JTAM (Jurnal Teori dan Aplikasi Matematika)
http://journal.ummat.ac.id/index.php/jtam
p-ISSN 2597-7512 | e-ISSN 2614-1175
Vol. 9, No. 3, July 2025, pp. 813-828

Spatial Clustering Analysis of Hand, Foot, and Mouth Disease in Jakarta using Local Indicator of Spatial Association Cluster Map and K-Means Clustering

Fatsa Vidyaningtyas Sabila[1]*, Yekti Widyaningsih[1]
[1]Universitas Indonesia, Indonesia
fatsa.vidyaningtyas@ui.ac.id

Article History:
Received : 28-03-2025
Revised : 24-05-2025
Accepted : 28-05-2025
Online : 01-07-2025

Keywords:
Hand Foot Mouth Disease;
K-Means Clustering;
Local Indicator of Spatial Association;
Moran's Index.

ABSTRACT

Hand, Foot, and Mouth Disease (HFMD) is a infectious disease characterized by ulcers and blisters, primarily affecting children. The objective of this quantitative study is to identify areas with the highest HFMD cases (hotspot areas) in Jakarta in 2024 and to classify areas (districts) based on the number of HFMD cases and variables associated with the disease. The analysis employs the Local Indicator of Spatial Association (LISA) Cluster Map to detect spatial hotspots and K-Means Clustering to group districts by HFMD cases and related variables. LISA is a univariate method for detecting hotspots based on the local Moran's Index that measures spatial dependence, whereas K-Means Clustering is a multivariate method for grouping individuals based on multiple variables. This study uses data from official government sources, including the number of HFMD cases, population density, average number of students per kindergarten, and average number of students per elementary school. The results of this study show that the LISA clustering reveals Kalideres and Cengkareng as High-High (H-H) clusters, while Tanah Abang, Menteng, and Senen form Low-Low (L-L) clusters. Makasar is classified as a Low-High (L-H) cluster. In contrast, the K-Means clustering groups districts into four clusters based on HFMD cases and related demographic factors, sorted in ascending order of HFMD cases. Areas with the lowest HFMD cases tend to have a moderate population density and fewer average number of students per kindergarten, while areas with the highest cases tend to have a lower population density but a higher average number of students per kindergarten. Areas classified as high cases HFMD by both methods, such as Cengkareng, should be prioritized for intervention. Cengkareng represents a district with the highest HFMD cases despite having a relatively low population density, along with a high average number of students per kindergarten and per elementary school.

https://doi.org/10.31764/jtam.v9i3.30339

——————————— ◆ ———————————

A. INTRODUCTION

HFMD is an infectious disease characterized by primary clinical symptoms such as ulcers and blisters on the hands, feet, and mouth. This disease is caused by a group of non-polio Enteroviruses, particularly viruses from the Human Enterovirus A (HEV-A) species, namely Coxsackievirus A16 (CA16) and Enterovirus 71 (EV71) (Meng et al., 2021). Most HFMD cases are mild. However, in some cases, especially in children under five years old, severe symptoms such as meningitis, encephalitis, and pulmonary edema may occur. In critical cases, the disease can progress rapidly and become fatal (Liu et al., 2024). The first HFMD outbreak was reported in 1957 in Toronto, Canada. Most patients recovered within a week with mild complications. In 1959, the term HFMD was first used to describe the disease with similar symptoms (Lei et al., 2015). Since then, HFMD has spread to various countries, including China, Australia, Brunei, Malaysia, Vietnam, and Indonesia (Susanti et al., 2014).

Research on HFMD transmission has become increasingly important as cases continue to rise in multiple countries, including Indonesia (Susanti et al., 2014). To gain a deeper understanding of the spatial distribution and dynamics of HFMD, spatial modeling and clustering approaches have been widely applied. The epidemiological features and spatial clusters of HFMD have been analyzed using various methods in different regions. Spatial autocorrelation techniques such as Moran's I and Local Indicators of Spatial Association (LISA) clustering are commonly used to detect disease hotspots. Moran's I measures the spatial autocorrelation of observations based on feature locations and attribute values to evaluate the overall data pattern (Liu et al., 2024). Meanwhile, Local Indicators of Spatial Association (LISA) identify local indicators of spatial correlation, which can be used to detect regions with significant spatial correlation and hotspot (Yang et al., 2022).

Machine learning clustering methods like K-Means clustering are applied to group areas based on similar characteristics. K-Means algorithm partitions a dataset into k clusters by grouping data points according to their numerical attributes (Syakur et al., 2018). K-means clustering is a type of partitional clustering algorithm that works by dividing a dataset into distinct clusters. This method works by minimizing the squared distance between each data point and the mean of its assigned cluster, ensuring that each point is grouped with the closest cluster center (Ikotun et al., 2023). K-Means is commonly applied in many fields of study because it is easy to use, efficient, and produces effective results (Jie et al., 2020).

Spatial analysis techniques have been widely employed to study the distribution of HFMD cases across various regions. Moran's I and LISA clustering was used to assess the spatial patterns of HFMD cases in Sarawak, Malaysia (Sham et al., 2014), Beijing, China (Qian et al., 2016), Vietnam (Phung et al., 2018), and Qinghai, China (Xu et al., 2018). Meanwhile, K-means clustering was utilized to analyze HFMD case clusters in China, using a dataset that included HFMD cases, meteorological factors, and demographic data (Meng et al., 2021).

This study proposes analysis of the spatial distribution of Hand, Foot, and Mouth (HFMD) in Jakarta in 2024. It involves spatial clustering using the Local Indicators of Spatial Association (LISA) cluster map and machine learning-based clustering of HFMD cases and demographic factors using K-Means Clustering. This study contributes to the existing literature by providing a focused spatial analysis of HFMD in Jakarta, Indonesia. While previous studies have primarily focused on regions such as China, Malaysia, and Vietnam. This study also integrates both spatial autocorrelation (Moran's I and LISA) and machine learning (K-Means) to explore not only spatial clusters but also the demographic characteristics associated with HFMD distribution. The findings are expected to provide valuable insights for targeted public health strategies in managing HFMD outbreaks in Jakarta.

