Journal of Data Science, Informatics and Computer Engineering (JDSICE),
Vol. 1, Issue 1
February 2026
ISSN: XXXX-YYYY

A Spatial-Temporal Framework for Hepatitis-A Vulnerability Analysis Using Hierarchical Clustering: Case Study of Pacitan Regency, Indonesia

Arna Fariza*, Fanny Maulana Rizky, Arif Basofi
Department of Informatics and Computer Engineering, Politeknik Elektronika Negeri Surabaya, Indonesia
**Corresponding author: Arna Fariza; email: arna@pens.ac.id*

Manuscript received July 24, 2026, revised September 24, 2026, accepted September 25, 2026
doi: zz.yyyyy/xxxxx.pp134-1399

Abstract – *Hepatitis A remains a significant public health concern in Indonesia, particularly in rural and coastal regions such as Pacitan Regency, where inadequate sanitation and contaminated water sources contribute to recurring outbreaks. Traditional mitigation strategies, including public hygiene campaigns and local sanitation initiatives, often lack the spatial precision and temporal depth required to identify and prioritize high-risk areas effectively. This study proposes a spatial-temporal vulnerability mapping framework using agglomerative hierarchical clustering to classify 12 sub-districts in Pacitan Regency based on four key indicators: population density, sanitation rate, clean water availability, and health indicators. The hierarchical clustering evaluation showed that Ward consistently achieved the highest Silhouette and Calinski–Harabasz scores across the three years, while Average obtained the lowest mean Davies–Bouldin Index of 0.61, demonstrating complementary strengths across the clustering evaluation metrics. The digital interface supports informed decision-making by health authorities and raises public awareness by delivering accessible insights to residents regarding the vulnerability status of their respective districts.*

Keywords: *Agglomerative hierarchical clustering, Hepatitis-A, Pacitan Regency, spatial-temporal visualization.*

I. Introduction

Hepatitis A is an infectious disease caused by the hepatitis A virus (HAV), which primarily affects the liver and is commonly transmitted through the fecal-oral route, particularly in areas with inadequate sanitation and contaminated water supplies [1]. In Indonesia, hepatitis A remains a public health concern, especially in rural and coastal regions, where periodic outbreaks have been reported in schools and communities [2].

Pacitan Regency in East Java, Indonesia, has experienced recurring Hepatitis A outbreaks, often escalating into public health emergencies requiring immediate intervention [3]. The 2019 outbreak was particularly severe, affecting 1,310 individuals across multiple subdistricts in Pacitan [4]. Studies identified the Kaligoro River as a primary source of contamination, worsened by communal food-sharing practices and inadequate sanitation in densely populated areas [3].

Identifying vulnerable regions over time and space is crucial for effective surveillance and intervention by local health authorities, including the Public Health Services of the Pacitan Regency. Traditional efforts to mitigate these outbreaks include promoting Clean and Healthy Living Behaviors (PHBS) and localized water sanitation initiatives, which have increased public awareness but lack the precision to identify high-risk areas effectively [5]. These methods often fail to provide actionable insights for targeted interventions, limiting their long-term effectiveness [6]. However, conventional epidemiological mapping often fails to capture the dynamic nature of disease spread, as it lacks the integration of both spatial and temporal dimensions [7].

Recent advances in geospatial analysis and machine learning offer new opportunities for health vulnerability assessment. In particular, hierarchical clustering has proven effective for uncovering patterns in complex health datasets and grouping areas with similar characteristics [8][9]. When integrated with spatial-temporal visualization techniques, this approach can provide actionable insights for targeted public health responses [10].

The study employs agglomerative hierarchical

clustering, a form of unsupervised machine learning, to group 12 sub-districts based on their similarity across the four criteria, consist of population density, sanitation rate, clean water availability rate, and health indicators. These factors represent a combination of environmental, demographic, and healthcare-related variables that collectively affect transmission potential and public health resilience. By calculating Euclidean distances and applying various linkage method, we uncover natural groupings that reflect varying degrees of vulnerability. This clustering is performed on a temporal basis, allowing us to observe changes in sub-district classification across multiple years. The results are then visualized spatial-temporal maps of hepatitis A vulnerability across the regency.

The contributions of this paper include: (1) the development of a vulnerability mapping approach tailored to hepatitis A, (2) the application of agglomerative hierarchical clustering to group 12 sub-district with similar risk profiles, and (3) the spatial-temporal visualization of clustering results to aid Public Health Services of the Pacitan Regency decision-making.

II. Related Works

Understanding the dynamics of disease outbreaks and identifying regions at varying levels of risk requires robust methodologies for analyzing spatial-temporal data. Clustering techniques are widely used in epidemiology to group geographical areas based on shared characteristics, enabling policymakers to target interventions more effectively. Among these techniques, hierarchical clustering has emerged as a versatile tool due to its ability to uncover relationships within complex, multidimensional datasets [11].

The application of hierarchical clustering in public health has expanded significantly over the past decade. It is particularly suitable for epidemiological analysis because it can form nested clusters, allowing researchers to conduct multi-level examinations of disease patterns [12]. Unlike partition-based methods such as k-means, hierarchical clustering does not require predefining the number of clusters, which is beneficial for exploratory data analysis [13].

