International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Mathematical Modelling of Engineering Problems
Vol. 12, No. 12, December, 2025, pp. 4282-4288
Journal homepage: http://iieta.org/journals/mmep

Comparative Analysis of Clustering Algorithms for Health Risk Profiling Based on Dietary and Physical Activity Patterns

Sri Mulyati*, Muhammad Harel, Hanuga Fathur Chaerulisma, Kurniawan Dwi Irianto

Department of Informatics, Faculty of Industrial Technology, Universitas Islam Indonesia, Yogyakarta 55281, Indonesia

Corresponding Author Email: mulya@uii.ac.id

https://doi.org/10.18280/mmep.121216

Received: 22 July 2025
Revised: 16 October 2025
Accepted: 12 November 2025
Available online: 31 December 2025

Keywords:
clustering, DBSCAN, K-Means, Hierarchical Clustering, food consumption, physical activity, health risk profiling, NCDs

ABSTRACT

Non-Communicable Diseases (NCDs) continue to rise in line with changes in people's consumption patterns and lifestyles, so a data-driven approach is needed to understand health risk segmentation. This study aims to classify food consumption behaviors and healthy lifestyles among the productive age group. Data were collected from 321 respondents through a structured survey that included eating habits, physical activity, as well as demographic and health factors. Three clustering algorithms were tested, namely K-Means, Hierarchical Clustering, and Density-Based Spatial Clustering of Applications with Noise (DBSCAN), with evaluation using Silhouette Score, Davies-Bouldin Index, and Dunn Index. The results showed that DBSCAN achieved the best performance (Silhouette Score = 0.416; Davies-Bouldin Index = 0.448; Dunn Index = 1.430), which indicates a separate cluster with a good degree of cohesiveness. In contrast, K-Means showed the lowest performance (Silhouette Score = 0.045; Davies-Bouldin Index = 2.936; Dunn Index = 0.251), while Hierarchical Clustering showed limited performance (Silhouette Score = 0.046; Davies–Bouldin Index = 2.956; Dunn Index = 0.239). For the K-Means analysis, the optimal number of clusters was determined to be k = 8 using the elbow method, which was subsequently consolidated. The cluster profiles identified three main groups: (1) individuals with healthy lifestyles, (2) moderate-risk individuals with high calorie consumption and low activity, and (3) high-risk individuals with poor diets and sedentary habits. These findings confirm that DBSCAN is effective in identifying patterns of health risks and can serve as the basis for designing more targeted promotive and preventive interventions to reduce the risk of NCDs.

1. INTRODUCTION

Non-Communicable Diseases (NCDs) are a group of diseases that are not caused by infection but are influenced by lifestyle, environmental, and genetic factors [1]. According to the World Health Organization (WHO), the prevalence of NCDs is increasing among young adults, a stage of life that should ideally be free from serious health problems. The *Global Burden of Disease Study* 2019 revealed that NCDs such as hypertension, diabetes, and cardiovascular disease have become the leading causes of morbidity and mortality worldwide, including in Indonesia [2]. Lifestyle factors, particularly food consumption patterns, adherence to a healthy diet, and physical activity, play a crucial role in shaping NCD risk. Previous research has also shown that the prevalence of risk factors such as obesity and hypertension is rising among young age groups [3-5]. A systematic analysis of the *Global Burden of Disease Study 2010* emphasized that NCDs are now among the leading causes of death globally, with significant increases among young adults [6, 7]. This highlights the urgency of profiling food consumption and lifestyle behaviors as a preventive approach to reduce long-term health burdens.

Food consumption patterns and adherence to healthy diets are essential in preventing obesity, diabetes, and cardiovascular diseases. However, these behaviors are often shaped by demographic characteristics, physical activity, and health awareness, making them complex to analyze. In Indonesia, mixed findings have emerged: one study reported no significant association between fast-food consumption and obesity in adolescents [8, 9], while others found that irregular diets and unhealthy food intake increased the risks of gastritis [10] and hypertension [11]. These studies suggest that poor diet quality and unbalanced nutrition remain important risk factors for NCDs. In addition, health literacy, particularly media health literacy, has been shown to influence behaviors such as diet, smoking, and physical activity, making it a potential preventive strategy among adolescents and young adults [12].

Beyond behavioral insights, data-driven approaches have been increasingly adopted. For example, Alosaimi et al. [11] showed that a combination of poor diet, low physical activity, and sedentary lifestyles significantly increases the risks of obesity and cardiovascular disease [13]. Recent analyses using European adolescent cohorts further reinforce this concern. Findings indicate that adolescents who adhere more closely to healthy dietary patterns such as the EAT Lancet recommendations are significantly more likely to achieve ideal cardiovascular health profiles [14]. Complementary evidence shows that eating behaviors driven by emotional stress, food interest, and motivational factors are strongly associated with higher consumption of energy-dense foods [15]. Nevertheless, current health systems remain more focused on disease detection after onset rather than proactive, data-driven prevention based on food consumption and lifestyle [11].

