Cluster-Based Analysis For Growth Of Respiratory infectious diseases in children Using K-Means And Morran's I: Case Study Of North Aceh

Mauliza[1*], Mutammimul Ula[2], Veri ilhadi[2], Ilham Sahputra[2], Rosya Afdelina[1], Muhammad Ikhsan[1]

[1]Medical Education, Faculty of Medicine, Malikussaleh University, Aceh, Indonesia;
[2]Information Systems Program, Faculty of Engineering, Malikussaleh University, Aceh, Indonesia

*Corresponding author: mauliza@unimal.ac.id

KEYWORDS

Infectious diseases, Growth pattern analysis, k-means clustering

ABSTRACT

Respiratory infectious diseases in children are a major public health problem that requires in-depth understanding for effective treatment and prevention of the spread of these diseases. This research proposes a cluster-based analysis approach using the K-Means and Morran's I model to understand the growth of respiratory infectious diseases in children in the North Aceh district. This research aims to identify disease distribution patterns and group areas with similar risks to avoid spreading in adjacent areas. The data used includes cases of respiratory infections in children for the last three years, starting from 2021 to 2023. The K-Means clustering model is used to group geographic areas that have similar patterns in the distribution of cases of respiratory infections in children. In contrast, Morran's I is used to measuring the level of spatial dependence between cases of respiratory infections in children in each region. The results showed that K-Means and Morran's I were able to identify spatial patterns of respiratory infectious diseases in children well. There are 3 clusters formed from the results of data categorization. Cluster 1 consists of 18 sub-districts characterized by having a small number of cases for all types of disease. Cluster 2 consists of 3 sub-districts characterized by a high number of cases of pulmonary tuberculosis, bronchitis, ARI, and a moderate number of cases of glandular tuberculosis. The cluster consists of 6 sub-districts characterized by high cases of pulmonary TB, low cases of glandular TB, and moderate cases of bronchitis and ispa. The spatial pattern formed from the Kmeans clustering results is in line with the results of Moran's I analysis, namely: cluster 1 has a dispersed pattern, and clusters 2 and 3 have a clustered pattern. Based on Moran's I analysis, cluster 1 is a cluster with negative spatial dependence on disease growth with an index value of -0.01091, while clusters 2 and 3 are clusters with positive spatial dependence on the growth and spread of respiratory infections with an index value of 0.675138 for each cluster 2 and 0.303281 for cluster 3.

Introduction

Infectious diseases and non-communicable diseases are two types of diseases that often cause health problems for people, especially children. Specifically, respiratory infections in children are detected from respiratory tract infections caused by syncytial virus air pollution, which has an impact on children [1] [2]. Respiratory infections in children require serious treatment for those affected, and treatment of the disease can involve hospitalization for the healing process [3]. According to a report published in 2019 by the World Health Organization (WHO), infectious respiratory infections in children identify previous events that are detrimental to public health in every region [4] [5]. Meanwhile, non-communicable diseases cannot be transmitted through air or human interaction [6].

Cluster analysis of the growth of infectious diseases in children often occurs in pulmonary TB infections, bronchitis, and tuberculosis to see areas affected by respiratory infections in children [7] [8]. This disease is ranked in the top 10 causes of death worldwide. This has a broad impact on the problem of social disadvantage in each region. In addition, many infectious diseases cause death [9] [10].

Geographic factors are important in determining growth patterns and the risk of infectious diseases in an area. Geographic factors include aspects such as climate, altitude, humidity, rainfall, and population in the incidence of infectious diseases [11]. Spatial analysis techniques can understand and solve problems related to the growth of disease spread, planning of areas of occurrence and distribution patterns between regions [12]. Based on the results of the analysis, the distribution pattern of dangerous infectious diseases spreads from children to the elderly, caused by viruses and bacteria that attack the lungs [13].

Local governments and national health services in various countries can see the growth patterns in the incidence of infectious diseases in the near future and take action to reduce the risk of infectious diseases through regulation. Effective prevention of infectious diseases is to strengthen the body's immune system [14]. The government and the Health Service collaborate in preventing infectious disease problems. Such as vaccination, socializing about the dangers of typhoid, conducting outreach, patient quarantine, and analyzing previous incidences of infectious diseases in each area [15] [16]. Therefore, various cluster methods with cluster-based spatial analysis approach with machine learning analysis models [17][18]. he research area in North Aceh Regency is one of the areas that will be the object of research to look at the growth of infectious diseases in children's respiratory tract [19].

The research analysis aims to look at the incidence of infectious diseases between regions to see the relationship between infectious diseases and various environmental factors. Includes regional geographic factors and weather condition factors in various fields. Next, the study model looks at regional growth using the k-means clustering model and morans spatial analysis, which can show distribution patterns originating from the Local Indicator of Spatial Association (LISA) [20]. The results of cluster analysis can be a reference in making important decisions in cluster-based analysis of the growth of infectious diseases. They can be seen based on clustering pattern analysis for each region where respiratory diseases spread [21] [22]. To improve performance analysis of clustering results in viewing the growth of each region [23][24]. Spatial Analysis can measure and identify spatial patterns of cancer in an area. spatial data analysis in viewing deaths due to infectious disease sufferers [25].

