Department of Digital Business
Journal of Artificial Intelligence and Digital Business (RIGGS)
Homepage: https://journal.ilmudata.co.id/index.php/RIGGS
Vol. 4 No. 2 (2025) pp: 937-944
P-ISSN: 2963-9298, e-ISSN: 2963-914X

Linkage Comparison in Agglomerative Hierarchical Clustering for Clustering Students' Knowledge of First Aid for Stroke Emergencies

Ivana Aikozora Putrinugroho, Mochammad Anshori, Wahyu Teja Kusuma
[1,2,3]Department of Informatics, Faculty of Science and Technology, Institut Teknologi, Sains, dan Kesehatan RS DR Soepraoen Kesdam V/BRW, Malang, Indonesia
[1]aikozoraivana@gmail.com, [2]moanshori@itsk-soepraoen.ac.id*, [3]wtkusuma@itsk-soepraoen.ac.id

Abstract

Stroke is a leading cause of disability and mortality worldwide, necessitating immediate and accurate first aid to mitigate severe outcomes. In Indonesia, limited public knowledge about stroke management, particularly among high school students, underscores the urgent need for targeted educational interventions. This study aims to evaluate students' understanding of stroke first aid and identify optimal methods for clustering educational data using Agglomerative Hierarchical Clustering (AHC). A validated questionnaire was distributed to 112 high school students, focusing on their knowledge of stroke symptoms, risk factors, and first-aid practices. Data preprocessing ensured quality and consistency before applying AHC with three linkage methods: Single Linkage, Complete Linkage, and Ward's method. The results were evaluated using Davies-Bouldin Index and Silhouette Coefficient to determine the most effective clustering approach. Ward's method outperformed other linkage methods, achieving superior cluster compactness and separation. Four clusters were identified, representing varying levels of knowledge, from basic understanding to high awareness of stroke and seizure management. These findings provide a foundation for designing tailored educational programs, addressing specific knowledge gaps, and enhancing firstaid preparedness. This study demonstrates the utility of machine learning in educational research and contributes to improving public health education. Future research should expand on these findings by incorporating diverse datasets and alternative clustering algorithms.

Keywords: *stroke, first aid, AHC, Single Linkage, Complete Linkage, Ward*

1. Introduction

Stroke is a critical medical condition that disrupts the blood flow to the brain, resulting in significant morbidity and mortality globally. According to the World Health Organization, stroke is the second leading cause of death worldwide, following ischemic heart disease, and it contributes substantially to long-term disability[1]. In Indonesia, stroke is the foremost cause of mortality, with 2019 data indicating an incidence rate of approximately 100 cases per 100,000 population [2]. The consequences of stroke, including physical, cognitive, and emotional impairments, place a significant burden on healthcare systems, families, and individuals. Early intervention through appropriate first aid during a stroke event has the potential to mitigate the severity of outcomes, emphasizing the importance of public awareness and education regarding effective first aid measures.

Stroke is generally categorized into ischemic and hemorrhagic types. Ischemic stroke, accounting for approximately 87% of all cases, occurs due to a blockage in the blood vessels supplying the brain. In contrast, hemorrhagic stroke results from the rupture of a blood vessel, leading to intracranial bleeding [3]. Both types of strokes necessitate immediate medical attention, and in the interim, first aid can significantly influence survival rates and recovery outcomes. The lack of public understanding about stroke symptoms—such as asymmetric facial drooping, sudden weakness, speech difficulties, and severe headache—as well as misconceptions about first aid measures, pose critical challenges [4]. Misguided practices, including inserting objects into a stroke victim's mouth or applying inappropriate interventions, can exacerbate the victim's condition [5]. Addressing these knowledge gaps is essential, particularly in communities with limited access to immediate professional medical assistance.