To address this issue, clustering methods have been widely applied to segment food consumption and lifestyle patterns. K-Means is popular due to its scalability and computational efficiency. López-Gil and Martínez-López [13] applied K-Means to children's macronutrient intake and identified three major dietary profiles [10]. Hierarchical Clustering provides the advantage of exploring data structures at multiple levels, as seen in workflows that combine Hierarchical Clustering and K-Means with principal component analysis for dietary pattern derivation in adult women [16]. Meanwhile, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) offers the ability to identify clusters of arbitrary shapes and isolate noise [17]. Using the sampling-based clustering enhancement introduced by de Moura Ventorim et al. [18], the method was applied to map obesity risks based on eating habits and physical activity, enabling the detection of vulnerable outlier populations. However, most existing studies rely on only one clustering method, limiting the comprehensiveness of segmentation results.

Although numerous studies have examined dietary and lifestyle factors associated with NCDs, few have systematically compared clustering algorithms for health-related profiling. Existing approaches tend to rely on a single method and thus fail to capture the variability of individual behaviors in food consumption and lifestyle. Yet each algorithm has distinct strengths and weaknesses that may influence segmentation quality.

This study seeks to fill that gap by developing a segmentation model using three clustering algorithms: K-Means, Hierarchical Clustering, and DBSCAN to determine which method produces the most cohesive and well-separated clusters for health behavior profiling. The novelty of this research lies in its systematic, machine learning based comparison of clustering techniques in the context of NCD prevention. By leveraging these methods, this study aims to identify distinct patterns of food consumption and lifestyle, thereby providing a foundation for more targeted promotive and preventive interventions, particularly for young adults at risk.

2. METHODOLOGY

2.1 Study population and survey design

This section describes the methodology employed in this study, including study design, data preprocessing, clustering approaches, and validation techniques. This study is a quantitative study that uses machine learning methods to analyze food consumption patterns and adherence to healthy diets in individuals aged 18–35 years. With a data-driven approach, this study aims to identify groups of individuals based on their food consumption habits and factors that affect their adherence to a healthy diet. This process includes segmentation of individuals using the K-Means method, which allows an analysis of the level of adherence to a healthy diet, taking into account demographic characteristics, physical activity, and health behaviors. To evaluate the effectiveness of the methods used, this study also compared the results of segmentation with the Hierarchical Clustering method and DBSCAN to assess the optimal model in grouping individuals based on their food consumption patterns. This analysis is carried out by calculating the Silhouette Score value. Each cluster formed will be further analyzed through cluster profiling to understand the unique characteristics of each segment, so as to provide in-depth insights into strategies that can be applied in increasing community adherence to a healthy lifestyle. With a better understanding of food consumption patterns and individual adherence to healthy diets, this research is expected to be a foundation for designing more effective and data-driven health education and intervention programs.

2.2 Data preprocessing

The raw dataset was cleaned to remove missing or inconsistent responses. Standardization was applied to numerical variables to ensure comparability across different scales. The dataset included six key health indicators: dietary diversity, physical activity, Body Mass Index (BMI), smoking habits, water sanitation, and access to healthcare. Each variable was normalized using a z-score transformation to reduce the influence of varying units of measurement. Composite health scores were then calculated using a weighted sum approach, where each indicator contributed proportionally based on its relevance to overall health outcomes. The assigned weights were: dietary diversity (0.25), physical activity (0.20), BMI (0.20), smoking habits (0.15), water sanitation (0.10), and healthcare access (0.10). This composite score captured holistic lifestyle dimensions and provided an integrated measure of individual health status. All variables and composite scores were subsequently used as inputs for the clustering procedures described as follows.

2.3 Clustering algorithms

This study adopted a comparative clustering approach to identify the most suitable segmentation technique for population health profiles. Three unsupervised learning algorithms, K-Means, Hierarchical Clustering, and DBSCAN, were applied using identical pre-processing and normalization procedures. The objective was not to prioritize one algorithm but to evaluate and compare their clustering performance to determine the optimal segmentation structure.

For K-Means, the algorithm partitions the dataset into k clusters by minimizing intra-cluster variance through iterative centroid updates. The optimal number of clusters (k) was determined using the Elbow Method, which analyzes the relationship between distortion scores and the number of clusters to identify the inflection point beyond which improvements become marginal.