This data must have information about geographic location based on coordinates or administrative area codes. Global indices, such as Moran's I, measure the extent to which variable values across regions show spatial interconnectedness. The results of the Spatial Autocorrelation analysis can determine the value and range of indexes in a variable where high values tend to be close to low values and vice versa [26][27].

This research explains the analysis of infectious diseases and the impact of infectious disease events in a region. Analysis of infectious diseases can be seen in each region with cluster results. The cluster results for an area can be used to see the growth pattern of infectious diseases in children and to see the distribution pattern of these infectious diseases [28] [29]. Clustering analysis can be used in various segmentation of growth patterns, distribution, anomaly detection, and event analysis to detect the level of danger and growth of disease in each region[30] [31]. Understanding cluster patterns in collecting data on the incidence of infectious diseases in the community in each region and relating to patient history in detecting distribution patterns and understanding growth patterns of infectious diseases[32].

Analysis of the seasonal growth of infectious diseases for each region affected by respiratory infectious diseases [33]. There is a correlation between the results regarding the interpretation of disease incidence data, such as the relationship between patterns of how growth is transmitted and how it is transmitted. This influences the incidence of infectious diseases in an area [34]. It is difficult to analyze the incidence of infectious diseases using various factors in the spread of infectious diseases, such as environmental, climate and geographic factors. Even if seen from regional cluster analysis, regional condition factors can influence the cluster growth of infectious diseases in a region and sub-district [35]. There is a clustering model process by calculating the distance between each data and the center of disease growth. This cluster model analysis can see the distribution obtained from the closeness of the growth of infectious disease incidence between regions that are related to each other [36] [37]. One solution for analyzing events uses a cluster-based spatial analysis approach. The cluster results for each region can show the grouping of events that occur in each region in the growth of an infectious disease [38].

By analyzing clusters in each region, this research found factors that are interrelated with the incidence between clusters and thus influence the growth pattern of infectious diseases [39]. Analysis of infectious disease incidence can be used on various incident factors in the fields of health, medical education, and biology. This approach has been applied in various fields, including business, education, and biology [17] [40][41].

Analysis of the growth of the spread of infectious diseases in several studies was carried out to collect data regarding the risk of disease spread through the level of interaction between patients affected by infectious and non-infectious diseases. The model used in clustering analysis can explore whether an infectious disease is contagious by looking at the transmission of infectious diseases with human interactions detected in each region [42]. Subsequent research presents a model for identifying potential outbreaks of infectious diseases based on groupings seen by genetic factors and evaluates models that are interrelated in these diseases that can be tested in actual practice. There is a cluster base in disease analysis based on sensors for various diseases [30]. Subsequent research applied cluster analysis to variables seen from the analysis of the incidence of infectious patients, predicting the high rate of spread of respiratory infections [33].Infection characteristics to identify the spread of disease through bloodstream infections and the main factors. Visual growth analysis can be seen effectively from time, region and infectious disease to the incidence of infectious disease in clusters for each region. [43] [44][45].

Methods

Data Collecting and Preprocessing

Data collection on pediatric infectious disease cases was conducted at Cut Mutia Hospital, North Aceh District. The data collected were all epidemic data related to infectious diseases in children that occurred in the last 3 years, starting from 2021 to 2023. The data obtained from the hospital is then cleaned and prepared according to the research needs. The data cleaning process is carried out with Python Programming using the Jupyter Notebook IDE. The important variables used in this study are the type of disease, number of cases, sub-district and time.

Clusterization

Grouping data on respiratory infections in children in the North Aceh district using the K-Means method based on grouping geographic areas based on their characteristics. The variables used are type of disease, number of cases, sub-district, and time. The K-Means algorithm functions to partition data into several clusters with similar distribution patterns of disease cases. This process includes data normalization, calculating the distance between each data point based on the attributes used as clustering parameters, determining the number of clusters, and grouping the data. Several stages are shown in Figure 1.

