Journal of Information System and Technology Research

journal homepage: https://journal.aira.or.id/index.php/jistr/

Spatial Clustering of Dengue Hemorrhagic Fever Risk Areas in Lamongan Regency Using the SKATER Method

Cahya Eka Melati[1*], Mohammad Idhom[2], Muhammad Nasrudin[3]

[1,2,3]Department of Data Science, Universitas Pembangunan Nasional "Veteran" Jawa Timur, Indonesia

ARTICLE INFO

Article history:

Received february 18, 2026
Accepted May 29, 2026
Available online May 31, 2026

Keywords:

Dengue Hemorrhagic Fever
Spatial Clustering
SKATER Method
Spatial Autocorrelation
Lamongan Regency

How to Cite :
C. E. Melati, M. Idhom, and M. Nasrudin, "Spatial clustering of dengue hemorrhagic fever risk areas in Lamongan Regency using the SKATER method," *Journal of Information System and Technology Research*, vol. 5, no. 2, pp. 219–230, May 2026

ABSTRACT

Dengue Hemorrhagic Fever (DHF) remains a major public health problem in Lamongan Regency, Indonesia, with unequal distribution across sub-districts. This study aims to identify and classify DHF-prone areas using Spatial 'K'luster Analysis by Tree Edge Removal (SKATER), a graph-based spatial clustering method. The study used cross-sectional secondary data at the sub-district level, including DHF Incidence Rate (IR), population density, rainfall, and percentage of adequate sanitation. Spatial autocorrelation was analyzed using Moran's Index, followed by weighted graph construction and Minimum Spanning Tree (MST) partitioning for cluster formation. Cluster quality was evaluated using the Sum of Squared Deviations (SSD) and Between-Cluster Sum of Squares (BSS). The Moran's I results showed significant spatial autocorrelation for all variables ($p < 0.05$). The five-cluster configuration produced better clustering performance, with lower SSD (49.84) and higher BSS (58.16) compared to the three-cluster configuration (SSD = 86.18; BSS = 21.82). The results revealed spatial variations in DHF vulnerability, ranging from very low to very high categories. These findings indicate that the SKATER method effectively identifies geographically contiguous and homogeneous DHF-prone areas to support spatially targeted DHF control planning in Lamongan Regency.

Corresponding Author:

Cahya Eka Melati,
Department of Data Science, Universitas Pembangunan Nasional "Veteran" Jawa Timur, Indonesia.
Email: 22083010090@student.upnjatim.ac.id

1. INTRODUCTION

Dengue Hemorrhagic Fever (DHF) is a viral infectious disease caused by the dengue virus and transmitted through the bite of the Aedes aegypti mosquito. The disease is commonly found in tropical and subtropical regions, including Indonesia. Clinically, DHF is characterized by sudden high fever, headache, muscle and joint pain, nausea, and vomiting. In severe cases, DHF may progress to dengue shock syndrome, which can be life-threatening [1]. Globally, the World Health Organization (WHO) reported a significant increase in dengue cases in 2024, with a total of 14,305,764 cases and 10,576 deaths across 112 countries, indicating that DHF remains a serious public health threat worldwide.

At the national level, Dengue Hemorrhagic Fever (DHF) remains one of the major public health concerns in Indonesia. As of May 2025, more than 56,000 DHF cases and 250 deaths have been reported across 456 regencies and municipalities, covering approximately 87% of the country's territory [2]. The high incidence rate indicates that DHF is endemic and requires comprehensive and region-based control efforts. Lamongan Regency, located in East Java Province, is one of the areas experiencing a significant increase in DHF cases. According to data from the Lamongan Regency Health

Office, the number of DHF cases in the first six months of 2025 reached 671 cases [3]. In addition to the increasing number of cases, the distribution of DHF across sub-districts shows noticeable disparities. Several sub-districts, such as Brondong and Solokuro, report relatively higher case numbers compared to others. These spatial disparities suggest the influence of geographical, environmental, and demographic factors on the level of DHF vulnerability [4]. However, existing studies and reports on DHF in Indonesia are still predominantly descriptive and have not explicitly identified spatial clustering patterns that consider both attribute similarity and geographical contiguity at the sub-district level. This limitation results in the lack of detailed spatial information for determining priority areas for intervention. Therefore, spatial analysis is required to better understand the clustering patterns of DHF-prone areas [5]. A spatial clustering approach is particularly relevant because it groups regions based on similar characteristics while simultaneously considering geographical contiguity among areas [6], thereby supporting more effective and targeted public health interventions.

Several previous studies have applied clustering and spatial analysis methods to identify DHF vulnerability patterns in Indonesia. Most of these studies employed non-spatial clustering approaches such as K-Means to classify regions based on DHF cases and related indicators [7]. Other studies mainly focused on hotspot detection, spatial correlation, or environmental factor analysis without generating geographically contiguous regional clusters. Although these approaches successfully identified vulnerability categories and spatial distribution patterns, they primarily emphasized attribute similarity and did not explicitly incorporate spatial adjacency relationships among neighboring regions. Consequently, the resulting clusters may consist of geographically separated areas, limiting their interpretability for region-based public health interventions. In addition, several previous studies mainly relied on epidemiological variables, while environmental and demographic factors influencing DHF transmission were not comprehensively integrated into the clustering framework [8].

Therefore, a research gap remains in developing a spatial clustering framework that integrates epidemiological, environmental, and demographic variables while simultaneously preserving geographical contiguity among regions at the sub-district level. To address this limitation, the present study applies the Spatial ‘K’luster Analysis by Tree Edge Removal (SKATER) method to identify homogeneous and geographically contiguous DHF-prone clusters in Lamongan Regency. Unlike conventional clustering approaches, the SKATER method explicitly incorporates spatial relationships among neighboring regions through graph partitioning, thereby producing more spatially coherent and interpretable regional groupings for targeted DHF control planning.

