Comparison Analysis of Hierarchical Clustering and K-Means Methods in Grouping Provinces in Indonesia Based on Dengue Hemorrhagic Fever (DHF) Cases

Alfidha Rahmah[1*], Nida Faoziatun Khusna[2], Safril Ahmadi Sanmas[3], Syifa Aulia[4], Shinta Amaria[5], Fatkhurokhman Fauzi[6]

[1,2,3,4,5,6] *Statistics Department, Faculty of Agricultural Science and Technology, Universitas Muhammadiyah Semarang, Indonesia*

*corr-author: alfidhrhm@gmail.com

Abstract - Indonesia, a tropical country, experiences climate variations that influence the spread of infectious diseases, including Dengue Hemorrhagic Fever (DHF). The increase in DHF cases necessitates clustering provinces based on their vulnerability to design effective mitigation strategies. This study compares two clustering methods: Hierarchical Clustering and K-Means Clustering. Within the hierarchical clustering analysis, five linkage methods were evaluated: Average Linkage, Complete Linkage, Single Linkage, Ward's Method, and Centroid Linkage. The best linkage method was identified using the cophenetic correlation coefficient, indicating that Average Linkage produced the most representative cluster structure, resulting in three distinct groups. For the K-Means method, the optimal number of clusters was determined using the Silhouette Coefficient, which also indicated three clusters. Clustering performance evaluation revealed that Average Linkage outperformed K-Means, with a higher Silhouette Score of 0.552. The resulting clusters categorized provinces into three risk groups: high-risk areas (e.g., DKI Jakarta), moderate-risk areas (e.g., West Java and East Java), and low-risk areas, comprising the remaining provinces in Indonesia.

Keywords: clustering, DHF, Hierarchical Clustering, K-Means

I. INTRODUCTION

Dengue Hemorrhagic Fever (DHF) continues to attract global attention due to its significant impact on public health. In Indonesia, DHF remains a major health issue, with an increasing number of cases reported annually. DHF is caused by the dengue virus, transmitted through the bite of infected Aedes aegypti or Aedes albopictus mosquitoes [1]. Common symptoms include high fever, skin rash, headache, muscle pain, nausea, and vomiting [2]. Several risk factors can increase an individual's likelihood of dengue virus infection and developing DHF, such as climatological conditions, sociodemographic characteristics, behavior, and environmental factors [3]. These risk factors are complex, and the chance of contracting DHF may involve a combination of several of these aspects. Therefore, preventive measures play a crucial role in reducing the spread of DHF.

Indonesia faces significant challenges in managing the continuous rise in DHF cases, highlighting the importance of thoroughly understanding the disease's transmission patterns. In this context, cluster analysis of provinces based on DHF cases can effectively detail and characterize the epidemiological features of the disease. Since the epidemiological dynamics of DHF vary from year to year, advanced analytical methods, such as clustering, are necessary to map provinces into distinct groups that reflect their unique epidemiological characteristics.

Cluster analysis partitions a set of data objects into subsets or clusters, where objects within a cluster share similar characteristics and differ from those in other clusters [4]. Data can be grouped into clusters using hierarchical and non-hierarchical approaches [5]. Hierarchical clustering begins by grouping objects with the highest similarity and progressively merges groups with decreasing similarity until a tree structure with defined levels or hierarchies is formed [6]. In contrast, non-hierarchical clustering employs partitioning or density-based techniques [5]. Research comparing both clustering methods is essential to evaluate their effectiveness on specific datasets.

Comparing the effectiveness of hierarchical and non-hierarchical methods, such as K-Means, is crucial for identifying the most suitable approach for a given dataset.

Hierarchical clustering produces a nested structure that can represent relationships among regions, whereas K-Means clustering groups regions based on similarity to centroids [7]. However, many previous studies have compared hierarchical and K-Means clustering without first evaluating the choice of linkage method used in hierarchical clustering. Hierarchical clustering provides several linkage methods—including single, complete, average, Ward's, and centroid linkage—each generating distinct cluster structures. Therefore, before comparing the overall performance of these clustering techniques, it is essential to determine the most suitable hierarchical linkage method. This study addresses this gap by calculating the cophenetic correlation coefficient for each hierarchical linkage method to identify the approach that best represents the structure of Dengue Hemorrhagic Fever (DHF) case data in Indonesia.