Despite the critical need for timely and accurate first aid during stroke emergencies, significant delays in response remain prevalent. Studies reveal that 83.9% of delays in providing first aid are attributable to insufficient knowledge among caregivers or bystanders [5]. Furthermore, educational deficits often result in harmful interventions. Research has demonstrated a correlation between education level and awareness of proper first aid techniques, underscoring the value of targeted educational initiatives [6]. High school students, given their formative stage of learning, represent a crucial demographic for disseminating stroke awareness and first aid training. Integrating such educational efforts into high school curricula could foster a generation better equipped to handle medical emergencies, particularly stroke-related incidents.

The development and evaluation of public knowledge about stroke first aid require robust analytical methodologies. Machine learning, particularly clustering techniques, offers a viable approach to categorize and assess knowledge levels among target populations. Agglomerative Hierarchical Clustering (AHC) is a prominent unsupervised learning algorithm that organizes data into hierarchies based on similarity metrics. This method's application has been explored across diverse domains, including healthcare, education, and public policy. For instance, AHC has been employed in physical fitness assessments in police academies, revealing its capacity to handle complex datasets [12]. Additionally, regional health policy planning in East Java utilized AHC to inform decisions on resource allocation and public health initiatives [7].

The success of AHC largely depends on the linkage method used to determine proximity between clusters. Common linkage methods include Single Linkage, which minimizes the distance between clusters; Complete Linkage, which maximizes the separation between clusters; and Ward's method, which minimizes the variance within clusters. Each method offers unique advantages and challenges. For example, Single Linkage is prone to chaining effects, which can distort clustering outcomes in noisy datasets [8]. Complete Linkage often produces distinct clusters but may overlook nuanced similarities between data points. In contrast, Ward's method emphasizes homogeneity within clusters, making it particularly suitable for balanced datasets [9]. Recent literature highlights the utility of AHC in educational contexts. Studies have applied AHC to classify students' physical fitness and academic performance, demonstrating its effectiveness in identifying distinct groups based on shared characteristics [10], [11]. However, limited research has focused on its application in assessing knowledge about health emergencies, such as stroke first aid. This gap presents an opportunity to explore AHC's potential in educational interventions aimed at improving public health outcomes.

Within the scope of stroke education, prior studies have examined various pedagogical and technological approaches to enhancing knowledge dissemination. For example, interactive workshops and digital platforms have been shown to significantly improve participants' understanding of stroke symptoms and first aid procedures [12]. However, the integration of machine learning techniques, particularly clustering methods like AHC, remains underexplored. While existing studies have validated AHC's efficacy in different domains, its application to health education—specifically in clustering students' knowledge about stroke first aid—is nascent.

This study aims to address these gaps by leveraging AHC to evaluate and categorize high school students' knowledge about stroke first aid. By comparing three linkage methods—Single Linkage, Complete Linkage, and Ward—this research seeks to identify the most effective approach for clustering knowledge levels. The findings will contribute to the development of targeted educational programs, enhancing students' preparedness to respond to stroke emergencies. Moreover, this study's novelty lies in its application of AHC in an educational context, providing insights into the algorithm's utility for public health education. The scope of this research includes assessing students' knowledge through validated questionnaires, preprocessing the data to ensure reliability, and applying rigorous evaluation metrics such as the Davies-Bouldin Index and Silhouette Coefficient to compare clustering outcomes. By addressing these objectives, the study aims to advance both the methodological and practical dimensions of stroke first aid education.

2. Research Methods

This study involves several processes as shown in the image below:

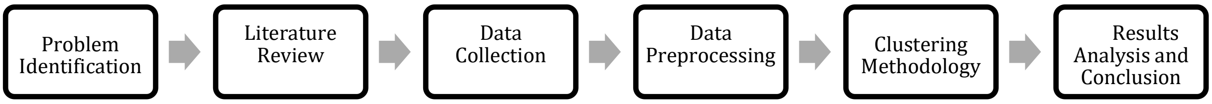

Figure 1. Research Methodology

Figure 1 above represents the methodology of the research that has been conducted. It consists of several stages, including problem identification, literature review, data collection through interviews and questionnaires, data preprocessing, data processing using clustering method, results analysis, and conclusion.