Administratively, Lamongan Regency consists of several sub-districts with diverse regional characteristics, including variations in population density, environmental conditions, and sanitation infrastructure. The regency encompasses coastal areas, lowlands, and regions dominated by agricultural activities, all of which may influence the spatial distribution of environmentally related diseases such as DHF [9]. This heterogeneity makes Lamongan Regency a relevant case study for identifying detailed spatial clustering patterns of DHF vulnerability at the sub-district level. DHF transmission is strongly associated with epidemiological, environmental, and demographic factors. Population density may increase the intensity of human–vector interaction, rainfall contributes to the formation of mosquito breeding sites, and inadequate sanitation conditions may support the development of Aedes aegypti larvae. In addition, the DHF Incidence Rate (IR) reflects the level of disease occurrence and regional vulnerability. Therefore, integrating these variables is important for comprehensively identifying spatial vulnerability patterns.

In this study, the Spatial ‘K’luster Analysis by Tree Edge Removal (SKATER) method is employed sub-districts in Lamongan Regency based on their level of DHF vulnerability. The SKATER method is selected because it explicitly incorporates spatial adjacency relationships and generates geographically contiguous and internally homogeneous clusters through graph partitioning and Minimum Spanning Tree (MST) construction [10][11]. Unlike conventional non-spatial clustering methods such as K-Means, SKATER preserves spatial connectivity among neighboring regions, making it particularly suitable for spatial epidemiological analysis.

The variables used in this study include DHF Incidence Rate (IR), population density, rainfall, and percentage of adequate sanitation, representing epidemiological, environmental, and demographic factors associated with DHF transmission [12]. The objective of this study is to identify and classify sub-districts in Lamongan Regency into homogeneous and spatially contiguous clusters based on DHF vulnerability indicators [5]. The contribution of this study lies in the integration of multiple DHF-related variables within a graph-based spatial clustering framework at the sub-district level. The novelty of this research is the application of the SKATER method to produce spatially contiguous DHF vulnerability clusters that provide more spatially coherent and interpretable regional groupings for targeted DHF control planning compared to conventional non-spatial clustering approaches.

2. RESEARCH METHOD

This study employs spatial analysis using a graph-based clustering approach to identify Dengue Hemorrhagic Fever (DHF)-prone areas in Lamongan Regency. The clustering method applied is the Spatial ‘K’luster Analysis by Tree Edge Removal (SKATER), which considers spatial adjacency among neighboring regions and generates clusters that are both geographically contiguous and homogeneous in terms of their characteristics [13]. This approach is utilized to identify spatial grouping patterns of regions based on their level of DHF vulnerability, thereby supporting more targeted and region-based DHF control planning.

2.1. Research Data and Variables

The data used in this study are secondary data collected for the year 2024. Data on Dengue Hemorrhagic Fever (DHF) cases were obtained from Lamongan Regency Health Office, while population density, adequate sanitation, and spatial data on sub-district administrative boundaries were obtained from Statistics Indonesia (BPS) of Lamongan Regency. The dataset covers all 27 sub-districts in Lamongan Regency, allowing the analysis to be conducted comprehensively at the sub-district administrative level as the spatial unit of analysis. The unit of analysis in this study is the sub-district administrative area [14]. Each sub-district is represented as a node in the graph structure of the Spatial 'K'luster Analysis by Tree Edge Removal (SKATER) method [15]. The data are cross-sectional, collected within the same observation year; therefore, the analysis focuses on spatial variation across regions rather than the temporal dynamics of DHF cases. The selection of research variables is based on theoretical and empirical considerations regarding factors influencing DHF transmission, including epidemiological, environmental, and demographic aspects. These variables were chosen because they conceptually contribute to DHF transmission risk through the interaction between human populations, physical environmental conditions, and climatic factors that support vector development [16]. Prior to analysis, the dataset underwent preprocessing procedures, including data cleaning and variable standardization using the z-score method to eliminate differences in measurement scales among variables. No missing values were identified in the dataset, and no significant outliers were detected during the preliminary data inspection process. The structure of the dataset used as the basis for the spatial clustering analysis is presented in the following table.

Table 1. Data Structure

Subdistrict	DHF IR	Population Density	Rainfall	Adequate Sanitation
Sukorame	95,81	578,4	245.3	88,03%
Bluluk	163,97	462,6	319,7	91,14%
Ngimbang	66,1	477,2	162,9	100%
...	...	...	...	...
Brondong	66,77	1084,4	743	100%

Source: Lamongan Regency Health Office (2024) and Statistics Indonesia (BPS) of Lamongan Regency (2024).

In Table 1, the Dengue Hemorrhagic Fever (DHF) Incidence Rate (IR) is defined as the number of DHF cases per 100,000 population in each sub-district during the observation year and serves as an indicator of disease occurrence and regional vulnerability. Population density is measured as the number of people per square kilometer and is associated with the intensity of contact between humans and the dengue vector. Rainfall represents climatic conditions that may influence the availability of mosquito breeding sites and consequently affect dengue transmission. Meanwhile, adequate sanitation is measured as the percentage of households with access to sanitation facilities that meet health standards; inadequate sanitation conditions may increase the risk of mosquito breeding and disease transmission. These four variables were selected because they represent epidemiological, demographic, and environmental factors associated with DHF risk and serve as the basis for the spatial clustering analysis.

2.2 Research Stages

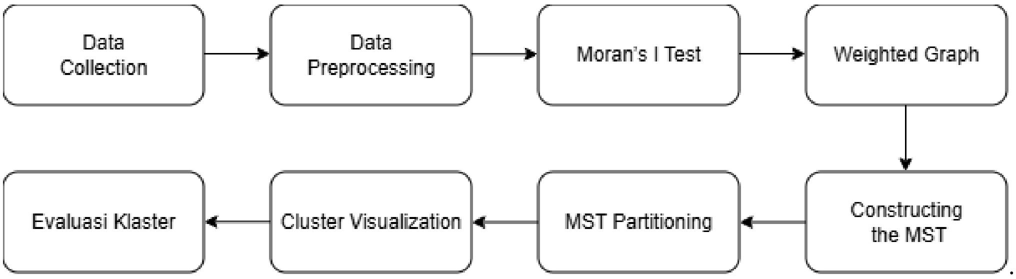

Figure 1. Research Workflow
(Source: Developed by the Author)