This study aims to compare the performance of Hierarchical Clustering and K-Means Clustering in grouping Indonesian provinces based on DHF cases. Both methods will be evaluated for their effectiveness using the silhouette score. The silhouette score measures how closely each data point fits within its cluster compared to other clusters and indicates how well-separated each cluster is. The silhouette score ranges from -1 to 1, with higher values representing better-defined clusters [8].

Several studies have previously investigated comparisons between clustering methods. Research by Jujun W. and Monika S. M. (2023), which clustered regencies/cities in North Sumatra based on crop productivity, showed that K-Means Clustering outperformed Average Linkage when the optimal number of clusters was two. Their study concluded that K-Means was more effective in distinguishing between low and high crop productivity, assigning low productivity to cluster 1 and high productivity to cluster 2 [9]. Another study conducted by [10], which analyzed malaria clusters in Papua Province using Single Linkage and K-Means methods, indicated that the Single Linkage method, with five clusters, yielded more accurate results than K-Means, which had nine clusters [10]. Additionally, a study by [11], comparing K-Means and Hierarchical Clustering methods to identify areas at risk of stunting, found that K-Means was more effective, producing two clusters with a Silhouette Coefficient of 0.48 and a Calinski-Harabasz index of 10.49 [11].

Building on these findings, this study not only compares the performance of Hierarchical and K-Means Clustering but also addresses a methodological gap by identifying the most suitable hierarchical linkage method before conducting evaluations. Through this approach, the study contributes to more accurate and reliable clustering of DHF cases, thereby supporting enhanced data-driven decision-making in public health planning.

II. METHOD

A. *Data Source*

This study uses secondary data obtained from the Indonesian Ministry of Health (Kementerian Kesehatan Republik Indonesia—Kemenkes RI), Statistics Indonesia (Badan Pusat Statistik—BPS), and the National Disaster Management Agency (Badan Nasional Penanggulangan Bencana—BNPB). The variables used in this research include the number of DHF cases, flood occurrences, population density, average temperature, average humidity, total rainfall, and sunshine duration. The dataset covers each Indonesian province over a one-year period.

B. *Research Method*

The steps taken in the research to cluster provinces in Indonesia based on DHF cases are as follows.

- Data collection and exploration.
- Descriptive analysis for each variable.
- Assessment of clustering assumptions, including multicollinearity testing to evaluate correlations among independent variables. Multicollinearity was assessed using the Variance Inflation Factor (VIF). If the VIF value is less than 10, multicollinearity is considered absent; otherwise, it is present. The VIF formula is shown in (1) [12].

$$VIF = \frac{1}{1-R^2} \tag{1}$$

  where R^2 is the coefficient of determination between the dependent and independent variables.
- Data standardization. The data were standardized using Z-score normalization, a common technique that transforms variables by centering them around the mean and scaling based on standard deviation [13]. This method preserves the distributional properties of the data and improves robustness against outliers, which is essential in distance-based methods such as hierarchical and k-means clustering [14]. The Z-score normalization formula is shown in (2)

$$Z = \frac{X-\mu}{\sigma} \tag{2}$$

  where X is the original value, μ is the mean value, and σ is the standard deviation.
- Hierarchical clustering.

- Identifying the most suitable clustering method by comparing cophenetic correlation coefficients. A value closer to 1 indicates a more optimal clustering result, whereas a value near 0 suggests poor clustering performance. The cophenetic correlation formula is shows in (3) [15]:

$$r_{coph} = \frac{\sum_{i<k}(d_{ik}-\bar{d})(d_{Cik}-\bar{d}_c)}{\sqrt{\left[\sum_{i<k}(d_{ik}-\bar{d})^2\right]\left[\sum_{i<k}(d_{Cik}-\bar{d}_c)^2\right]}} \quad (3)$$

where d_{ik} is the Euclidean distance between object i and object k, $\bar{d}$ is the average of d_{ik}, d_{Cik} is the cophenetic distance between object i and object k, and $\bar{d}_c$ is the average of d_{Cik}.