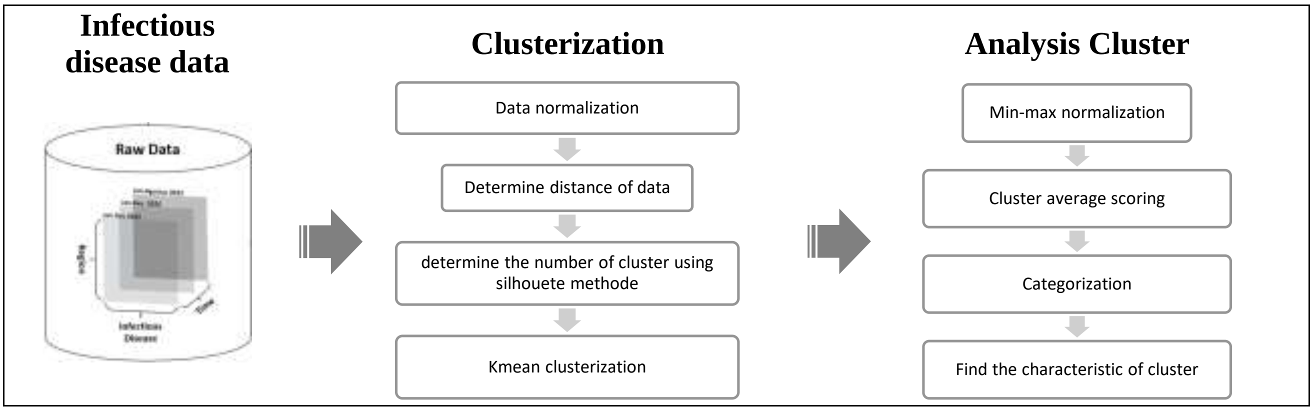


Figure 1. Clusterization stage

Each cluster that has been formed is then analyzed to obtain the characteristics of each cluster. This process involves 4 stages including: (1) min-max normalization to convert case data into a range of 1-100, (2) cluster average scoring to determine the average value of features in each cluster, (3) categorization to determine the category of case severity in a particular sub-district, and (4) find the characteristic of cluster to determine the inherent characteristics of the cluster that has been formed.

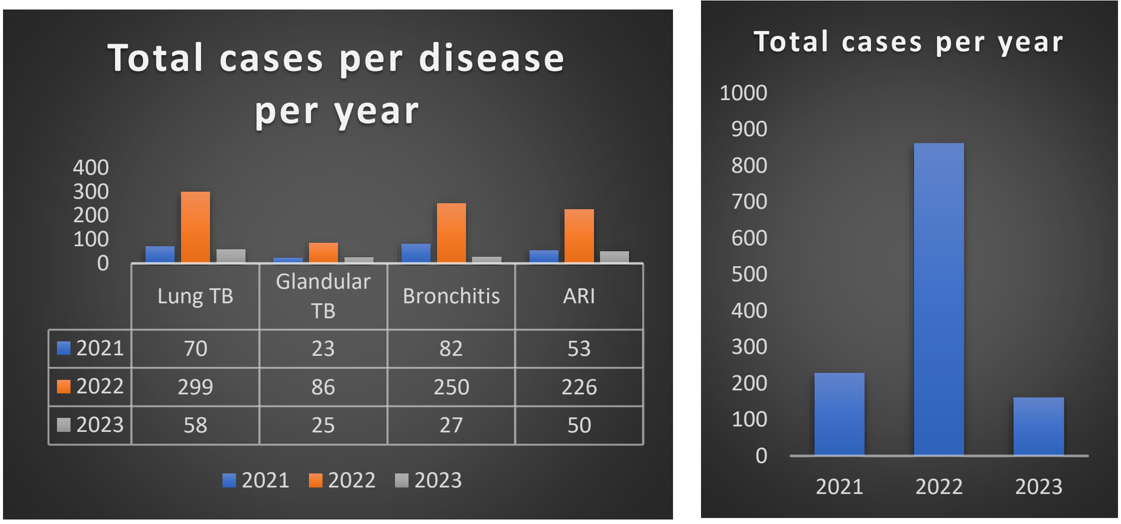
Spasial Analysis

Spatial analysis was conducted using Moran's I method. This method is applied to each cluster formed with Kmean at the clustering stage. This stage produces an index value that explains the spatial dependence of disease spread/growth in each cluster.

Result and Discussion

Infectious disease data During 2021 - 2023

In this study, based on the medical records of the Cut Meutia Hospital in North Aceh, 1249 data on cases of respiratory infectious diseases in children that occurred during the period 2021 to 2023 were obtained. This amount of data is the amount of data that has been cleaned and ready to be used in research. This amount of data consists of 4 diseases including: Lung TB, Glandular TB, Bronchitis, and ARI. Figure 1 shows the number of cases of each disease that occur each year.



	Lung TB	Glandular TB	Bronchitis	ARI
2021	70	23	82	53
2022	299	86	250	226
2023	58	25	27	50

Figure 2: Total cases per year

The cases that occurred were spread across 27 sub-districts in North Aceh. The number of cases occurring in each sub-district varies each year for each type of disease. Details of the distribution of the number of cases per sub-district for all types of diseases for 3 years are shown in Table 1.

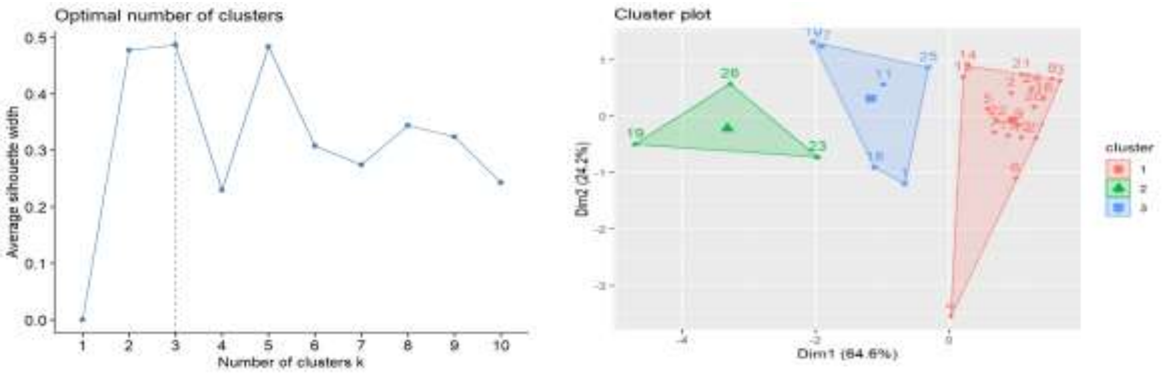
Table 1. Respiratory infectious disease data

