Comparison of the Application of Weighted Cosine Similarity and Minkowski Distance Similarity Methods in Stroke Diagnostic Systems

Joko Purwadi[1], Rosa Delima[2], Argo Wibowo[3], Angelina Rumuy[4]
Informatics Department, Universitas Kristen Duta Wacana, Yogyakarta, Indonesia[1, 2, 4]
Information System, Universitas Kristen Duta Wacana, Yogyakarta, Indonesia[3]

***Abstract*—Stroke is a critical medical condition requiring prompt intervention due to its multifaceted symptoms and causes influenced by various factors, including psychological aspects and the patient's lifestyle or daily habits that impact risk factors. The recovery process involves consistent medical care and lifestyle adjustments tailored to the individual case. Expert Systems, a scientific field focused on studying and developing diagnostic systems, can employ the Case-based Reasoning method to identify the type of stroke based on similarities with prior patient cases, considering specific causes and symptoms. This study utilizes the Weighted Cosine, Jaccard Coefficient, and Minkowski Distance methods to assess the similarity of stroke cases. The evaluation is based on input data such as patient causes or symptoms and risk factors from medical records. The analysis of case similarity and solutions involves applying the Weighted Cosine, Jaccard Coefficient, and Minkowski Distance methods, with a defined threshold value. The highest similarity values from previous patient cases are selected for each method. The test outcomes suggest that employing the Minkowski Distance method with a threshold value of 75 and an r value of three or four yields the highest levels of accuracy, recall, and precision. The Minkowski Distance achieves an accuracy and recall rate of more than 88 percent with 100 percent precision.**

Keywords*—Expert system; stroke; case-based reasoning; Minkowski Distance; jaccard coefficient; weighted cosine; threshold; accuracy; diagnosis

I. Introduction

Stroke is an emergency disease that must be treated immediately to minimize brain damage due to lack of oxygen and nutrients. Stroke can cause paralysis and death for patients. According to [1], stroke is an non-communicable condition arising from blockage, constriction, or hemorrhage within the brain's blood vessels, resulting in a diminished blood flow to the brain.

The development of information technology specifically for intelligent systems or artificial intelligence (AI) has had a positive impact on progress in the field of medicine and health. One of the applications of intelligent systems to support disease diagnosis is the development of an Expert System. According to reference [2][3], the expert system is a system based on knowledge that utilizes expert knowledge to address a particular issue. The essential components required for constructing the expert system include a knowledge base, inference engine, working memory, and user. There has been a number of systems developed in the medical field including [4] [5] [6] [7] [8] [9] [10] [11].

The utilization of techniques in creating the expert system is highly varied, with one example being the Case-Based Reasoning (CBR) approach. CBR involves retrieving cases from past occurrences, subsequently reusing and adapting them in new situations. [12] [13].

In the academic realm, expert systems play a crucial role in the learning process, especially in the field of stroke-related studies. The medical histories of individuals who have experienced strokes can be employed as a point of reference when diagnosing new cases, utilizing the Case-Based Reasoning (CBR) method. In this CBR process, successfully resolved issues are archived for potential use in future problem-solving scenarios. Conversely, if a problem persists without resolution, the case is identified and stored to prevent similar errors in the future [14]. The CBR method comprises four key stages of problem-solving: retrieve, reuse, revise, and retain.

Retrieve is taking back cases most similar or relevant to the new case [15]. Meanwhile, reuse is the process of reusing information and knowledge from old cases as solutions for new cases. The old case, which has a similarity value above the threshold value and which has the highest value, is reused as a solution to solve the new case. Revise is the revision process that involves reviewing and revising the proposed solution. In this process, information is re-evaluated to address problems that arise in new cases. After that, the system will generate a solution for the new problem [16], and retaining is the process of storing new cases that have been successfully resolved and have found solutions to the database so that they can be reused as solutions for new cases in the future.

There has been many fields that apply CBR to solve problems. In the geographical field, Dou, et al. [17] conducted research to detect landslides using CBR. Tempola and colleagues [18] conducted a study investigating the use of Case-Based Reasoning (CBR) to assess the qualification of students for scholarship awards.

In the CBR system, the calculation of case similarity at the retrieve phase becomes a very important part. This calculation is the basis for determining the level of document similarity. There are several similarity calculation metrics, such as Minkowski Distance similarity and weighted cosine similarity.

The Minkowski Distance represents a generalized version of the Euclidean Distance and Manhattan Distance approaches. [19]. The main difference lies in the value of r, which is a power constant in the *Minkowski Distance* method. Meanwhile, *Weighted Cosine Similarity* calculates the similarity between two objects based on the size of the cosine angle [20]. The primary objective of this study is to assess and compare the effectiveness of the Minkowski Distance Similarity and Weighted Cosine Similarity methods in achieving the highest accuracy for the stroke diagnosis system.