B. METHODS

1. Study Area

This study proposes clustering analysis of HFMD in Jakarta in 2024. Jakarta is the capital city of Indonesia, centered at approximately 6.2088° South latitude and 106.8456° East longitude (Abidin et al., 2015). The area of Jakarta consists of 5 municipality and 1 regency, they are West Jakarta, East Jakarta, Center Jakarta, South Jakarta, North Jakarta, and Kepulauan Seribu. Jakarta has 44 districts across the area. In this study, districts of North Kepulauan Seribu and South Kepulauan Seribu are excluded from this study due to their geographical separation and lack of contiguity with the remaining 42 districts of Jakarta. Therefore, the analysis includes only 42 districts of Jakarta.

2. Dataset

This study utilizes data on HFMD cases and supporting demographic factors collected from the official government sources. All datasets used are publicly accessible and obtained from institutions responsible for maintaining reliable and up-to-date records. The primary data used in this study consists of the number of HFMD cases reported in 42 districts of Jakarta from January 2021 to December 2024. The additional data utilized in this study includes population density, average number of students per kindergarten, and average number of students per elementary school. Population density is obtained by dividing the population size of each district by its area. The average number of students per kindergarten is calculated by dividing the total number of kindergarten students by the number of kindergartens in each district. Similarly, the average number of students per elementary school is calculated by dividing the total number of elementary school students by the number of elementary schools in each district. This study uses quantitative methods, such as LISA clustering and K-Means Clustering. Table 1 shows the detailed information of variables in this study and Figure 1 shows the flowchart of this study.

Table 1. Overview of the data in this study

No	Variable	Initial	Unit	Data Source
1	Number of HFMD cases	HFMD	person	Surveillance, Epidemiology, and Immunization Division of the Jakarta Health Office website (Dinas Kesehatan Provinsi DKI Jakarta, 2025)
2	Population density	PD	person/km^2	Jakarta Civil Registry Office website (Dinas Kependudukan dan Pencatatan Sipil Provinsi DKI Jakarta, 2025)
3	Average number of students per kindergarten	ASK	person/school	Education Data Center of the Directorate General of Early Childhood, Primary, and Secondary Education website (Kemendikdasmen, 2025)
4	Average number of students per elementary school	ASE	person/school	

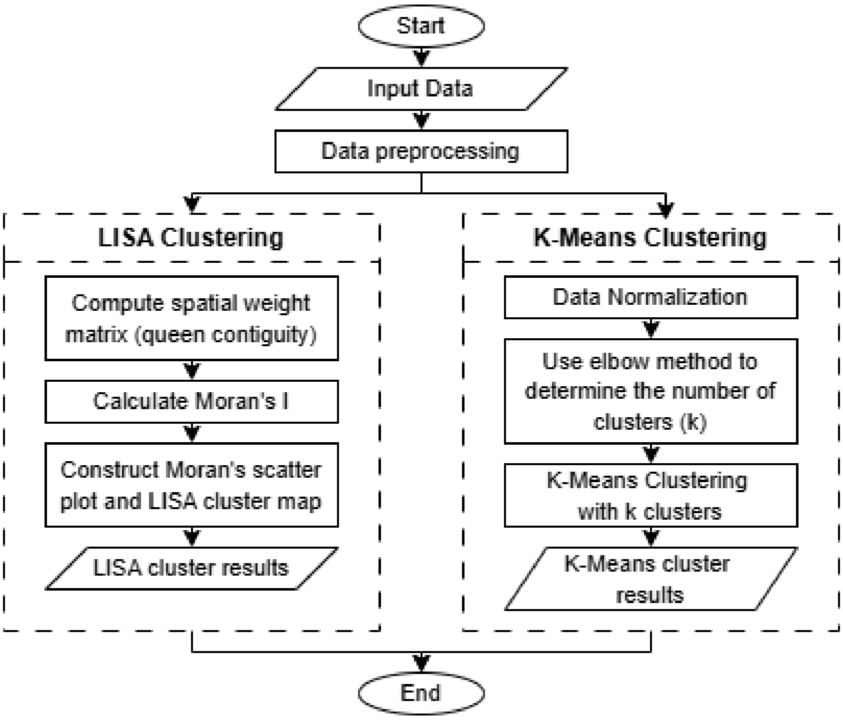


Figure 1. Flowchart of this study

3. Spatial Weight Matrix

Spatial weight matrix is a square matrix that shows how locations are related to each other. Each element of this matrix, denoted as w_{ij}, represents the weight or relationship between locations i and location j among n locations (Suryowati et al., 2018).

$$\boldsymbol{W} = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{13} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{pmatrix} \tag{1}$$

This matrix describes how locations are arranged and whether nearby locations affect each other, which is called spatial dependence. One common method to define spatial relationships is contiguity, which identifies connections between spatial units based on their relative positions (Lesage, 1999). There are several types of contiguity, including linear, rook, bishop, and queen. This study uses queen contiguity, where regions sharing either a common edge or a vertex are considered neighbors. In this case, $w_{ij} = 1$ if location i and location j share a common edge or vertex, and $w_{ij} = 0$ otherwise. For better interpretability, the spatial weight matrix is row-standardized so that the sum of each row equals one (Anselin, 2001). Spatial weight matrix is used to determine Moran's I and LISA clusters.

4. Moran's I

Spatial autocorrelation analysis is a statistical method used to explore the distribution of data within a geographical space. This analysis consists of two aspects, global spatial autocorrelation analysis and local spatial autocorrelation analysis. Global spatial autocorrelation analysis evaluates the overall degree of clustering or dispersion of a phenomenon across an entire region, whereas local spatial autocorrelation analysis identifies and measures localized patterns of clustering or dispersion within a geographical space. Moran's Index, or Moran's I, is used to measure the strength of global spatial autocorrelation (Liu et al., 2024). The value of Moran's I is expressed as follows (Saputro et al., 2021):

$$I = \frac{n \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\left(\sum_{j=1, j \neq i}^{n} w_{ij}\right) \sum_{i=1}^{n} (x_i - \bar{x})^2} \quad (2)$$

where n is the number of observations, x_i is the value of a variable x at location i, $\bar{x}$ is the mean value of variable x of n observations, and w_{ij} is the spatial weight of location i relative to location j, obtained from spatial weight matrix.