The research stages are illustrated in Figure 1. The study begins with the collection of Dengue Hemorrhagic Fever (DHF) case data and spatial data of sub-district administrative boundaries. The next stage involves data preprocessing, which includes data cleaning and variable standardization using the z-score method to eliminate scale differences among variables. After the data are prepared for analysis, a global spatial autocorrelation test using Moran's I is conducted to identify the presence of spatial dependency patterns in each research variable. Subsequently, spatial adjacency relationships among sub-districts are determined using the k-nearest neighbors (KNN) approach, where neighbors are defined based on the distance between sub-district centroids [17]. This adjacency structure serves as the basis for constructing a weighted graph. In the weighted graph, each node represents a sub-district, while edge weights are calculated using the Euclidean distance between standardized variables to measure the level of dissimilarity among regions. From the constructed weighted graph, a Minimum Spanning Tree (MST) is generated as the fundamental structure for clustering. The MST is then partitioned using the SKATER method to produce several clusters of DHF-prone areas. Finally, the clustering results are visualized in the form of thematic maps to illustrate the spatial patterns of regional grouping.

2.3. Spatial Analysis Methods

The analytical methods employed in this study include the global spatial autocorrelation test using Moran's I and the evaluation of cluster quality using the Sum of Squared Deviations (SSD) and the Between-Cluster Sum of Squares (BSS) [18]. Moran's I is used to identify the presence and direction of spatial autocorrelation for each research variable.

Mathematically, Moran's I statistic is formulated as follows [19]:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} w_{ij}(x_i - \bar{x})\,(x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{1}$$

Where:
n : Number of spatial units
x_i : Value of the variable in region-i
$\bar{x}$: Mean value of the variable
w_{ij} : Spatial weight between region-i and region-j

$$S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \tag{2}$$

The value of Moran's I ranges from −1 to +1. A positive value indicates spatial clustering, whereas a negative value indicates a dispersed spatial pattern [19]. The evaluation of cluster quality is conducted using the Sum of Squared Deviations (SSD) and the Between-Cluster Sum of Squares (BSS). The SSD measures the level of homogeneity within clusters, while the BSS measures the degree of separation between clusters [15].

The Sum of Squared Deviations (SSD) is mathematically defined as follows [20]:

$$SSD = \sum_{k=1}^{K} \sum_{i \in C_k} \|x_i - \bar{x}_k\|^2 \tag{3}$$

Where:
K : Number of clusters
C_k : Set of regions belonging to cluster-k
x_i : Data vector of region- i
x_k : Mean vector of cluster- k

The Between-Cluster Sum of Squares (BSS) is mathematically defined as follows:

$$BSS = \sum_{k=1}^{K} n_k \|\bar{x}_k - \bar{x}\|^2 \tag{4}$$

Where:
n_k : Number of regions in cluster- k
$\bar{x}$: Overall mean vector of the dataset

2.4. *Spatial 'K'luster Analysis by Tree Edge Removal* (SKATER)

Spatial 'K'luster Analysis by Tree Edge Removal (SKATER) is a graph-based spatial clustering method that groups regions by considering both attribute similarity and spatial adjacency among neighboring areas. Unlike non-spatial clustering methods, SKATER ensures that regions within the same cluster remain geographically contiguous [21]. The initial stage of the SKATER method involves constructing a weighted graph that represents spatial adjacency relationships among regions. In this study, adjacency is determined using the k-nearest neighbors (KNN) approach based on the distance between sub-district centroids [22]. Each node in the graph represents a sub-district, while edges connect neighboring sub-districts. The edge weights are calculated using the Euclidean distance between standardized variable vectors, namely the Dengue Hemorrhagic Fever (DHF) Incidence Rate (IR), population density, rainfall, and the percentage of adequate sanitation. These weights reflect the degree of dissimilarity between regions, where smaller weight values indicate greater similarity in characteristics.

From the constructed weighted graph, a Minimum Spanning Tree (MST) is then generated. The MST is a tree structure that connects all nodes with the minimum total edge weight without forming cycles. In the context of spatial clustering, the MST serves to simplify the graph structure while preserving the most essential spatial relationships among regions, ensuring that all areas remain connected through the most relevant characteristic-based links [23]. The final stage of the SKATER method involves partitioning the MST to form regional clusters [18]. The partitioning process is carried out by removing a number of edges with the largest weights, as these edges represent the most significant dissimilarities between connected regions. Cutting these edges produces several connected components, each representing a spatial cluster. Through this approach, the resulting clusters are homogeneous in terms of their attributes while remaining geographically contiguous[24].

3. RESULTS AND DISCUSSION

This section presents the results of the spatial analysis conducted to identify Dengue Hemorrhagic Fever (DHF)-prone areas in Lamongan Regency using the SKATER method, along with a discussion of the findings. The results are presented in the form of tables, graphs, and thematic maps.

3.1. Moran's I Test

Table 2 presents the results of the Moran's I test, which aims to identify the presence and pattern of spatial autocorrelation for each variable. The Moran's I value indicates the direction of spatial correlation, while the p-value indicates its statistical significance.

Table 2. Moran's I Test Results

No.	Variable	Moran's I	P-value
1.	DHF Incidence Rate (IR)	0,1795	0,032
2.	Population Density	-0,1580	0,029
3.	Rainfall	0,4822	0,001
4.	Adequate Sanitation	0,2348	0,016

(Source: Developed by the Author)

Based on Table 2, all research variables have p-values less than 0.05, indicating the presence of statistically significant spatial autocorrelation among sub-districts in Lamongan Regency. This result confirms that the distribution of the observed variables is not random but spatially dependent, thereby justifying the use of spatial-based analytical approaches. Furthermore, the direction and pattern of spatial autocorrelation are determined based on the value of Moran's Index (I). A positive Moran's I value ($I > 0$) indicates positive spatial autocorrelation with a clustered pattern, whereas a negative Moran's I value ($I < 0$) indicates negative spatial autocorrelation with a dispersed pattern [25][26]. Based on these results, the variables DHF Incidence Rate (IR), rainfall, and percentage of adequate sanitation exhibit positive Moran's I values, suggesting a tendency for regions with similar characteristics to cluster spatially.