- Comparing hierarchical methods, including average, complete, single, Ward, and centroid linkage. These methods differ in how they measure the distance between clusters. The formula for the Average Linkage method is shown in (4) [15]:

$$d_{(ij)k} = \frac{\sum_a \sum_b d_{ab}}{n_{(ij)} n_k} \quad (4)$$

where d_{ab} is the distance between an object in the cluster (ij) and an object in the cluster k, $n_{(ij)}$ is the number of objects in the cluster (ij), and n_k is the number of objects in the cluster k.

- Performing clustering using the selected clustering method.
- Conducting a characteristic analysis of the clustering results.

• K-Means clustering.

- Determining the optimal number of clusters using the silhouette method, which measures how closely related an object is to its cluster compared to other clusters. The silhouette coefficient formula is shown in (5) [16].

$$s(i) = \frac{b(i)-a(i)}{max\{a(i),b(i)\}} \quad (5)$$

where $s(i)$ is the average distance of the data point i to all other points in the same cluster.

- Performing clustering using K-Means algorithm.
- Conducting a characteristic analysis of the resulting clusters.

• Comparing the results of hierarchical and K-Means clustering using the silhouette coefficient. A higher average silhouette value (closer to 1) indicates better clustering quality, whereas a value approaching -1 suggests poor cluster separation. The criteria for interpreting the silhouette coefficient are presented in Table I [17].

III. RESULT AND DISCUSSION

The research results are derived from analyses using descriptive analysis and cluster analysis.

A. Descriptive Analysis

A descriptive statistical approach was employed to understand the initial characteristics of the data. Table II presents the mean, standard deviation, minimum, and maximum values for each variable.

Based on Table II, the average number of flood occurrences per province in Indonesia in 2022 was 45, with a minimum of 3 and a maximum of 186 events. Population density across provinces ranged from 10.00 to 16,158 people per square kilometer. The average temperature was 28°C, accompanied by an average humidity of 80.40%. Sunshine duration varied between 3.74 and 7.99 hours. Total rainfall averaged 2,898.32 mm, with a range from 879.40 to 4,950.50 mm. The average DHF incidence cases per province was 4,235. The highest number of cases occurred in West Java (365,964 cases), while the lowest was in Maluku (96 cases). The distribution of DHF cases in Indonesia in 2022 is illustrated in the Fig. 1 Fig. 1 displays the distribution of DHF cases across Indonesia, categorized into three groups. The group with the highest number of reported DHF cases (⩾ 36,594) consists solely of West Java Province. The next group, with moderate DHF cases ranging from 8,138 to 36,594, includes North Sumatra, DKI Jakarta, Central Java, and East Java. The remaining provinces fall into the low-case group, with fewer than 8,138 cases.

B. Multicollinearity Test

To evaluate the presence of multicollinearity among the independent variables, the Variance Inflation Factor (VIF) was calculated using (1). The results are presented in Table III.

The VIF values for each variable did not exceed 10, indicating the absence of multicollinearity among the variables. Therefore, the Euclidean distance metric is suitable for use in this analysis.

TABLE I
SILHOUETTE COEFFICIENT CRITERIA

Silhouette Coefficient	Evaluation Criteria
0.7 < SC < 1.0	Strong Structure
0.5 < SC < 0.7	Medium Structure
0.25 < SC < 0.5	Weak Structure
SC ≤ 0.25	No Structure

TABLE II
DESCRIPTIVE ANALYSIS

Variable	Mean	Standard Deviation	Min	Max
Flood Occurrences	45.03	47	3	186
Population Density	751.15	2751.23	10	16158
Average Temperature	28	0.92	24.99	29.41
Average Humidity	80.4	3.25	73.66	86.33
Sunshine Duration	4.99	0.85	3.74	7.79
Total Rainfall	2898.32	837.35	879.4	4950.5
Number of DHF Cases	4235.47	6566.35	96	36594

C. Hierarchical Clustering

1) Determination of Clustering Method: The cophenetic correlation coefficient was used to determine the most appropriate clustering method among various hierarchical clustering approaches. A coefficient value approaching 1 indicates a better representation of the data structure by the clustering solution. The coefficient was calculated using (3), and the results are presented in Table IV.