Another research was conducted by Adawiyah [21] regarding the use of the *Minkowski Distance* for the system to detect premature baby birth. A system accuracy testing was carried out using 20 test data, seven data with a normal diagnosis and 13 data with a premature diagnosis. From the test, there was two data obtained whose results are not appropriate because the value is below the *threshold*, which is ≤ 60%. The accuracy of the system is 90% in detecting premature births.

In 2022, Mubarak, et al. conducted a research on CBR for the diagnosis of malaria using the *Minkowski Distance Similarity* method where testing was carried out with 25 test data and 58 training data which showed a system accuracy value of 92% with a *threshold* of 80% [22].

A research on case-based reasoning to diagnose malnutrition in children aged 0 - 5 years by applying the Cosine Similarity method [23] was conducted by Soinbala, et al. in 2019. The system's accuracy and validity were evaluated through testing with 40 new cases, comparing the system's diagnostic results with those provided by experts. The test outcomes reveal an 80% accuracy rate when employing an 80% threshold.

Zainuddin and colleagues conducted a study in 2016 [24], concentrating on Case-Based Reasoning (CBR) for diagnosing strokes. They employed the K-Nearest Neighbor algorithm with an 80% threshold value. The research, based on 15 test cases, reported a system accuracy of 93.3%, consistent with expert diagnoses. In a separate study, Warman and team [25] investigated the use of expert systems in identifying diseases in rice plants. They utilized CBR with the K-Nearest Neighbor method for distance calculation. The evaluation of system sensitivity and accuracy involved 52 test data points with a threshold value of 70%. The findings indicated a system sensitivity of 100% and an accuracy rate of 82.69%.

This study represents a continuation of the research conducted by Nelson et al. in 2018. In their investigation, Nelson et al. developed an expert system for diagnosing strokes using a Case-Based Reasoning (CBR) approach. The system employs the Jaccard Coefficient method for calculating case similarity. The research utilized the Siriraj score as a distinguishing factor between ischemic and hemorrhagic types of strokes, incorporating dense indexing for enhanced efficiency [26]. The system underwent testing with 45 cases as test data and utilized 135 cases as a case base. The findings revealed that a threshold value of 0.7 resulted in superior sensitivity and accuracy compared to threshold values of 0.8, 0.9, and 1. The system demonstrated a sensitivity level of 89.88% and accuracy of 81.67% with indexing and 84.44% without indexing. Further research was carried out using the same dataset with the Minkowski Distance similarity calculation method [27]. The research results show that Minkowski Distance provides a better accuracy rate of 88.89% compared to the Jaccard Coefficient method.

The research carried out is a continuation of research [26] [27]. The focus of this research is to compare the level of similarity test between the Minkowski Distance Similarity and Weighted Cosine Similarity methods in the diagnosis of stroke patients. This research wants to find out whether Weighted Cosine Similarity can increase the accuracy of the system in diagnosing stroke. This research contributes to increasing the effectiveness of expert systems in diagnosing stroke.

II. Method

The process of developing the system involves several stages, commencing with needs analysis, followed by system design, program code implementation, and culminating in system testing, as depicted in Fig. 1.

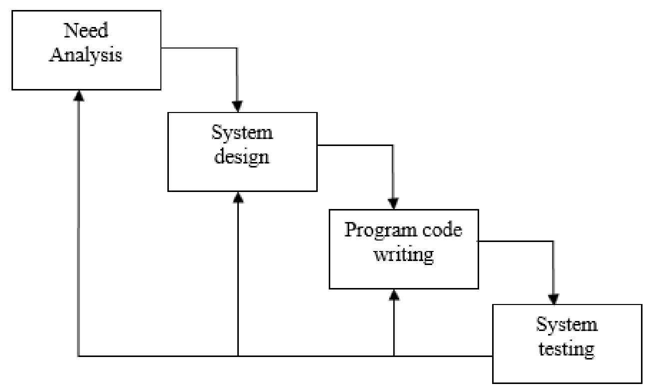

Fig. 1. The system development stages.

A. Data

The data used in this study was sourced from Nelson et al.'s 2018 research [26], specifically the study titled "Case-Based Reasoning for Stroke Diseases Diagnosis." It encompasses medical records extracted from patients who had experienced strokes and were treated at Dr. Soetarto DKT Hospital in Yogyakarta during the period of 2015-2016. The data were categorized into four types of stroke based on both the cause and anatomical pathology, namely embolic stroke, thrombotic stroke, subarachnoid hemorrhage stroke, and intracerebral hemorrhage stroke. We have limitations in terms of test data and future work has the opportunity to carry out better and more complete tests.

