Contents Lists Available at: https://journal2.unusa.ac.id/

Medical Technology and Public Health Journal

Journal Home Page: https://journal2.unusa.ac.id/index.php/MTPHJ

Research Article

Regional Cluster Analysis in East Java Province Based on Non-Communicable Disease Using Fuzzy C-Means

Dwi Handayani[1*] | *Abdul Hakim Zakkiy Fasya*[1] | *Mursyidul Ibad*[1] | *Tamara Nur Budiarti*[2]

[1]*Department of Public Health, Faculty of Health, Universitas Nahdlatul Ulama Surabaya, Indonesia*
[2]*Department of Occupational Health and Safety, Politeknik Kesehatan Kerta Cendekia, Indonesia*

Corresponding Author:
Dwi Handayani, Department of Public Health, Faculty of Health, Universitas Nahdlatul Ulama Surabaya, Indonesia.
Email: handayani.dwi@unusa.ac.id

DOI: *10.33086/mtphj.v8i2.5568*

Article History:
Received, January 11th, 2024
Revised, February 13th, 2024
Accepted, March 06th, 2024
Available Online: September 1st, 2024

Please cite this article as:
Handayani, Dwi, et. al, "Regional Cluster Analysis in East Java Province Based on Non-Communicable Disease Using Fuzzy C-Means" Register: Medical Technology and Public Health Journal, Vol. 8, No. 2, pp. 122-129, 2024

ABSTRACT

East Java has a prevalence of NCD exceeding the national average. Limited information on the disease clustering makes less optimal policy-making for preventing and controlling. This study aims to cluster areas and identify NCD risk factors. The study used secondary data from Basic Health Research in 2018. The data was analyzed using FCM. There were 3 clusters resulting from this research. In Cluster 1, the average proportion of all NCD's and 33.33% of the NCD's risk factors exceeds the average proportion of the province. In Cluster 2, the average proportion of 1 NCD and 77.78% of the NCD's risk factors exceeds the average proportion of the province. In Cluster 3, all NCD's have an average proportion less than the average proportion in the province and 22.22% of the NCD's risk factors have an average proportion exceeding the average proportion in the province. Cluster 1 is the cluster with the average proportion of NCD's exceeding the average proportion of NCD's in the province and the highest among the other clusters. Cluster 2 is a cluster with an area with risk factors for NCD's with the highest average proportion of the different clusters.

Keywords: *NCDs, analysis cluster, fuzzy c-means*

INTRODUCTION

The rapid development of the era towards the industrial and technological revolution brought socio-economic changes to people's habit patterns. This change has resulted in an epidemiological transition, namely that morbidity and mortality are not only influenced by high infections of communicable diseases but also by an increase in the prevalence of non-communicable diseases. Previously, infectious diseases and malnutrition were the leading causes of death (Kabudula et al., 2017).

In the modern era, like today, there is a shift in disease patterns in society, which places non-communicable diseases as a high contributor to death due to lifestyle changes. Non-communicable diseases are disorders of human physiological functions that last a long time and manifest a combination of genetic factors, physiological conditions, environment, and lifestyle. Non-

communicable diseases are also often referred to as chronic diseases because they can cause the appearance of symptoms over a long period (World Health Organization, 2021b).

The World Health Organization (WHO) stated that 41 million deaths worldwide are caused by non-communicable diseases each year. Types of infectious diseases that contribute the most to the world's death rate include coronary heart disease (17,900,000), cancer (9,300,000), and diabetes (1,500,000) (World Health Organization, 2021b). Accozrding to the Basic Health Research data 2018, during 2013-2018, Indonesia experienced a significant spike in the prevalence of NCD, which was more than 34%. Some common types of NCDs include allergies, diabetes, rheumatism, depression, hypertension, stroke, asthma, and chronic lung disease, which are frequently experienced by the community (Kemenkes RI, 2019).

East Java Province is a province with NCD prevalence exceeding the national prevalence. Cases reached 44.74% of the total health problems in 1990 and jumped to 75.1% in 2017. The high spike in NCD cases in East Java is often associated with the poor quality of people's lifestyles. According to the survey, less than 20% of the population adopts a healthy lifestyle. Ideally, the community's application of a healthy lifestyle in an area to suppress the spread of the disease must reach 80% (BKKBN & Kemenkes RI, 2018).