Fig. 1 Distribution of DHF cases in each province in indonesia

TABLE III
MULTICOLLINEARITY TEST RESULTS

Variable	VIF
Flood Occurrences	3.159316
Population Density	1.406733
Average Temperature	1.784259
Average Humidity	2.137254
Sunshine Duration	1.192499
Total Rainfall	1.148927
Number of DHF Cases	2.925648

TABLE IV
COPHENETIC CORRELATION RESULTS

Average	Complete	Single	Ward	Centroid
0.9301497	0.9026534	0.8947319	0.4860531	0.9273629

Table IV shows that the Average Linkage method yielded the highest cophenetic correlation coefficient compared to other hierarchical methods. Therefore, it was selected as the most suitable method for clustering DHF-related data.

2) Clustering Results with the Average Linkage Method: The clustering results for provinces in Indonesia based on DHF cases in 2022, using the Average Linkage method, are illustrated in Fig. 2

.

As shown in Fig. 2, the provinces were grouped into three clusters. The first cluster consists of one province, the second includes two provinces, and the third comprises 31 provinces, as listed in the Table V.

Following the clustering process, a characteristic analysis was conducted to examine the mean values of each variable within the clusters. The characteristics of each cluster are summarized in Table VI.

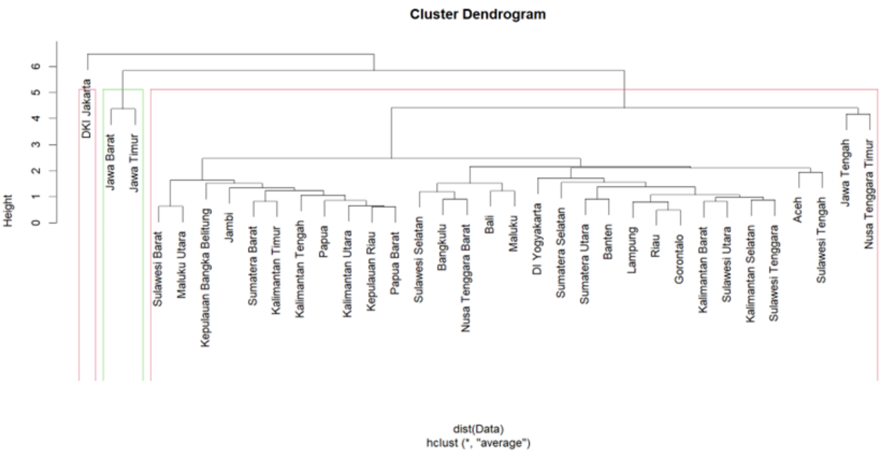

Fig. 2 Dendrogram cluster average linkage

TABLE V
CLUSTER RESULTS WITH THE AVERAGE LINKAGE METHOD

Cluster	Number of Members	Provinces
1	1	DKI Jakarta
2	2	Jawa Barat, Jawa Timur
3	31	Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Bangka Belitung, Jawa Tengah, DI Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Selatan, Sulawesi Tengah, Sulawesi Tenggara, Sulawesi Barat, Gorontalo, Papua, Kepulauan Riau, Maluku, Maluku Utara, Papua Barat