The value of Moran's I ranges from -1 to 1, where a value close to 1 indicates positive spatial autocorrelation, a value close to -1 indicates negative spatial autocorrelation, and a value near 0 indicates the absence of spatial autocorrelation (Liu et al., 2024). Moran's I can be visualized using a Moran scatter diagram. This diagram is divided into four quadrant regions. Quadrant I is high-high (H-H), representing areas with high observation values surrounded by other areas with high observation values. Quadrant II is low-high (L-H), indicating areas with low observation values but surrounded by areas with high observation values. Quadrant III is low-low (L-L), represents areas with low observation values surrounded by other low-value areas. Quadrant IV is high-low (H-L), represents areas with high observation values surrounded by areas with low observation values. A Moran scatter diagram in which most areas or objects fall within the L-L and H-H quadrants tends to indicate positive spatial autocorrelation. Conversely, if most areas or objects fall within the L-H and H-L quadrants, negative spatial autocorrelation is likely to be observed (Saputro et al., 2021).

5. Local Indicator of Spatial Association (LISA) Cluster

Local Indicator of Spatial Association (LISA) is used to evaluate the tendency of spatial correlation at each observation location (Saputro et al., 2021). The LISA value is used to identify the presence of significant spatial correlation as well as its outliers. The LISA value is visualized by using LISA cluster Map. The LISA value for each location is expressed as follows (Saputro et al., 2021):

$$L_i = \frac{(x_i - \bar{x}) \sum_{j=1}^{n} w_{ij} (x_i - \bar{x})}{\sum_{j=1}^{n} \left(x_j - \bar{x}\right)^2} \quad (3)$$

where n is the number of observations, x_i is the value of a variable x at location i, $\bar{x}$ is the mean value of variable x of n observations, and w_{ij} is the spatial weight of location i relative to location j.

6. K-Means Clustering

K-Means method divides a dataset into k clusters, grouping data points based on numerical attributes (Syakur et al., 2018). Before the clustering process, the elbow method is applied to identify the optimal number of clusters by evaluating clustering compactness using the total within-cluster sum of squares as the assessment criterion (Meng et al., 2021). The total within-cluster sum of squares, denotes as J, is calculated using Equation 4 (Syakur et al., 2018).

$$J = \sum_{j=1}^{k} \sum_{x_i \in S_j} \left\| x_i - c_j \right\|^2 \tag{4}$$

K-Means method proceeds through the following iterative steps below (Syakur et al., 2018).

a. Determine the number of clusters k and the maximum number of iterations.
b. Randomly initialize k initial centroids $C = \{c_1, c_2, .., c_k\}$ as the initial cluster centers.
c. Assign each data point x_i to the nearest centroid c_j based on Euclidean distance in Equation (5).

$$d(x_i, c_j) = \sqrt{\left(x_{i1} - c_{j1}\right)^2 + \left(x_{i2} - c_{j2}\right)^2} \tag{5}$$

d. Update each centroid c_j using Equation (6) for each cluster as the mean of all data points assigned to it, where S_j is the number of observations in cluster j.

$$c_j = \frac{1}{|S_j|} \sum_{x_i \in S_j} x_i \tag{6}$$

e. If the centroids remain unchanged or the maximum number of iterations is reached, the process stops. Otherwise, return to step 3.

In this study, K-Means clustering is applied to a dataset consists of four variables, such as number of HFMD cases, population density, average number of students per kindergarten, and average number of students per elementary school.

7. Principal Component Analysis for Biplot

Principal Component Analysis (PCA) is a dimensionality reduction technique widely used to transform high-dimensional data into a lower-dimensional representation while preserving variance. It achieves this by computing new orthogonal variables, called principal components (PC), which are linear combinations of the original variables (Jollife & Cadima, 2016). Biplot is a graphical representation of PCA that displays both the observations and the variables in a single plot. It is constructed from a low-rank approximation of the data matrix using the first two or three principal components. Biplots can be utilized to visualize K-Means clustering results by representing both the clustered observations and the relationships between variables in a reduced-dimensional space. Observations are color-coded according to their K-

Means clusters (Markos et al., 2019). The length of a variable vector in biplot indicates its contribution to the principal components, with longer arrows representing stronger influence. The cosine of the angle between any two vectors representing variables is the coefficient of correlation between those variables. Similarly, the cosine of the angle between any vector representing a variable and the axis representing a given principal component is the coefficient of correlation between those two variables (Jollife & Cadima, 2016).

C. RESULT AND DISCUSSION

1. Analysis of Spatial and Temporal Patterns

The temporal distribution of Hand, Foot, and Mouth Disease (HFMD) cases in Jakarta over the past four years (2021-2024) was analyzed to study the trend of the disease. Based on Figure 2, HFMD cases in Jakarta are increasing every year. The highest total cases was in 2024, they are 417 cases in Jakarta. The increase of HFMD cases has occurred in Indonesia, particularly after the COVID-19 pandemic, following the rise in public mobility after the termination of Large-Scale Social Restrictions (PSBB) in 2021 (Suni, 2024).

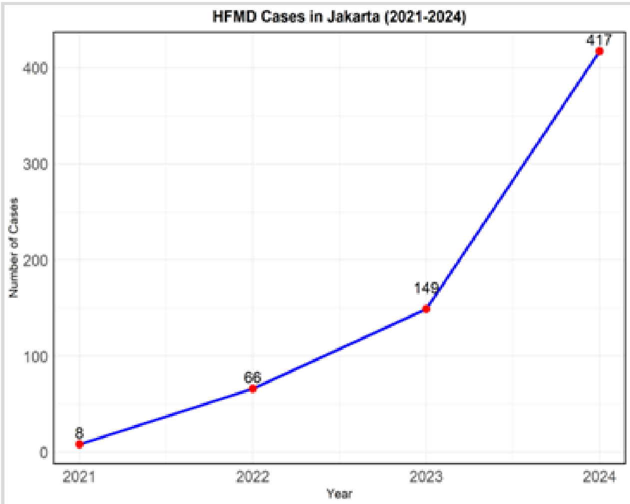


Figure 2. The number of HFMD cases in Jakarta (2021-2024)

The spatial autocorrelation test was conducted using Moran's I statistic to examine the spatial dependence of HFMD cases, with spatial weight matrix based on queen contiguity. Moran's I was calculated using the *spdep* package in R Studio. The results for each year were summarized in Table 2 below.