Various efforts have been made to suppress the increase in NCDs in East Java. One of the embodiments of NCD's prevention and control strategy and its risk factors is the implementation of Integrated Guidance Pos (Posbindu)-based NCD risk factor monitoring surveillance. Posbindu NCD is a community-based health effort (UKBM) under the guidance of the Community Health Center (Puskesmas) with a systematic and continuous analysis of risk factors for NCD. The Posbindu NCD aims to collect data, process data, and deliver epidemiological information to health program organizers so that they can carry out countermeasures effectively and efficiently (Indonesia Ministry of Health, 2014).

If left unchecked, the high spike in NCD cases in East Java is feared to threaten the quality of human resources and reduce life expectancy. Therefore, innovation is needed to determine NCD's prevention and control strategy. Regional clustering is one of the efforts to support the determination of the intervention strategy. Regional clustering is a technique of grouping regions based on the most similar characteristics (Suyanto, 2017). Regional clustering aims to identify areas with the most similar characteristics so that the policy formulation process is on target. Regional clustering can be done using the Fuzzy C-Means method.

Fuzzy C-Means (FCM) is a method of clustering data into a cluster based on the degree of membership of each cluster (Sadaaki et al., 2008). In contrast to k-Means, FCM can group data points far from a cluster's center into that cluster (Suyanto, 2017). Clustering with FCM results in a cluster average closer to the initial data average than clustering with K-Means (Firdaus et al., 2021). Therefore, this study aims to cluster areas based on the average proportion of non-communicable diseases using the Fuzzy C-Means method and identify NCD risk factors in each cluster.

MATERIAL AND METHODS

This study is an ecological analysis focusing on a specific region. This type of research is non-reactive research, where the research subject is unaware of being the research subject (there is no interaction between the researcher and the research subject). The data source comes from the aggregated data of the 2018 Basic Health Research report. The research population is all regencies/cities in East Java Province. Sampling was done by taking all regencies/cities to be included as research samples (total sampling).

The variables in this study include the average proportion of non-communicable diseases and the average proportion of risk factors for non-communicable diseases. NCDs involved in the study included hypertension, diabetes mellitus, and central obesity. The risk of NCD included the prevalence of smoking activity, lack of physical activity, consumption of sweet foods, consumption of sugary drinks, consumption of salty foods, consumption of fatty foods, consumption of burned foods, not consuming fruits and vegetables, and passive smokers.

Data analysis was conducted to classify areas in the East Java Province based on the average proportion of NCD and to describe the risk factors for NCD in each cluster formed. The data will be analyzed using a fuzzy C-Means clustering technique. The stages in FCM include inputting the data to be grouped, determining the number of clusters (c), the rank for the partition matrix (w), the maximum iteration (MaxIter), the most minor expected error (ε), the initial objective function (P0=0), and the initial iteration (t=1); generate random numbers μik, i=1,2,…,n; k=1,2,…,c as elements of the initial partition matrix U; calculate the center of the k-cluster: Vkj, with k=1,2,…,c; dan j=1,2,…,m; calculate the objective function at the t-th iteration; calculate partition matrix changes; and ends by checking the stop condition, namely when ($|Pt - Pt{-}1| < \varepsilon$) or (t > MaxIter). Clusters in this study are divided into three, i.e., Cluster 1, Cluster 2, and Cluster 3.

RESULTS AND DISCUSSION

Regencies/Cities in East Java Province consist of 38 regions. The results of regional clustering based on the average proportion of NCD (hypertension, diabetes mellitus, and central obesity) can be seen in Table 1.

Table 1. Clustering of Regions Based on Prevalence of Non-Communicable Diseases

Indicators	Average Proportion			
	Province	Cluster 1	Cluster 2	Cluster 3
Hypertension	34,6	35,7*	34,9	34,1
Diabetes Mellitus	2,6	3,3*	2,4	1,5
Central obesity	30,4	34,7*	28,4	22,3

Data source: Basic Health Research, 2018

**Highest average proportion*

The order of the average proportion of hypertension from the highest based on Table 1 is Cluster 1, Cluster 2, and Cluster 3. The order of the average proportion of diabetes mellitus from the highest based on Table 1 is Cluster 1, Cluster 2, and Cluster 3. Table 1 shows the average proportion of central obesity from the highest, namely Cluster 1, Cluster 2, and Cluster 3.