TABLE VI
CLUSTER CHARACTERISTICS BASED ON MEAN VALUES

Variable	*Cluster 1*	*Cluster 2*	*Cluster 3*
Flood Occurrences	11	169.5	38.10
Population Density	16158	1095.5	231.94
Average Temperature	28.47	25.57	28.15
Average Humidity	77.21	81.55	80.42
Sunshine Duration	3.85	5.21	5.01
Total Rainfall	2136.3	3282.55	2898.16
Number of DHF Cases	8138	24941.5	2773.71

Cluster 1 consists solely of DKI Jakarta, characterized by an exceptionally high population density (16,158 people/km²) and a moderate number of DHF cases (8,138 cases in 2022), despite experiencing relatively few flood occurrences. This indicates that densely populated urban areas may still face significant DHF risk due to increased exposure to dengue-carrying mosquitoes and limited space for effective vector control [18]. Cluster 2, which includes West Java and East Java, recorded the highest average number of DHF cases (24,941.5), accompanied by the highest levels of rainfall and flood events. These findings align with previous studies showing that excessive rainfall and frequent flooding contribute to increased mosquito breeding and heightened DHF transmission risk [19]. Although the population density in this cluster is lower than in Cluster 1, it remains substantial, indicating the combined impact of environmental and demographic factors. Cluster 3, comprising 31 provinces, reported the lowest average number of DHF cases (2,773.71) and the lowest population density. Despite high rainfall, these provinces experienced fewer flood events, possibly due to better water management practices. The relatively low incidence of DHF in Cluster 3 suggests that both environmental resilience and reduced human exposure contribute to the prevention of the disease.

D. K-Means Clustering

1) Determination of the Number of Clusters: Determining the number of clusters is necessary before performing non-hierarchical clustering. This can be done randomly or by using specific methods, such as the Silhouette Coefficient. The Silhouette Coefficient combines cohesion and separation approaches to help identify the optimal number of clusters Based on Fig. 3, the highest silhouette value is observed at k=3. Therefore, the optimal number of clusters to be formed is 3.

2) Clustering Results with the K-Means Method: Three clusters were formed to categorize provinces in Indonesia based on DHF case factors. The results are shown in Fig. 4. Based on the clustering of provinces shown in Fig. 4, three clusters of provinces have been formed. The first cluster has two members, the second has 16 members, and the third has 16 members, as listed in the Table VII.

Next, the average of each variable in each cluster is calculated to observe the characteristics of each cluster, which can be seen in Table VIII.