DOI: https://doi.org/10.31004/riggs.v4i2.564

a. Problem Identification
The study begins with identifying the problem of insufficient knowledge about stroke first aid among high school students. Interviews with medical experts and a literature review revealed that delays in administering first aid contribute significantly to adverse outcomes in stroke cases [8]. Furthermore, misconceptions about first aid practices were identified, emphasizing the need for targeted educational interventions. The problem identification phase underscores the critical need to evaluate and improve the knowledge levels of students regarding stroke emergencies.

b. Literature Review
The literature review stage is a method of gathering information from various scholarly sources. In this stage, the researcher collects information from various journals on relevant topics. During this stage, the researcher identifies research gaps, which serve as the initial step for the research being conducted. This is done to obtain references and information for the research.

c. Data Collecting
Primary data for this study were collected through a validated questionnaire distributed to students at SMAN 2 Pasuruan. The questionnaire, developed in consultation with stroke specialists, comprised 20 items covering topics such as stroke definition, symptoms, first aid steps, and risk factors. The questions also addressed specific scenarios, such as recognizing and managing seizures associated with stroke. Before distribution, the questionnaire underwent expert validation to ensure its relevance and clarity. Data collection yielded 114 responses, which were subsequently preprocessed to remove duplicate entries and irrelevant attributes, reducing the dataset to 112 valid responses. The questionnaire consists of 20 questions shown at Table 1 below.

Tabel 1. List of questionnaires

No	Question
1	Risk Factors
2	Management of Seizures
3	Importance of Managing Seizures
4	Level of Importance of First Aid
5	First Aid Steps
6	Pre-Hospitals Steps
7	Relationship Between Seizures and Stroke
8	First Aid Training
9	Curiosity About First Aid
10	Risk Factors
11	Stroke Symptoms
12	Encountering a Patient
13	Sign of Seizures
14	Actions During a Seizure
15	Stroke Information in Schools
16	Time of Assistance
17	Knowledge of Seizures
18	Steps for Seizure Assistance
19	Confidence in Providing First Aid
20	Suggestions for Education

Based on Table 1, there are 20 questions on the topics of stroke, education, and knowledge regarding the first aid management for stroke patients. These questions will be used as part of a questionnaire to be filled out by the respondents.

d. Data Preprocessing

DOI: https://doi.org/10.31004/riggs.v4i2.564

Data preprocessing involved multiple steps to prepare the dataset for clustering analysis. First, duplicate responses were identified and removed to ensure data integrity. Attributes such as respondent names, timestamps, and email addresses were excluded to maintain respondent anonymity and focus on relevant variables. Data type formatting was applied to standardize the representation of categorical and numerical attributes. These preprocessing steps were critical to ensuring the reliability and accuracy of the clustering results[13].

e. Clustering Methodology

The study employed Agglomerative Hierarchical Clustering (AHC), an unsupervised machine learning technique, to group respondents based on their knowledge levels. AHC operates by iteratively merging data points into clusters based on proximity metrics. The proximity matrix was calculated using the Euclidean distance formula, ensuring an accurate representation of similarities and differences among data points [21]. The formula for calculating the distance is as follows Equation (1):

$$\mathrm{d}(x, y) = \sqrt{\sum_{i=1}^{n} (x - y)^2} \quad (1)$$

where $d(x, y)$ is distance between x and y data points.

In determining the most optimal number of clusters in hierarchical clustering, you need to look at the longest hierarchical line, this is called linkage[14].Three linkage methods were compared:

Single Linkage: This method minimizes the distance between the closest points of two clusters. While efficient, it is prone to the chaining effect, leading to elongated clusters [9].

$$d(xy)z = min(dx{,}z;\ dy{,}z) \quad (2)$$

Equation (2) is the formula of single linkage The distance between the combined cluster (XY) and cluster Z is the minimum value of the distances between every point in X and Z, and every point in Y and Z.

