Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6, 4518-4528
2024
Publisher: Learning Gate
DOI: 10.55214/25768484.v8i6.2984

Implementation of ward's agglomerative hierarchical clustering model to detect pulmonary tuberculosis endemic areas in Aceh Utara regency

Mutammimul Ula[1*], Mauliza[2], Ar Razi[3], Ilham Sahputra[4], Muhammad Abdullah Ali[5], Yumna Rilasmi Said[6]
[1,4,5,6]Department of Information Systems, Faculty of Engineering, Universitas Malikussaleh, Aceh, Indonesia.
mutammimul@unimal.ac.id (M.U.)
[2]Department of Medical Education, Faculty of Medicine, Universitas Malikussaleh, Aceh, Indonesia.
[3]Department of Informatics Engineering, Faculty of Engineering, Universitas Malikussaleh, Aceh, Indonesia.

Abstract: This study aims to detect pulmonary Tuberculosis (TB) endemic areas in Aceh Utara Regency based on altitude, population density, and the number of TB cases. The Agglomerative Hierarchical Clustering (AHC) algorithm was used to cluster 27 subdistricts, each measured by these three factors. The clustering results divided the subdistricts into three main clusters with distinct characteristics. Cluster 1 consists of subdistricts with low altitude, high population density, and relatively high numbers of TB cases, identifying this area as having the highest risk of TB endemicity. Cluster 2 includes areas with moderate population density and TB case numbers, while Cluster 3 consists of subdistricts at higher altitudes with fewer TB cases. The clustering results were evaluated using three key metrics: Silhouette Score, Davies-Bouldin Index, and Dunn Index, which indicated that the clustering model performed well, although some subdistricts were positioned near the cluster boundaries. This research provides valuable information for health authorities to prioritize interventions and allocate resources to areas most in need of TB management.
Keywords: *AHC, Cluster evaluation, Clustering, Pulmonary TB, Ward.*

1. Introduction

Pulmonary Tuberculosis (TB) is an infectious disease that remains a public health issue in Indonesia [1]. According to the World Health Organization (WHO), Indonesia is one of the countries with the highest TB burden in the world [2], with Aceh Utara being one of the regions with high prevalence [3]. This disease is transmitted through the air, and environmental and socioeconomic factors can contribute to its spread [4] [5]. In this context, factors such as altitude, population density, and the number of cases are crucial to consider when detecting and mapping pulmonary TB endemic areas [6] [7].

The size of an area is related to accessibility and the distribution of health resources, while altitude can influence disease spread patterns through changes in climate and habitat [8]. Population density is a critical factor that can increase the risk of TB transmission, as frequent interactions among individuals in densely populated areas can heighten the likelihood of infection [9] [10]. Therefore, understanding how these three factors interact can assist in detecting areas at high risk for pulmonary TB [11] [12].

Clustering is an effective technique for grouping data based on specific characteristics [13] [14]. One commonly used clustering method is the Agglomerative Hierarchical Clustering (AHC) algorithm [15]. This method can divide data into interrelated clusters, allowing for the identification of patterns that may not be evident in traditional analysis [16] [15]. This study aims to explore and analyze existing data using the AHC algorithm to identify clusters of pulmonary TB endemic areas in Aceh Utara Regency, [17], based on altitude, population density, and the number of cases [18]. The results of this study are expected to provide useful information for policymakers in designing more effective and targeted interventions to combat pulmonary TB in the region.

* Correspondence: mutammimul@unimal.ac.id

2. Literature Review

2.1. Pulmonary Tuberculosis and Its Impact in Indonesia

Pulmonary Tuberculosis (TB) is an infection caused by the Mycobacterium tuberculosis bacteria, which attacks the respiratory system and can lead to death if not properly treated [18]. TB remains one of the leading causes of morbidity and mortality worldwide, including in Indonesia, which ranks second after India in the number of TB cases, with approximately 845,000 new cases diagnosed in 2022 [2]. The high prevalence of pulmonary TB in Indonesia is influenced by various factors, including population density and access to healthcare services [19]. dditionally, environmental factors such as the size of the area and altitude, as well as demographic factors like population density, significantly affect the spread of pulmonary TB [20]. A deep understanding of these factors is crucial for identifying endemic TB areas and focusing public health interventions [2] [20].

2.2. Factors of Population Density, Area Size, and Altitude in the Spread of Pulmonary TB

Research shows that areas with high population density have a greater risk for the spread of pulmonary TB [21], as this condition allows for rapid disease transmission, especially in urban areas that often face space limitations and lower access to healthcare services [22]. Increased population density is directly related to a higher prevalence of pulmonary TB, particularly in areas that receive less attention regarding health interventions [23] [24]. This factor is crucial in the spread of TB because densely populated areas typically have more social interactions, which increases the likelihood of transmission [20]. Studies in various countries have also shown that increased population density is directly related to a rise in TB cases [20]. Additionally, the size of an area is often associated with the spread of diseases, including TB. Some studies have found that larger areas carry a higher risk of disease transmission, especially in regions with limited access to healthcare services [25] [26]. Altitude can also influence TB prevalence, where lower-altitude regions tend to have conditions more conducive to TB transmission due to higher temperatures and humidity [27].