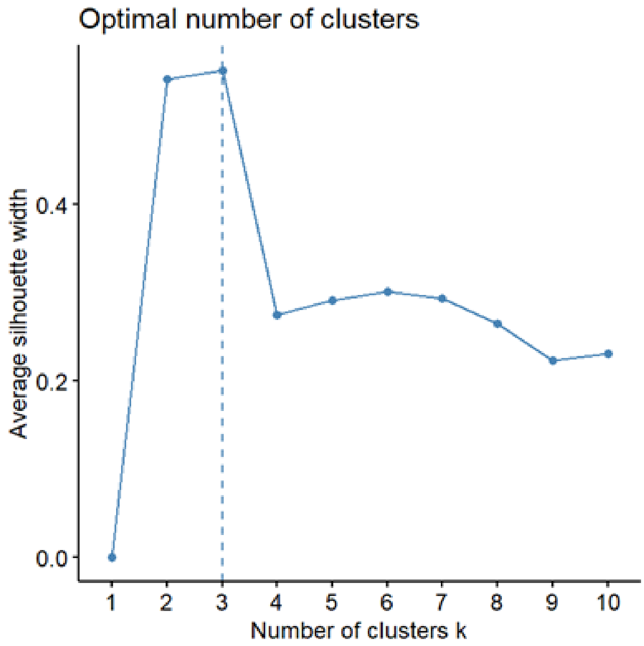

Fig. 3 Silhouette graph

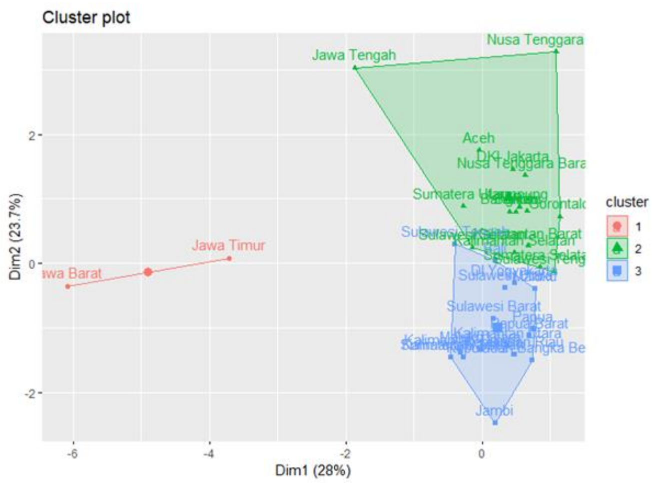

Fig. 4 Cluster plot k-means

TABLE VII
RESULTS OF CLUSTERING USING THE K-MEANS METHOD

Cluster	Number of Members	Province
1	1	*Jawa Barat, Jawa Timur*
2	16	*Aceh, Sumatera Utara, Riau, Sumatera Selatan, Bengkulu, Lampung, DKI Jakarta, Jawa Tengah, Banten, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Selatan, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo*
3	16	*Sumatera Barat, Jambi, Bangka Belitung, Kepulauan Riau, DI Yogyakarta, Bali, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua*

TABLE VIII
CLUSTER CHARACTERISTICS BASED ON MEAN VALUES

Variable	Cluster 1	Cluster 2	Cluster 3
Flood Occurrences	169.5	51.81	22.69
Population Density	1095.5	1270.25	189
Average Temperature	25.57	28.73	27.59
Average Humidity	81.55	77.78	82.86
Sunshine Duration	5.21	5.05	4.9
Total Rainfall	3282.55	2868.79	2879.83
Number of DHF Cases	24941.5	3891	1991.69

As shown in Table VIII, Cluster 1 recorded the highest average DHF cases (24941.5), along with the highest flood occurrences and rainfall. This suggests a significant correlation between heavy rainfall, flooding, and increased DHF risk. Cluster 2 has a moderate average of DHF (3891) and higher temperatures, but lower humidity, indicating that DHF risk is not solely driven by flood events, but also by climatic interactions. Cluster 3 shows the lowest DHF cases (1991.69), lowest population density, and lowest flood occurrences, which may reflect lower transmission potential due to lower human–mosquito contact and less environmental stress.

E. *Comparison of the Average Linkage and K-Means Method*

The two clustering methods, K-means and Average Linkage, were evaluated using the Silhouette Coefficient, calculated using Equation (5), to assess the quality of the clustering results. Table IX presents the evaluation results for clustering quality with $k = 3$.

From the silhouette score calculation in Table IX with k = 3, the Silhouette Coefficient for the Average Linkage method is 0.552. This indicates that the clustering results produced by the Average Linkage strong clustering quality. In contrast, the K-Means method yielded a score of 0.263, which is lower than that of the Average Linkage method. Therefore, the comparison shows that the Average Linkage method provides more optimal clustering results than the K-Means method.

These findings are consistent with previous research that applied similar clustering approaches. A study by [20], which clustered provinces based on maternal health services, also found that the Average Linkage method outperformed K-Means in terms of internal validation metrics. This similarity suggests that Average Linkage tends to deliver more reliable clustering outcomes, particularly in regional classification problems involving health-related indicators. The consistency of these results across different studies reinforces the robustness of hierarchical methods like Average Linkage in capturing meaningful patterns in public health data [20].