Complete Linkage: This approach maximizes the distance between the farthest points of two clusters, resulting in more compact and distinct clusters [9].

$$d(xy)z = max(dx{,}z;\ dy{,}z) \quad (3)$$

Equation (3) is the formula of complete linkage The distance between the combined cluster (XY) and cluster Z is the maximum value of the distances between every point in X and Z, and every point in Y and Z.

Ward's Method: By minimizing the increase in variance within clusters during merging, this method emphasizes intra-cluster homogeneity [15].

$$ESS = \sum_{i=1}^{n} \|x_i - \mu\|^2 \quad (4)$$

Equation (4) is the formula of ward linkage where n shows amount data in a cluster, x is data and μ is means inside a cluster.

Two evaluation metrics were used to assess the quality of clustering results:

- Davies-Bouldin Index (DBI): DBI measures the ratio of intra-cluster distances to inter-cluster distances[14]. Lower DBI values indicate better clustering quality, with distinct and compact clusters [16].

  $$DBI = \frac{1}{n} \Sigma_{i=1}^{n} = \frac{max}{x \neq y} \left(\frac{S_x + S_y}{d_{xy}} \right) \quad (5)$$

  Equation (5) shows the formula of DBI, where n is num of cluster, S_x, S_y is average distance x and y; and d_{xy} is distance between center of cluster x and y.
- Silhouette Coefficient (SC): SC evaluates how similar an object is to its cluster compared to other clusters. Values range from -1 to 1, with higher values indicating better-defined clusters[17].

  $$SC = \frac{y - x}{\max xy} \quad (6)$$

  Equation (6) shows the formula of DBI, where x is the average distance between a data point and other points in a cluster (cohesion) and y is Average distance between a data point and a point in the nearest cluster (separation).

After applying AHC using each linkage method, the resulting clusters were evaluated using DBI and SC. The analysis focused on identifying the optimal number of clusters and the linkage method that produced the most meaningful grouping.

DOI: https://doi.org/10.31004/riggs.v4i2.564

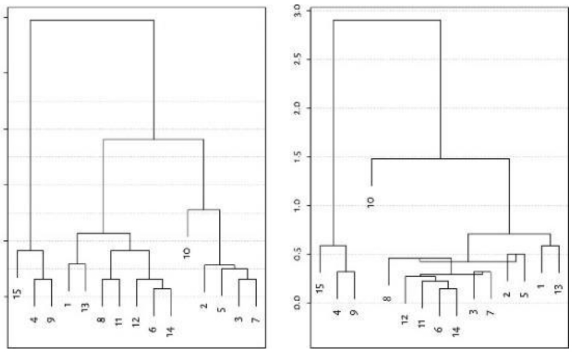

Figure 2. Visualization of AHC with Dendrogram

Visualizations, including dendrograms, were generated to interpret the hierarchical relationships among clusters. Each linkage method's strengths and limitations were analyzed to determine its suitability for the dataset. Figure 2 is a visualization of the dendrogram. In the dendrogram, each node represents a cluster, while the leaf nodes are referred to as single clusters [18][19].

The final stage of the methodology involved labeling the clusters based on their characteristics. The analysis identified distinct patterns in students' responses, categorizing them into levels such as high understanding of stroke and seizures, moderate awareness of risk factors, and minimal knowledge of first aid. These labels provided actionable insights for designing targeted educational interventions. Summary This study's methodological design integrates robust data collection, preprocessing, clustering, and evaluation techniques to achieve its objectives. The comparative analysis of AHC linkage methods and the use of standardized evaluation metrics ensure the reliability and validity of the findings. By addressing knowledge gaps and enhancing first aid education, the study contributes to improving public health outcomes for stroke emergencies.

3. Results and Discussions

The initial dataset consisted of 114 responses collected from students at SMAN 2 Pasuruan. After data preprocessing, which included removing duplicate entries and incomplete responses, the dataset was refined to 112 valid responses. Preprocessing steps, such as duplicate removal, data formatting, and attribute selection, ensured the data's quality and relevance for clustering analysis. The finalized dataset served as the basis for implementing Agglomerative Hierarchical Clustering (AHC).