No	sub-districts	2021				2022				2023				Total Cases			
		p1	p2	p3	p4	p1	p2	p3	p4	p1	p2	p3	p4	p1	p2	p3	p4
1	Baktiya	1	2	1	2	22	9	5	10	6	1	2	5	29	13	8	41
2	Baktiya Barat	0	0	2	1	9	2	4	2	0	0	0	1	9	2	6	12
3	Banda Baro	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
4	Cot Girek	2	4	0	1	6	19	6	8	1	1	0	2	9	23	6	18
5	Dewantara	5	0	4	0	6	4	5	2	1	0	3	1	12	7	12	14
6	Geureudong Pase	0	0	0	0	6	10	1	4	0	0	0	0	6	10	1	10
7	Kuta Makmur	4	0	1	1	11	5	0	2	2	1	0	2	17	5	1	20
8	Langkahan	0	0	0	0	1	0	2	1	0	0	0	0	1	0	2	2
9	Lapang	1	0	0	0	10	3	1	0	1	2	0	1	12	3	1	12
10	Lhoksukon	8	0	15	5	30	0	16	18	7	0	1	0	45	1	32	68
11	Matangkuli	1	1	4	2	25	1	21	11	1	1	0	4	27	2	25	40
12	Muara Batu	7	1	0	2	0	2	0	2	0	3	1	0	7	4	1	11
13	Meurah Mulia	5	1	9	9	0	0	11	0	0	0	1	3	5	2	21	14
14	Nibong	2	0	4	2	14	0	9	4	0	0	0	3	16	0	13	22
15	Nisam	0	4	1	1	6	2	3	3	0	0	1	2	6	7	5	10
16	Nisam Antara	0	0	0	0	3	0	2	1	0	2	0	0	3	0	2	4
17	Paya Bakong	4	0	4	5	31	0	21	23	7	0	0	1	42	0	25	70
18	Pirak Timur	0	5	1	0	18	6	13	17	17	0	0	1	35	11	14	52
19	Samudera	16	5	6	3	29	5	53	35	4	2	6	8	49	16	65	87
20	Sawang	0	0	0	0	4	0	3	1	1	3	0	0	5	0	3	6
21	Lhoksukon	8	0	15	5	30	0	16	18	7	0	1	0	45	1	32	68
22	Matangkuli	1	1	4	2	25	1	21	11	1	1	0	4	27	2	25	40
23	Muara Batu	7	1	0	2	0	2	0	2	0	3	1	0	7	4	1	11
24	Samudera	16	5	6	3	29	5	53	35	4	2	6	8	49	16	65	87
25	Tanah Luas	8	0	9	8	20	2	27	41	2	2	6	6	30	8	42	79
26	Tanah Pasir	0	0	2	0	1	2	2	0	0	4	1	0	1	3	5	1
Sub Total Kasus		70	23	82	53	299	86	250	226	58	25	27	50	427	136	359	706
Total		228				861				160				1628			

Information: P1 : TBC Paru, P2 : TBC Kelenjar , P3 = Bronkities, P4= ISPA

Clusterization

Data clustering is done based on 4 features, namely: f1 (total cases of pulmonary tuberculosis per district), f2 (total cases of glandular tuberculosis per district), f3 (total cases of bronchitis per district), and f4 (total cases of ARI per district). The recommended number of clusters based on the silhoute calculation is 3 clusters as shown in Figure 3(a) and the results of clustering areas using Kmeans are shown in Figure 3(b):



(a) (b)

Figure 3. silhouette and cluster plot

In Figure 3(b), the numbers contained in the cluster plot are sub-district numbers with an explanation that can be seen in Table 1. Based on the clustering results, cluster 1 consists of 18 sub-districts, Cluster 2 consists of 3 sub-districts and cluster 3 consists of 6 sub-districts. The following is a list of clusters and their sub-districts.

Table 2. Results of region clustering

Cluster	Sub-districts
1	Baktiya Barat, Banda Baro, Cot Girek, Dewantara, Geureudong Pase, Kuta Makmur, Langkahan, Lapang, Muara Batu, Meurah Mulia, Nibong, Nisam, Nisam Antara, Sawang, Seunudon, Simpang Keramat, Syamtalira Bayu, Tanah Pasir
2	Samudera, Syamtalira Aron, Tanah Luas
3	Baktiya, Tanah Jambo Aye, Lhoksukon, Matangkuli, Paya Bakong, Pirak Timur

Cluster Analysis

- Min-max normalization

The total case data of each disease in all sub-districts is normalized into values with a range of 1-100 using the Min-max normalization method. This is done to simplify the categorization process. The normalization results are shown in the following table:

Table 3. Min-max normalization

No	District	F1	F2	F3	F4	Cluster
1	Baktiya	18.36735	8.33333333	9.2307692	7.272727	3
2	Baktiya Barat	0	0	0	1.818182	1
3	Banda Baro	18.36735	100	9.2307692	20	1
4	Cot Girek	24.4898	16.6666667	18.461538	5.454545	1
5	Dewantara	12.2449	41.6666667	1.5384615	7.272727	1
6	Geureudong Pase	34.69388	25	1.5384615	9.090909	1
…	…	…	…	…		
…	…	…	…	…		
21	Lhoksukon	10.20408	4.16666667	32.307692	21.81818	3
22	Matangkuli	32.65306	0	20	16.36364	3
23	Muara Batu	12.2449	25	7.6923077	10.90909	1
24	Samudera	100	50	100	83.63636	2
25	Tanah Luas	71.42857	45.8333333	21.538462	32.72727	2

26	Paya Bakong	73.46939	4.16666667	16.923077	12.72727	1

- *Cluster average scoring*

Based on the data in Table 2, the average value for each feature in each cluster is calculated. The results obtained at this stage are shown in the following table:

Table 4. Cluster average scoring

Cluster	Average (f1)	Average (f2)	Average (f3)	Average (f4)
1	31.85941043	21.2962963	14.95726496	16.76767677
2	39.45578231	20.83333333	38.97435897	30.90909091
3	29.93197279	18.75	27.69230769	33.93939394

- *Categorization*

Feature categorization was done into 3 groups, namely: (i) Low with a mean feature value < 34, (ii) Medium with a mean feature value < 77, and (iii) High with a mean feature value >= 77. The categorization results based on these rules are shown in the following table:

Table 5. Categorization

Cluster	(f1) Lung TB	(f2) Glandular TB	(f3) Bronchitis	(f4) ARI
1	Low	Low	Low	Low
2	High	Medium	High	High
3	High	Low	Medium	Medium

- *Find the characteristics*

Based on the categorization results, the characteristics for each cluster are obtained as follows:

1. Cluster 1 is a cluster with a low number of cases for all types of diseases.
2. Cluster2 is a cluster with a high number of cases for Lung TB, Bronchitis, and ARI, while for the case of Glandular TB the number of cases is moderate.
3. Cluster 3 is a cluster with a high number of cases for Lung TB, a low number of cases for Glandular TB, and a medium number of cases for Bronchitis and ARI.

Spasial Analysis

Based on the results of regional clustering shown in table 2, a spatial map of disease distribution patterns is formed as follows:

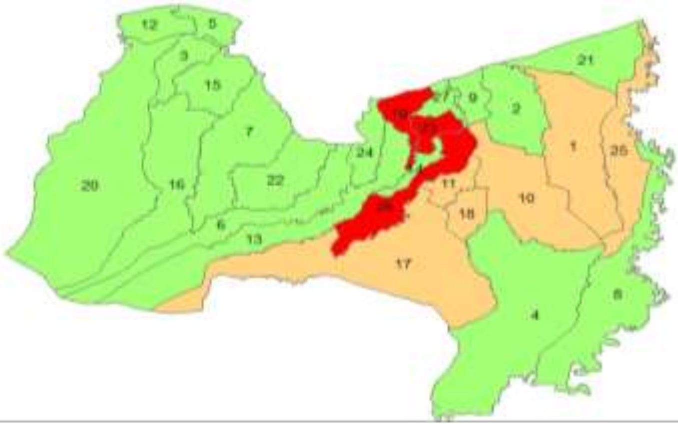


Figure 4. Spatial map of clustering results

The green color on the map is cluster 1, the red color on the map is cluster 2, and the orange color on the map is cluster 3. The number on the map is the sub-district number as described in table 1. Spatial analysis was conducted to see the spatial influence on disease spread. The analysis was conducted using the Moran's I method, the results of the analysis are shown in the following table.

Table 6. Analisis Moran's I

Cluster	District	Moran's I	Information
1	Baktiya Barat, Banda Baro, Cot Girek, Dewantara, Geureudong Pase, Kuta Makmur, Langkahan, Lapang, Muara Batu, Meurah Mulia, Nibong, Nisam, Nisam Antara, Sawang, Seunudon, Simpang Keramat, Syamtalira Bayu, Tanah Pasir	-0.01091	Ketergantungan Spasial Negative terhadap penyebaran penyakit Pola Dispered
2	Samudera, Syamtalira Aron, Tanah Luas	0.675138	Ketergantungan Spasial Positif terhadap penyebaran penyakit Pola Clustered Kuat
3	Baktiya, Tanah Jambo Aye, Lhoksukon, Matangkuli, Paya Bakong, Pirak Timur	0.303281	Ketergantungan Spasial Positif Pola Clustered Lemah

Discussion

In this study, we conducted a cluster-based analysis to investigate the growth/spread pattern of respiratory infectious diseases in children in North Aceh district. Our data shows that there are 4 types of respiratory infectious diseases in children that occur frequently and spread across 27 sub-districts, including: Lung TB, Glandular TB, Bronchities, and ARI. The number of cases varies for each year and type of disease.