In contrast, population density shows a negative Moran's I value, indicating a dispersed spatial pattern among sub-districts. These findings suggest that DHF incidence and environmental factors such as rainfall and sanitation are spatially concentrated in certain areas, potentially reflecting localized environmental and socio-ecological conditions that support disease transmission. This result is consistent with previous studies reporting that dengue transmission is strongly influenced by environmental variables and tends to exhibit spatial clustering patterns. However, most prior studies primarily focused on identifying spatial dependence using global indices, without further classifying regions into homogeneous and spatially contiguous groups [26]. Therefore, this study extends previous work by not only confirming the presence of spatial autocorrelation but also applying spatial clustering to generate more detailed and interpretable regional groupings of DHF vulnerability.

3.2. Construction of the Weighted Graph

At this stage, a weighted graph is constructed to represent the spatial adjacency relationships among sub-districts in Lamongan Regency. In this graph, each node represents a sub-district, while edges represent spatial adjacency relationships determined using the k-nearest neighbors (KNN) approach based on the distance between regional centroids. The KNN approach is applied to ensure that connections are established only between geographically proximate sub-districts, thereby preserving the spatial structure of the region. The weight assigned to each edge is calculated using the Euclidean distance between standardized variable vectors, namely the Dengue Hemorrhagic Fever (DHF) Incidence Rate (IR), population density, rainfall, and the percentage of adequate sanitation. These weights represent the degree of dissimilarity between neighboring sub-districts. Smaller weight values indicate that the connected sub-districts have relatively similar characteristics, whereas larger weight values indicate greater differences in their attributes. The resulting weighted graph serves as the basis for constructing the Minimum Spanning Tree (MST) in the clustering process using the SKATER method [24].

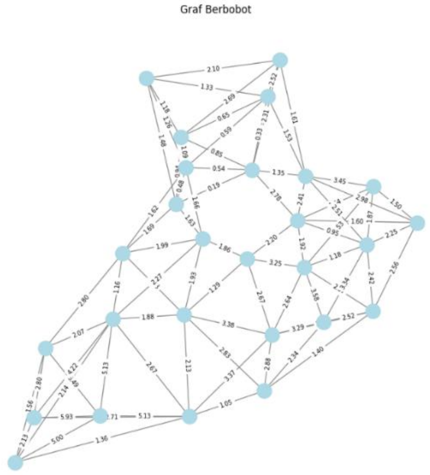

Figure 2. Weighted Graph
(Source: Developed by the Author)

Figure 2 illustrates the weighted graph, where each node represents a sub-district and each edge represents spatial adjacency. The edge weight reflects the level of dissimilarity between connected regions, where smaller weights indicate higher similarity. The purpose of this graph is to model spatial relationships while preserving geographical proximity. The graph shows that sub-districts are primarily connected to nearby regions, ensuring that spatial structure is maintained. This approach is in line with spatial clustering studies that emphasize the importance of incorporating geographical relationships into clustering models, as ignoring spatial adjacency may produce unrealistic groupings.

3.3. Construction of the Minimum Spanning Tree (MST)

The Minimum Spanning Tree (MST) is a tree structure derived from a weighted graph that connects all nodes with the minimum total edge weight without forming cycles. The MST ensures that each node in the graph is connected through a unique path, thereby eliminating redundant connections that may arise in the original graph structure. In the context of spatial analysis, the MST serves to simplify complex interregional relationships while preserving the most essential spatial connectivity structure. In this study, the MST is used to simplify the weighted graph representing adjacency relationships among sub-districts, allowing interregional relationships to be represented more efficiently while remaining informative. The construction of the MST is based on edge weights calculated using the Euclidean distance between standardized variables for neighboring sub-districts. Through this approach, the MST first connects sub-districts with the highest degree of similarity in characteristics, while simultaneously minimizing the total dissimilarity across the network. The resulting MST structure ensures that all sub-districts remain connected within a single spatial network through the most relevant characteristic-based links. Therefore, the MST functions as the fundamental framework of the SKATER method. The clustering process is subsequently performed by removing selected edges from the MST to separate regions with the most significant characteristic differences, while maintaining geographical contiguity within each cluster [27].

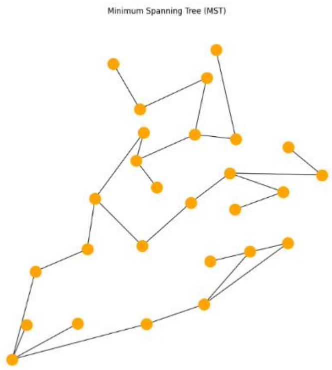

Figure 3. Minimum Spanning Tree (MST)
(Source: Developed by the Author)

Figure 3 presents the visualization of the constructed Minimum Spanning Tree (MST), which is derived from the weighted graph by selecting edges with the minimum total weight while maintaining connectivity among all sub-districts. Compared to the initial weighted graph, the MST structure appears simpler due to the significant reduction in the number of edges. This simplification removes redundant connections while preserving the most important spatial relationships, allowing clearer identification of the primary connections between sub-districts. The MST can be interpreted as a backbone network that highlights the strongest similarity relationships among regions, where shorter edges indicate higher similarity between connected sub-districts. This structure facilitates the identification of spatial patterns and serves as a crucial foundation for the subsequent clustering process.

The MST acts as the main structural framework in the SKATER method, where clusters are formed by removing edges with the largest weights to separate regions exhibiting the greatest differences in characteristics. Thus, the MST not only simplifies the spatial network but also directly determines how clusters will be formed in the next stage. These findings are consistent with previous studies indicating that MST-based approaches are effective in reducing network complexity while preserving spatial structure, making them particularly suitable for spatially constrained clustering methods such as SKATER. Consequently, the construction of the MST represents a key stage linking the weighted graph formation to the cluster partitioning stage in the SKATER method, ensuring that the resulting clusters remain geographically contiguous and internally homogeneous.

3.4. Partitioning of the Minimum Spanning Tree (MST)

The partitioning of the Minimum Spanning Tree (MST) constitutes the core stage of the Spatial ‘K’luster Analysis by Tree Edge Removal (SKATER) method. This stage aims to form regional clusters by dividing the MST structure into several connected components [24]. The process is carried out by removing specific edges from the MST, thereby splitting the originally fully connected network into multiple subgraphs, each representing a cluster. In the context of spatial clustering, MST partitioning is intended to group regions based on similarity in attribute characteristics while maintaining geographical contiguity within each cluster. This property represents a key advantage of the SKATER method compared to non-spatial clustering approaches, as each resulting cluster is guaranteed to consist of geographically adjacent regions.