For Hierarchical Clustering, an agglomerative approach was implemented using Ward's linkage and Euclidean distance to iteratively merge clusters that minimize within-cluster variance. The resulting dendrogram was analyzed to determine the appropriate number of clusters that best represented the data structure.

For DBSCAN, clustering was based on data point density rather than distance metrics. Two key parameters, eps (ε) and min_samples, were adjusted experimentally and set to 0.45 and 5, respectively. These values provided the most effective separation between dense regions and noise points, allowing

DBSCAN to detect irregularly shaped clusters and handle outliers effectively.

To assess clustering quality, three internal validation indices were calculated for each algorithm. (1) Silhouette Score, which measures cohesion and separation between clusters; (2) Davies–Bouldin Index (DBI), which evaluates intra- versus inter-cluster similarity (lower values indicate better performance); and (3) Dunn Index, which assesses compactness and distinctness between clusters.

The algorithm achieving the most favorable combination of these indices was regarded as the optimal segmentation model, serving as the basis for interpreting population health profiles into meaningful and actionable categories.

2.4 Validation and cluster profiling

The performance of each clustering algorithm was evaluated using multiple internal validation metrics: Silhouette Score, Davies–Bouldin Index, and Dunn Index. These metrics provided a comprehensive assessment of compactness, separation, and inter-cluster distance.

After validation, cluster profiling was conducted to describe the socio-demographic, behavioral, and dietary characteristics of each identified segment. This analysis aimed to provide actionable insights into designing tailored interventions to improve adherence to healthy lifestyles and prevent NCD risks.

2.5 Data collection

Data were collected through a structured questionnaire distributed to 321 respondents aged 18–35 years. The instrument was designed to capture four major dimensions: food consumption patterns, adherence to a healthy diet, physical activity, and demographic/health-related factors. The key attributes used for clustering are summarized in Table 1.

Table 1. Comparison of clustering algorithm performance

No.	Attributes	Column Names in Dataset	Description
1	Sports frequency	Physical_Activity_1	Number of weekly exercise sessions, used to measure activity frequency.
2	Duration of exercise	Physical_Activity_2	Average minutes per exercise session, assessing adequacy of exercise.
3	Dietary barriers	Diet_Barriers	Frequency of difficulties in maintaining a healthy diet.
4	Monthly food expenditure	Food_Consumption_Cost	Average monthly spending on food, reflecting purchasing power for diet.
5	Location of residence	Residence	Urban/rural classification, to capture geographical variations.
6	Education level	Education	Highest level of education, reflecting nutrition awareness.
7	Compliance with health guidelines	Diet_Adherence	Frequency of following health recommendations (e.g., low salt, low sugar).

The questionnaire in this study was designed to evaluate various aspects related to food consumption patterns and factors that affect the health of respondents. First, food consumption patterns are measured based on the frequency of consumption of foods containing saturated fats, sugar, salt, and fruit and vegetable intake, to evaluate food consumption habits. Second, adherence to a healthy diet was assessed by asking the extent to which respondents followed the recommendations of a balanced diet, the level of difficulty in implementing it, and the factors that support or hinder their adherence. Third, the demographic and health factors of the respondents were studied, including age, gender, education level, disease diagnosis, and consumption cost needs. Fourth, physical activity is analyzed based on the frequency and duration of exercise or physical activity carried out in the past week. The questionnaire also included the types of physical activity that were most often performed, such as walking, running, swimming, cycling, or group sports, as well as respondents' perceptions of the adequacy of their physical activity in maintaining health. Finally, dietary and health-related behavioral factors were reviewed to understand the extent to which respondents were aware of the impact of diet on their health. This includes anxiety levels about the consequences of an unhealthy diet, perceptions of the difficulty or cost of implementing a healthy diet, as well as psychological barriers that may affect their eating habits.

Based on the results of demographic data analysis from 321 respondents, the majority are female. The last most reported level of education was Senior High School. Most of the respondents live in private houses or in dormitories/boarding houses. The majority of respondents do not smoke and have never checked themselves at a health facility related to NCDs. In addition, most respondents also reported never being diagnosed with NCDs such as diabetes, hypertension, or high cholesterol. Age data is not analyzed due to inconsistent formatting in the age column fill-in.

3. RESULT AND DISCUSSION

3.1 K-Means clustering

The optimal number of clusters was determined using the elbow method. As shown in Figure 1, the distortion score decreased consistently as the number of clusters (k) increased. The inflection point was observed at $k = 8$, with a distortion score of approximately 6960.193. This point represents the optimal number of clusters, where the rate of decrease in distortion score begins to flatten significantly. Selecting this elbow point ensures a balance between clustering accuracy and model simplicity, thereby preventing overfitting and ensuring effective segmentation of data [19-21].