Based on the data obtained, we conducted regional clustering based on the frequency of cases of the four diseases for three years. As a result, 3 regional clusters with distinctive characteristics were formed. Cluster 1 was formed by 18 subdistricts with a low number of cases for all four diseases. Cluster 2 was formed by 3 sub-districts with a high number of cases for Lung TB, Bronchitis, ARI, and a medium number of cases for Glandular TB. Cluster 3 is formed by 6 sub-districts with a medium number of cases for Bronchitis, ARI, a low number of cases for Glandular TB disease, and a high number of cases for Lung TB disease.

The spatial map formed from the clusterization results shows a neat pattern in each cluster formed. Clusters 2 and 3 form a clustering pattern, each sub-district that is a member of clusters 2 and 3 are sub-districts that border each other. while cluster 1 forms a disperse pattern.

The results of spatial analysis using Moran's I indicate that clusters 2 and 3 are clusters with positive spatial dependence, which means that the growth / spread of disease is influenced by spatial factors. While cluster 1 is a cluster with negative spatial dependence, which means that the growth/spread of the disease is not influenced by spatial factors.

If you look at the spatial pattern formed on the map shown in Figure 6, the sub-district area represented by numbers 2, 9, 21 and 27 is an area that is very potential for an increase in the number of infectious disease cases, because the 4 sub-districts are crushed by 2 clusters with a high number of cases of child infectious diseases, and based on the results of Moran's I calculation, the two clusters have a clustered pattern.

Conclusion

Based on the results of Kmeans and Moran's I analysis, the growth/distribution of respiratory infectious diseases in children in North Aceh is divided into 2 patterns, namely: Clustered pattern with positive spatial dependence and dispered with negative spatial dependence. The clustered pattern occurs in clusters 2 and 3, while the dispered pattern occurs in cluster 1. Based on the observation of the map of the results of regional clustering and Moran's I for clusters 2 and 3, it is concluded that sub-districts with number 2 (Baktiya Barat), number 9 (Lapang), number 21 (Seunudon), and number 27 (Tanah pasir) are sub-districts that have great potential for the spread of infectious diseases in children, because the existence of the four sub-districts is crushed by two clusters with positive spatial dependence on the spread of infectious diseases in children.

Ethics approval

This research was approved by the Research Ethics Committee of Malikussaleh University and the Medical Records Rumah Sakit Cut Meutia, Lhokseumawe, Indonesia.

Acknowledgments

The author would like to thank the Research and Community Service Institute of Malikussaleh University and the staff of the Medical Records Department of Cut Meutia Hospital Lhokseumawe, Indonesia for their assistance during the research

Competing interests

All the authors declare that there are no conflicts of interest.*”*

Underlying data

All data underlying the results is available as part of the article and no additional source data is required.

References

[1] X. Wang *et al.*, “Global burden of respiratory infections associated with seasonal influenza in children under 5 years in 2018: a systematic review and modelling study,” *Lancet Glob. Heal.*, vol. 8, no. 4, pp. e497–e510, Apr. 2020, doi: 10.1016/S2214-109X(19)30545-5.

[2] D. Mondal and P. Paul, “Effects of indoor pollution on acute respiratory infections among under-five children in India: Evidence from a nationally representative population-based study,” *PLoS One*, vol. 15, no. 8, p. e0237611, Aug. 2020, doi: 10.1371/journal.pone.0237611.

[3] Y. Li *et al.*, “National burden estimates of hospitalisations for acute lower respiratory infections due to respiratory syncytial virus in young children in 2019 among 58 countries: a modelling study,” *Lancet Respir. Med.*, vol. 9, no. 2, pp. 175–185, Feb. 2021, doi: 10.1016/S2213-2600(20)30322-2.

[4] S. Sugita *et al.*, “Use of a Comprehensive Polymerase Chain Reaction System for Diagnosis of Ocular Infectious Diseases,” *Ophthalmology*, vol. 120, no. 9, pp. 1761–1768, 2013, doi: 10.1016/j.ophtha.2013.02.020.

[5] W. H. Organization, “World Health Organization Global Health Estimates 2016: Deaths by Cause, Age, Sex, by Country and by Region,” *World Health Organization*. 2016.

[6] J. E. Prynn and H. Kuper, “Perspectives on Disability and Non-Communicable Diseases in Low- and Middle-Income Countries, with a Focus on Stroke and Dementia,” *Int. J. Environ. Res. Public Health*, vol. 16, no. 18, p. 3488, 2019, doi: 10.3390/ijerph16183488.

[7] L. A. F. Batista, K. J. S. Silva, L. M. da Costa e Silva, Y. F. de Moura, and F. C. R. Zucchi, “Tuberculosis: A granulomatous disease mediated by epigenetic factors,” *Tuberculosis*, vol. 123, p. 101943, Jul. 2020, doi: 10.1016/j.tube.2020.101943.

[8] P. Winskill, A. B. Hogan, J. Thwing, L. Mwandigha, P. G. T. Walker, and B. Lambert, “Health inequities and clustering of fever, acute respiratory infection, diarrhoea and wasting in children under five in low- and middle-income countries: a Demographic and Health Surveys analysis,” *BMC Med.*, vol. 19, no. 1, p. 144, Dec. 2021, doi: 10.1186/s12916-021-02018-0.

[9] B. Jang, M. Lee, and J. W. Kim, “PEACOCK: A Map-Based Multitype Infectious Disease Outbreak Information System,” *IEEE Access*, vol. 7, pp. 82956–82969, 2019, doi: 10.1109/ACCESS.2019.2924189.