The clustering of regions in East Java Province based on the prevalence of NCDs in Table 1 shows that the average proportion of all non-communicable diseases in Cluster 1 exceeds those in East Java Province. While in Cluster 2, there is 1 NCD with an average proportion exceeding the average proportion of non-communicable diseases in East Java Province, the rest does not exceed the average proportion of non-communicable diseases in East Java Province but is higher than in Cluster 3. In Cluster 3, all non-communicable diseases have an average proportion less than the average in East Java Province. The list of regencies/cities in each cluster is presented in Table 2.

Table 2. List of Regencies/Cities in Cluster 1, Cluster 2, and Cluster 3

Cluster 1	Cluster 2	Cluster 3
Trenggalek Regency	Pacitan Regency	Tulungagung Regency
Blitar Regency	Ponorogo Regency	Bangkalan Regency

Cluster 1	Cluster 2	Cluster 3
Banyuwangi Regency	Kediri Regency	Sampang Regency
Pasuruan Regency	Malang Regency	Pamekasan Regency
Sidoarjo Regency	Lumajang Regency	Sumenep Regency
Jombang Regency	Jember Regency	
Nganjuk Regency	Bondowoso Regency	
Lamongan Regency	Situbondo Regency	
Gresik Regency	Probolinggo Regency	
Kediri City	Mojokerto Regency	
Blitar City	Madiun Regency	
Malang City	Magetan Regency	
Pasuruan City	Ngawi Regency	
Mojokerto City	Bojonegoro Regency	
Madiun City	Tuban Regency	
Surabaya City	Probolinggo City	
	Batu City	

The results of regional clustering based on Table 2 show that there are 16 regencies/cities in Cluster 1, 17 regencies/cities in Cluster 2, and 5 regencies in Cluster 3. The mapping of Regencies/Cities in East Java Province based on the type of cluster can be seen in Figure 1.