The partitioning process involves identifying and removing edges in the MST with the largest weights. In this study, edge weights are calculated using the Euclidean distance between standardized variable vectors, namely the Dengue Hemorrhagic Fever (DHF) Incidence Rate (IR), population density, rainfall, and the percentage of adequate sanitation. Larger edge weights indicate greater dissimilarity between two connected sub-districts. Therefore, removing edges with the largest weights logically separates the most dissimilar regions, resulting in clusters that are more internally homogeneous. Mathematically, if the desired number of clusters is k, then $k-1$ edges must be removed from the MST. This removal produces k connected components, each representing a regional cluster. Through this approach, each cluster satisfies two primary characteristics: attribute homogeneity and spatial contiguity.

3.5. Determination of the Optimal Number of Clusters

Determining the number of clusters is a crucial step in the spatial clustering process, as the selected number directly influences the interpretation of regional grouping patterns. In the SKATER method, the number of clusters is determined through the partitioning of the Minimum Spanning Tree (MST) by removing a specific number of edges. However, the selection of the optimal number of clusters should not rely solely on the technical aspect of edge removal it must also be supported by a quantitative evaluation of cluster quality.

In this study, a comparison was conducted between three-cluster and five-cluster configurations to identify the most representative grouping structure. The selection of these two scenarios was exploratory in nature, aiming to examine variations in vulnerability levels under a relatively simple grouping structure (three clusters) and a more detailed grouping structure (five clusters). The quality of clustering in each scenario was subsequently evaluated based on measures of within-cluster homogeneity and between-cluster separation.

To determine the most optimal number of clusters, cluster quality was evaluated using the Sum of Squared Deviations (SSD) and the Between-Cluster Sum of Squares (BSS). The SSD reflects the degree of homogeneity within clusters, indicating how small the data variation is within each cluster. In contrast, the BSS measures the degree of separation between clusters, indicating how distinct the clusters are from one another. A lower SSD value combined with a higher BSS value indicates better clustering quality, as it reflects compact clusters internally and well-separated clusters externally.

Table 3. Cluster Evaluation Results

Cluster Configuration	SSD	BSS
3 Clusters	86.179866	21.820134
5 Clusters	49.842175	58.157825

(Source: Developed by the Author)

The evaluation results show that the five-cluster configuration yields a lower SSD value (49.84) compared to the three-cluster configuration (86.18), and a higher BSS value (58.16) compared to the three-cluster configuration (21.82). These findings indicate that the five-cluster solution produces more internally homogeneous clusters and clearer separation between clusters.

Therefore, the five-cluster configuration was selected as the most optimal representation of DHF vulnerability levels in Lamongan Regency. Based on this evaluation, subsequent analysis focuses on the five-cluster results as the best representation of spatial vulnerability patterns. The classification of vulnerability levels for each cluster is determined based on the relative combination of the clustering variables, particularly the DHF Incidence Rate (IR) and supporting environmental factors, rather than relying solely on the absolute value of a single variable.

Table 4. Mean Values of Variables in Each Cluster

Cluster	DHF IR	Population Density	Rainfall	Adequate Sanitation
1	56.21	792.47	1589.34	89.56
2	66.1	4772	1629	100
3	80.02	1071.88	1.45	76.34
4	163.97	472.6	3197	91.14
5	46.88	1006	1724	100

(Source: Developed by the Author)

Based on the mean values of the variables in each cluster, sub-districts in Lamongan Regency can be classified into five levels of Dengue Hemorrhagic Fever (DHF) vulnerability: very high, high, moderate, low, and very low. The classification is derived from the relative combination of epidemiological and environmental indicators rather than from a single variable.:

Cluster 1 : Is categorized as a low vulnerability area. The DHF Incidence Rate (56.21) is relatively low, accompanied by low population density and fairly good sanitation conditions. These characteristics suggest that although DHF transmission risk is present, it remains relatively lower compared to clusters with moderate to high vulnerability.

Cluster 2 : Represents a moderate vulnerability area. This cluster exhibits a moderately high DHF incidence rate (66.10) and very high population density. High population density may increase the intensity of human–vector contact, thereby elevating transmission risk. However, the excellent sanitation condition (100%) likely mitigates this risk, resulting in an overall moderate vulnerability level.

Cluster 3 : Is classified as a high vulnerability area. It shows a relatively high DHF incidence rate (80.02) and the lowest percentage of adequate sanitation (76.34) among all clusters. The combination of elevated incidence and suboptimal sanitation conditions indicates a greater potential for sustained DHF transmission due to environmental and behavioral factors.

Cluster 4 : Is categorized as a very high vulnerability area. This cluster records the highest DHF incidence rate (163.97), accompanied by extremely high rainfall (3197 mm). Such climatic conditions create a highly favorable environment for mosquito breeding, significantly increasing transmission risk, despite relatively low population density and adequate sanitation levels.

Cluster 5 : Represents a very low vulnerability area. It has the lowest DHF incidence rate (46.88), supported by excellent sanitation coverage (100%) and relatively moderate population density. These characteristics indicate environmental conditions that are more conducive to effective DHF control, resulting in the lowest transmission risk among all clusters.

The classification of vulnerability levels in this study was based on the relative ranking of cluster profiles derived from the SKATER clustering results. The classification considered the combined characteristics of the Dengue Hemorrhagic Fever (DHF) Incidence Rate (IR), population density, rainfall, and adequate sanitation. Clusters exhibiting higher DHF incidence rates together with environmental conditions favorable for dengue transmission, such as high rainfall and lower sanitation coverage, were assigned higher vulnerability categories. Conversely, clusters characterized by lower incidence rates and more favorable environmental conditions were assigned lower vulnerability categories. Therefore, the vulnerability labels (very high, high, moderate, low, and very low) were determined through the comparative interpretation of cluster mean values rather than through predefined statistical thresholds or a composite vulnerability index.