[10] W. Qiu, C. Chu, A. Mao, and J. Wu, “The Impacts on Health, Society, and Economy of SARS and H7N9 Outbreaks in China: A Case Comparison Study,” *J. Environ. Public Health*, vol. 2018, pp. 1–7, 2018, doi: 10.1155/2018/2710185.

[11] M. Ahmadi, A. Sharifi, S. Dorosti, S. J. Ghoushchi, and N. Ghanbari, “Investigation of effective climatology parameters on COVID-19 outbreak in Iran,” *Sci. Total Environ.*, vol. 729, p. 138705, 2020, doi: 10.1016/j.scitotenv.2020.138705.

[12] I. Franch-Pardo, B. M. Napoletano, F. Rosete-Verges, and L. Billa, “Spatial analysis and GIS in the study of COVID-19. A review,” *Sci. Total Environ.*, vol. 739, p. 140033, 2020, doi: 10.1016/j.scitotenv.2020.140033.

[13] 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.

[14] M. Ula, Mursyidah, Y. Hendriana, and R. Hardi, “An expert system for early diagnose of vitamins and minerals deficiency on the body,” in *2016 International Conference on Information Technology Systems and Innovation (ICITSI)*, Oct. 2016, pp. 1–6. doi: 10.1109/ICITSI.2016.7858225.

[15] I. Area, H. Batarfi, J. Losada, J. J. Nieto, W. Shammakh, and Á. Torres, “On a fractional order Ebola epidemic model,” *Adv. Differ. Equations*, vol. 2015, no. 1, p. 278, 2015, doi: 10.1186/s13662-015-0613-5.

[16] Mauliza and J. Fitriany, “Typhoid Fever Profiles at Cut Meutia Hospital, North Aceh, Indonesia, in 2016-2017,” 2018, pp. 395–400. doi: 10.1108/978-1-78756-793-1-00031.

[17] L. Wang and J. T. Wu, “Characterizing the dynamics underlying global spread of epidemics,” *Nat. Commun.*, vol. 9, no. 1, p. 218, 2018, doi: 10.1038/s41467-017-02344-z.

[18] M. U. G. Kraemer *et al.*, “Utilizing general human movement models to predict the spread of emerging infectious diseases in resource poor settings,” *Sci. Rep.*, vol. 9, no. 1, p. 5151, 2019, doi: 10.1038/s41598-019-41192-3.

[19] K. D. N. Indonesia, “Peraturan Menteri Dalam Negeri Nomor 137 Tahun 2017 tentang Kode Dan Data Wilayah Administrasi Pemerintahan,” *Indonesia, Kementerian Dalam Negeri*. Kementerian Dalam Negeri, 2017.

[20] C. Mergenthaler, M. Gurp, E. Rood, and M. Bakker, “The study of spatial autocorrelation for infectious disease epidemiology decision-making: a systematized literature review,” *CABI Rev.*, vol. 2022, 2022, doi: 10.1079/cabireviews202217018.

[21] 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.

[22] E. Kwiyolecha *et al.*, “Patterns of viral pathogens causing upper respiratory tract infections among symptomatic children in Mwanza, Tanzania,” *Sci. Rep.*, vol. 10, no. 1, p. 18490, Oct. 2020, doi: 10.1038/s41598-020-74555-2.

[23] Mutammimul Ula, Ananda Faridhatul Ulva, Mauliza, Muhammad Abdullah Ali, and Yumna Rilasmi Said, “Application Of Machine Learning In Predicting Children’s Nutritional Status With Multiple Linear Regression Models,” *MULTICA Sci. Technol. J.*, vol. 2, no. 2, pp. 124–130, Oct. 2022, doi: 10.47002/mst.v2i2.363.

[24] E. J. de Fortuny, D. Martens, and F. Provost, “Predictive Modeling With Big Data: *Is Bigger Really Better* ?,” *Big Data*, vol. 1, no. 4, pp. 215–226, 2013, doi: 10.1089/big.2013.0037.

[25] S. Sukarna, W. Sanusi, and H. Hardiono, “Analisis Moran’s I, Geary’s C, dan Getis-Ord G pada Penerapan Jumlah Penderita Kusta di Kabupaten Gowa,” *J. Math. Comput. Stat.*, vol. 2, no. 2, p. 151, May 2020, doi: 10.35580/jmathcos.v2i2.12577.

[26] K. Vilinová, “Spatial Autocorrelation of Breast and Prostate Cancer in Slovakia,” *Int. J. Environ. Res. Public Health*, vol. 17, no. 12, p. 4440, 2020, doi: 10.3390/ijerph17124440.

[27] M.-Y. Kim and S.-W. Lee, “Regression Tree Analysis for Stream Biological Indicators Considering Spatial Autocorrelation,” *Int. J. Environ. Res. Public Health*, vol. 18, no. 10, p. 5150, 2021, doi: 10.3390/ijerph18105150.

[28] X. Rodó *et al.*, “Climate change and infectious diseases: Can we meet the needs for better prediction?,” *Clim. Change*, vol. 118, no. 3–4, pp. 625–640, 2013, doi: 10.1007/s10584-013-0744-1.