Table 2. Moran's I for HFMD Cases in Jakarta

Year	Moran's I	p-value
2021	0.0225	0.2704
2022	0.3282	0.000008*
2023	0.0163	0.2282
2024	0.2054	0.0039*

The Moran's I values and their corresponding p-values indicate that spatial autocorrelation was present in 2022 and 2024, as their p-values were statistically significant (p-value < 0.05). In 2022, Moran's I was 0.3282 with p-value of 0.000008, showing the presence of spatial dependence of HFMD cases. This suggests that in 2022, neighboring regions tended to experience similar levels of HFMD incidence, possibly due to shared environmental conditions or social behaviors.

Similarly, in 2024, Moran's I was 0.2054 with p-value of 0.0039, also showing the lower presence of spatial dependence of HFMD cases than 2022. This suggests that in 2024, neighboring regions still tended to have similar HFMD incidence levels, but the clustering was less appeared, possibly reflecting more diffuse transmission or the impact of health interventions. In contrast, the years 2021 and 2023 did not have significant spatial autocorrelation, as their p-values were above the significance threshold, indicating that HFMD cases in those years were more randomly distributed across the study area. The 2024 dataset was chosen for the clustering analysis because it satisfies the assumption of spatial autocorrelation, which is essential for spatial clustering. Moreover, using the 2024 data ensures that the analysis reflects the most recent epidemiological conditions, offering timely and relevant insights into the spatial distribution of HFMD cases in Jakarta, as shown in Figure 3.

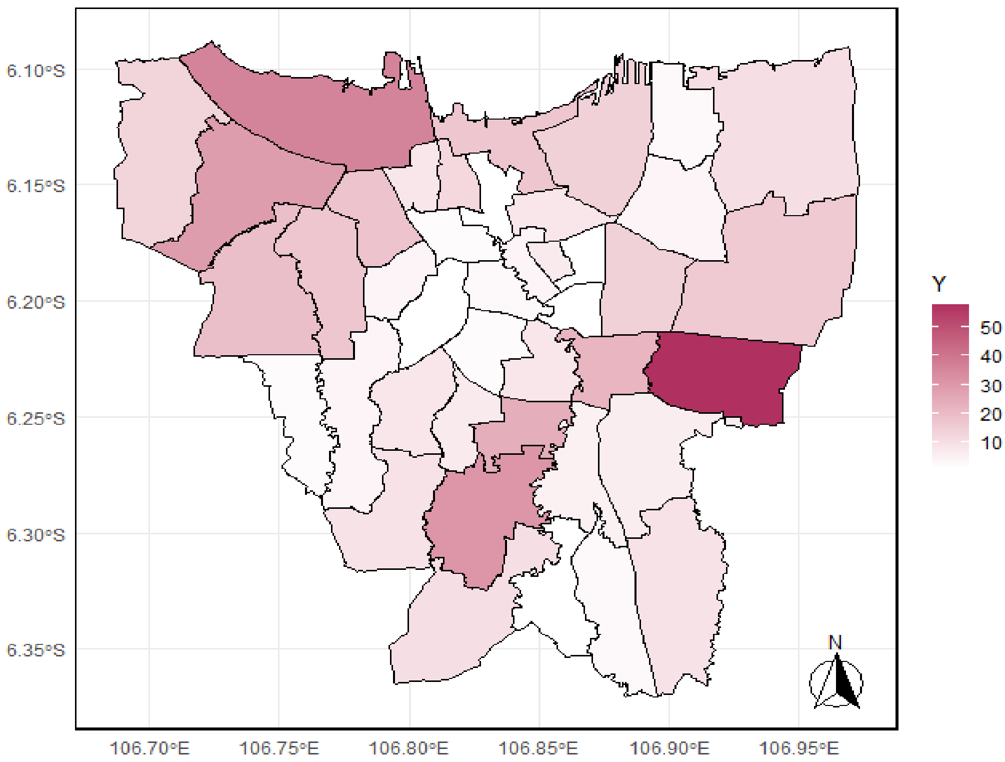
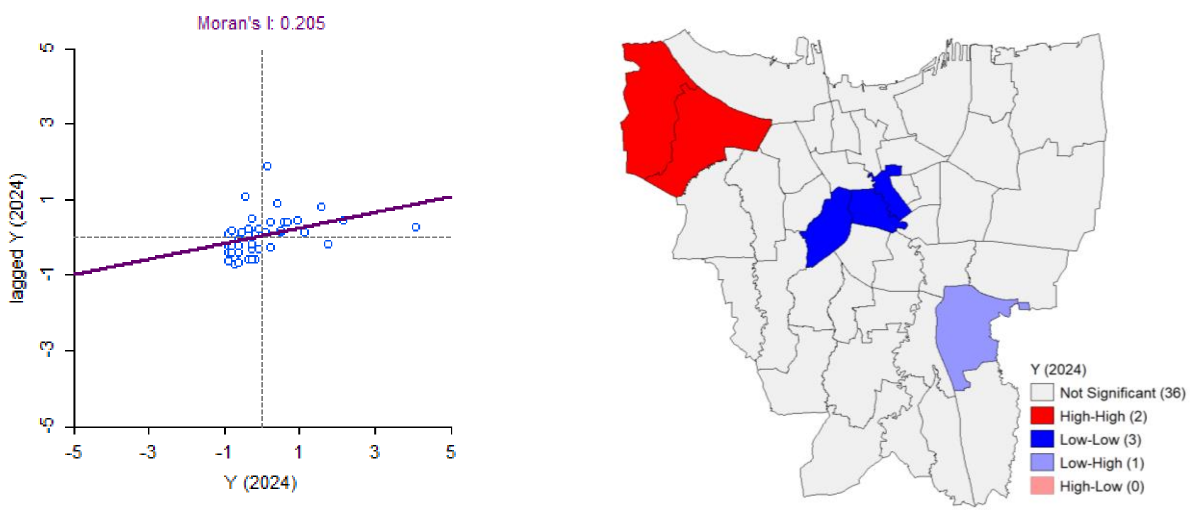


Figure 3. Spatial distribution of HFMD Cases in Jakarta (2024)

2. LISA Cluster Analysis

The Moran scatter plot and LISA cluster map were visualized to further explore the spatial dependence in 2024, illustrating the relationship between local and global spatial autocorrelation. Moran scatter plot and LISA cluster map were visualized using *Geoda* version 1.22, as shown in Figure 4.