The findings of this study are generally consistent with previous research on the spatial distribution of dengue fever. The very high vulnerability observed in Cluster 4 is associated with the highest DHF incidence rate and extremely high rainfall levels. This result supports previous studies reporting that rainfall plays a significant role in increasing mosquito breeding habitats by creating standing water environments suitable for the development of Aedes aegypti larvae. Consequently, areas with higher rainfall tend to experience greater dengue transmission risk.

Similarly, the high vulnerability identified in Cluster 3 is associated with relatively high DHF incidence and the lowest level of adequate sanitation among all clusters. This finding is in agreement with previous empirical studies suggesting that inadequate sanitation conditions may contribute to the persistence of mosquito breeding sites and increase the likelihood of dengue transmission. Furthermore, the moderate vulnerability observed in Cluster 2 highlights the importance of population density as a contributing factor to disease spread, since densely populated areas may increase human–vector interaction and facilitate transmission. These findings indicate that dengue vulnerability is influenced by a combination of epidemiological, environmental, and demographic factors rather than by a single determinant.

The identified vulnerability patterns have important implications for public health planning and dengue prevention programs in Lamongan Regency. Sub-districts classified as very high vulnerability areas, particularly those in Cluster 4, should be prioritized for intensive vector control programs, environmental management, and enhanced disease surveillance. Areas in Cluster 3 may benefit from interventions focused on improving sanitation infrastructure and strengthening community participation in mosquito breeding site elimination activities. For Cluster 2, preventive strategies should emphasize population-based surveillance and public awareness campaigns due to the high population density. Meanwhile, sub-districts classified as low and very low vulnerability areas should continue routine monitoring and preventive measures to maintain favorable environmental conditions and prevent future increases in dengue transmission risk.

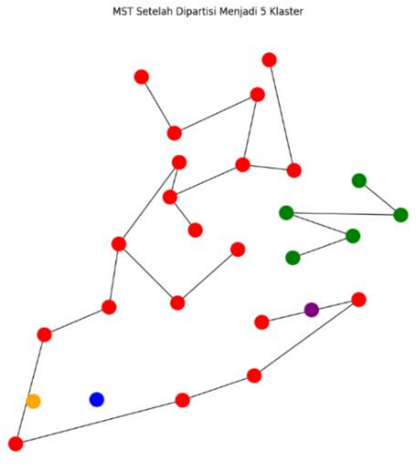

Figure 4. MST Partition Graph (5 Clusters)
(Source: Developed by the Author)

The results of partitioning the Minimum Spanning Tree (MST) into five clusters are presented in Figure 4, illustrating the connectivity structure among sub-districts after removing edges with the largest weights. Each connected component in the figure represents a spatial cluster, where sub-districts within the same component are grouped based on similarity in characteristics. Figure 4 can be interpreted by observing that connected nodes indicate regions with relatively similar attributes, while the separation between components reflects differences between clusters. This demonstrates that the partitioning process successfully separates regions with high dissimilarity while maintaining spatial connectivity within each cluster. The clustering results are further visualized in the form of a thematic map in Figure 5 to display the spatial distribution of DHF-prone clusters in Lamongan Regency. Figure 5 provides a geographic representation of the clusters, allowing clearer identification of the spatial location and distribution of each cluster. Through this map, differences in regional grouping patterns can be more easily observed, highlighting areas with similar levels of DHF vulnerability and supporting more effective spatial interpretation of the clustering results.

Table 5. DHF Vulnerability Cluster Classification of Sub-Districts in Lamongan Regency

Cluster	Sub-District	Vulnerability Category
1	Brondong, Paciran, Solokuro, Laren, Karanggeneng, Kalitengah, Maduran, Sekaran, Pucuk, Sukodadi, Babat, Kedungpring, Sugio, Kembangbahu, Mantup, Sambeng, Sarirejo, Modo.	Low vulnerability
2	Karangbinangun, Turi, Lamongan, Deket, and Glagah.	Moderate vulnerability
3	Ngimbang.	High vulnerability
4	Bluluk and Sukorame.	Very High vulnerability
5	Tikung.	Very Low vulnerability

(Source: Developed by the Author)

Based on Table 5, the majority of sub-districts in Lamongan Regency are classified into Cluster 1, which represents areas with low DHF vulnerability. In contrast, only a small number of sub-districts belong to the higher vulnerability categories, namely Cluster 3 and Cluster 4. These results indicate the presence of spatial variation in DHF vulnerability among sub-districts in Lamongan Regency. The classification also demonstrates that regions with similar epidemiological, environmental, and demographic characteristics tend to form geographically contiguous clusters. To provide a clearer representation of the spatial distribution of the identified clusters, the clustering results are further visualized in the form of a thematic map, as presented in Figure 5.

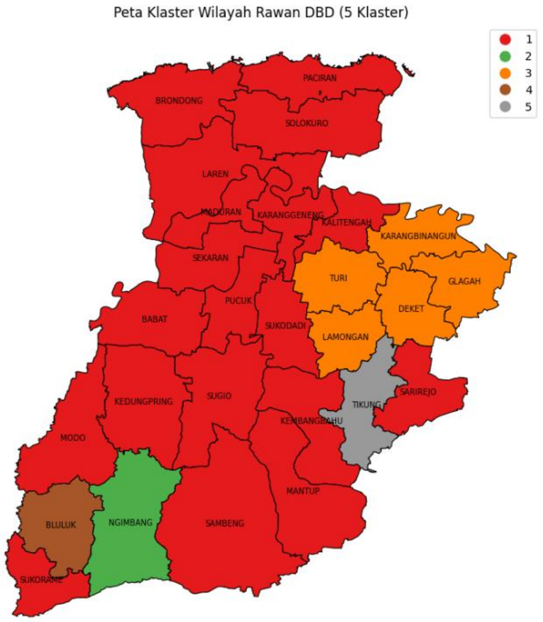

Figure 5. Cluster Map
(Source: Developed by the Author)

Figure 5 presents the visualization of the cluster map of Dengue Hemorrhagic Fever (DHF)-prone areas in Lamongan Regency, which is divided into five clusters. Each cluster is represented by a different color to illustrate the level of vulnerability based on the results of the analysis using the SKATER method. In general, the map shows the presence of spatial variation in DHF vulnerability across sub-districts. It can be observed that most areas are dominated by a particular cluster, indicating that many sub-districts share relatively similar characteristics. Meanwhile, several other areas form smaller and more isolated clusters, suggesting the presence of distinct characteristics compared to surrounding regions. The spatial distribution pattern shown on the map also indicates that areas with similar levels of vulnerability tend to form geographically contiguous clusters rather than being randomly distributed. This finding confirms that the factors influencing DHF incidence exhibit spatial dependence across regions. Overall, this map visualization facilitates the identification of spatial patterns of DHF-prone areas in Lamongan Regency, thereby providing valuable insights for determining priority areas and supporting more targeted and effective intervention planning.