The blue curve in the graph represents the distortion score, which consistently decreases as the number of clusters increases, indicating improved compactness within clusters. Meanwhile, the green dashed line indicates the computation fit time required for each clustering process, showing the computational cost as k increases. The vertical dashed line highlights the identified elbow point at $k = 8$, which is selected as the optimal cluster number for further analysis.

Although the elbow method identified $k = 8$ as the optimal point based on the distortion score, subsequent clustering analysis consolidated the eight clusters into three clinically and practically meaningful main groups based on health profiles. These three groups are: (1) healthy group comprising individuals with optimal health values, (2) moderate-risk group showing health indicators requiring attention, and (3) high-risk group with critical health conditions. This consolidation approach maintains the granularity of *k* = 8 in detailed analysis while providing a practical interpretation applicable to the public health context.

3.2 Hierarchical Clustering and DBSCAN

For the Hierarchical Clustering method, the agglomerative approach with the Ward linkage criterion was applied [22]. The dendrogram in Figure 2 illustrates the merging process of clusters. The vertical axis represents the linkage distance, which reflects the level of dissimilarity between merged clusters. By applying a cutoff threshold (red dashed line), the optimal number of clusters was determined to be four to five. These clusters highlight groups of individuals with similar food consumption patterns and lifestyle behaviors.

For DBSCAN, the density-based approach was implemented with parameters eps = 0.5 and min_samples = 5 (tuned experimentally to balance compactness and noise detection). Unlike K-Means and Hierarchical Clustering, DBSCAN does not require pre-specifying the number of clusters [23]. Instead, it identifies dense regions in the data and isolates outliers. This property is particularly suitable for food consumption data, which often exhibit irregular shapes and heterogeneous distributions.