Hierarchical clustering operates by iteratively merging (agglomerative) or splitting (divisive) clusters, generating a tree-like structure known as a dendrogram, which visually represents data relationships. In the context of Hepatitis A vulnerability mapping, this technique provides a useful framework for visualizing relative risk levels across districts, making it easier to identify high-priority intervention areas. Agglomerative clustering, in particular, is favored in public health studies. It begins with each data point as an individual cluster and successively merges the closest pairs until all points belong to a single cluster [14].

The linkage criterion plays a vital role in determining how clusters are formed. Ward's method, which minimizes the variance within clusters, is widely used for its robustness and is effective in spatial epidemiology applications [15]. Average linkage, on the other hand, emphasizes internal cohesion of clusters [13], while complete and single linkage methods are useful in identifying extreme clustering scenarios—either compact or widely dispersed data groups [14].

This research adopts agglomerative hierarchical clustering with various linkage methods to comprehensively assess vulnerability patterns. Ward's linkage is utilized to produce compact clusters, average linkage for balanced grouping, and complete/single linkage to explore sensitivity to spatial outliers. Each method has demonstrated effectiveness in previous disease clustering studies, including mapping dengue fever and cholera outbreaks [16][17].

Feature selection is also essential to clustering success. Epidemiological variables such as population density, sanitation coverage, clean water availability, and health indicators are selected as clustering inputs. These factors align with prior studies emphasizing the importance of combining socio-environmental and health-related indicators for accurate disease vulnerability modelling [18][19]. To ensure compatibility across variables, standardization and normalization techniques are applied—a best practice in hierarchical clustering [20].

Finally, integrating hierarchical clustering with spatial-temporal visualization enhances the ability to detect and communicate patterns of disease vulnerability. Studies show that clustering paired with GIS-based visualization improves understanding of geographical disparities in health risks and supports evidence-based public health planning [21].

III. Proposed Methodology

This research aims to assess the risk of Hepatitis A across various districts in Pacitan by analyzing key health and environmental indicators. The system developed for this purpose employs a hierarchical clustering algorithm to classify districts based on the risk levels associated with Hepatitis A, considering factors such as population density, sanitation, water availability, and existing health conditions. The system design can be seen in Figure 1.

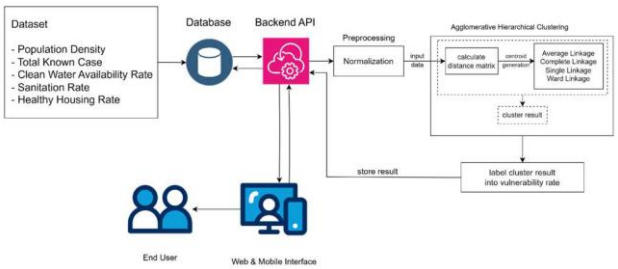


Fig. 1. System Design

The system dynamically updates as new data is entered, ensuring that the classifications remain current. These risk

classifications are visually represented in a web-based dashboard, allowing users to interact with and explore the data. The general workflow begins with the collection of relevant data from official sources, followed by preprocessing and clustering using the hierarchical clustering method. The resulting clusters categorize the districts into different risk levels—such as low, medium, or high—and these categories are represented using visualizations on a web interface for ease of interpretation.

A. Dataset Formation

The dataset used in this study was obtained from the Public Health Services of the Pacitan Regency. It comprises records from the years 2017, 2018, and 2019, detailing key indicators for 21 sub-districts. These indicators include:

- Population density, representing the number of individuals residing in a specific area;
- Sanitation rate, indicating the proportion of households with adequate sanitation facilities;
- Clean water availability rate, showing the percentage of households with access to clean water; and
- Health indicators, specifically including the number of Hepatitis A cases reported during the years in question.

This dataset is comprehensive and allows for an annual analysis of risk levels, reflecting changes in environmental and health conditions across the district.

To better understand the relationships between the features used in the clustering process, we conducted a Pearson correlation analysis. Pearson's correlation coefficient measures the linear relationship between two variables, with values ranging from -1 to 1. A coefficient of 1 indicates a perfect positive correlation, -1 indicates a perfect negative correlation, and 0 indicates no linear relationship.

Pearson's correlation is particularly useful in this context as it helps identify potential multicollinearity issues, where highly correlated features might distort the clustering results, particularly for algorithms like K-means that rely on distance-based metrics. By understanding the feature correlation, we can determine whether dimensionality reduction techniques, such as Principal Component Analysis (PCA), should be applied to reduce redundancy and improve the clustering process. The Pearson correlation coefficients between the features were calculated, and a heatmap matrix was generated to visually inspect the strength and direction of relationships. This allowed us to identify pairs of features that exhibit strong correlations and to evaluate the overall structure of the feature set.

Figure 2 presents the correlation matrix among the features used in the analysis. Several key observations can be identified. The Sanitation Level and Clean Water features exhibit a strong positive correlation (r = 0.54), indicating that areas with better sanitation levels tend to have better access to clean water. Similarly, the Population Density and Healthy Housing features show a significant positive correlation (r = 0.60), suggesting that areas with higher population densities tend to have better housing conditions. In contrast, the Total Case and Population Density features show a very weak correlation (r = -0.053), as do the Clean Water and Population Density features (r = 0.0091), indicating a minimal relationship between these variables.