A B

Figure 4. (A) Moran scatter plot of HFMD Cases in Jakarta (2024); and (B) LISA cluster map of HFMD Cases in Jakarta (2024)

Moran scatter plot of HFMD cases in Jakarta in 2024 is shown in Figure 4A. The Moran's I value of 0.205 suggests a weak pattern of similar case levels clustering together. This means areas with high HFMD cases are slightly more likely to be near other high-case areas (H-H), and low-case areas tend to be near other low-case areas (L-L). Some regions show the opposite pattern, where high-case areas are surrounded by low-case areas (H-L) and (L-H), but these cases are less common. To further analyze local spatial clustering, the LISA cluster map was visualized to identify significant (H-H), (L-L), (H-L), and (L-H) cluster areas.

The LISA cluster map in Figure 4B shows how HFMD cases are spread across districts of Jakarta in 2024. Kalideres and Cengkareng are identified as High-High (H-H) clusters, indicating high HFMD cases in these areas and their surrounding regions. Kalideres and Cengkareng are located in West Jakarta. On the other hand, Low-Low (L-L) clusters are found in Tanah Abang, Menteng, and Senen, indicating low HFMD cases in these areas and their surrounding regions. Tanah Abang, Menteng, and Senen are located in Central Jakarta. Makasar is identified a Low-High (L-H) cluster, where HFMD cases are low in Makasar, while the surrounding areas have a high number of cases. Makasar is located in East Jakarta. In conclusion, The LISA clustering pattern in 2024 indicates that West Jakarta (specifically Kalideres and Cengkareng) requires targeted intervention due to high concentrations of HFMD cases, while Central and East Jakarta (particularly Menteng, Senen, Tanah Abang, and Makasar) may offer insights into effective prevention strategies that contribute to maintaining lower case numbers.

3. K-Means Clustering Analysis

K-Means clustering was implemented to the number of HFMD cases, population density, average number of students per kindergarten, and average number of students per elementary school. K-Means clustering was implemented using the *factoextra* package in R Studio. Before clustering process, the data is normalized to ensure uniform scaling and to improve clustering accuracy. The optimal number of clusters was determined using the Elbow Method, where the within-cluster sum of squares (WCSS) from Equation 4 was plotted against the number of clusters. The point where the rate of decrease in WCSS slows down and forms an elbow, it indicates the optimal number of clusters. Based on this method that shown in Figure 5, four clusters were selected for this study.

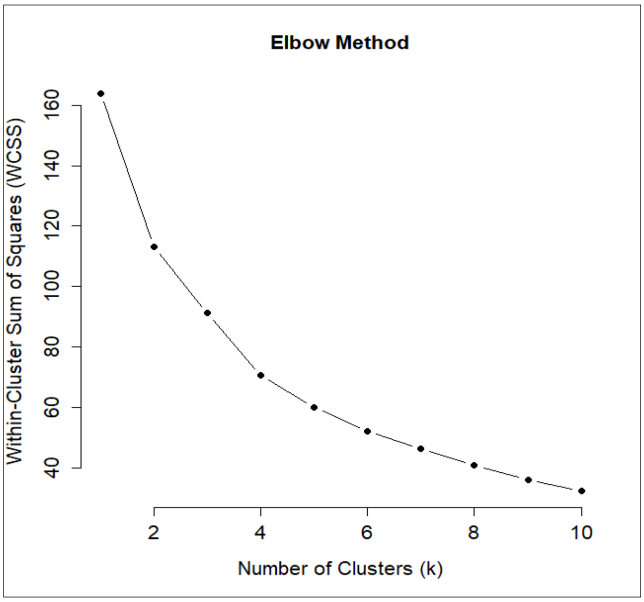


Figure 5. Determination of the optimal number of clusters using elbow method

The summary of each cluster with rounded average values is presented in Table 3, while the district names of each cluster are listed in Table 4. The clusters are sorted based on the average of HFMD cases. Cluster 1 represents districts with the lowest average HFMD cases, a middle-range population density, the lowest average number of students per kindergarten, and the lowest average number of students per elementary school. Cluster 2 represents districts with slightly higher HFMD cases and the highest population density, with a slightly greater average number of students per kindergarten and per elementary school than Cluster 1. Cluster 3 represents districts with higher HFMD cases but has the lowest population density, along with the highest average number of students per kindergarten and per elementary school. Cluster 4 is distinct, with the highest HFMD cases despite relatively low population density, and a higher average number of students per kindergarten and per elementary school than Clusters 1 and 2. This pattern is also visualized in Figure 6.

Table 3. Summary of K-Means Clustering Results for HFMD Cases and Related Factors

Cluster	Average of HFMD	Average of PD	Average of ASK	Average of ASE
1	7	17,834	33	267
2	8	39,485	34	283
3	9	16,784	40	388
4	39	15,935	41	361

Table 4. Summary of K-Means Clustering Results for HFMD Cases and Related Factors

Cluster	Districts
1	Cempaka Putih, Gambir, Grogol Petamburan, Kebayoran Baru, Kelapa Gading, Mampang Prapatan, Menteng, Pancoran, Sawah Besar, Senen, Setiabudi, Tamansari, Tanah Abang, Tanjung Priok, Tebet
2	Jatinegara, Johar Baru, Kemayoran, Matraman, Palmerah, Tambora
3	Cakung, Cilandak, Cilincing, Cipayung, Ciracas, Jakagarsa, Kalideres, Kebayoran Lama, Kebon Jeruk, Kembangan, Koja, Kramat Jari, Makasar, Pademangan, Pasar Rebo, Pesanggrahan, Pulo Gadung
4	Cengkareng, Duren Sawit, Pasar Minggu, Penjaringan