2.3. Clustering Algorithms in Epidemiological Analysis

The Agglomerative Hierarchical Clustering (AHC) algorithm is a method frequently used to group data based on similarities or distances between data points [27]. This method is effective in identifying patterns and structures within complex data [28]. The advantage of AHC lies in its ability to produce hierarchical representations that provide a deeper understanding of the cluster structures formed [15]. In the context of pulmonary TB, AHC can assist in identifying areas at high risk based on demographic and epidemiological factors. Several previous studies have applied AHC to analyze health data, including mapping the spread of pulmonary TB in various regions [29]. In the field of epidemiology, clustering can be used to identify disease spread patterns, including pulmonary TB [30]. In this research context, AHC can be utilized to cluster districts in North Aceh based on altitude, population density, and the number of cases, thus facilitating the identification of TB endemic areas [31].

2.4. Application of AHC in Pulmonary TB Health Research

The application of clustering methods in the health sector has proven successful in identifying areas that require further intervention [30]. A study shows that clustering analysis can aid in planning more efficient disease control strategies [28]. Therefore, this research aims to apply AHC to detect endemic areas of pulmonary TB in North Aceh Regency, considering factors such as altitude, population density, and the number of cases.

The application of AHC in pulmonary TB health research shows promising results [30]. For instance, in a study, AHC was used to identify clusters of areas with high prevalence of respiratory diseases, including pulmonary TB [28]. This research highlights the importance of using demographic and health data to better understand the spread of TB and to focus health interventions in at-risk areas.

In Indonesia, the application of clustering algorithms in the context of pulmonary TB is still limited. However, by utilizing existing data, mapping endemic areas of pulmonary TB can enhance the effectiveness of health programs [32], especially in terms of resource allocation and the development of more targeted prevention strategies [33]. Therefore, this research aims to apply AHC to detect endemic areas of pulmonary TB in North Aceh Regency, considering population density, altitude, and the number of cases as the main variables.

2.5. Comparison with Other Clustering Algorithms

In addition to AHC, various other clustering algorithms such as K-Means and DBSCAN have been applied for mapping endemic areas of infectious diseases [34]. KK-Means groups data based on a predetermined number of clusters and is effective for data with a clear distribution, while DBSCAN can identify irregularly shaped clusters and handle outliers more effectively [35]. However, AHC has the advantage of hierarchical analysis, which can identify complex relationships between areas that may not be detected by other algorithms. Therefore, AHC was chosen in this study to map TB endemic areas based on population density and mortality rates, with the hope of providing deeper insights for public health interventions.

3. Methods

3.1. Research Design

This study employs a quantitative research design with a descriptive approach. The aim is to analyze and identify endemic areas of pulmonary TB in Aceh Utara Regency based on the factors of altitude, population density, and the number of cases using the Agglomerative Hierarchical Clustering (AHC) algorithm. Figure 1 illustrates the stages of the research process.

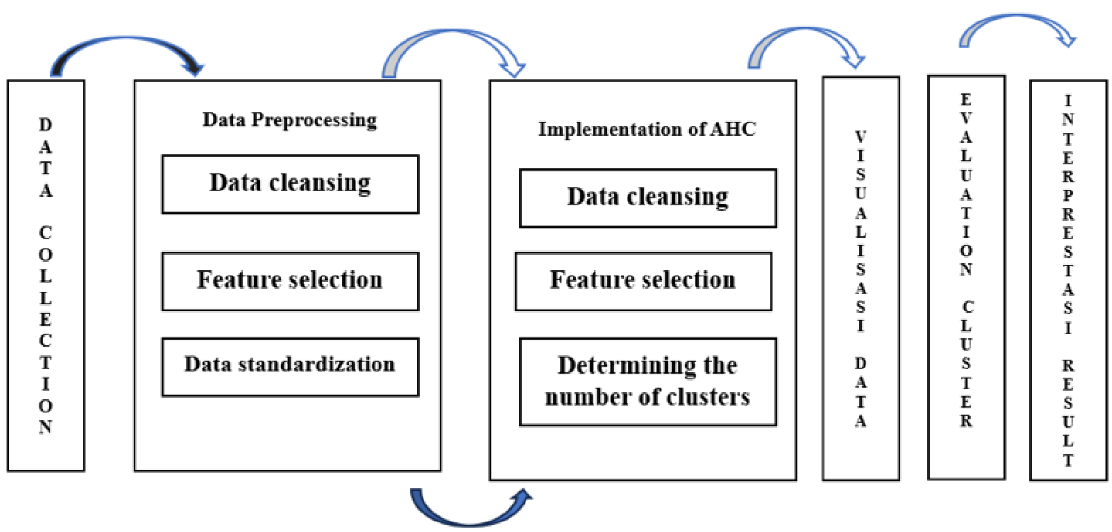

Figure 1.
Research stages.