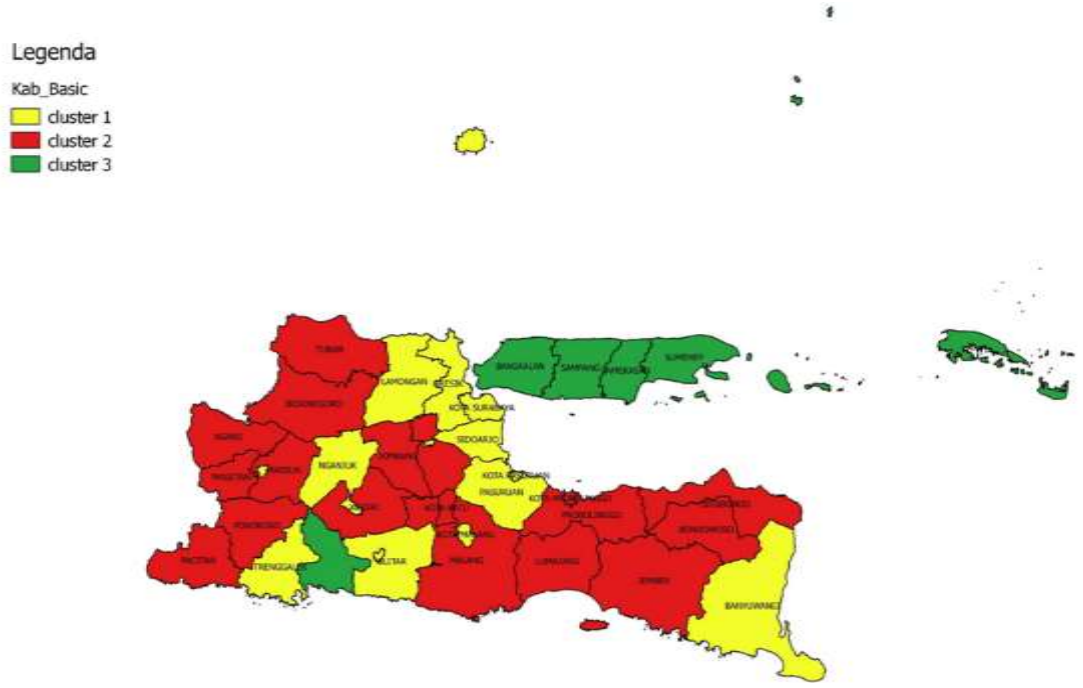

Fig. 1. The Mapping of Regencies/Cities in East Java Province Based on The Type of Cluster

Mapping of regencies/cities in East Java Province based on cluster type based on Figure 1 shows that areas included in Cluster 1 are colored yellow, areas included in Cluster 2 are colored red, and areas included in Cluster 3 are colored green. The mapping based on Figure 1 shows that areas in red tend to dominate, followed by yellow and green.

Identification of NCD risk factors includes the prevalence of unhealthy lifestyles in the community as measured by the prevalence of smoking activity, lack of physical activity, consumption of sweet foods, consumption of sugary drinks, consumption of salty foods, consumption of fatty foods, consumption of burned foods, not consuming fruits and vegetables, and passive smoking in each cluster are presented in Table 3.

Table 3. Identification of NCD Risk Factors in Each Cluster

Indicators	Average Proportion			
	Province	Cluster 1	Cluster 2	Cluster 3
Smoking habits	28,1	26,8	29,7*	26,1
Lack of physical activity	26,5	26,9	23,6	31,4*
Consumption of sweet food	31,3	31,5	32,5*	22,1
Consumption of sweet drinks	57,0	58,5	63,8*	34,5
Consumption of salty food	28,8	23,7	35,8*	17,8
Consumption of fatty foods	48,6	46,6	62,6*	25,9
Consumption of burnt food	2,8	2,4	3,2*	2,5
Do not eat fruit and vegetables per day of the week	8,6	7,5	7,3	15,5*
Passive smokers	30,8	27,5	37,8*	18,4

Data source: Basic Health Research, 2018
**Highest average proportion*

Identification of NCD risk factors in Cluster 1 based on Table 3 shows that 33.33% of the measurement variables have an average proportion that exceeds the average proportion in East Java Province but is lower than the average proportion in Cluster 2, and 66.67% of the measurement variable has an average proportion less than the average proportion in East Java Province. Identification of NCD risk factors in Cluster 2 based on Table 3 shows that 77.78% of the measurement variables have an average proportion exceeding the average proportion in East Java Province and the highest among values in other clusters, and 22.22% of the measurement variables have the average proportion is less than the average proportion in East Java Province. Identification of NCD risk factors in Cluster 3 based on Table 2 shows that 22.22% of the measurement variables have an average proportion exceeding the average proportion in East Java Province but lower than the average proportion in Cluster 2, and 77.78% of the measurement variable has an average proportion less than the average proportion in East Java Province.

The clustering of regencies/cities in East Java Province through the Fuzzy C Means method is a form of essential identification and preliminary studies in policy-making related to infectious disease control and disease risk. Regional clustering aims to find similarities between several regions and collect them into one large group. The final result of regional clustering is to obtain homogeneous conditions between regions in one group and heterogeneous conditions between groups (Suyanto, 2017). The clustering results can be used as a reference in determining the priority of problem-solving so that the management of health problems is carried out appropriately, effectively, and efficiently. Fuzzy C Means is a clustering technique based on unsupervised learning, namely a machine learning technique to identify patterns or groups in data that were previously very random, unlabeled, and unclassified (Albayrak & Amasyali, 2003; Bouchefry K El, 2020). The clustering results show three clusters with 16 regencies/cities in Cluster 1, 17 regencies/cities in Cluster 2, and 5 regencies/cities in Cluster 3.

The World Health Organization (WHO) targets a 25% reduction in the risk of death from non-communicable diseases by 2025. An essential step in managing non-communicable disease control is suppressing disease risk factors. It is crucial for the government to monitor the prevalence of non-communicable diseases and associated risk factors within the community. In order to reduce the prevalence of non-communicable diseases, the WHO recommends a comprehensive approach,

involving policy formation, health promotion, and the provision of behavior change support to target communities. This approach should involve all sectors, including health, finance, transportation, education, agriculture, and others (World Health Organization, 2021b).