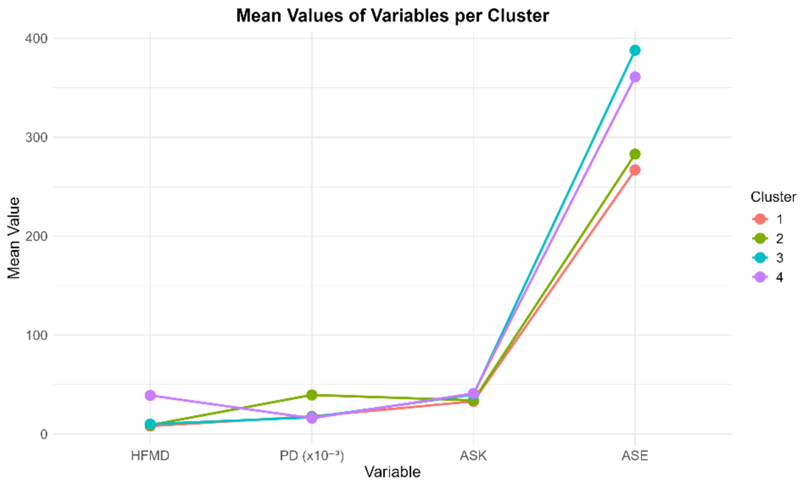
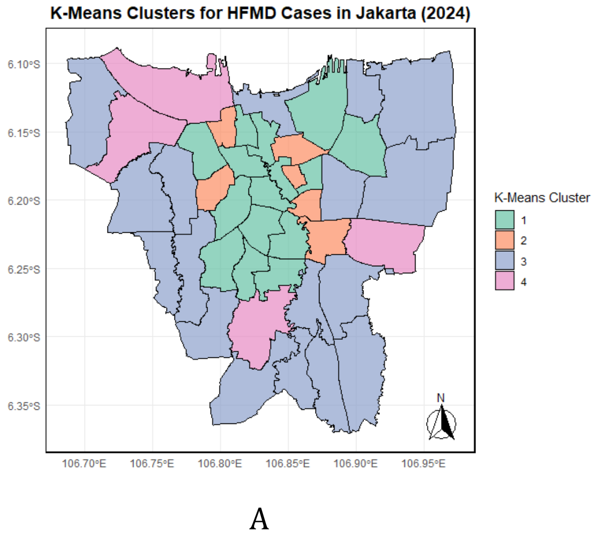


Figure 6. Mean values of each variables for each K-Means cluster

Table 5. Characteristics of districts in Cluster 4

No	Districts	HFMD	PD	ASK	ASE
1	Cengkareng	29	21921	38	325
2	Duren Sawit	58	19666	41	372
3	Pasar Minggu	31	14970	36	349
4	Penjaringan	36	7181	49	399

Table 5 presented the characteristics of Cluster 4, which consists of subdistricts with the highest average number of HFMD cases. The table shows that these districts tend to have much higher HFMD case numbers compared to others. Additionally, these subdistricts also have a higher average number of students per elementary school (ASE) than districts in other clusters, as shown in Figure 7.



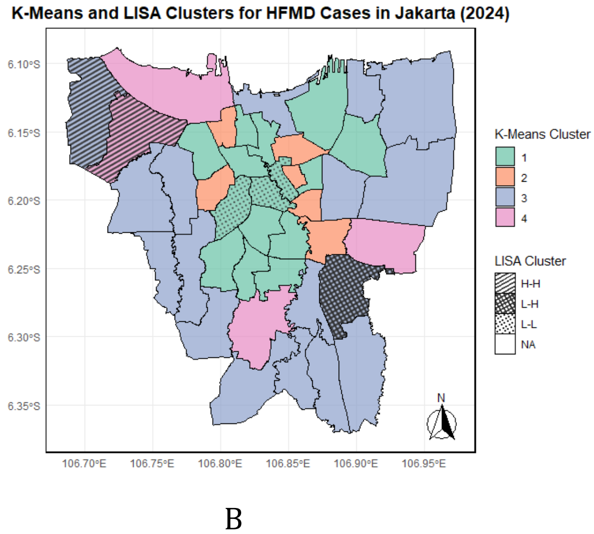


Figure 7. K-Means clusters for HFMD cases in Jakarta (2024): (A) K-Means clustering results; and (B) K-Means clusters overlaid with LISA clusters.

K-Means clustering results were also visualized using map in Figure 7B. LISA classified Kalideres and Cengkareng as High-High (H-H) clusters, meaning these areas have high HFMD cases and are surrounded by other high-case areas. K-Means Cluster 4 also included Cengkareng but adds Duren Sawit, Pasar Minggu, and Penjaringan to the high-HFMD case group. This suggests that these additional areas, though not marked as H-H in LISA, have characteristics that align with high HFMD cases. LISA classified Tanah Abang, Menteng, and Senen as Low-Low (L-L) clusters, meaning they have few HFMD cases and are surrounded by other low-case areas. K-Means Cluster 1 included Menteng, Senen, and Tanah Abang, along with other districts like Cempaka Putih, Gambir, and Kelapa Gading. This indicates that these areas not only have low HFMD cases but also share similar demographic characteristics. LISA classified Makasar as a Low-High (L-H) cluster, meaning it has low cases despite being surrounded by high-case areas. K-Means classified Makasar in Cluster 3, which has slightly higher HFMD cases but the lowest population density.

Both the LISA and K-Means clustering methods provide insights into the spatial distribution of HFMD cases in Jakarta, but they focus on different aspects. The LISA cluster is based only on HFMD case number and highlights localized spatial dependencies, while K-Means clustering highlights additional demographic factors, such as population density, average number of students per kindergarten, and average number of students per elementary school. Areas classified as high-HFMD in both methods, such as Cengkareng, should be prioritized for intervention efforts. Districts consistently identified as low-case areas, including Menteng, Senen, and Tanah Abang, may provide insights into factors contributing to lower case numbers. Variations in clustering across districts indicate that HFMD risk is influenced not only by spatial dependence but also by demographic conditions.