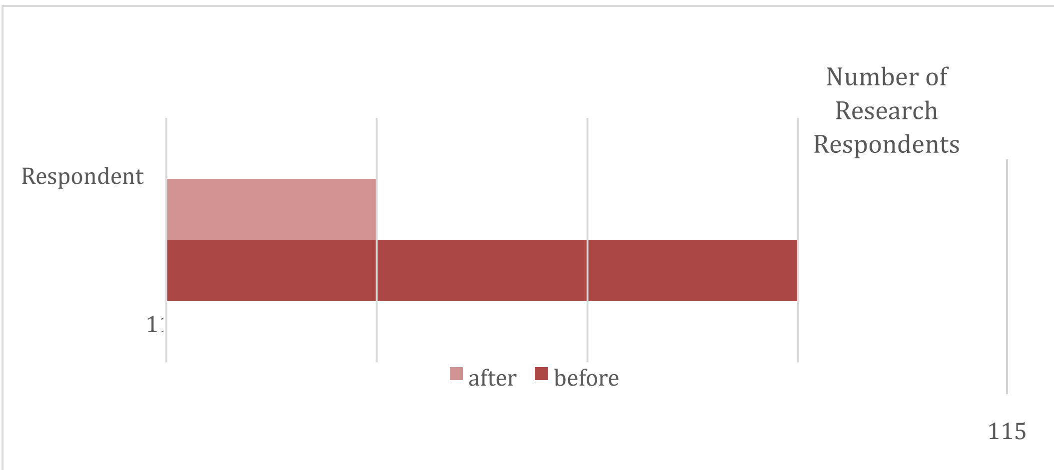

Figure 3. Comparison of the Number of Respondents Before and After Data Preprocessing

Referring to Figure 3 above, the number of respondents in this study reached 114 students. However, after data preprocessing, including data cleaning such as removing duplicate data, the number of respondents decreased to 112 students. This reduction resulted in a decrease of 2 students. The data with 112 students will now undergo clustering using the AHC method.

To determine the best cluster, the researcher conducted a comparison of linkage methods in AHC. Three linkage methods—Single Linkage, Complete Linkage, and Ward—were applied to the dataset using the AHC algorithm.

DOI: https://doi.org/10.31004/riggs.v4i2.564

Each method generated distinct clusters, which were visualized through dendrograms to better understand the grouping patterns. The results are summarized below

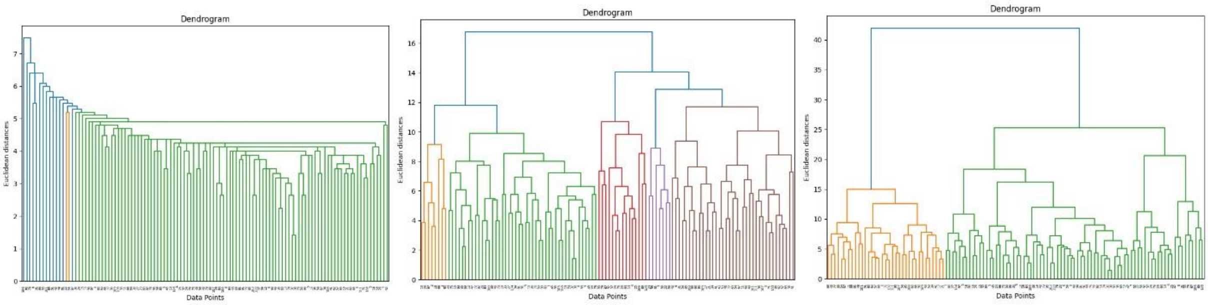

Figure 4. Dendrogram of single linkage (left), complete linkage (center), ward (right)

Figure 4 (left), the dendrogram for Single Linkage revealed elongated clusters caused by the chaining effect, where data points were sequentially linked based on the minimum distances between clusters. This method resulted in less compact clusters, which reduced the interpretability and effectiveness of the grouping.