[29] G. M. Vazquez-Prokopec *et al.*, “Using GPS Technology to Quantify Human Mobility, Dynamic Contacts and Infectious Disease Dynamics in a Resource-Poor Urban Environment,” *PLoS One*, vol. 8, no. 4, p. e58802, 2013, doi: 10.1371/journal.pone.0058802.

[30] S. Jung, J. Moon, and E. Hwang, “Cluster-Based Analysis of Infectious Disease Occurrences Using Tensor Decomposition: A Case Study of South Korea,” *Int. J. Environ. Res. Public Health*, vol. 17, no. 13, p. 4872, 2020, doi: 10.3390/ijerph17134872.

[31] Y. Asbar, S. Biby, J. Simarmata, Abror, A. Pratama, and M. Ula, “Analysis of the application of data mining clustering model in the selection of the best study program based on student interests at the University of Aceh Province,” 2022, p. 090005. doi: 10.1063/5.0106950.

[32] L. Goscé and A. Johansson, “Analysing the link between public transport use and airborne transmission: mobility and contagion in the London underground,” *Environ. Heal.*, vol. 17, no. 1, p. 84, 2018, doi: 10.1186/s12940-018-0427-5.

[33] A. Zhang *et al.*, “The Collapse of Infectious Disease Diagnoses Commonly Due to Communicable Respiratory Pathogens During the Coronavirus Disease 2019 Pandemic: A Time Series and Hierarchical Clustering Analysis,” *Open Forum Infect. Dis.*, vol. 9, no. 7, Jul. 2022, doi: 10.1093/ofid/ofac205.

[34] N. C. Grassly and C. Fraser, “Seasonal infectious disease epidemiology,” *Proc. R. Soc. B Biol. Sci.*, vol. 273, no. 1600, pp. 2541–2550, 2006, doi: 10.1098/rspb.2006.3604.

[35] S. S. Morse, “Factors in the emergence of infectious diseases,” *Emerg. Infect. Dis.*, vol. 1, no. 1, pp. 7–15, 1995, doi: 10.3201/EID0101.950102.

[36] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, “K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data,” *Inf. Sci. (Ny).*, vol. 622, pp. 178–210, 2023, doi: 10.1016/j.ins.2022.11.139.

[37] K. P. Sinaga and M.-S. Yang, “Unsupervised K-Means Clustering Algorithm,” *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.

[38] R. Indrakumari, T. Poongodi, and S. R. Jena, “Heart Disease Prediction using Exploratory Data Analysis,” *Procedia Comput. Sci.*, vol. 173, pp. 130–139, 2020, doi: 10.1016/j.procs.2020.06.017.

[39] H. N. Surura, M. Mauliza, and J. Fitriany, “Hubungan Riwayat Status Imunisasi Bacille Calmette-Guérin (Bcg) Dengan Kejadian Tuberkulosis (Tb) Pada Anak Di Badan Layanan Umum Daerah Rumah Sakit Umum Cut Meutia Aceh Utara Tahun 2015,” *AVERROUS J. Kedokt. dan Kesehat. Malikussaleh*, vol. 3, no. 2, p. 57, Feb. 2018, doi: 10.29103/averrous.v3i2.441.

[40] H. Nen-Fu, H. I-Hsien, L. Chia-An, C. Hsiang-Chun, T. Jian-Wei, and F. Tung-Te, “The Clustering Analysis System Based on Students’ Motivation and Learning Behavior,” in *2018 Learning With MOOCS (LWMOOCS)*, 2018, pp. 117–119. doi: 10.1109/LWMOOCS.2018.8534611.

[41] A. H. Durán, T. M. Greco, B. Vollmer, I. M. Cristea, K. Grünewald, and M. Topf, “Protein interactions and consensus clustering analysis uncover insights into herpesvirus virion structure and function relationships,” *PLOS Biol.*, vol. 17, no. 6, p. e3000316, 2019, doi: 10.1371/journal.pbio.3000316.

[42] X. Xiao, A. J. van Hoek, M. G. Kenward, A. Melegaro, and M. Jit, “Clustering of contacts relevant to the spread of infectious disease,” *Epidemics*, vol. 17, pp. 1–9, 2016, doi:

10.1016/j.epidem.2016.08.001.

[43] R. M. McCloskey and A. F. Y. Poon, “A model-based clustering method to detect infectious disease transmission outbreaks from sequence variation,” *PLOS Comput. Biol.*, vol. 13, no. 11, p. e1005868, 2017, doi: 10.1371/journal.pcbi.1005868.

[44] M. C. V. Guilamet, M. Bernauer, S. T. Micek, and M. H. Kollef, “Cluster analysis to define distinct clinical phenotypes among septic patients with bloodstream infections,” *Medicine (Baltimore).*, vol. 98, no. 16, p. e15276, 2019, doi: 10.1097/MD.0000000000015276.

[45] Y. Zhou, K. Gu, and T. Huang, “Unsupervised Representation Adversarial Learning Network: from Reconstruction to Generation,” in *2019 International Joint Conference on Neural Networks (IJCNN)*, 2019, pp. 1–8. doi: 10.1109/IJCNN.2019.8852395.