4. CONCLUSION

This study identified spatial patterns of Dengue Hemorrhagic Fever (DHF) vulnerability in Lamongan Regency using the Spatial 'K'luster Analysis by Tree Edge Removal (SKATER) method. The results of Moran's I analysis revealed significant spatial autocorrelation for DHF Incidence Rate (IR), rainfall, and adequate sanitation, indicating clustered spatial patterns, while population density exhibited a dispersed pattern. These findings confirm that DHF vulnerability is spatially dependent and justify the application of spatially constrained clustering approaches. The clustering evaluation demonstrated that the five-cluster configuration was superior to the three-cluster configuration, as indicated by a lower Sum of Squared Deviations (SSD = 49.84) and a higher Between-Cluster Sum of Squares (BSS = 58.16). This result suggests that the five-cluster solution provides greater within-cluster homogeneity and clearer separation between clusters. Based on the identified cluster profiles, the sub-districts of Lamongan Regency were classified into five levels of DHF vulnerability, namely very high, high, moderate, low, and very low vulnerability. The very high vulnerability cluster was characterized by the highest DHF incidence rate and rainfall levels, whereas the very low vulnerability cluster exhibited the lowest incidence rate and favorable environmental conditions. Methodologically, this study demonstrates the usefulness of the SKATER method in producing geographically contiguous and internally homogeneous clusters, thereby providing a more spatially meaningful classification of DHF vulnerability compared with non-spatial clustering approaches. From a practical perspective, the resulting vulnerability map can support local health authorities in identifying priority areas for surveillance, vector control, sanitation improvement, and other dengue prevention programs. Nevertheless, this study has several limitations. First, the analysis was based on cross-sectional data from a single observation period; therefore, it cannot capture temporal variations or long-term dynamics of DHF transmission. Second, the clustering model only incorporated four variables, namely DHF incidence rate, population density, rainfall, and adequate sanitation, which may not fully represent all factors influencing dengue vulnerability. Consequently, the findings should be interpreted as a spatial classification of vulnerability patterns rather than evidence of causal relationships. Future studies are recommended to utilize multi-year datasets to examine temporal changes in DHF vulnerability, incorporate additional environmental and socio-economic variables, and perform external validation of the clustering results. Such improvements may provide a more comprehensive understanding of dengue vulnerability and strengthen the applicability of spatial clustering for public health planning.

5. ACKNOWLEDGEMENT

The authors express sincere gratitude to Universitas Pembangunan Nasional "Veteran" Jawa Timur for the academic support and research facilities provided during the completion of this study. Appreciation is also extended to the Lamongan Regency Health Office and Statistics Indonesia (BPS) Lamongan Regency for providing the secondary data used in this research. Their contribution was essential in supporting the spatial analysis and the development of the clustering model in this study.

6. REFERENCES

[1] S. Anggraini, "Jurnal Epidemiologi Kesehatan Indonesia Analisis Potensial Kejadian Luar Biasa DBD Berdasarkan Kasus DBD 5 Tahun Terakhir 2019-2024 : Studi Ekologi di Kota Lubuklinggau Provinsi Sumatera Selatan Analisis Potensial Kejadian Luar Biasa DBD Berdasarkan Kasus," vol. 9, no. 2, 2025, doi: https://doi.org/10.7454/epidkes.v9i2.1151.

[2] D. Jaminan Sosial Tri Rini Puji Lestari, "MENJAGA KOMITMEN NASIONAL PENGENDALIAN DBD MENUJU NOL KEMATIAN TAHUN 2030," 2025.

[3] F. Reggie, "Kasus DBD Lamongan Meningkat, 6 Bulan 671 Kasus," RADIO REPUBLIK INDONESIA. [Online]. Available: https://rri.co.id/tuban/daerah/1709555/kasus-dbd-lamongan-meningkat-6-bulan-671-kasus

[4] R. Wulandari, R. Hamidy, and B. Bayhakki, "Pengaruh sanitasi lingkungan dan perilaku keluarga terhadap kejadian penyakit demam berdarah dengue (DBD) di Kecamatan Benai Kabupaten Kuantan Singingi," SEHATI J. Kesehat., vol. 1, no. 2, pp. 57–62, 2021, https://doi.org/10.52364/sehati.v1i2.9.

[5] V. A. H. Ho, S. P. M. Wijayanti, and D. S. S. Rejeki, "Analisis spasial terhadap kejadian demam berdarah dengue (DBD) di Indonesia: A systematic review," Holistik J. Kesehat., vol. 19, no. 5, pp. 1012–1018, 2025, doi: https://doi.org/10.33024/hjk.v19i5.1180.

[6] L. I. Harlyan and F. I. & R. E. , Abu Bakar Sambah, "Klasterisasi Spasial Keragaman Spesies Tuna di Perairan

Selatan Jawa," vol. 5066, 2021, https://doi.org/10.22146/jfs.58917

[7] Z. A. Rabbani, A. Avisa, I. Budiawan, and R. Pakpahan, "Penerapan Algoritma K-Means untuk Pengelompokan Kerentanan Wilayah terhadap Kasus DBD di Kota Bandung perkembangbiakan nyamuk . Faktor-faktor ini menyebabkan distribusi kasus DBD di Kota lebih spesifik untuk mendukung kebijakan kesehatan masyarakat . 2024," no. September, 2025.