Figure 4 (center), the Complete Linkage method produced more distinct clusters with clearly separated boundaries. The use of maximum distances during the clustering process prevented premature merging of distant data points, resulting in better-defined clusters. However, this approach still exhibited slight imbalances in cluster sizes.

Figure 4 (right), Ward's method demonstrated the most balanced and homogeneous clustering outcomes. By minimizing the total within-cluster variance, this approach produced compact clusters with consistent sizes. This method's dendrogram showcased stable clustering patterns, making it the most suitable choice for the given dataset.

The clustering outcomes were evaluated using the Davies-Bouldin Index (DBI) and Silhouette Coefficient (SC), as shown in Table 2. These metrics provided insights into the compactness and separation of the clusters formed by each linkage method.

Table 2. Comparison of linkage methods

n_cluster	Single Linkage		Complete Linkage		Ward	
	DBI	SC	DBI	SC	DBI	SC
2	0.6987	0.1807	2.010	0.1763	1.5972	0.2120
3	0.6720	0.1514	1.855	0.1961	1.9517	0.2073
4	0.6546	0.1379	1.7555	0.1984	1.5726	0.2136
5	0.7694	0.1245	1.7030	0.1922	1.5506	0.2093

Based on Table 2, the DBI values obtained from the single linkage, complete linkage, and ward methods are all less than optimal. The values are far from 0, indicating that they are not optimal. The comparison results show that the number of clusters 4 in Single Linkage produced the lowest value compared to other cluster numbers, which is 0.6546. This indicates that clustering with 4 clusters provides better results in terms of greater inter-cluster distance and tighter data distribution within clusters. On the other hand, the number of clusters 2 in Complete Linkage has the highest DBI value, which is 2.010, indicating that clustering with 2 clusters has the lowest quality compared to the other cluster numbers and methods.

In addition, the calculations with the Silhouette Coefficient also produced less than optimal values. The values obtained are less than 1, indicating that they are not optimal. Among the three comparisons, the Ward method has the highest value compared to the Single Linkage and Complete Linkage methods. This value is found with 4 clusters, which is 0.2136. This value is higher compared to other cluster numbers and methods. This indicates that the points within the clusters are more distinct from other clusters. Meanwhile, the value for the Single Linkage method with 5 clusters, which is 0.1245, is lower than both other methods.

DOI: https://doi.org/10.31004/riggs.v4i2.564

With the calculations to evaluate the AHC linkage methods, namely DBI and SC, the single linkage, complete linkage, and ward methods all have less than optimal values for both. However, for the evaluation with DBI and SC, the ward method performs better than the single linkage and complete linkage methods. The comparison of the three linkage methods can be seen in the table below:

Tabel 3. Comparison of the best evaluation on each lingkage of AHC

Linkage Method	DBI Evaluation	n_cluster DBI	SC Evaluation	n_cluster SC
Single Linkage	0.6546	4	0.1807	2
Complete Linkage	1.7030	5	0.1984	4
Ward	1.5506	5	0.2136	4

Based on Table 3, the DBI and SC evaluation values for the Ward method are better compared to single linkage and complete linkage. This is because the values obtained from the evaluations using both DBI and SC do not show any poor results. It can also be seen in the Ward method dendrogram, which displays more balanced cluster sizes.

After reviewing the evaluation results, the researcher determined the number of clusters based on the evaluation using SC. This is because the results from the SC evaluation were more optimal compared to the DBI evaluation. It is considered more optimal because the difference between the optimal SC value and the evaluation results of this study is smaller than the difference in the evaluation using DBI. Therefore, the number of clusters determined in this study is 4 clusters. Subsequently, an analysis was performed to assign labels to each cluster. Labeling was done by identifying the characteristics of each cluster. This analysis is based on the patterns within the clusters, such as students' understanding of stroke, understanding of first aid for seizures, and awareness of stroke risk factors. These labels help the researcher provide educational recommendations for stroke first aid to students.