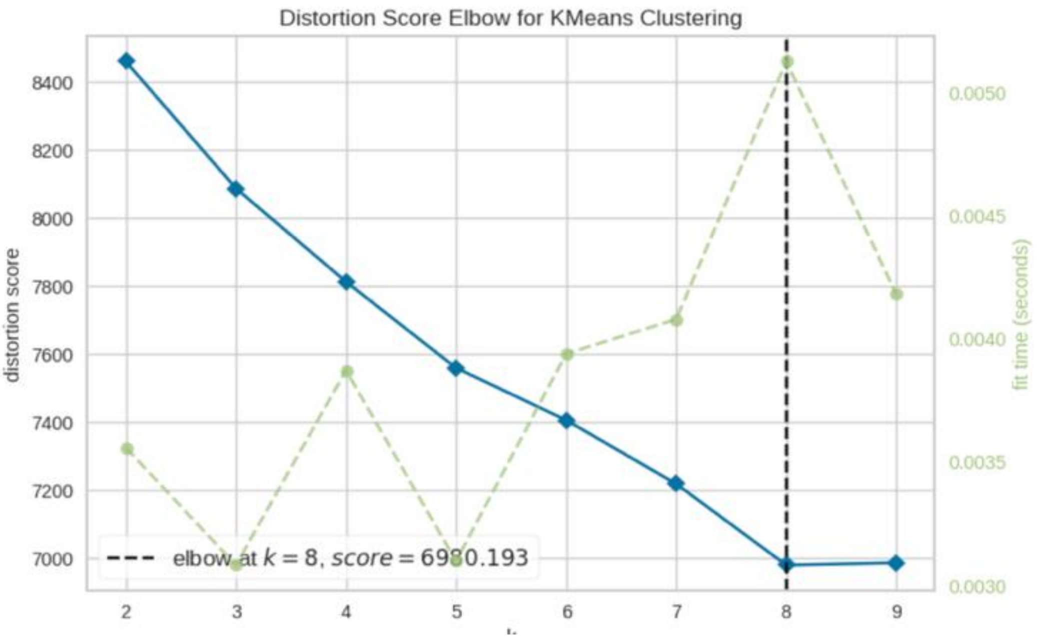

Figure 1. The elbow method graph for determining the optimal

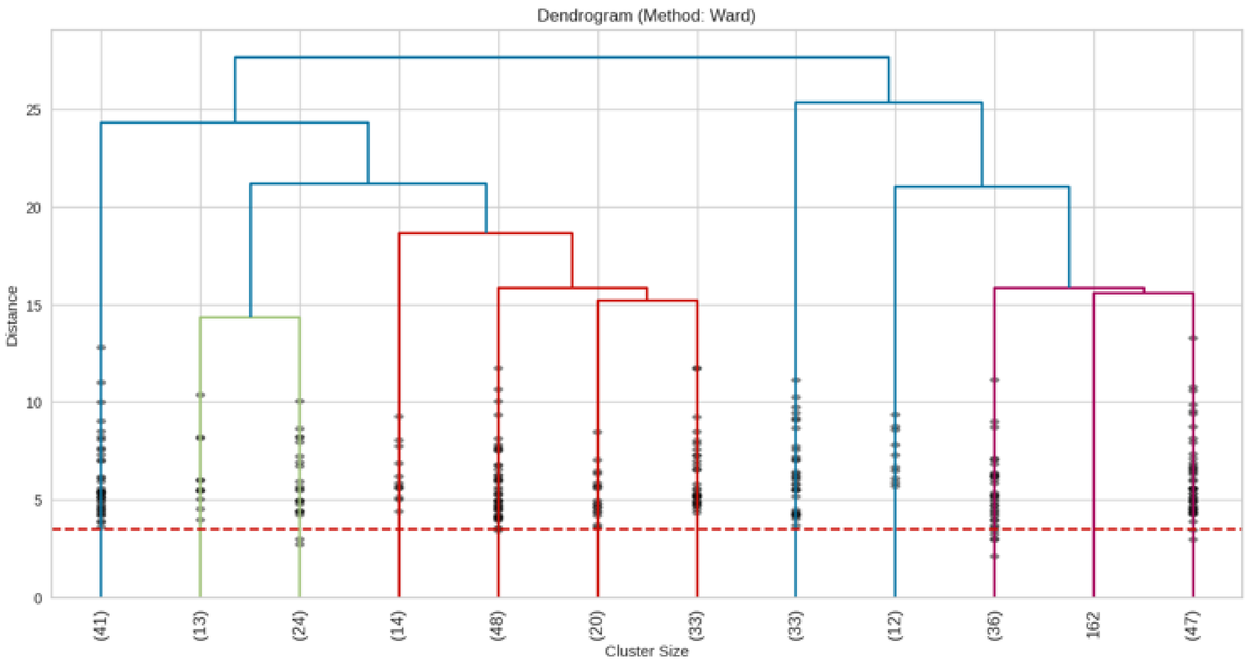

Figure 2. Dendrogram results of Hierarchical Clustering using the Ward method

Table 2. Comparison of clustering algorithm performance

Algorithm	Silhouette Score	DBI	Dunn Index
K-Means	0.045034	2.935749	0.250674
Hierarchical	0.046421	2.956123	0.238539
DBSCAN	0.416001	0.447664	1.429872

3.3 Performance evaluation of clustering algorithms

To evaluate the clustering results, three internal validation metrics were used: Silhouette Score, DBI, and Dunn Index. These metrics assess cohesion and separation within clusters. The consolidated comparison is presented in Table 2, ensuring consistency across methods.

The evaluation results demonstrate that DBSCAN outperforms both K-Means and Hierarchical Clustering, consistent with previous findings showing DBSCAN's robustness in handling complex datasets [24, 25]. With a Silhouette Score of 0.414, DBSCAN shows well-separated clusters, while its DBI of 1.062 indicates compact groupings with clear separation. The Dunn Index of 1.136 further supports the robustness of DBSCAN clusters.

In contrast, Hierarchical Clustering indicates limited performance. Its Silhouette Score (0.046) and Dunn Index (0.239) indicate weak cluster separation and a considerable degree of overlap between clusters, although its performance is marginally better than that of K-Means. Meanwhile, K-Means exhibits the poorest clustering quality, with a very low Silhouette Score (0.045) and the highest Davies–Bouldin Index (2.936), suggesting unclear and overlapping cluster structures. In comparison, DBSCAN clearly outperforms the other algorithms, achieving a substantially higher Silhouette Score (0.416), a low Davies–Bouldin Index (0.448), and the highest Dunn Index (1.430), reflecting well-separated and compact clusters.

From a methodological perspective, this result aligns with computational studies indicating that datasets with high variability require clustering approaches capable of capturing sub-spatial structures and irregular cluster shapes, for which density-based methods such as DBSCAN are particularly effective [26].

Overall, the comparison across the three algorithms highlights the strength of DBSCAN in capturing the complexity of food consumption and lifestyle patterns, particularly due to its ability to detect outliers and adapt to clusters of arbitrary shapes. These findings are consistent with previous studies, such as Alosaimi et al. [11], and are further supported by technical reviews showing DBSCAN's superior adaptability in handling irregular behavioral data [27], which reported that clustering methods can effectively identify combined patterns of diet and physical activity [28]. Similarly, the HELENA Study confirmed that unhealthy dietary habits are often linked with low physical activity levels [11]. Furthermore, evidence from the PeNSE 2015 Survey demonstrated that adolescents with poor diets and insufficient physical activity face an elevated risk of obesity and hypertension from an early age. Together, these results suggest that machine learning-based clustering, especially DBSCAN, provides a more accurate foundation for designing targeted and data-driven health interventions aimed at reducing the risk of NCDs among young adults.