[8] A. P. Pasaribu et al., "Spatio-temporal patterns of dengue incidence in Medan City, North Sumatera, Indonesia," Trop. Med. Infect. Dis., vol. 6, no. 1, 2021, https://doi.org/10.3390/tropicalmed6010030.

[9] M. T. Handayani, M. Raharjo, and T. Joko, "Pengaruh Indeks Entomologi dan Sebaran Kasus Demam Berdarah Dengue di Kabupaten Sukoharjo," J. Kesehat. Lingkung. Indones., vol. 22, no. 1, pp. 46–54, Feb. 2023, https://doi.org/10.14710/jkli.22.1.46-54

[10] G. E. Dirgantara, I. Novianti, and R. D. Bekti, "SPATIAL CLUSTER UNTUK PENGELOMPOKKAN WILAYAH SETIAP PROVINSI DI INDONESIA BERDASARKAN KARAKTERISTIK KESENJANGAN EKONOMI (SPATIAL CLUSTER FOR GROUPING REGIONS IN EVERY PROVINCE IN INDONESIA BASED ON THE CHARACTERISTICS OF ECONOMIC DISPARITY)," pp. 430–437, 2019.

[11] I. Kharisudin, A. Nur, and U. Suchaini, "Analisis Klaster Spasial Data Infrastruktur Kesehatan Desa di Kabupaten Purworejo Menggunakan Metode SKATER," vol. 6, pp. 775–781, 2023.

[12] H. D. Nurkhairiyyah, A. Zaidiah, and H. N. Irmanda, "Analisis Daerah Potensi Persebaran Demam Berdarah Dengue di DKI Jakarta Menggunakan K-Means," vol. 9, no. 1, pp. 34–42, 2024, https://doi.org/10.33633/joins.v9i1.7050

[13] N. Shela Abdila, R. Fitriani, and M. Liswansyah Pratama, "CLUSTERING WITH SKATER METHODS AND UTILIZATION OF LISA ON UNEMPLOYMENT RATE," J. Math. App, vol. 19, no. 4, pp. 2633–2646, 2025, https://doi.org/10.30598/barekengvol19iss4pp2633-2646.

[14] M. Zareba, H. Dlugosz, T. Danek, and E. Weglinska, "Big-Data-Driven Machine Learning for Enhancing Spatiotemporal Air Pollution Pattern Analysis," Atmosphere (Basel)., vol. 14, no. 4, 2023, https://doi.org/10.3390/atmos14040760.

[15] N. Amoroso *et al.*, "PSI Spatially Constrained Clustering : The Sibari and Metaponto Coastal Plains," pp. 1–16, 2023.

[16] M. J. Membrive-Jiménez et al., "Burnout in nursing managers: A systematic review and meta-analysis of related factors, levels and prevalence," Int. J. Environ. Res. Public Health, vol. 17, no. 11, pp. 1–10, 2020, https://doi.org/10.3390/ijerph17113983.

[17] M. Rizki, A. Hermawan, and D. Avianto, "Optimization of Hyperparameter K in K-Nearest Neighbor Using Particle Swarm Optimization," vol. 12, no. 1, pp. 71–79, 2024.

[18] Y. N. Dili, E. R. Wulan, and F. Ilahi, "SOLUSI OPTIMAL DENGAN PENDEKATAN MINIMUM SPANNING TREE (MST) MENGGUNAKAN ALGORITMA," pp. 44–50.

[19] N. Hidayati, R. Amalia, and W. S. Eko, "Analisis Spasial Kejadian Demam Berdarah Dengue (DBD) di Kabupaten Bantul Tahun 2022," *Indones. J. Public Heal.*, vol. 18, no. 4, pp. 27–33, 2023, [Online]. Available: https://jurnal.unimus.ac.id/index.php/jkmi,jkmi@unimus.ac.id

[20] K. Nasional and M. Xx, "SPATIAL CLUSTER ING DENGAN METODE SKATER (K'LUSTER ANALYSIS BY TREE EDGE REMOVAL) UNTUK PENGELOMPOKAN SEBARAN COVID-19 DI KABUPATEN TULUNGAGUNG".

[21] Z. HARISQA, "PENERAPAN METODE SKATER DALAM PENGELOMPOKAN SEBARAN TEMPAT WISATA Di PULAU BAWEAN SKRIPSI OLEH: ZIYADA HARISQA NIM.19610085 PROGRAM STUDI MATEMATIKA FAKULTAS SAINS DAN TEKNOLOGI UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG 2024," Universitas Islam Negeri Maulana Malik Ibrahim Malang, 2024.

[22] S. N. Bakri *et al.*, "Analisis klasifikasi Algoritma K-Nearest Neighboar (K-NN) pada struktur Daerah di Kota Medan," pp. 182–193, 2025.

[23] H. Du, D. Lu, Z. Wang, C. Ma, X. Shi, and X. Wang, "Fast clustering algorithm based on MST of representative points," vol. 20, no. July, pp. 15830–15858, 2023, https://doi.org/10.3934/mbe.2023705

[24] R. M. Assunção, M. C. Neves, G. Câmara, and C. Da Costa Freitas, "Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees," Int. J. Geogr. Inf. Sci., vol. 20, no. 7, pp. 797–811, Aug. 2006, https://doi.org/10.1080/13658810600665111

[25] H. T. S. Nurul Fat'Ha, "Identifikasi Autokorelasi Spasial Pada Penggangguran di Jawa Timur Menggunakan Indeks Moran," vol. 8, no. 2, pp. 89–92, 2020, https://doi.org/10.26740/mathunesa.v8n2.p89-92.

[26] P. C. Latue and H. Rakuasa, "Spatial Analysis of Landscape Suitability of Ambon City for Settlement Using Geographic Information System," J. Ris. Multidisiplin dan Inov. Teknol., vol. 1, no. 02, pp. 59–69, 2023, doi: https://jesa.aks.or.id/jesa/article/view/13.

[27] S. Dhouib, "Innovative method to solve the minimum spanning tree problem: The Dhouib-Matrix-MSTP (DM-MSTP)," Results Control Optim., vol. 14, no. September 2023, p. 100359, 2024, doi: https://doi.org/10.1016/j.dsm.2024.01.005