Table 4. Number of cluster data

Cluster	**Cluster Category**	**Number of Data**
1	High Understanding of Stroke and Seizures	55
2	High Understanding of Stroke	35
3	Medium Understanding of Stroke and Risk	13
4	Basic Understanding of Seizures	9

From Table 4, it can be seen that the amount of data in the cluster with a high understanding of stroke and seizures is larger compared to the amount of data in other clusters. However, it is still expected to provide additional material for first aid to complement the students' knowledge and involve students in this cluster in group activities with students from other clusters. This way, they will receive more education on stroke first aid.

4. Conclusion

This study evaluated high school students' knowledge of first aid for stroke and compared linkage methods in Agglomerative Hierarchical Clustering (AHC) to determine the optimal approach for educational data analysis. Among the methods evaluated, Ward's method demonstrated superior performance, producing compact, homogeneous clusters with a Davies-Bouldin Index (DBI) of 1.5506 and a Silhouette Coefficient (SC) of 0.2136. These results highlight Ward's suitability for datasets requiring balanced groupings, making it a valuable tool for categorizing educational needs. The analysis revealed four distinct clusters representing varying levels of knowledge about stroke and seizures. These clusters provide actionable insights for designing targeted educational interventions. For instance, students with basic knowledge could benefit from visual aids and hands-on training, while those with higher knowledge could participate in peer education programs. Such tailored approaches have the potential to enhance first-aid preparedness and public health outcomes. By integrating machine learning into educational research, this study contributes to the broader application of clustering algorithms in health education. Future research should explore diverse datasets and alternative clustering

DOI: https://doi.org/10.31004/riggs.v4i2.564

techniques to validate these findings and expand their applicability. Overall, this study underscores the importance of targeted education in improving stroke first-aid literacy and fostering community resilience.

Acknowledgment

All praise and gratitude to God Almighty for His blessings and love that are always bestowed upon me, allowing me to complete this research. I would like to express my gratitude to my mother, brother, and sister who have always provided unwavering support. To my supervising lecturers, Vera, Rosyida, Furqon, and Zidan, thank you for the companionship and assistance provided throughout this process. Lastly, to my father in heaven, thank you for the unforgettable experiences that made this research possible, even though we are now separated by the world.

Reference

[1] R. S. Segall and S. Takashashi, "Data Visualization of Big Data for Predictive and Descriptive Analytics for Stroke, COVID-19, and Diabetes," *Int. J. Big Data Anal. Healthc.*, vol. 8, no. 1, 2023, doi: 10.4018/ijbdah.331996.

[2] W. H. Organisation, "WHO: Stroke, Cerebrovascular accident," 2019.

[3] Y. Haiga, I. Prima Putri Salman, and S. Wahyuni, "Perbedaan Diagnosis Stroke Iskemik dan Stroke Hemoragik dengan Hasil Transcranial Doppler di RSUP Dr. M. Djamil Padang," *Sci. J.*, vol. 1, no. 5, 2022, doi: 10.56260/sciena.v1i5.72.

[4] I. Rahmawati and L. M. Palupi, "DETEKSI DINI GEJALA STROKE SEBAGAI BAGIAN DARI PROGRAM BINA KELUARGA LANSIA DI KABUPATEN MALANG," *LINK*, vol. 16, no. 1, 2020, doi: 10.31983/link.v16i1.5669.

[5] T. Wahyuni and R. N. Martawinarti, "HUBUNGAN PENGETAHUAN KELUARGA TERHADAP PERTOLONGAN PERTAMA PADA PASIEN STROKE DIKUMPEH," *J. Ilmu-Ilmu Kesehat.*, vol. 9, no. 1, 2023, doi: 10.52741/jiikes.v9i1.62.

[6] D. Febryani, E. Rosalina S, and W. H. Susilo, "Hubungan Antara Pengetahuan, Usia, Tingkat Pendidikan Dan Pendapatan Kepala Keluarga Dengan Perilaku Hidup Bersih Dan Sehat Pada Tatanan Rumah Tangga Di Kecamatan Kalideres Jakarta Barat," *Carolus J. Nurs.*, vol. 3, no. 2, 2021, doi: 10.37480/cjon.v3i2.74.