4. Biplot Analysis

A biplot based on Principal Component Analysis (PCA) for dimensionality reduction is used to visualize the clustering results of K-Means clustering. The PCA-based biplot is shown in Figure 8. The x-axis represents Dim1 (first dimension), which explains 44.6% of the total variation in the data. The y-axis represents Dim2 (second dimension), which explains 22.1% of the total variation. The biplot explains 66.7% of the total variation in the data. Each color represents a different cluster, and the ellipses show how the data points within each cluster are spread out. The arrows indicate the influence and direction of the original variables (HFMD, PD, ASK, ASE) in this reduced space. HFMD is linked to the negative side of Dim1 and is most associated with Cluster 4. PD is positioned on the positive side of Dim1 and is closely related to Cluster 2. ASK and ASE are positively correlated and strongly influence Cluster 3, meaning this cluster has higher values for these two variables. The average ASE in Cluster 3 is 388, which is higher than the ASE values in the other clusters. Variables with arrows in similar directions are more correlated, while those in opposite directions are negatively correlated. It is evident that HFMD does not show any correlation with the other variables.

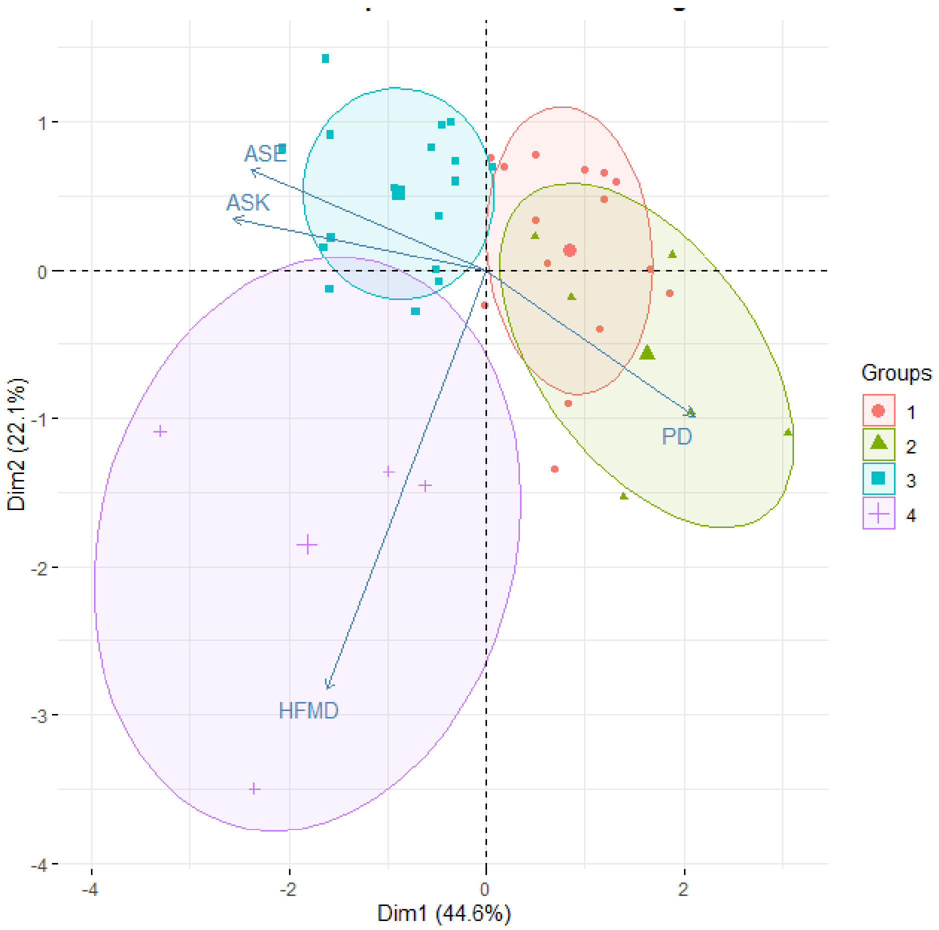


Figure 8. PCA-based Biplot for K-Means clusters

D. CONCLUSION AND SUGGESTIONS

The LISA and K-Means clustering results are complementary rather than contradictory. LISA and K-Means clustering complement each other by capturing different but connected aspects of HFMD spread. LISA focuses on spatial autocorrelation, highlighting areas where HFMD cases cluster geographically and revealing local patterns of disease transmission between neighboring districts. LISA provides insights into the spatial dependence of HFMD cases, showing how cases spread across neighboring districts. Meanwhile, K-Means clustering reveals how demographic factors influence case distribution.

Areas identified as high-HFMD in both methods (Cengkareng) should be prioritized for interventions. Cengkareng represent a district with the highest HFMD cases despite relatively low population density, and a high average number of students per kindergarten and per elementary school. Low-case areas identified by both methods (Menteng, Senen, Tanah Abang) may offer insights into contributing factors to lower case numbers. These areas represent districts with the lowest average HFMD cases, a middle-range population density, the lowest average number of students per kindergarten, and the lowest average number of students per elementary school. The differentiation in some districts suggests that HFMD risk is not only driven by spatial dependence but also by demographic factors. For effective disease control, both spatial clustering (LISA) and demographic clustering (K-Means) should be considered when designing intervention strategies. Although differences in study areas limit direct comparison with previous research, this study supports the importance of integrating spatial and demographic factors to better understand and control HFMD transmission.

The PCA-based biplot further shows how the data is grouped into clusters and how different factors influence these groupings. HFMD cases is mainly linked to cluster 4, while PD (population density) is linked to cluster 1 and cluster 2 , meanwhile ASK (average of students per kindergarten) and ASE (average of students per elementary school) are linked to cluster 3, showing distinct patterns in how these variables interact. Additionally, ASK and ASE are closely correlated, meaning that the average of student per kindergarten and per elementary school tend to increase together.

Future studies could improve the analysis by incorporating additional environmental and healthcare-related variables. Environmental variables such as temperature, humidity, and air quality can affect virus survival and human exposure. Meanwhile, healthcare factors such as the availability of medical facilities and vaccination coverage may have a crucial role in early detection and controlling the spread of the disease. Incorporating these factors into spatial models would provide a clearer picture of disease dynamics. Additionally, exploring temporal trends through time-series analysis could offer further insights into the dynamics of the disease transmission because it allows for the detection of seasonal patterns and changes over time.

ACKNOWLEDGEMENT

The authors declare that there are no specific acknowledgments for this study.