The World Health Organization WHO targets a decrease in the prevalence of hypertension by 33% in 2030. The Indonesian government, through the preparation of the 2015-2019 National Medium-Term Development Plan (RPJMN), states that in 2019, the prevalence of hypertension must be reduced to 23.4% from 25.8 % in 2013. The results showed that the average proportion of hypertension in Cluster 1 was 35.7%, Cluster 2 was 34.9%, and Cluster 3 was 34.1%. Based on the global targets set by WHO and Indonesia's national targets, the average proportion of hypertension in areas in Cluster 1, Cluster 2, and Cluster 3 has yet to reach the target (Bappenas, 2014; WHO, 2021a).

The results showed that the average proportion of diabetes mellitus in areas in Cluster 1 was 3.3%, Cluster 2 was 2.4%, and Cluster 3 was 1.5%. WHO targets an increase of 0% in the prevalence of diabetes mellitus in 2025, meaning that there will be no additional cases of diabetes mellitus in 2025 (World Health Organization, 2022). The results showed that the average proportion of central obesity in Cluster 1 reached 34.7%, Cluster 2 reached 28.4%, and Cluster 3 reached 22.3%. The Government of Indonesia, through the Ministry of Health, has set a policy in the form of a target for reducing obesity in the population aged >18 years, not more than 21.8 by 2024, which is contained in the Regulation of the Minister of Health of the Republic of Indonesia Number 21 of 2020 concerning the Strategic Plan of the Ministry of Health for the Year 2020-2024. Compared with this target, the average proportion of obesity in Cluster 1, Cluster 2, and Cluster 3 has not yet reached the national target (Kemenkes RI, 2020).

Many studies have been conducted to identify risk factors for the emergence of non-communicable diseases. The high prevalence of smokers, lack of physical activity in the community, and low levels of HDL fat could increase the prevalence of non-communicable diseases in an area (Rarau et al., 2017). An increase in physical activity and an improvement in the quality of food consumption can reduce the prevalence of non-communicable diseases such as cancer, coronary heart disease, stroke, diabetes mellitus, and dementia (Peters et al., 2019).

Passive smoking is a term that refers to those who are exposed to secondhand smoke. The literature states that as a passive smoker, a person has a significant enough chance of suffering from atherothrombosis, which can develop into cardiovascular disease (Digiacomo et al., 2018). In addition, male passive smokers have an increased risk of developing lung cancer and colorectal cancer. In contrast, female passive smokers have an increased risk of developing breast cancer and cervical cancer (Poirier et al., 2019).

The Government of Indonesia, through the Ministry of Health, has established several policies to reduce the proportion of risk factors for non-communicable diseases, such as setting a target to reduce the prevalence of smoking among adolescents by 5.4% in 2029, reducing the prevalence of unhealthy food consumption, increasing the prevalence of healthy foods, and increasing physical activity (Kemenkes RI, 2020). The results showed that the average proportion of risk factors for non-communicable diseases was still relatively high based on the three clusters formed in East Java Province.

Cluster 2 is a cluster with an area with risk factors for non-communicable diseases with the highest average proportion of the other two clusters. Suppose it is associated with the average proportion of non-communicable diseases. In that case, it can be seen that Cluster 1, with the highest average proportion of non-communicable diseases, has a lower average proportion of non-

communicable disease risk factors than Cluster 2. This condition, of course, requires further analysis in the following study. The limitation of this research is that the identification of the risk of non-communicable diseases is only done descriptively without going through statistical analysis, so there are no known factors that have a significant effect on the proportion of non-communicable diseases. In addition, this study did not perform clustering involving spatial conditions.

CONCLUSION AND SUGGESTION

Three clusters are formed with 16 Regencies/Cities in Cluster 1, 17 Regencies/Cities in Cluster 2, and 5 in Cluster 3. Cluster 1 is the cluster with the average proportion of NCDs exceeding the average NCDs in East Java Province and the highest among the other two clusters. Cluster 2 is a cluster with an area with risk factors for NCDs with the highest average proportion of the other two clusters.