The eight clusters obtained from K-Means were analyzed based on their mean scores across key health indicators, including dietary diversity, physical activity, BMI, smoking habits, sanitation, and healthcare access. Clusters with similar health characteristics were merged into three meaningful groups: healthy, moderate risk, and high risk to enhance interpretability. The same grouping logic was consistently applied to the Hierarchical and DBSCAN outputs to ensure comparable health profile segmentation across algorithms.

4. EVALUATION AND ANALYSIS

To assess the clustering results, three internal validation metrics were applied: Silhouette Score, DBI, and Dunn Index. A higher Silhouette Score indicates better separation between clusters, a lower DBI reflects more compact and well-separated clusters, and a higher Dunn Index denotes greater inter-cluster distance relative to intra-cluster compactness. These metrics collectively provide a comprehensive view of clustering quality.

The results indicate that DBSCAN consistently outperforms the other two algorithms across all three metrics. Its Silhouette Score of 0.414 indicates a reasonably well-defined cluster structure, the DBI of 0.447 confirms compact and distinct clusters, and the Dunn Index of 1.430 highlights strong inter-cluster separation. This implies that DBSCAN is particularly effective for handling heterogeneous data distributions and irregularly shaped clusters often found in behavioral and lifestyle data.

Hierarchical Clustering achieved moderate performance, with a Silhouette Score of 0.164 and Dunn Index of 0.513. These values suggest that while clusters were formed, significant overlap remained between groups. Nevertheless, Hierarchical Clustering performed slightly better than K-Means in separating non-uniform segments [29], making it a secondary but less optimal option compared to DBSCAN.

K-Means produced the lowest clustering quality, as evidenced by a Silhouette Score of 0.137 and a high DBI of 2.936, indicating weak separation and poor compactness. These results confirm that K-Means, which assumes spherical clusters of similar size, is less suitable for datasets with complex or non-linear structures such as food consumption and lifestyle behaviors.

To validate the robustness of these findings, statistical tests were performed across multiple runs (n = 30) for each algorithm. One-way ANOVA revealed statistically significant differences in Silhouette Scores between the algorithms ($p < 0.01$). Post-hoc t-tests confirmed that DBSCAN's performance was significantly higher than both K-Means and Hierarchical Clustering, while the difference between K-Means and Hierarchical was not statistically significant. These results reinforce the superiority of DBSCAN for this dataset and support its use in identifying meaningful consumption and lifestyle clusters.

Taken together, these findings indicate that clustering performance is strongly influenced by the underlying assumptions of each algorithm. Distance-based methods, such as K-Means and Hierarchical Clustering, show limitations when applied to heterogeneous lifestyle data with non-uniform distributions, whereas density-based approaches are more capable of capturing meaningful structures within such data. From a practical perspective, this implies that DBSCAN provides a more reliable foundation for health risk profiling and supports the development of targeted, data-driven interventions aimed at addressing diverse patterns of food consumption and physical activity.

5. CONCLUSIONS

The segmentation revealed three distinct groups: a low-risk cluster characterized by balanced diets and sufficient physical activity, a moderate-risk cluster with high calorie intake but low activity levels, and a high-risk cluster marked by unhealthy eating habits and sedentary lifestyles. For K-Means specifically, the optimal clustering structure was determined at $k = 8$ using the elbow method, which was then consolidated to represent these three clinically meaningful groups. These findings offer a practical framework for public health agencies and policymakers to design targeted interventions. For instance, low-risk groups may benefit from reinforcement programs to maintain healthy behaviors, while moderate- and high-risk groups require structured campaigns, improved access to healthier food options, and integrated preventive services.

Nevertheless, this study has certain limitations. The sample size was relatively small, the reliance on self-reported data may lead to recall errors or social desirability bias that could affect the accuracy of the findings, and issues in age formatting reduced the precision of demographic analysis. These constraints highlight the need for cautious interpretation of the results and suggest that broader datasets with more robust collection methods are necessary to strengthen the findings.

Future research should consider integrating supervised learning models to predict risk levels, employing real-time monitoring systems, and incorporating data from wearable technologies to capture more granular lifestyle behaviors. Such enhancements would improve the accuracy of segmentation and support more personalized, data-driven health interventions to mitigate the growing burden of NCDs.