TABLE IX
EVALUATION RESULTS OF CLUSTERING QUALITY USING THE SILHOUETTE COEFFICIENT

Method	Silhouette Coefficient
Average Linkage	0.5524718
K-Means	0.2629804

IV. CONCLUSION

Based on the analysis results, the provinces in Indonesia were grouped into three clusters using both Hierarchical and K-Means methods. The selection of the hierarchical clustering approach was based on the highest Cophenetic Correlation Coefficient value (0.9301), which was achieved by the Average Linkage method, outperforming four other hierarchical methods. Subsequently, the Average Linkage method was compared with the K-Means method to evaluate the quality of the resulting clusters. From the evaluation results using the silhouette score, it can be concluded that the Average Linkage method outperformed K-Means method in grouping Indonesian provinces based on DHF cases, achieving a silhouette score of 0.552. In the Average Linkage clustering, the first cluster consists of DKI Jakarta, the second cluster includes West Java and East Java, and the third cluster comprises all remaining provinces. Meanwhile, the K-Means method produced a different grouping: the first cluster includes West Java and East Java, the second cluster includes 16 provinces (Aceh, North Sumatra, Riau, South Sumatra, Bengkulu, Lampung, DKI Jakarta, Central Java, Banten, West Nusa Tenggara, East Nusa Tenggara, West Kalimantan, South Kalimantan, South Sulawesi, Southeast Sulawesi, and Gorontalo), and the third cluster includes the remaining 15 provinces. This study has a limitation in that it only utilizes data from a single year (2022), thus providing only a snapshot of the epidemiological condition at one point in time, without capturing long-term trends or seasonal variations. Therefore, future research is recommended to use multi-year data to analyze the temporal dynamics and trends of DHF cases. It is also advisable to integrate clustering with spatial or time-series analysis to obtain deeper insights into the spread patterns of DHF across provinces.

ACKNOWLEDGEMENT

This research was made possible through the support of the Statistical Research in Health and Environment Group, Undergraduate Statistics Study Program, Universitas Muhammadiyah Semarang, and all parties who provided assistance and support throughout the research process. The study was also funded by PKM-AI, which contributed to its successful implementation.

REFERENCES

[1] M. Kanan, M. Naffaa, A. Alanazi, F. Nasser, A. A. Alsaiari, M. Almehmadi, A. Assiry, H. Muzafar, H. Katam, A. Arar, S. M. B. Asdaq, Abida, M. Imran, and T. Dzinamarira, "Genetic variants associated with dengue hemorrhagic fever. A systematic review and meta-analysis," *J Infect Public Health*, vol. 17, no. 4, pp. 579–587, 2024, doi: 10.1016/j.jiph.2024.02.001.

[2] World Health Organization (WHO), "Dengue and Severe Dengue." [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue

[3] R. Pakaya, D. Daniel, P. Widayani, and A. Utarini, "Spatial model of Dengue Hemorrhagic Fever (DHF) risk: scoping review," *BMC Public Health*, vol. 23, no. 1, pp. 1–16, 2023, doi: 10.1186/s12889-023-17185-3.

[4] N. Shahid, "Comparison of hierarchical clustering and neural network clustering: an analysis on precision dominance," *Sci Rep*, vol. 13, no. 1, pp. 1–11, 2023, doi: 10.1038/s41598-023-32790-3.

[5] L. C. R. Antunes, A Comparison Between Hierarchical and Non-Hierarchical Software Clustering, vol. 1, no. 1. Association for Computing Machinery.

[6] E. Balbi, P. Cianfarra, L. Crispini, S. Tosi, and G. Ferretti, "Hierarchical-agglomerative clustering analysis of geomorphic features applied to tectonic investigation of terrestrial planets: An example from Claritas Fossae, Mars," *Icarus*, vol. 420, no. June, p. 116197, 2024, doi: 10.1016/j.icarus.2024.116197.

[7] M. Nazari, A. Hussain, and P. Musilek, "Applications of Clustering Methods for Different Aspects of Electric Vehicles," *Electronics (Basel)*, vol. 12, no. 4, p. 790, Feb. 2023, doi: 10.3390/electronics12040790.

[8] V. Michalakopoulos, E. Sarmas, I. Papias, P. Skaloumpakas, V. Marinakis, and H. Doukas, "A machine learning-based framework for clustering residential electricity load profiles to enhance demand response programs," *Appl Energy*, vol. 361, p. 122943, May 2024, doi: 10.1016/j.apenergy.2024.122943.