ACKNOWLEDGEMENT

The authors would like to thank Universitas Nahdlatul Ulama Surabaya for financially supporting the entire research process until the publication.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest in this research.

REFERENCES

Albayrak, S., & Amasyali, M. F. (2003). *Fuzzy C-Means Clustering on Medical Diagnostic Systems*. International Twelfth Turkish Symposium on Artificial Intelligence and Neural Network.

Bappenas. (2014). *The Medium-Term National Development Plan (RPJMN) 2015–2019*. Jakarta

BKKBN & Kemenkes RI. (2018). *The 2017 Indonesia Demographic and Health Survey*. Jakarta

Bouchefry K El, R. de S. (2020). *Knowledge Discovery in Big Data from Astronomy and Earth Observation*. Elsevier.

Digiacomo, S., Jazayeri, M., Barua, R., & Ambrose, J. (2018). Environmental Tobacco Smoke and Cardiovascular Disease. *Int J Environ Res Public Health*, *16*(1), 96.

Firdaus, H. S., Nugraha, A. L., Sasmito, B., & Awaluddin, M. (2021). Perbandingan Metode Fuzzy C-Means Dan K-Means Untuk Pemetaan Daerah Rawan Kriminalitas Di Kota Semarang. *Elipsoida : Jurnal Geodesi Dan Geomatika*, *4*(01), 58–64. https://doi.org/10.14710/elipsoida.2021.9219

Kemenkes RI. (2014). *Technical Guideline for Posbindu-Based NCD Risk Factor Surveillance*. Directorate General of Disease Control and Environmental Health.

Kemenkes RI. (2019). *The 2018 East Java Province Basic Health Survey (Riskesdas)*. Publishing Agency for Health Research and Development.

Kemenkes RI. (2020). *Regulation of the Indonesia Minister of Health Number 21 of 2020 concerning the Strategic Plan of the Ministry of Health for 2020-2024*. Jakarta

Kabudula, C. W., Houle, B., Collinson, M. A., Kahn, K., Gómez-Olivé, F. X., Clark, S. J., & Tollman, S. (2017). Progression of the epidemiological transition in a rural South African setting: findings from population surveillance in Agincourt, 1993-2013. *BMC Public Health*, *17*(1), 1–15. https://doi.org/10.1186/s12889-017-4312-x

Peters, R., N, E., Peters, J., Beckett, N., Booth, A., Rockwood, K., & Anstey, K. (2019). Common risk factors for major noncommunicable disease, a systematic overview of reviews and commentary: the implied potential for targeted risk reduction. *Ther Adv Chronic Dis*, *10*.

Poirier, A., Ruan, Y., Grevers, X., Walter, S., Villeneuve, P., Friedenreich, C., & Brenner, D. (2019). Estimates of the current and future burden of cancer attributable to active and passive tobacco smoking in Canada. *Prev Med*, *122*, 9–19.

Rarau, P., Vengiau, G., Gouda, H., Phuanukoonon, S., Kevau, I. H., Bullen, C., Scragg, R., Riley, I., Marks, G., Umezaki, M., Morita, A., Oldenburg, B., McPake, B., & Pulford, J. (2017). Prevalence of non-communicable disease risk factors in three sites across Papua New Guinea: A cross-sectional study. *BMJ Global Health*, *2*(2). https://doi.org/10.1136/bmjgh-2016-000221

Sadaaki, M., Hidetomo, I., & Katsuhiro, H. (2008). *Algorithms for Fuzzy Clustering*. Springer Berlin

Suyanto. (2017). *Data Mining: For Data Classification and Clustering*. Informatika.

WHO. (2021a). *Hypertention*. https://www.who.int/news-room/fact-sheets/detail/hypertension

WHO. (2021b). *Non Communicable Diseases*. https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases

WHO. (2022). *WHO diabetes targets: accelerating progress towards 2030*. WHO. https://idf.org/our-activities/advocacy-awareness/advocacy-activities/advocacy-events/who-diabetes-targets-accelerating-progress-towards-2030