These observations suggest that some features are highly correlated, which could potentially affect the quality of the clustering. For instance, the strong correlation between Sanitation Level and Clean Water suggests that these features may convey similar information, potentially leading to redundancy in the data and influencing the performance of distance-based clustering methods. In contrast, weakly correlated features such as Total Case and Population Density could provide unique information for the clustering process.

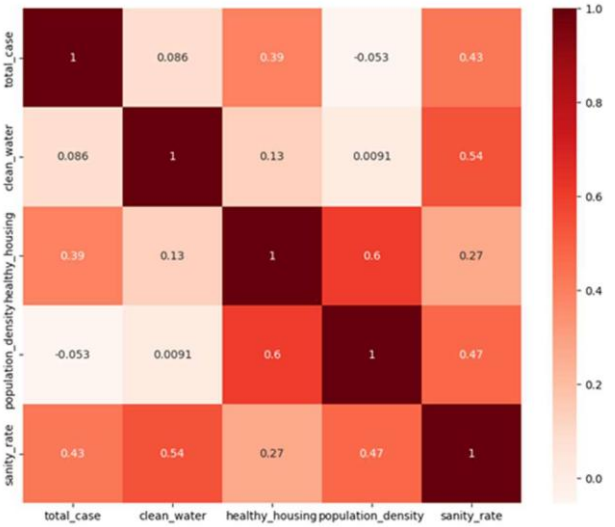


Fig. 2. Heatmap of Feature Similarity Correlation Matrix

B. Pre-processing

Before clustering, the dataset undergoes a preprocessing stage to ensure that all variables are standardized and suitable for analysis. This involves applying min-max normalization, which scales each variable to a uniform range, typically [0, 1]. Normalization is essential for balancing the influence of variables with differing units or scales, ensuring that no single indicator disproportionately affects the clustering results. The normalization process is performed using the formula:

$$X_{normalized} = \frac{X - X_{min}}{X - X_{min}} \quad (1)$$

where X represents the original value, and X_{min} and X_{max} are the minimum and maximum values of the variable in the dataset, respectively. After normalization, the dataset is ready for clustering, with all variables contributing equally to the risk classification.

C. Agglomerative Hierarchical Clustering for Hepatitis-A Vulnerability Assessment

The clustering process in this study utilizes a hierarchical clustering algorithm to group data based on similarities. The objective is to identify patterns or clusters within the dataset that can inform decision-making processes for stakeholders. This process is carried out on the preprocessed dataset, which has been normalized to ensure uniformity in data scale. The workflow of agglomerative hierarchical clustering for assessing Hepatitis-A vulnerability is as follows:

1) Calculates the distance matrix

Initially, the algorithm calculates the distance matrix, a pairwise comparison of data points based on their features. The Euclidean distance is employed as the metric to measure similarities or dissimilarities between the data points. For two data xi and xj, the Euclidean distance d is calculated as:

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^{n} (x_{i,k} - x_{j,k})^2} \quad (2)$$

where n is the number of features, and $x_{i,k}$ and $x_{j,k}$ are the values of the k-th feature for data points x_i and x_j, respectively.

2) Iterative merging of clusters

Following the computation of the distance matrix, the hierarchical clustering algorithm is applied. This study employs the agglomerative method, where clusters are iteratively merged based on their proximity. The Various linkage criteria are explored in this research to assess their impact on cluster formation. These criteria include:

- Ward's: Minimizes the variance within clusters, leading to compact and homogeneous groupings.
- Single Linkage: Considers the shortest distance between data points in different clusters, resulting in elongated and chain-like clusters.
- Complete Linkage: Uses the largest distance between data points in different clusters, producing compact and spherical clusters.
- Average Linkage: Computes the average distance between all pairs of points from two clusters, balancing the characteristics of single and complete linkage.

Among these criteria, Ward's method is selected as the primary linkage criterion for the following reasons. First, Ward's method minimizes the total within-cluster variance at each merging step, which tends to produce clusters of relatively equal size and high internal homogeneity. This characteristic is particularly suitable for the present study, where the objective is to classify sub-districts into distinct vulnerability levels (high, medium, and low) that are internally consistent and clearly separable. Second, unlike single linkage, which is prone to chaining effects, and complete linkage, which is sensitive to outliers, Ward's method provides a more balanced and stable clustering structure, especially when the features have been normalized. Third, the variance-minimization principle of Ward's method aligns well with the composite nature of the vulnerability indicators used in this study, which combine multiple normalized variables (e.g., sanitation, clean water access, population density, and healthy housing) into a single grouping framework. To ensure robustness, however, the other linkage criteria are still computed and compared, and their resulting dendrograms are evaluated alongside the Ward-based solution

3) Dendrogram Construction

Each linkage criterion provides a unique perspective on the dataset's structure, and their application helps evaluate the robustness of clustering results. To identify the optimal number of clusters, a dendrogram is generated for each linkage method. The dendrogram visually represents the hierarchical relationships between clusters, and a predefined threshold is used to determine the final cluster count. This threshold is chosen based on domain-specific knowledge and the overall goals of the analysis. In addition to the visual inspection of the dendrogram, the optimal number of clusters is validated using the silhouette score, which measures the cohesion within clusters and the separation between clusters. The combination of dendrogram inspection and silhouette analysis ensures that the selected cluster count is both statistically and contextually justified.

The clustering results are visualized using a dendrogram, which graphically represents the merging process and provides insights into the hierarchical relationships between clusters. By setting a predefined threshold or "cut-off" on the dendrogram, the optimal number of clusters is determined. This threshold is selected based on the dataset's characteristics and the specific requirements of the research.