3.2. Data and Data Sources

The data used in this study includes all districts in North Aceh Regency, totaling 27 districts. The data utilized in this research comprises:

1. Altitude (meters): Altitude data is obtained from publications issued by the North Aceh BPS (Central Bureau of Statistics).
2. Population Density (people/km²): This data is sourced from population census data published by the Aceh Utara BPS.
3. Number of Cases: Data is collected from Cut Meutia Hospital.

3.3. Data Analysis

Data analysis is conducted in several steps:

1. Data Preprocessing: The obtained data will be cleaned of missing values and standardized to ensure it is on the same scale [36]. The following is the formula used:

$$Z = \frac{X - \mu}{\sigma} \qquad (1)$$

Z is the standardized data value (z-score), X is the original data value, μ is the mean of all data values, and σ is the standard deviation of the data.

2. Application of the AHC Algorithm: AHC is applied to cluster the districts based on the four factors under study. The clustering process is carried out with the following steps:

a) Creating a Distance Matrix: Calculate the distance between districts using the Euclidean method. The formula is as follows:

$$d(p,q) = \sqrt{\sum_{i=1}^{n} (q_1 - p_1)^2} \qquad (2)$$

$d(p,q)$ is the Euclidean distance between two points p and q, p_i and q_i are the coordinates of points p and q in the i-th dimension,

b) Clustering Process: Using the agglomerative method to hierarchically group the data. Each district will be treated as an individual cluster and then merged based on their proximity in distance [33].
c) Determination of the Number of Clusters: The number of clusters will be determined using a dendrogram to visualize the clustering results and select the appropriate cutoff point.

3. The clustering results will be visualized using a dendrogram to facilitate understanding of the structure of the formed clusters.
4. Cluster Evaluation: The evaluation of the clustering results is conducted using the Silhouette Score, Davies-Bouldin Index, and Dunn Index metrics to assess the quality of the formed clusters (Rousseeuw, 1987). Below are some formulas used:

a) Silhouette Score

$$S(i) = \frac{b(i) - a(i)}{\max\,(a(i), b(i))} \qquad (3)$$

$S(i)$ is the Silhouette value for the i-th data point, $a(i)$ is the average distance from the i-th data point to all other data points in the same cluster (internal cluster cohesion), and $b(i)$ is the average distance from the i-th data point to all points in the nearest cluster that is not its own (distance to the nearest cluster).

b) Davies-Bouldin Index

$$DBI = \frac{1}{n} \sum_{i=1}^{n} max_{i \neq i} \frac{(s_i + s_j)}{d_{i,j}} \qquad (4)$$

n represents the total number of clusters, s_i a is the spread or average distance between the data points in cluster i and the centroid of cluster i (intra-cluster distance), and, $d_{i,j}$ is the distance between the centroids of cluster i and cluster j (inter-cluster distance). The term $max_{i \neq i} \frac{(s_i + s_j)}{d_{i,j}}$ calculates the worst-case ratio between clusters i and j, and this value is maximized across all cluster pairs to assess how well-separated the clusters are.

c) Dunn Index

$$D = \frac{min_{1 \le i < j \le k}\, d(C_i, C_j)}{max_{1 \le l \le k}\, \Delta(C_l)} \qquad (5)$$

$d(C_i, C_j)$ represents the minimum distance between two clusters C_i and C_j (the closest distance between points in these two clusters). $\Delta(C_l)$ is the spread (diameter) of cluster (C_l), and k which measures the maximum distance between any two points within the same cluster.

5. Interpretation of Results: Each cluster formed will be analyzed to identify which districts are classified as endemic areas for pulmonary TB. The assessment criteria include high population density and a significant number of deaths due to pulmonary TB..

4. Result and Discussion

4.1. Data Collection Results

The data used in this study covers 27 sub-districts across various regions. Each sub-district was measured based on three main features: Elevation (meters above sea level), Population Density (number of residents per km²), and the Number of TB Cases. These factors were selected as they can influence the spread of infectious diseases such as TB, where elevation affects environmental conditions, population density influences transmission potential, and the number of cases serves as an indicator of endemic levels. The data collection results show significant variation among the sub-districts, with elevation ranging from 5 meters to 362 meters, population density from 22 to 1,188 people/km², and TB cases ranging from 2 to 32 cases per sub-district. Table 1 presents the detailed data collected in this study.