B. System Planning

The expert system is developed utilizing the Case-Based Reasoning (CBR) method, integrating the Minkowski Distance similarity method to evaluate similarities between newly entered cases and existing ones. Users input information in the form of the patient's personal data, symptoms, and risk factors. Subsequently, the system calculates local and global similarity values between the newly entered case data (user-provided data) and the cases stored in the case base. The case exhibiting the highest similarity, surpassing a predetermined threshold, is applied as the solution for new cases. In instances where the similarity value falls below the threshold, the case is retained in the case base for expert review. Conversely, if the similarity exceeds the threshold, the system generates an output indicating the type of stroke affecting the patient. The use case diagram for stroke diagnosis in the expert system is depicted in Fig. 2.

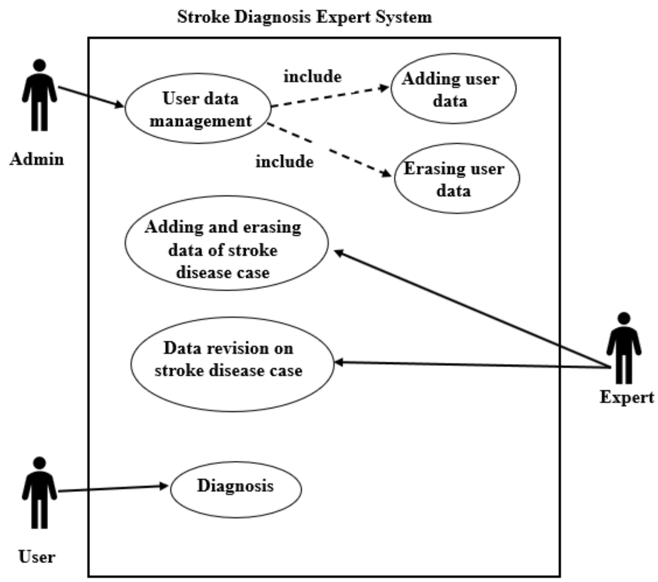

Fig. 2. The expert system usecase diagram for stroke diagnosis.

The goal of similarity measurement is to evaluate how closely two objects resemble each other. The determination of the similarity value involves calculating two values: the local similarity value and the global similarity value.

1) Local similarity: The aim of similarity measurement is to quantify how much two objects resemble each other. Calculations for local similarity are conducted to obtain similarity values by comparing the attributes of a problem with those of a case. The local similarity is determined based on the characteristics of the data and its features [28].

- Numeric data type

$$f(s,t) = 1 - \frac{|s-t|}{R} \qquad (1)$$

Here, *s* and *t* represent the values of the features under comparison, and *R* denotes the range of values associated with these features.

- Boolean data type

$$f(s,t) = \begin{cases} 1, if\ s = t \\ 0, if\ s \neq t \end{cases} \qquad (2)$$

where s, t {true, false}

2) Global similarity: Global similarity is utilized for assessing the similarity between problems and cases on a case base. This study will compare the accuracy of systems using Minkowski Distance Similarity and Weighted Cosine Similarity.

- Minkowski distance simmilarity [29]

$$E(C_i, C_j) = \left[\frac{\sum_{k=1}^{n} w_k^r * |d_k(C_{ik}, C_{jk})|^r}{\sum_{k=1}^{n} w_k^r}\right]^{1/r} * T(C_j) * \frac{n(C_i, C_j)}{n(C_i)} \qquad (3)$$

Here, $E(C_i, C_j)$ is the global similarity between target case (C_i) and source case (C_j), meanwhile w_k is the weight value of attribute k; $d_k(C_{ik}, C_{jk})$ is the local similarity value between target case attribute to k and source case attribute to k, and r is a Minkowski factor (positive integer); $T(C_j)$ is The confidence level of the case in the case base, $n(C_i, C_j)$ is the total attributes of the target case (C_i) that appear in the source case (C_j), and $n(C_i)$ is the total number of attributes in the target case (C_i)

- Weighted Cosine Similarity

$$\text{Weighted Cosine Similarity} = \frac{\sum_{i=1}^{n} w_i x_i y_i}{\sqrt{\sum_{i=1}^{n} w_i {x_i}^2}\ \sqrt{\sum_{i=1}^{n} w_i {y_i}^2}} \qquad (4)$$

Here, w_i is the weight value of attribute i, x_i is the value of local similarity for the first object (*target case*), and y_i is the local similarity value of the second object (*source case*)

C. Implementation

The system will be developed as web-based software using Hypertext Markup Language (HTML)/Cascading Style Sheets (CSS) and Hypertext Preprocessor (PHP), with Apache serving as the webserver and MySQL handling the database.