4) Cluster Interpretation and Visualization

Once the clusters are identified, they are labeled according to their characteristics and further analyzed to extract meaningful patterns. Assign each sub-district to a cluster representing high, medium, and low vulnerability sub-districts. Visualize the clusters on a map to show spatial distribution.

IV. Result and Discussion

In this section, evaluation of the clustering results, as well as visualization of the user interface, is performed.

A. Spatial-Temporal Hepatitis-A Vulnerability

The results of the clustering process using agglomerative hierarchical clustering produced the

vulnerability levels of Hepatitis A for each sub-district in Pacitan Regency for the years 2017–2019, as shown in Table I. Table I presents a comparison of clustering outcomes based on four linkage criteria: Ward, single, average, and complete.

The table shows that some sub-districts, such as Tulakan, Pacitan, Bandar, and Donorojo, consistently appeared in the high vulnerability cluster in earlier years but shifted to lower vulnerability levels by 2019. Conversely, sub-districts like Ngadirojo and Sudimoro demonstrated an increase in vulnerability levels over time.

The comparison across different linkage methods reveals a high degree of consistency in cluster assignments, particularly under the Ward and Average linkage methods. This consistency suggests that the clustering results are robust and reliable for identifying spatial-temporal patterns of disease vulnerability across multiple years. These findings offer valuable insight into the dynamic distribution of Hepatitis A risk, providing a foundation for more targeted interventions and resource allocation by the local health authorities.

B. Hierarchical Clustering Evaluation

Three clustering evaluation metrics—Silhouette Score, Davies-Bouldin Index, and Calinski-Harabasz Index—were used to assess the quality of clustering for the dataset.

1) Silhouette Coefficient

The Silhouette Coefficient was used to assess the quality of clustering by evaluating both the cohesion and separation of clusters. The score ranges from -1 to 1, where a higher value indicates that the data points are well clustered, with a clear separation between different clusters. A score near 1 indicates that the data points are well matched to their own cluster and poorly matched to neighboring clusters, while a negative value suggests that the points may have been assigned to the wrong cluster. Table II presents the Silhouette Coefficient scores for the clustering results across the years 2017, 2018, and 2019 using different clustering methods.

The Silhouette Coefficient results indicate that the Ward method consistently achieved the highest clustering quality across all three years, with an average score of 0.34, suggesting relatively cohesive and moderately well-separated clusters. This suggests that the Ward method is particularly effective in creating distinct clusters. In contrast, the Single method showed consistently lower scores, especially in 2017 and 2019, reflecting poor separation and cohesion between clusters. The Average and Complete methods exhibited moderate performance, with scores generally falling between 0.4 and 0.6. While the Complete method saw a slight improvement in 2019, it still did not outperform the Ward method in terms of clustering quality. Overall, these observations highlight that the Ward method is the most reliable for achieving well-defined clusters, while the Single method should be avoided due to its poor performance. The Average and Complete methods showed more moderate results, with the Complete method showing some improvement over time, though not enough to surpass the Ward method.

TABLE I
HIERARCHICAL CLUSTER RESULT OF HEPATITIS-A VULNERABILITY

Sub District	2017				2018				2019			
	Ward	***Single***	***Average***	***Complete***	***Ward***	***Single***	***Average***	***Complete***	***Ward***	***Single***	***Average***	***Complete***
Arjosari	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Low	Low	Low	Low
Bandar	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium	**High**	**High**	**High**	**High**
Donorojo	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium	**High**	**High**	**High**	**High**
Kebonagung	Medium	Medium	Medium	Medium	Low	Medium	Low	Low	Medium	Medium	Medium	Medium
Nawangan	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium
Ngadirojo	Low	Low	Low	Low	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium
Pacitan	**High**	**High**	**High**	**High**	**High**	**High**	**High**	**High**	Low	Low	Low	Low
Pringkuku	Medium	Medium	Medium	Medium	Low	Medium	Low	Low	Medium	Medium	Medium	Medium
Punung	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium
Sudimoro	Low	Low	Low	Low	Low	Low	Low	Low	Medium	Medium	Medium	Medium
Tegalombo	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium
Tulakan	**High**	**High**	**High**	**High**	**High**	**High**	**High**	**High**	Low	Low	Low	Low

TABLE II
SILHOUETTE COEFFICIENT EVALUATION

Linkage Criteria	Silhoutte Coefficient Score		
	2017	*2018*	*2019*
Ward	**0.46**	0.24	0.33
Single	0.30	**0.27**	0.32
Average	0.45	**0.27**	0.24
Complete	0.45	0.21	**0.29**

2) Davies-Bouldin Index

The Davies-Bouldin Index (DBI) is a metric used to evaluate the clustering quality by calculating the ratio of within-cluster scatter to between-cluster separation. A lower DBI value indicates better clustering, as it implies that the clusters are compact and well-separated from each other. Conversely, higher values suggest poor separation and compactness, meaning the clusters are less distinct and more dispersed. Table III presents the Davies-Bouldin Index scores for the clustering results across the years 2017, 2018, and 2019 using different clustering methods.

