item/global-strategic-preparedness--readiness-and-response-plan-for-dengue-and-other-aedes-borne-arboviruses

Rahmah, A., Faoziatun Khusna, N., Sanmas, S. A., Aulia, S., Amaria, S., & Fauzi, F. (2025). *Comparison Analysis of Hierarchical* (Vol. 13).

Rahman, A. ur. (2022). Geo-Spatial Disease Clustering for Public Health Decision Making. *Informatica*, *46*(6). https://doi.org/10.31449/inf.v46i6.3827

Salim, M. F., Satoto, T. B. T., & Danardono. (2025). Predicting spatio-temporal dynamics of dengue using INLA (integrated nested laplace approximation) in Yogyakarta, Indonesia. *BMC Public Health*, *25*(1). https://doi.org/10.1186/s12889-025-22545-2

Sena, B. F., Herrera, B. B., Martins, D. B. G., & Lima Filho, J. L. (2025). Geospatial clustering reveals dengue hotspots across Brazilian municipalities, 2024. *Frontiers in Public Health*, *13*. https://doi.org/10.3389/fpubh.2025.1620914

Shen, Y., Ren, Z., Fan, J., Xiao, J., Zhang, Y., & Liu, X. (2025). Fine-Scale Risk Mapping for Dengue Vector Using Spatial Downscaling in Intra-Urban Areas of Guangzhou, China. *Insects*, *16*(7). https://doi.org/10.3390/insects16070661

Thi-Quynh Nguyen, & Thi-Hien Cao. (2023). GIS-based infectious disease mapping: A case study of hotspots of dengue virus in Ho Chi Minh City, Vietnam. *World Journal of Biology Pharmacy and Health Sciences*, *16*(1), 239–247. https://doi.org/10.30574/wjbphs.2023.16.1.0455

Wongoutong, C. (2024). The impact of neglecting feature scaling in k-means clustering. *PLoS ONE*, *19*(12). https://doi.org/10.1371/journal.pone.0310839

*Amiq Fahmi