D. Test Design

During system testing, the confusion matrix method is utilized to produce accuracy, recall (sensitivity), and precision values. The confusion matrix [30] acts as a concise result table, presenting the counts of true and false test data. This matrix facilitates a comparison between the actual values and the predicted results, allowing for the calculation of accuracy, prediction, and recall values, as depicted in Table I.

TABLE I. SYSTEM TESTING MATRIC

Predicted Values	Actual Values		
		Positive	Negative
	Positive	TP-True Positive	FP-False Negative
	Negative	FN-False Negative	TN-True Negative

The formula for calculating accuracy, precision, and recall [30] can be seen in equations five to seven.

- Accuracy: The extent of accuracy exhibited by the model in accurately performing the classification.

$$accuracy = \frac{TP+TN}{Total} \qquad (5)$$

Precision: The degree of accuracy between the requested data and the predicted results from the model.

$$precision = \frac{TP}{TP+FP} \qquad (6)$$

- Recall: The system's effectiveness in retrieving information.

$$recall = \frac{TP}{TP+FN} \quad (7)$$

The test uses data from a research [31] in the form of medical record data from stroke patients during 2015-2016 at Dr Soetarto DKT Hospital, Yogyakarta consisting of 180 cases, where 30% of the cases, namely 54 cases, will be used as test data. The system underwent testing with various threshold values, specifically from 0.6 to 0.95. To find the highest accuracy value, the system will be tested using two different distance calculation methods, such as Minkowski Distance Similarity and Weighted Cosine Similarity methods. In a system that implements Minkowski Distance Similarity, a test is carried out on the Minkowski rank (r) to get the most optimal r value using different r values, started with r = 1 and continuing to increase by 1 until the resulting accuracy value does not show a significant difference.

III. Results

A. Stroke Diagnosis System

Research produces a system that can be used to diagnose stroke. The system has several interfaces, including an interface for carrying out diagnosis (see Fig. 3), an interface for displaying diagnosis results (see Fig. 4), and an interface for the system revision process (see Fig. 5).

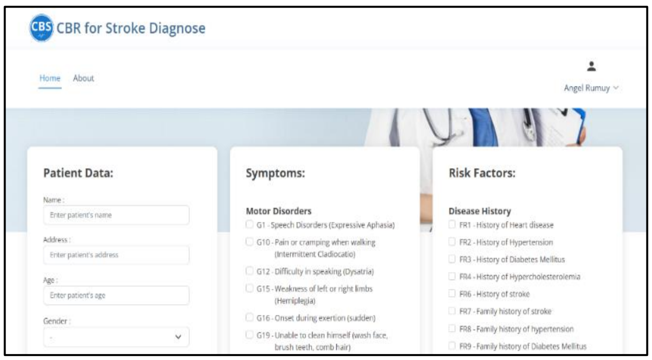

Fig. 3. Diagnosis page [27].

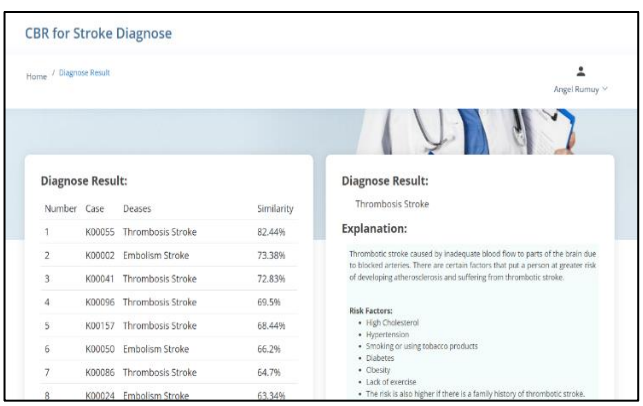

Fig. 4. Diagnosis result page [27].

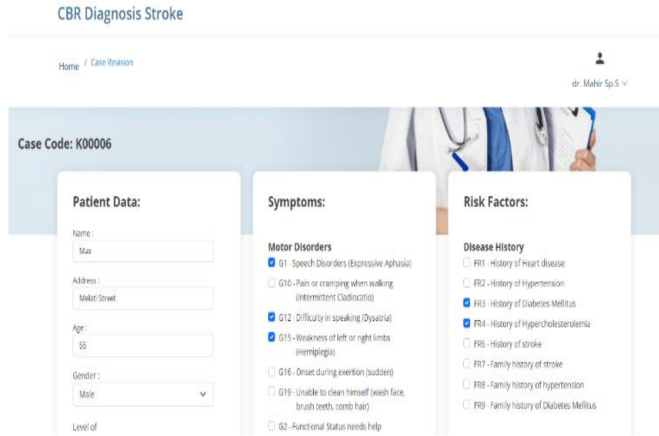

Fig. 5. Expert revise page [27].