TABLE III
DAVIES-BOULDIN INDEX EVALUATION

Linkage Criteria	Davies-Bouldin Index Score		
	2017	*2018*	*2019*
Ward	0.95	1.24	0.86
Single	**0.45**	**0.79**	0.70
Average	0.53	**0.79**	**0.51**
Complete	0.53	1.29	0.97

The Davies–Bouldin Index (DBI) results indicate that the Average linkage method achieved the best overall clustering performance across the three years, with the lowest mean DBI score of 0.61. The Average method also achieved the lowest DBI score in 2019 (0.51) and tied with the Single method in 2018 (0.79). In 2017, the Single linkage method obtained the lowest DBI score (0.45). In contrast, the Ward method yielded the highest mean DBI score of 1.02, indicating relatively less compact and less well-separated clusters compared with the other linkage methods. The Complete method produced an intermediate overall performance, with a mean DBI score of 0.93. Overall, the results suggest that the Average linkage method provided the most favorable clustering quality based on the DBI criterion.

3) Calinski-Harabasz Index

The Calinski-Harabasz Index (CHI) evaluates clustering quality by measuring the ratio of the sum of between-cluster dispersion to within-cluster dispersion. A higher CHI value indicates better clustering, as it suggests that the clusters are well-separated and distinct. In contrast, a lower CHI value implies less clear separation between the clusters, suggesting that the clustering is not as effective. Table IV presents the Calinski-Harabasz Index scores for the clustering results across the years 2017, 2018, and 2019 using different clustering methods.

TABLE IV
CALINSKI-HARABASZ INDEX EVALUATION

Linkage Criteria	Calinski-Harabasz Index Score		
	2017	*2018*	*2019*
Ward	**7.61**	3.99	**4.68**
Single	3.31	3.69	4.01
Average	7.06	3.69	2.58
Complete	7.06	**4.20**	4.35

The Calinski-Harabasz Index results indicate that the Ward method achieved the highest CHI scores across all three years, demonstrating the best clustering performance in terms of separation and compactness (with scores of 7.61 in 2017, 3.99 in 2018, and 4.681 in 2019). The Single method had the lowest CHI scores, suggesting weaker clustering quality due to poor separation between clusters (with scores of 3.31 in 2017, 3.69 in 2018, and 4.01 in 2019). The Complete and Average methods performed similarly, with the Complete method slightly outperforming the Average method in 2018 and 2019 (scores of 4.20 in 2018 vs 3.69, and 4.35 in 2019 vs 2.58). Nevertheless, none of the methods surpassed the Ward method, which consistently delivered the best results. Based on these findings, the Ward method is the most effective for generating well-separated and compact clusters, while the Single method should be avoided due to its inferior clustering results.

Overall, the hierarchical clustering evaluation reveals that the clustering performance varied across linkage criteria and evaluation metrics. The Ward method consistently achieved the highest Silhouette Coefficient and Calinski–Harabasz Index scores across the three years, indicating relatively stronger cluster cohesion and separation according to these two criteria. However, the Average linkage method obtained the lowest mean Davies–Bouldin Index score (0.61), indicating more favorable cluster compactness and separation according to the DBI criterion. The Single method generally exhibited lower Silhouette and Calinski–Harabasz scores, although it achieved the lowest DBI value in 2017 and was tied with the Average method in 2018. These findings indicate that no single linkage criterion consistently outperformed the others across all evaluation metrics. Overall, the Ward and Average methods demonstrated complementary strengths, with Ward showing stronger performance in terms of cluster separation and cohesion based on the Silhouette and Calinski–Harabasz indices, while Average provided more favorable results according to the Davies–Bouldin Index.

C. User Interface

The system provides two user-facing interfaces: a web-based dashboard for administrators and a mobile

application for residents of Pacitan Regency. Both interfaces are built upon a three-tier architecture consisting of a data layer, a processing layer, and a presentation layer. The data layer manages the preprocessed dataset and spatial boundary files, the processing layer executes the analytical pipeline (correlation analysis, agglomerative hierarchical clustering with Ward's linkage, and cluster labeling), and the presentation layer delivers the visualization and interaction components to end users.

The web-based interface is designed for administrators to manage the dataset, perform re-clustering, and inspect the clustering results. Its main functions include: (i) uploading and editing datasets, (ii) triggering the re-clustering process, (iii) visualizing clustering results on an interactive map, and (iv) exporting the resulting vulnerability classification. As shown in Figure 3, administrators can create and edit datasets through a structured form, while Figure 4 presents the dashboard that displays clustering outcomes and vulnerability levels. Figure 5 further illustrates the spatial visualization of clustering results, highlighting areas with different vulnerability levels, and Figure 6 presents a graph showing clustering results for each district across all years simultaneously, providing a comprehensive temporal view of vulnerability.