[7] A. Septianingsih, "PEMETAAN KABUPATEN KOTA DI PROVINSI JAWA TIMUR BERDASARKAN TINGKAT KASUS PENYAKIT MENGGUNAKAN PENDEKATAN AGGLOMERATIF HIERARCHICAL CLUSTERING," *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 3, no. 2, 2022, doi: 10.46306/lb.v3i2.139.

[8] K. B, "A Comparative Study on K-Means Clustering and Agglomerative Hierarchical Clustering," *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 5, 2020, doi: 10.30534/ijeter/2020/20852020.

[9] A. M. Jarman, "Hierarchical Cluster Analysis: Comparison of Single Linkage, Complete Linkage, Average Linkage and Centroid Linkage," *Int. J. Georg. Soutern Univ.*, vol. 1, no. 1, 2020.

[10] Z. Muttaqin, "IMPLEMENTASI UNSUPERVISED LEARNING PADA NILAI JASMANI KESAMAPTAAN SEKOLAH POLISI NEGARA DENGAN METODE CLUSTERING ANALYSIS," *PROSISKO J. Pengemb. Ris. dan Obs. Sist. Komput.*, vol. 10, no. 1, 2023, doi: 10.30656/prosisko.v10i1.6269.

[11] A. Saxena *et al.*, "A review of clustering techniques and developments," *Neurocomputing*, vol. 267, 2017, doi: 10.1016/j.neucom.2017.06.053.

[12] E. Burghardt, D. Sewell, and J. Cavanaugh, "Agglomerative and divisive hierarchical Bayesian clustering, *Comput. Stat. Data Anal.*, vol. 176, 2022, doi: 10.1016/j.csda.2022.107566.

[13] M. M. Muttaqin, Wahyu Wijaya Widiyanto, A. W. Green Ferry Mandias, Stenly Richard Pungus, S. A. H. Wiranti Kusuma Hapsari, E. F. B. Aslam Fatkhudin, Pasnur, and N. S. Mochammad Anshori, Suryani, *Pengenalan Data Mining*, no. July. 2023.

[14] T. Pipit Muliyah, Dyah Aminatun, Sukma Septian Nasution, Tommy Hastomo, Setiana Sri Wahyuni Sitepu, *Data Science dan Pembelajaran Mesin*, vol. 7, no. 2. 2020.

[15] V. Vijaya, S. Sharma, and N. Batra, "Comparative Study of Single Linkage, Complete Linkage, and Ward Method of Agglomerative Clustering," in *Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing: Trends, Prespectives and Prospects, COMITCon 2019*, 2019. doi: 10.1109/COMITCon.2019.8862232.

[16] Mardhiatul Azmi, Atus Amadi Putra, Dodi Vionanda, and Admi Salma, "Comparison of the Performance of the K-Means and K-Medoids Algorithms in Grouping Regencies/Cities in Sumatera Based on Poverty Indicators," *UNP J. Stat. Data Sci.*, vol. 1, no. 2, 2023, doi: 10.24036/ujsds/vol1-iss2/25.

[17] H. Belyadi and A. Haghighat, "Unsupervised machine learning: clustering algorithms," *Mach. Learn. Guid. Oil Gas Using Python*, pp. 125–168, Jan. 2021, doi: 10.1016/B978-012-821929-4.00002-0.

[18] A. Serra and R. Tagliaferri, "Unsupervised Learning: Clustering," *Encycl. Bioinforma. Comput. Biol. ABC Bioinforma.*, vol. 1–3, pp. 350–357, Jan. 2019, doi: 10.1016/B978-012-809633-8.20487-1.

[19] M. Gagolewski, A. Cena, S. James, and G. Beliakov, "Hierarchical clustering with OWAbased linkages, the Lance–Williams formula, and dendrogram inversions," *Fuzzy Sets Syst.*, vol. 473, p. 108740, Dec. 2023, doi: 10.1016/J.FSS.2023.108740.

DOI: https://doi.org/10.31004/riggs.v4i2.564