REFERENCES

Abidin, H. Z., Andreas, H., Gumilar, I., & Brinkman, J. J. (2015). Study on the risk and impacts of land subsidence in Jakarta. *Proceedings of the International Association of Hydrological Sciences*, *372*, 115–120. https://doi.org/10.5194/piahs-372-115-2015

Anselin, L. (2001). *Spatial Econometrics*. In Baltagi, BH (Ed.), *A companion to theoretical econometrics*, Blackwell, Malden, MA.

Dinas Kependudukan dan Pencatatan Sipil Provinsi DKI Jakarta. (2025). *Profil Perkembangan Kependudukan DKI Jakarta*. https://kependudukancapil.jakarta.go.id/profile_perkembangan_adminduk/

Dinas Kesehatan Provinsi DKI Jakarta. (2025). *Seksi Surveilans Epidemiologi dan Imunisasi*. https://surveilans-dinkes.jakarta.go.id/sarsbaru/dashboard.php

Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, *622*, 178–210. https://doi.org/10.1016/j.ins.2022.11.139

Jie, C., Jiyue, Z., Junhui, W., Yusheng, W., Huiping, S., & Kaiyan, L. (2020). Review on the Research of K-means Clustering Algorithm in Big Data. *2020 IEEE 3rd International Conference on Electronics and Communication Engineering, ICECE 2020*, (pp. 107–111). https://doi.org/10.1109/ICECE51594.2020.9353036

Jollife, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. In *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065). Royal Society of London. https://doi.org/10.1098/rsta.2015.0202

Kemendikdasmen. (2025). *Data Pokok Pendidikan (DAPODIK)*. https://dapo.dikdasmen.go.id/sp/1/010000

Lei, X., Cui, S., Zhao, Z., & Wang, J. (2015). Etiology, pathogenesis, antivirals and vaccines of hand, foot, and mouth disease. In *National Science Review, 2(3), 268–284.* Oxford University Press. https://doi.org/10.1093/nsr/nwv038

Lesage, J. P. (1999). *The Theory and Practice of Spatial Econometrics*. http://www.econ.utoledo.edu

Liu, J., Wang, H., Zhong, S., Zhang, X., Wu, Q., Luo, H., Luo, L., & Zhang, Z. (2024). Spatiotemporal Changes and Influencing Factors of Hand, Foot, and Mouth Disease in Guangzhou, China, From 2013 to 2022: Retrospective Analysis. *JMIR Public Health and Surveillance*, *10*. https://doi.org/10.2196/58821

Markos, A., D'enza, A. I., & van de Velden, M. (2019). Beyond tandem analysis: Joint dimension reduction and clustering in R. *Journal of Statistical Software*, *91*, Article 10. https://doi.org/10.18637/jss.v091.i10

Meng, D., Xu, J., & Zhao, J. (2021). Analysis and prediction of hand, foot and mouth disease incidence in China using Random Forest and XGBoost. *PLoS ONE*, *16*(12), e0261629. https://doi.org/10.1371/journal.pone.0261629

Phung, D., Nguyen, H. X., Nguyen, H. L. T., Do, C. M., Tran, Q. D., & Chu, C. (2018). Spatiotemporal variation of hand-foot-mouth disease in relation to socioecological factors: A multiple-province analysis in Vietnam. *Science of the Total Environment*, *610–611*, 983–991. https://doi.org/10.1016/j.scitotenv.2017.08.158

Qian, H., Huo, D., Wang, X., Jia, L., Li, X., Li, J., Gao, Z., Liu, B., Tian, Y., Wu, X., & Wang, Q. (2016). Detecting spatial-temporal cluster of hand foot and mouth disease in Beijing, China, 2009-2014. *BMC Infectious Diseases*, *16*(1), 206. https://doi.org/10.1186/s12879-016-1547-6

Saputro, D. R. S., Widyaningsih, Y., Widyaningsih, P., Sutanto, & Widiastuti. (2021). Spatio-temporal patterns of dengue hemorrhagic fever (DHF) cases with local indicator of spatial association (LISA) and cluster map at areas risk in Java-Bali Indonesia. *AIP Conference Proceedings*, *2326(1),* 020027. https://doi.org/10.1063/5.0040334

Sham, N. M., Krishnarajah, I., Akma Ibrahim, N., & Lye, M.-S. (2014). Temporal and spatial mapping of hand, foot and mouth disease in Sarawak, Malaysia. *Geospatial Health, 8*(2), 503–507. https://doi.org/10.4081/gh.2014.39

Suni, N. S. P. (2024). Mewaspadai lonjakan kasus hand, foot, mouth disease pascalibur Lebaran 2024. *Info Singkat Pusat Analisis Keparlemenan DPR RI, 16*(8/II), 1–8. https://berkas.dpr.go.id/pusaka/files/info_singkat/Info%20Singkat-XVI-8-II-P3DI-April-2024-1957.pdf

Suryowati, K., Bekti, R. D., & Faradila, A. (2018). A Comparison of Weights Matrices on Computation of Dengue Spatial Autocorrelation. *IOP Conference Series: Materials Science and Engineering*, *335*(1), 012052. https://doi.org/10.1088/1757-899X/335/1/012052

Susanti, N., Purnamawati, S., Setiawaty, V., & Biomedis dan Teknologi Dasar Kesehatan Balitbangkes, P. (2014). *Deteksi Penyebab dan Sebaran Kasus Kejadian Luar Biasa Hand Foot and Mouth Diseases (HFMD) Tahun 2008-2012*.

Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018). Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster. *IOP Conference Series: Materials Science and Engineering*, *336*(1), 012017. https://doi.org/10.1088/1757-899X/336/1/012017

Xu, L., Shi, Y., Rainey, J. J., Zhang, Z., Zhang, H., Zhao, J., Li, Y., Rao, H., Li, Y., Liao, Q., & Ma, Y. (2018). Epidemiological features and spatial clusters of hand, foot, and mouth disease in Qinghai Province, China, 2009-2015. *BMC Infectious Diseases*, *18*(1), 351. https://doi.org/10.1186/s12879-018-3509-7

Yang, D. L., Li, W., Pan, M. H., Su, H. X., Li, Y. N., Tang, M. Y., & Song, X. K. (2022). Spatial analysis and influencing factors of pulmonary tuberculosis among students in Nanning, during 2012–2018. *PLoS ONE*, *17*(5), e0268472. https://doi.org/10.1371/journal.pone.0268472