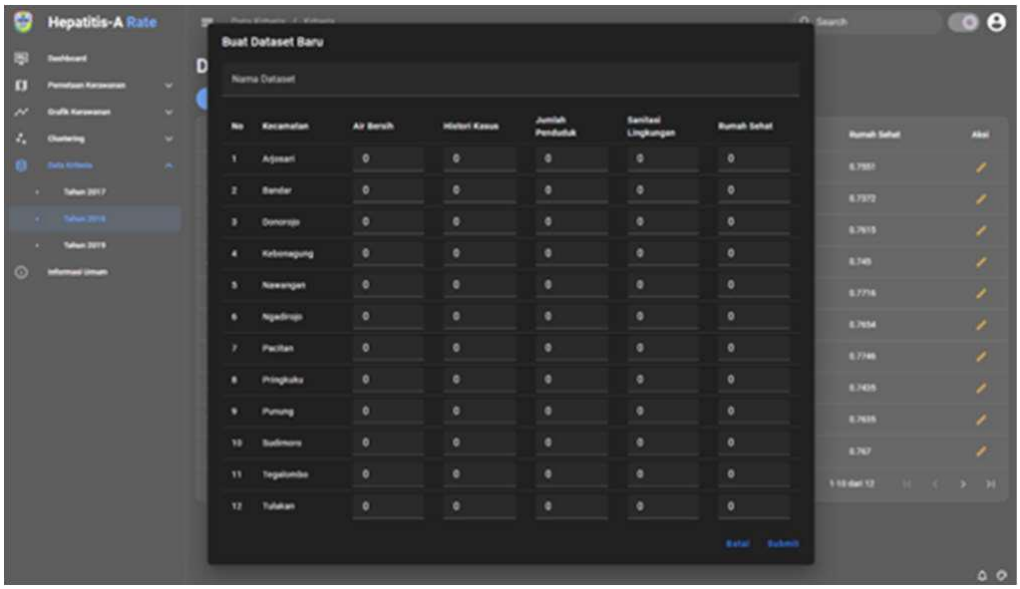


Fig. 3. Web interface for creating and editing datasets

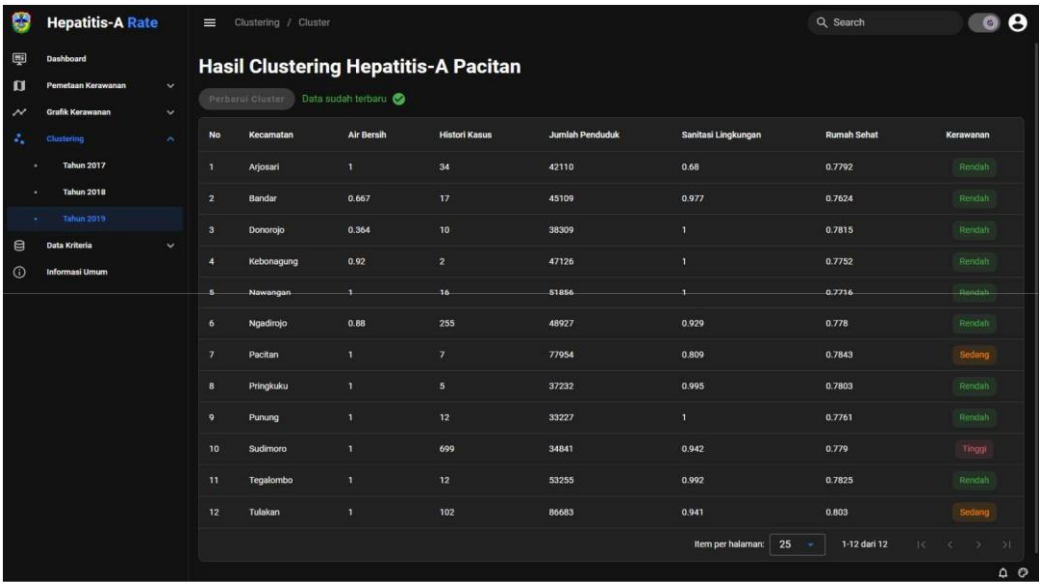


Fig. 4. Web dashboard displaying clustering results and vulnerability levels

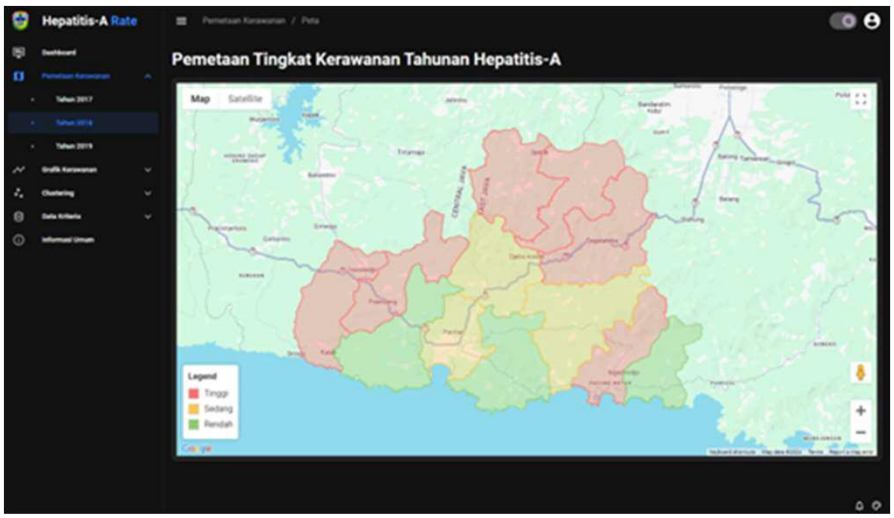


Fig. 5. Web interface showing the clustering results with vulnerability levels in each area