REFERENCES

[1] Vos, T., Lim, S.S., Abbafati, C., Abbas, K.M., et al. (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. The Lancet, 396(10258): 1204-1222. https://doi.org/10.1016/S0140-6736(20)30925-9

[2] Sudikno, Izwardy, D., Eko, S., Irawan, I.R., Aditianti, Sandjaja. (2021). Prevalence and risk factors of stunting in children aged 0-23 months in Indonesia. In the 8th International Conference on Public Health, Solo, Indonesia, p. 169. http://doi.org/10.26911/ab.maternal.icph.08.2021.43

[3] Al-Mawali, A., Jayapal, S.K., Morsi, M., Al-Shekaili, W., Pinto, A.D., Al-Kharusi, H., Al-Harrasi, A., Al-Balushi, Z., Idikula, J. (2021). Prevalence of risk factors of non-communicable diseases in the Sultanate of Oman: STEPS survey 2017. PLoS One, 16(10): e0259239. https://doi.org/10.1371/journal.pone.0259239

[4] Danaei, G., Andrews, K.G., Sudfeld, C.R., Fink, G., et al. (2016). Risk factors for childhood stunting in 137 developing countries: A comparative risk assessment analysis at global, regional, and country levels. PLoS Medicine, 13(11): e1002164. https://doi.org/10.1371/journal.pmed.1002164

[5] Lozano, R., Naghavi, M., Foreman, K., Lim, S., et al. (2012). Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: A systematic analysis for the Global Burden of Disease Study 2010. The Lancet, 380(9859): 2095-2128.

[6] Feng, Y.F., Sun, D., Sun, X.R., Guo, Q.Q., Zhang, J., Li, Y.Z. (2025). Global burden of noncommunicable diseases attributable to modifiable behavioral risks among adolescents and young adults aged 10–24 years, 1990–2021. BMC Medicine, 23(1): 636. https://doi.org/10.1186/s12916-025-04463-7

[7] Akseer, N., Mehta, S., Wigle, J., Chera, R., Brickman, Z.J., Al-Gashm, S., Sorichetti, B., Vandermorris, A., Hipgrave, D.B., Schwalbe, N., Bhutta, Z.A. (2020). Non-communicable diseases among adolescents: Current status, determinants, interventions and policies. BMC Public Health, 20(1): 1908. https://doi.org/10.1186/s12889-020-09988-5

[8] Firmansyah, Ahsan, E.D., Zulaekah, S., Puspitasari, D.I. (2025). Frequency of students fast food consumption, physical activity, and nutritional status at the Faculty of Health Sciences, Muhammadiyah University of Surakarta. Indonesian Journal of Human Nutrition, 12(1): 41-51. https://doi.org/10.21776/ub.ijhn.2025.012.01.4

[9] Agustina, R., Rianda, D., Setiawan, E.A. (2022). Relationships of child-, parents-, and environment-associated determinants with diet quality, physical activity, and smoking habits among Indonesian urban adolescents. Food and Nutrition Bulletin, 43(1): 44-55. https://doi.org/10.1177/03795721211046145

[10] Prasetyaningsih, E., Duru, E.P., Novitasari, E., Patrisia, I., Surbakti, J.F. (2021). The description of eating patterns and risk for gastritis in students at a private university in Western Indonesia. Nursing Current: Jurnal Keperawatan, 9(1): 48-55. https://doi.org/10.19166/nc.v9i1.3456

[11] Alosaimi, N., Sherar, L.B., Griffiths, P., Pearson, N. (2023). Clustering of diet, physical activity and sedentary behaviour and related physical and mental health outcomes: A systematic review. BMC Public Health, 23: 1572. https://doi.org/10.1186/s12889-023-16372-6

[12] Maugeri, A., Barchitta, M., Favara, G., La Mastra, C., La Rosa, M.C., Magnano San Lio, R., Agodi, A. (2023). The application of clustering on principal components for nutritional epidemiology: A workflow to derive dietary patterns. Nutrients, 15(1): 195. https://doi.org/10.3390/nu15010195

[13] López-Gil, J.F., Martínez-López, M.F. (2024). Clustering of dietary patterns associated with health-related quality of life in Spanish children and adolescents. Nutrients, 16(14): 2308. https://doi.org/10.3390/nu16142308

[14] Cacau, L.T., Hanley-Cook, G.T., Vandevijvere, S., Leclercq, C., et al. (2024). Association between adherence to the EAT-Lancet sustainable reference diet and cardiovascular health among European adolescents: The HELENA Study. European Journal of Clinical Nutrition, 78: 202-208. https://doi.org/10.1038/s41430-023-01379-4