4.2. Data Normalization

Before the clustering process is carried out, the data is normalized using the Standardization method to ensure that each feature has a uniform scale. This is important because each feature has different units (meters, people/km², and case numbers), and without normalization, features with a larger scale, such as elevation, could dominate the clustering process. Normalization is done by subtracting the mean from each feature and dividing it by the standard deviation, resulting in each feature having a mean of zero and a standard deviation of one. The normalization results allow the clustering algorithm to treat all features equally during the cluster formation process.

ISSN: 2576-8484
Vol. 8, No. 6: 4518-2528, 2024
DOI: 10.55214/25768484.v8i6.2984

Table 1.
Dataset of pulmonary TB cases in Aceh Utara regency.

NO	District	Elevation (Meters above sea level)	Population density (People/km²)	Number of cases	No	District	Elevation (Meters above sea level)	Population density (People/km²)	Number of cases
1	Baktiya	10	249	17	15	Meurah Mulia	25	110	9
2	Lhoksukon	7	212	27	16	Nibong	12	251	9
3	Samudera	19	655	19	17	Nisam	25	181	11
4	Seunuddon	9	263	21	18	Pirak Timu	11	134	12
5	Syamtalira Aron	8	708	17	19	Syamtalira Bayu	7	296	11
6	Tanah Jambo Aye	9	276	32	20	Tanah Pasir	6	497	6
7	Tanah Luas	13	861	25	21	Cot Girek	151	109	10
8	Dewantara	8	1188	9	22	Geuredong Pase	195	22	2
9	Muara Batu	7	857	10	23	Kuta Makmur	168	187	4
10	Baktiya Barat	8	239	8	24	Nisam Antara	362	166	4
11	Banda Baro	13	199	3	25	Paya Bakong	113	38	13
12	Langkahan	64	154	5	26	Sawang	198	104	6
13	Lapang	6	460	2	27	Simpang Kramat	165	138	3
14	Matang Kuli	5	345	5					

Table 2.
Normalized dataset.

No	District	Elevation (Meters above sea level)	Population density (People/km²)	Number of cases	No	District	Elevation (Meters above sea level)	Population Density (People/km²)	Number of cases
1	Baktiya	-0.5705	-0.28531	0.749533	15	Meurah Mulia	-0.39986	-0.77739	-0.2687
2	Lhoksukon	-0.60463	-0.4163	2.022325	16	Nibong	-0.54775	-0.27823	-0.2687
3	Samudera	-0.46812	1.151997	1.004092	17	Nisam	-0.39986	-0.52604	-0.01414
4	Seunuddon	-0.58188	-0.23575	1.25865	18	Pirak Timu	-0.55913	-0.69243	0.113137
5	Syamtalira Aron	-0.59326	1.339626	0.749533	19	Syamtalira Bayu	-0.60463	-0.11892	-0.01414
6	Tanah Jambo Aye	-0.58188	-0.18973	2.658721	20	Tanah Pasir	-0.61601	0.59265	-0.65054
7	Tanah Luas	-0.53638	1.881272	1.767767	21	Cot Girek	1.033566	-0.78093	-0.14142
8	Dewantara	-0.59326	3.038907	-0.2687	22	Geuredong Pase	1.534127	-1.08893	-1.15966
9	Muara Batu	-0.60463	1.867111	-0.14142	23	Kuta Makmur	1.226965	-0.5048	-0.9051
10	Baktiya Barat	-0.59326	-0.32071	-0.39598	24	Nisam Antara	3.433985	-0.57915	-0.9051
11	Banda Baro	-0.53638	-0.46232	-1.03238	25	Paya Bakong	0.601263	-1.03229	0.240416
12	Langkahan	0.04382	-0.62163	-0.77782	26	Sawang	1.568257	-0.79864	-0.65054
13	Lapang	-0.61601	0.461664	-1.15966	27	Simpang Kramat	1.192836	-0.67827	-1.03238
14	Matang Kuli	-0.62739	0.054545	-0.77782					

4.3. Results of Regional Clusters

After the data was normalized, the Agglomerative Hierarchical Clustering (AHC) algorithm was applied to detect endemic regions of tuberculosis (TB). This algorithm uses Ward's linkage method to minimize variance within clusters. The results of the clustering are shown in the following figure:

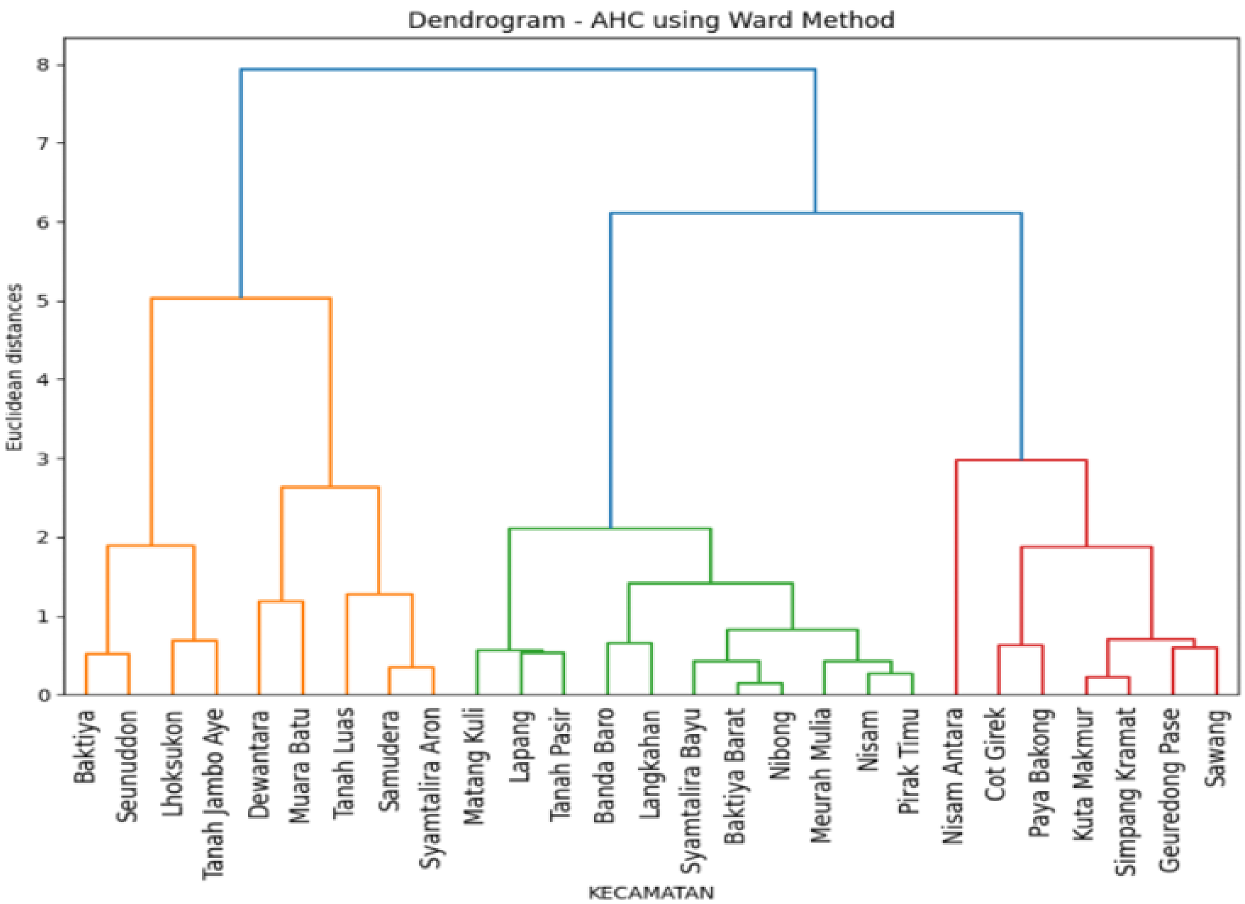

Figure 2.
Dendrogram of the clustering results for the regions.

Based on the clustering results, the sub-districts are divided into three main clusters, with the members of each cluster shown in Table 3.

Table 3.
Clustering results of regions.

Cluster	District
1	Baktiya, Lhoksukon, Samudera, Seunuddon, Syamtalira Aron, Tanah Jambo Aye, Tanah Luas, Dewantara, Muara Batu.
2	Baktiya Barat, Banda Baro, Langkahan, Lapang, Matang Kuli, Meurah Mulia, Nibong, Nisam, Pirak Timu, Syamtalira Bayu, Tanah Pasir.
3	Cot Girek, Geuredong Pase, Kuta Makmur, Nisam Antara, Paya Bakong, Sawang, Simpang Kramat

Each cluster has distinct characteristics, as described below:

DOI: 10.55214/25768484.v8i6.2984

1) Cluster 1: Areas with low elevation (5-19 meters), medium to high population density (249-1188 people/km²), and a relatively high number of TB cases (9-32 cases). This cluster is dominated by areas at high risk of transmission due to the dense population.
2) Cluster 2: Districts with low to moderate elevation (5-25 meters), lower population density (110-497 people/km²), and a varying number of TB cases, but generally lower compared to Cluster 1 (2-11 cases). These areas exhibit more moderate environmental and demographic conditions.
3) Cluster 3: Districts with higher elevation (113-362 meters), varying population density, and a relatively low number of TB cases (2-13 cases). Areas in this cluster tend to have more challenging environmental conditions, which may influence the rate of disease transmission.

4.4. Evaluation of Clusters

The quality evaluation of the clustering results is conducted using three main metrics: Silhouette Score, Davies-Bouldin Index (DBI), and Dunn Index.