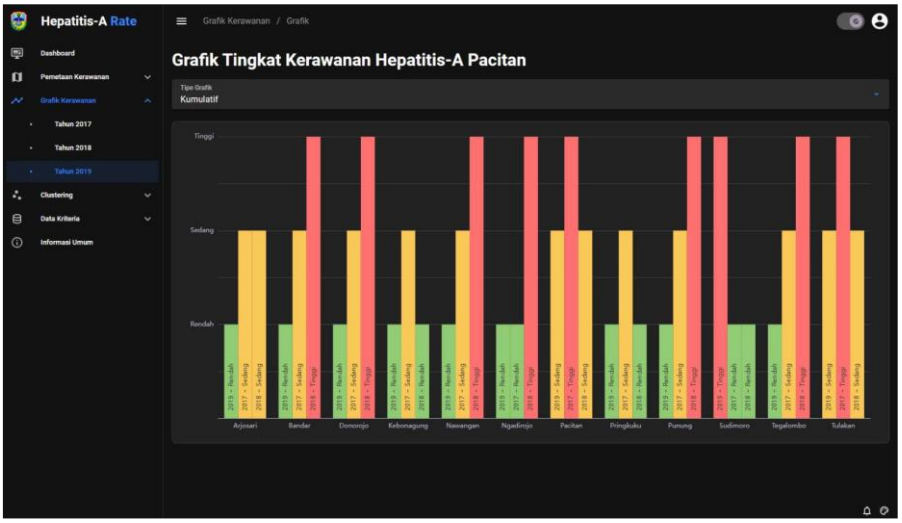


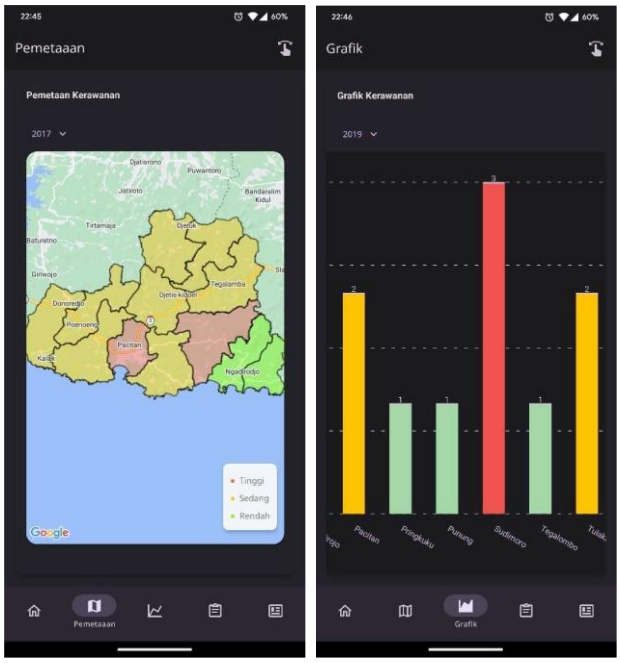
Fig. 6. Graph displaying the clustering results for each district across all years

The mobile application targets residents of Pacitan Regency and is designed to communicate vulnerability status in an accessible format. Its main functions include: (i) displaying the vulnerability level of the user's area based on the latest clustering results, (ii) providing temporal graphs of clustering outcomes per district per year, and (iii) delivering real-time updates through a simplified, non-technical interface. Figure 7(a) shows the mobile interface displaying area vulnerability data, while Figure 7(b) presents the per-year clustering graph for each district, allowing residents to track changes in vulnerability over time.

V. Conclusion

This study developed a spatial-temporal framework for assessing Hepatitis A vulnerability across 12 sub-districts in Pacitan Regency using agglomerative hierarchical clustering over three consecutive years (2017–2019). The results demonstrate that vulnerability patterns varied across sub-districts and years, indicating the dynamic nature of Hepatitis A vulnerability and the importance of continuous spatial-temporal surveillance. Four linkage criteria—Ward, Single, Average, and Complete—were evaluated using the Silhouette Coefficient, Davies–Bouldin Index, and Calinski–Harabasz Index. The results indicate that Ward generally provided stronger cluster cohesion and separation based on the Silhouette and Calinski–Harabasz indices, whereas Average achieved a more favorable mean Davies–Bouldin Index of 0.61. These findings demonstrate that clustering performance depends on the evaluation criterion and that no single linkage method consistently outperformed the others across all metrics. The study further developed a web-based administrative dashboard and mobile application interface to provide accessible visualizations of Hepatitis A vulnerability for health administrators and the public. These digital tools can facilitate the interpretation and dissemination of spatial-temporal vulnerability information to support epidemiological surveillance and public awareness. Future research should incorporate additional real-time and contextual risk factors, investigate predictive models for vulnerability

forecasting, and evaluate the framework across other regions to assess its generalizability and applicability.



(a) (b)

Fig. 7. (a) Mobile app interface displaying area vulnerability data; (b) Mobile app graph showing the clustering results for each district per year

Acknowledgements

This work was supported by Politeknik Elektronika Negeri Surabaya.

References

[1] M. Migueres, S. Lhomme, J. Izopet, "Hepatitis A: epidemiology, high-risk groups, prevention and research on antiviral treatment," Viruses, vol. 13 no. 10, pp. 1900, 2021.

[2] Indonesian Ministry of Health, "Profil Kesehatan Indonesia Tahun 2022," Jakarta, Indonesia: Kemenkes RI, 2023. [Online]. Available: https://pusdatin.kemkes.go.id

[3] S. Retyono, S. S. Rahardjo, B. Murti, "Multilevel analysis: Village does not have contextual effect on hepatitis A outbreak in Pacitan, East Java, Indonesia," Journal of Epidemiology and Public Health, vol. 5 no. 2, pp. 195-206, 2020.