[15] Maneschy, I., Moreno, L.A., Ruperez, A.I., Jimeno, A., et al. (2022). Eating behavior associated with food intake in European adolescents participating in the HELENA Study. Nutrients, 14(15): 3033. https://doi.org/10.3390/nu14153033

[16] Maugeri, A., Barchitta, M., Favara, G., La Mastra, C., La Rosa, M.C., Magnano San Lio, R., Agodi, A. (2023). The application of clustering on principal components for

nutritional epidemiology: A workflow to derive dietary patterns. Nutrients, 15(1): 195. https://doi.org/10.3390/nu15010195

[17] Deng, D.S. (2020). DBSCAN clustering algorithm based on density. In 2020 7th International Forum on Electrical Engineering and Automation (IFEEA), Hefei, China, pp. 949-953. https://doi.org/10.1109/IFEEA51475.2020.00199

[18] de Moura Ventorim, I., Luchi, D., Rodrigues, A.L., Varejão, F.M. (2021). BIRCHSCAN: A sampling method for applying DBSCAN to large datasets. Expert Systems with Applications, 184: 115518. https://doi.org/10.1016/j.eswa.2021.115518

[19] Herdiana, I., Kamal, M.A., Triyani, T., Estri, M.N., Renny, R. (2025). A more precise elbow method for optimum K-means clustering. arXiv preprint arXiv:2502.00851. https://doi.org/10.48550/arXiv.2502.00851

[20] Nainggolan, R., Perangin-angin, R., Simarmata, E., Tarigan, A.F. (2019). Improved the performance of the K-Means Cluster using the Sum of Squared Error (SSE) optimized by using the elbow method. Journal of Physics: Conference Series, 1361: 012015. https://doi.org/10.1088/1742-6596/1361/1/012015

[21] Syakur, M., Khotimah, B.K., Rochman, E.M.S., Satoto, B.D. (2018). Integration K-Means clustering method and elbow method for identification of the best customer profile cluster. IOP Conference Series: Materials Science and Engineering, 336: 012017. https://doi.org/10.1088/1757-899X/336/1/012017

[22] Theotista, G., Febe, M., Ryan, M.S. (2025). Analysing market dynamics: Revealing obscured patterns in LQ45 stocks (2021-2023) using Ward'S hierarchical clustering. Barekeng: Jurnal Ilmu Matematika dan Terapan, 19(1): 163-172. https://doi.org/10.30598/barekengvol19iss1pp163-172

[23] Chen, F.Y. (2021). An improved DBSCAN algorithm for adaptively determining parameters in multi-density environment. In 2021 2nd International Conference on Artificial Intelligence and Information Systems, Chongqing, China, pp.1-4. https://doi.org/10.1145/3469213.3470400

[24] Jeena, S., Chaudhary, A., Thakur, A. (2023). Implementation & analysis of online retail dataset using clustering algorithms. In 2023 4th International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, pp. 1-6. https://doi.org/10.1109/ICIEM59379.2023.10166552

[25] Oliveira, M., Marçal, A.R.S. (2023). Clustering LiDAR data with K-means and DBSCAN. In Proceedings of the 12th International Conference on Pattern Recognition Applications and Methods ICPRAM, Lisbon, Portugal, pp. 822-831. https://doi.org/10.5220/0011667000003411

[26] Thejaswini, M.S., Kumar, G.H., Aradhya, V.N.M. (2025). Hybrid dimensionality reduction model for real-time EEG-based emotion recognition: A combined subspace approach using principal component analysis and independent component analysis. Mathematical Modelling of Engineering Problems, 12(5): 1771-1788. https://doi.org/10.18280/mmep.120532

[27] Kulkarni, O., Burhanpurwala, A. (2024). A survey of advancements in DBSCAN clustering algorithms for big data. In 2024 3rd International Conference on Power Electronics and IoT Applications in Renewable Energy and Its Control (PARC), Mathura, India, pp. 106-111. https://doi.org/10.1109/PARC59193.2024.10486339

[28] Ottevaere, C., Huybrechts, I., Benser, J., De Bourdeaudhuij, I., et al. (2011). Clustering patterns of physical activity, sedentary and dietary behavior among European adolescents: The HELENA study. BMC Public Health, 11: 328. https://doi.org/10.1186/1471-2458-11-328

[29] López, Y.G., Vega, J.T.V., Rosillo, F.F., Alayo, E.M.C., Ríos, M.A.C., Huatangari, L.Q., Mendoza, M.M. (2024). Predicting the shelf life of cup chocolate using the Arrhenius model based on peroxide value. Mathematical Modelling of Engineering Problems, 11(2): 517-522. https://doi.org/10.18280/mmep.110224