1) Silhouette Score: The obtained value is 0.397, which indicates that most districts are clustered well; however, some districts are on the border between clusters. This value suggests a fairly good separation of clusters but not perfect, where some districts could potentially be grouped more optimally.
2) Davies-Bouldin Index (DBI): The DBI value is 0.879, indicating that the formed clusters are compact and separated from each other. A DBI value close to 1 shows fairly distinct clusters, with moderate intra-cluster variation.
3) Dunn Index: The generated Dunn Index is 0.439, which indicates that the distance between clusters is still quite good, although it could be improved to make the clusters more compact and further separated from each other.

Overall, the evaluation results indicate that the formed clusters are quite good, although there are still some districts that are on the borders between clusters and have characteristics similar to other clusters.

4.5. Interpretation of Results

Based on the results of clustering and evaluation, Cluster 1 can be considered as the area with the highest risk of TB endemicity, characterized by a high population density and a relatively larger number of cases. Areas in this cluster need more attention in efforts to prevent and manage TB due to the high potential for transmission. Cluster 2 represents districts with a moderate endemic risk, where the population density is lower than in Cluster 1, but the number of cases still requires monitoring. This area may have environmental and social factors that influence the disease's transmission rate. Meanwhile, Cluster 3 consists of areas with higher elevations and a lower number of TB cases. This suggests that physical environments, such as altitude and greater distances from population centers, may play a role in limiting disease spread. Efforts in this cluster could focus on prevention to avoid an increase in cases.

5. Conclusion

Based on the clustering results using the Agglomerative Hierarchical Clustering (AHC) algorithm, districts in North Aceh Regency are divided into three main clusters based on the features of elevation, population density, and the number of TB cases.

1. This finding provides guidance in identifying areas with different levels of TB endemic risk. Cluster 1 includes districts with high population density and a larger number of TB cases, indicating areas with a faster potential for disease spread that require prioritized health interventions. Cluster 2 consists of areas with moderate demographic and epidemiological characteristics, where the number of TB cases is relatively controlled, although there is still a potential for transmission that should be monitored. Cluster 3 includes areas with higher elevations and fewer cases, suggesting that geographical factors such as altitude may play a role in limiting the spread of TB.

2. The evaluation of clustering results with a Silhouette Score of 0.397, Davies-Bouldin Index of 0.879, and Dunn Index of 0.439 indicates that the clustering model has produced fairly good clusters, although there is room for improvement in separating some districts that are on the border between clusters.
3. Overall, this research provides important insights into detecting areas with a higher risk of TB endemicity, allowing prevention and management strategies to be focused on the areas most in need. This clustering approach can also help health authorities allocate resources more efficiently and effectively.

Copyright:

References

[1] S. D. Pralambang and S. Setiawan, "Faktor Risiko Kejadian Tuberkulosis di Indonesia," *J. Biostat. Kependudukan, dan Inform. Kesehat.*, vol. 2, no. 1, p. 60, Nov. 2021, doi: 10.51181/bikfokes.v2i1.4660.
[2] G. W. H. Organization., "World Health Organization," *Global Tuberculosis Report 2020*, 2020.
[3] M. Ula, A. Zulfikri, A. F. Ulva, and R. A. Rizal, "Penerapan Machine Learning Clustering K-Means dan Linear Regression Dalam Penentuan Tingkat Resiko Tuberkulosis Paru," *Indones. J. Comput. Sci.*, vol. 12, no. 1, Feb. 2023, doi: 10.33022/ijcs.v12i1.3162.
[4] E. A. Wikurendra, G. Nurika, Y. G. Tarigan, and A. A. Kurnianto, "Risk Factors of Pulmonary Tuberculosis and Countermeasures: A Literature Review," *Open Access Maced. J. Med. Sci.*, vol. 9, no. F, pp. 549–555, Nov. 2021, doi: 10.3889/oamjms.2021.7287.
[5] P. Narasimhan, J. Wood, C. R. MacIntyre, and D. Mathai, "Risk Factors for Tuberculosis," *Pulm. Med.*, vol. 2013, pp. 1–11, 2013, doi: 10.1155/2013/828939.
[6] Y. Ito *et al.*, "Risk factors for nosocomial tuberculosis transmission among health care workers," *Am. J. Infect. Control*, vol. 44, no. 5, pp. 596–598, May 2016, doi: 10.1016/j.ajic.2015.11.022.
[7] K.-S. Chen, T. Liu, R.-R. Lin, Y.-P. Peng, and G.-C. Xiong, "Tuberculosis transmission and risk factors in a Chinese antimony mining community," *Int. J. Tuberc. Lung Dis.*, vol. 20, no. 1, pp. 57–62, Jan. 2016, doi: 10.5588/ijtld.15.0215.
[8] X. Wu, Y. Lu, S. Zhou, L. Chen, and B. Xu, "Impact of climate change on human infectious diseases: Empirical evidence and human adaptation," *Environ. Int.*, vol. 86, pp. 14–23, Jan. 2016, doi: 10.1016/j.envint.2015.09.007.
[9] S. S. Triana Srisantyorini, Putri Nabilla, Dadang Herdiansyah, Dihartawan Dihartawan, Fini Fajrini, "Analisis Spasial Kejadian Tuberkulosis di Wilayah DKI Jakarta Tahun 2017-2019," *JKK J. Kedokt. dan Kesehat.*, vol. 18, no. 2, pp. 131–138, 2022, doi: https://doi.org/10.24853/jkk.18.2.131-138.
[10] H. NN, "Faktor Resiko Tuberkulosis pada penduduk urban di Kalianyar, Jakarta Barat," *J. Kedokt. dan Kesehat. Publ. Ilm. Fak. Kedokt. Univ. Sriwij.*, vol. 7, no. 1, pp. 1–8, Feb. 2020, doi: 10.32539/JKK.V7I1.8854.
[11] Y. Ji *et al.*, "Screening for pulmonary tuberculosis in high-risk groups of diabetic patients," *Int. J. Infect. Dis.*, vol. 93, pp. 84–89, Apr. 2020, doi: 10.1016/j.ijid.2020.01.019.
[12] R. McNerney *et al.*, "Tuberculosis Diagnostics and Biomarkers: Needs, Challenges, Recent Advances, and Opportunities," *J. Infect. Dis.*, vol. 205, no. suppl_2, pp. S147–S158, May 2012, doi: 10.1093/infdis/jir860.
[13] M. Mauliza, M. Ula, I. Sahputra, and R. Afdelina, "Fuzzy C-means Cluster Pattern Analysis and Ward Model Mapping in Viewing the Growth of Infectious and Non-Infectious Diseases Children in North Aceh," in *2023 International Workshop on Artificial Intelligence and Image Processing (IWAIIP)*, Dec. 2023, pp. 237–242. doi: 10.1109/IWAIIP58158.2023.10462743.
[14] W. Fuadi, I. Sahputra, D. Fariadi, and Agustinar, "Application of the K-Medoids Clustering Model and Borda Decision Group in Viewing Flood Area Groupings in North Aceh Regency," in *2023 International Workshop on Artificial Intelligence and Image Processing (IWAIIP)*, Dec. 2023, pp. 266–271. doi: 10.1109/IWAIIP58158.2023.10462874.
[15] M. Ula, Tsania Asha Fadilah Daulay, Richki Hardi, Sujacka Retno, Angga Pratama, and Ilham Sahputra, "Density Based Spatial Clustering of Applications and Spatial Pattern Analysis In Mapping the Distribution of ISPA Disease in Bireuen Regency," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 7, no. 3, pp. 733–742, Jun. 2023, doi: 10.29207/resti.v7i3.4936.
[16] M. Sadli, Fajriana, W. Fuadi, Ermatita, and I. Pahendra, "Machine Learning Approach for Electrical Load Forecasting Using Support Vector Regression," *J. Phys. Conf. Ser.*, vol. 1361, no. 1, p. 012065, Nov. 2019, doi: 10.1088/1742-6596/1361/1/012065.
[17] P. Garikapati, K. Balamurugan, T. P. Latchoumi, and R. Malkapuram, "A Cluster-Profile Comparative Study on Machining AlSi7/63% of SiC Hybrid Composite Using Agglomerative Hierarchical Clustering and K-Means," *Silicon*, vol. 13, no. 4, pp. 961–972, Apr. 2021, doi: 10.1007/s12633-020-00447-9.
[18] T. Li, A. Rezaeipanah, and E. M. Tag El Din, "An ensemble agglomerative hierarchical clustering algorithm based on

clusters clustering technique and the novel similarity measurement," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 6, pp. 3828–3842, Jun. 2022, doi: 10.1016/j.jksuci.2022.04.010.

[19] F. Fahdhienie, M. Mudatsir, T. F. Abidin, and N. Nurjannah, "Risk factors of pulmonary tuberculosis in Indonesia: A case-control study in a high disease prevalence region," *Narra J*, vol. 4, no. 2, p. e943, Aug. 2024, doi: 10.52225/narra.v4i2.943.

[20] K. E. Mokwena, "The Africa Region Misses the Deadline for the Millennium Development Goals: What Now?," *J. Trop. Dis.*, vol. 03, no. 04, 2015, doi: 10.4172/2329-891X.1000174.

[21] N. A. Mohidem, M. Osman, Z. Hashim, F. M. Muharam, S. Mohd Elias, and R. Shaharudin, "Association of sociodemographic and environmental factors with spatial distribution of tuberculosis cases in Gombak, Selangor, Malaysia," *PLoS One*, vol. 16, no. 6, p. e0252146, Jun. 2021, doi: 10.1371/journal.pone.0252146.