[4] R. Lukmitarani, M. Ardiyah, R. Azizah, "The outbreak of hepatitis A: Evidence in Pacitan District," Journal Kesehatan Lingkungan, vol. 13 no. 1, pp. 35, 2021.

[5] J. Rosenthal, R. E. Arku, J. Baumgartner, J. Brown, T. Clasen, J. N. Eisenberg, P. Hovmand, P. Jagger, D.A. Luke, A Quinn, G.N. Yadama, "Systems science approaches for global environmental health research: enhancing intervention design and implementation for household air pollution (HAP) and water, sanitation, and hygiene (WASH) program,". Environmental Health Perspectives, vol. 128 no. 10, pp. 105001, 2020.

[6] E. Hamunyela, M. Hipondoka, F. Persendt, H. S. Nghiyalwa, C. Thomas, K. Matengu, "Spatio-temporal characterization of surface water dynamics with Landsat in endorheic Cuvelai-Etosha Basin (1990–2021)," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 191, pp. 68-84, 2022.

[7] Y. Wang, X. Chen, X., F. Xue, "A review of Bayesian spatiotemporal models in spatial epidemiology," ISPRS International Journal of Geo-Information, vol. 13 no. 3, pp. 97, 2024.

[8] T. J. Loftus, B. Shickel, J. A. Balch, P. J. Tighe, K. L. B. Abbott, Fazzone, E.M. Anderson, J. Rozowsky, T. Ozrazgat-Baslanti, Y. Ren, S.A. Berceli, "Phenotype clustering in health care: a narrative review for clinicians," Frontiers in artificial intelligence, vol. 5, pp. 842306, 2022.

[9] Q. An, S. Rahman, J. Zhou, J. J. Kang, "A comprehensive review on machine learning in healthcare industry: classification, restrictions, opportunities and challenges," Sensors, vol. 23 no. 9, pp. 4178, 2023.

[10] D. Kim, B. Cánovas-Segura, M. Campos, J. M. Juarez, "Visualization of Spatial–Temporal Epidemiological Data: A Scoping Review," Technologies, vol. 12 no. 3, pp. 31, 2024.

[11] T. Alasalı, Y. Ortakcı, "Clustering Techniques in Data Mining: A Survey of Methods, Challenges, and Applications," Computer Science, vol. 9 no. 1, pp. 32-50, 2024.

[12] Z. Khorrami, M. Rezapour, K. Etemad, S. Yarahmadi, S. Khodakarim, A. Mahdavi Hezaveh, M. Kameli, N. Khanjani, "The patterns of non-communicable disease multimorbidity in Iran: a multilevel analysis," Scientific Reports, vol. 10 no. 1, pp. 3034, 2020.

[13] X. Ran, Y. Xi, Y. Lu, X. Wang, Z. Lu, "Comprehensive survey on hierarchical clustering algorithms and the recent developments," Artificial Intelligence Review, vol. 56 no. 8, pp. 8219-8264, 2023.

[14] S. Modak, "Finding groups in data: an introduction to cluster analysis: authored by Leonard Kaufman and Peter J. Rousseeuw," John Wiley and Sons, 2005, ISBN: 0-47-1-73578-7, 2024.

[15] N. Randriamihamison, N. Vialaneix, P. Neuvial, "Applicability and interpretability of Ward's hierarchical agglomerative clustering with or without contiguity constraints," Journal of Classification, vol 38 no. 2, pp. 363-389, 2021.

[16] H. Al-Nefaie, A. Alsultan, R. Abusaris, "Temporal and spatial patterns of dengue geographical distribution in Jeddah, Saudi Arabia," Journal of Infection and Public Health, vol. 15 no. 9, pp. 1025-1035, 2022.

[17] L. M. Mageto, N. Mutono, G. Aboge, P. Gathura, E. Okunga, A. Muange, C. Mbae, S.M. Thumbi, S. Kariuki, "Spatial and temporal analysis of cholera cases in high-risk areas of Kenya. "medRxiv, pp. 2023-10, 2023.

[18] D. S. S. Rejeki, N. Nurhayati, B. Aji, "A spatiotemporal analysis of dengue hemorrhagic fever in Banyumas, Indonesia," Int J Public Health Sci, vol. 10 no. 2, pp. 231, 2021.

[19] M. Widawati, P. W. Dhewantara, R. Anasi, T. Wahono, R. Marina, I. P. Pertiwi, A.A. Wibowo, A. Ruliansyah, M.U. Riandi, D. Widiastuti, E.P. Astuti, "An investigation of geographical clusters of leptospirosis during the outbreak in Pangandaran, West Java, Indonesia," Geospatial health, vol. 18 no. 2, 2023.

[20] G. James, D. Witten, T. Hastie, R. Tibshirani, J. Taylor, "Statistical learning. In An introduction to statistical learning: With applications in Python (pp. 15-67)," Cham: Springer International Publishing, 2023.

[21] B. Oinam, V. Anand, R. K. Kajal, "A spatiotemporal geographic information system-based assessment of human immunodeficiency virus/acquired immune deficiency syndrome distribution in Manipur, India," Indian Journal of Public Health, vol. 65 no. 4, pp. 362-368, 2021..