[9] J. Wijaya and M. S. Manurung, "Regency / City Clustering in North Sumatra Based on Food Crop Productivity," *Journal of Analytical Research, Statistics and Computation*, vol. 2, no. 2, pp. 36–51, 2023.

[10] A. M. Sroyer, S. A. Mandowen, and F. Reba, "Analisis Cluster Penyakit Malaria Provinsi Papua Menggunakan Metode Single Linkage Dan K-Means," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 7, no. 3, pp. 147–154, Jan. 2022, doi: 10.25077/TEKNOSI.v7i3.2021.147-154.

[11] I. Indra, N. Nur, Muh. Iqram, and N. Inayah, "Perbandingan K-Means dan Hierarchical Clustering dalam Pengelompokan Daerah Beresiko Stunting," *INOVTEK Polbeng - Seri Informatika*, vol. 8, no. 2, p. 356, Nov. 2023, doi: 10.35314/isi.v8i2.3612.

[12] A. Mahmudan, "Clustering of District or City in Central Java Based COVID-19 Case Using K-Means Clustering," *Jurnal Matematika, Statistika dan Komputasi*, vol. 17, no. 1, pp. 1–13, Aug. 2020, doi: 10.20956/jmsk.v17i1.10727.

[13] C. Wongoutong, "The impact of neglecting feature scaling in k-means clustering," *PLoS One*, vol. 19, no. 12, p. e0310839, Dec. 2024, doi: 10.1371/journal.pone.0310839.

[14] T. Zhu and Y. Han, "Enhancing Beer Recommendations through Clustering: A Comparison of Hierarchical and K-means Clustering Methods on Normalized Data," *in Proceedings of the 2nd International Conference on Mathematical Statistics and Economic Analysis, MSEA* 2023, May 26–28, 2023, Nanjing, China, EAI, 2023. doi: 10.4108/eai.26-5-2023.2334332.

[15] Iis, I. Yahya, Gusti Ngurah A. Wibawa, Baharuddin, Ruslan, and L. Laome, "Penggunaan Korelasi Cophenetic Untuk Pemilihan Metode Cluster Berhierarki Pada Mengelompokkan Kabupaten/Kota Berdasarkan Jenis Penyakit di Provinsi Sulawesi Tenggara Tahun 2020," *Prosiding Seminar Nasional Sains Dan Terapan (Sinta)*, no. April, pp. 1–16, 2022.

[16] K. Pratama Simanjuntak and U. Khaira, "Hotspot Clustering in Jambi Province Using Agglomerative Hierarchical Clustering Algorithm," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 1, no. 1, pp. 7–16, 2021.

[17] I. F. Ashari, E. Dwi Nugroho, R. Baraku, I. Novri Yanda, and R. Liwardana, "Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta," *Journal of Applied Informatics and Computing*, vol. 7, no. 1, pp. 89–97, 2023, doi: 10.30871/jaic.v7i1.4947.

[18] Y. S. Amelinda, R. A. Wulandari, and A. Asyary, "The effects of climate factors, population density, and vector density on the incidence of dengue hemorrhagic fever in South Jakarta Administrative City 2016-2020: an ecological study," *Acta Biomedica*, vol. 93, no. 6, 2022, doi: 10.23750/abm.v93i6.13503.

[19] S. Istiqamah, A. Arsin, and S. Sirajuddin, "Average rainfall effect on dengue hemorrhagic fever in Kendari City, Indonesia in 2014-2018," *EAS Journal of Parasitology and Infectious Diseases*, vol. 1, no. 5, pp. 98–102, 2019, doi: 10.36349/EASJPID.2019.v01i05.003.

[20] A. Azzahra and A. W. Wijayanto, "Comparison of Agglomerative Hierarchical and K-Means in Grouping Provinces Based on Maternal Health Services," *SISTEMASI*, vol. 11, no. 2, p. 481, May 2022, doi: 10.32520/stmsi.v11i2.1829.