[22] P.-C. Lai, C.-T. Low, W.-S. C. Tse, C.-K. Tsui, H. Lee, and P.-K. Hui, "Risk of tuberculosis in high-rise and high density dwellings: An exploratory spatial analysis," *Environ. Pollut.*, vol. 183, pp. 40–45, Dec. 2013, doi: 10.1016/j.envpol.2012.11.025.

[23] S. Lu, Y. Wu, Z. Mao, and X. Liang, "Association of Formal and Informal Social Support With Health-Related Quality of Life Among Chinese Rural Elders," *Int. J. Environ. Res. Public Health*, vol. 17, no. 4, p. 1351, Feb. 2020, doi: 10.3390/ijerph17041351.

[24] M. Ula, R. Hardi, and I. Hipiny, "An Improved Structure for Academic Information Services through AI Chatbots," *J. Eng. Sci. Technol. Rev.*, vol. 16, no. 5, pp. 164–173, 2023, doi: 10.25103/jestr.165.20.

[25] M. Ula, A. Pratama, Y. Asbar, W. Fuadi, R. Fajri, and R. Hardi, "A New Model of The Student Attendance Monitoring System Using RFID Technology," *J. Phys. Conf. Ser.*, vol. 1807, no. 1, p. 012026, Apr. 2021, doi: 10.1088/1742-6596/1807/1/012026.

[26] W. Sun *et al.*, "A Spatial, Social and Environmental Study of Tuberculosis in China Using Statistical and GIS Technology," *Int. J. Environ. Res. Public Health*, vol. 12, no. 2, pp. 1425–1448, Jan. 2015, doi: 10.3390/ijerph120201425.

[27] C. Linhart *et al.*, "Continued increases in blood pressure over two decades in Samoa (1991–2013); around one-third of the increase explained by rising obesity levels," *BMC Public Health*, vol. 18, no. 1, p. 1122, Dec. 2018, doi: 10.1186/s12889-018-6016-2.

[28] Satriani, Anwar, and Suriah, "Spatial Pattern and Risking Factors of Tuberculosis Disease Occurrence in Barru Regency," in *Proceedings of the International Conference on Healthcare Service Management 2018*, Jun. 2018, pp. 141–148. doi: 10.1145/3242789.3242824.

[29] M. Ula, M. -, and I. Sahputra, "Optimization of Multilayer Perceptron Hyperparameter in Classifying Pneumonia Disease Through X-Ray Images with Speeded-Up Robust Features Extraction Method," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 10, 2022, doi: 10.14569/IJACSA.2022.0131025.

[30] J. F. Broekmans *et al.*, "European framework for tuberculosis control and elimination in countries with a low incidence," *Eur. Respir. J.*, vol. 19, no. 4, pp. 765–775, Apr. 2002, doi: 10.1183/09031936.02.00261402.

[31] "Comparison of clustering in tuberculosis using fuzzy c-means and k-means methods," *Commun. Math. Biol. Neurosci.*, 2022, doi: 10.28919/cmbn/7335.

[32] M. Tamba, Muhathir, Badriana, M. Ula, and I. Sahputra, "Classification of Autism Histogram of Oriented Gradient (HOG) Feature Extraction with Support Vector Machine (SVM) Method," in *2023 International Conference on Modeling & E-Information Research, Artificial Learning and Digital Applications (ICMERALDA)*, Nov. 2023, pp. 144–148. doi: 10.1109/ICMERALDA60125.2023.10458209.

[33] D. Abdullah, K. Jayaraman, and S. B. M. Kamal, "A Conceptual Model of Interactive Hotel Website: The Role of Perceived Website Interactivity and Customer Perceived Value Toward Website Revisit Intention," *Procedia Econ. Financ.*, vol. 37, pp. 170–175, 2016, doi: 10.1016/S2212-5671(16)30109-5.

[34] S. Sirsikar and K. Wankhede, "Comparison of Clustering Algorithms to Design New Clustering Approach," *Procedia Comput. Sci.*, vol. 49, pp. 147–154, 2015, doi: 10.1016/j.procs.2015.04.238.

[35] J. Wu, A. E. Hassan, and R. C. Holt, "Comparison of clustering algorithms in the context of software evolution," in *21st IEEE International Conference on Software Maintenance (ICSM'05)*, 2005, pp. 525–535. doi: 10.1109/ICSM.2005.31.

[36] A. Bouguettaya, Q. Yu, X. Liu, X. Zhou, and A. Song, "Efficient agglomerative hierarchical clustering," *Expert Syst. Appl.*, vol. 42, no. 5, pp. 2785–2797, Apr. 2015, doi: 10.1016/j.eswa.2014.09.054.