On the diagnosis page, the user enters the symptoms experienced by a patient. This page is the input page for the system. Based on input from the user, the system calculates the similarity of the user input with the case dataset that the system already has. Instances of case representation are illustrated in Table II. When the system identifies similar cases (with similarity exceeding the threshold value), it will present a diagnosis results page. In the absence of cases resembling the symptoms entered by the user, the system will store the case data, and experts can review new cases through the system's revision page.

TABLE II. Example of Case Representation [26]

Base Case		
Patient Code:		K00007
General Condition:		
1	Age	60
2	Gender	Male
3	Awareness	Compo Mentis
Symptom:		
G1	Confusion	No
G3	Trouble balancing	No
Gn	n-th symptom	…
Risk Factor:		
FR1	History of heart disease	No
FR2	History of hypertension	Yes
FRn	n-th risk factor	…
Diagnosis:	Embolism Stroke	

C. System Testing Results

System testing involves the calculation of accuracy, precision, and recall using Eq. (5), (6), and (7). The testing process utilizes 30% of the case data as test data, consisting of 54 cases. To enhance efficiency, an automation script is employed. This script logs into the system, automatically inputs patient data, symptoms, and risk factors based on the test data, and records the system results.

The test outcomes for the system utilizing the Minkowski Distance Similarity method, along with the Confusion Matrix, are detailed in Table III. This table depicts different levels of accuracy, sensitivity, and recall corresponding to each threshold value and *r* value. The peak accuracy is achieved with a threshold value of 75 and *r* values of three and four, resulting in an accuracy rate of 88.89%.

Table IV displays the outcomes of system testing utilizing the Weighted Cosine Similarity method and the confusion matrix. The highest accuracy is achieved with threshold values of 75 and 80, yielding an accuracy percentage of 83.33%. The accuracy value signifies the system's ability to diagnose correctly, with higher accuracy indicating more precise diagnosis results or solutions provided by the system.

TABLE III. MINKOWSKI DISTANCE SIMILARITY SYSTEM TEST RESULTS

Threshold	Nilai R	Accuracy (%)	Recall (%)	Precision (%)
5	1	72,22	78	90,7
10	1	72,22	78	90,7
15	1	72,22	78	90,7
20	1	72,22	78	90,7
25	1	72,22	78	90,7
30	1	72,22	78	90,7
35	1	72,22	78	90,7
40	1	74,07	78	92,86
45	1	74,07	74	97,37
50	1	81,48	80	100
55	1	85,19	84	100
60	1	85,19	84	100
65	2	85,19	84	100
70	2	87,04	86	100
75	3 & 4	88,89	88	100
80	8	77,78	76	100
85	12	62,96	60	100
90	12	38,89	34	100
95	12	37,04	32	100

TABLE IV. WEIGHTED COSINE SIMILARITY SYSTEM TEST RESULTS

Threshold	Accuracy (%)	Recall (%)	Precision (%)
5	75,93	82	91,11
10	75,93	82	91,11
15	75,93	82	91,11
20	75,93	82	91,11
25	75,93	82	91,11
30	75,93	82	91,11
35	75,93	82	91,11
40	75,93	82	91,11
45	75,93	82	91,11
50	77,78	82	93,18
55	77,78	82	93,18
60	77,78	82	93,18
65	79,63	82	95,35
70	81,48	82	97,62
75	83,33	82	100
80	83,33	82	100
85	81,48	80	100
90	44,44	40	100
95	24,07	18	100

IV. DISCUSSION

Accuracy calculates all actual predicted values without specificity for each label, so a higher accuracy doesn't necessarily indicate good performance in predicting specific labels. Therefore, recall and precision values are crucial. Recall assesses the system's success in retrieving information, with higher values indicating better identification of positive cases.

Fig. 6 depicts the system's recall rate at a threshold of more than or equal 75 using the Minkowski Distance method, which surpasses the recall rates of the system using the Jaccard Coefficient method without indexing and the Weighted Cosine method. The Minkowski Distance method achieves the highest recall value at 88%. Precision, a metric measuring the accuracy of positive predictions, is highest in the Minkowski Distance method when applying a threshold value of more than or equal 50, reaching 100% (see Fig. 7).

Across the three tested methods, the Minkowski Distance approach with a threshold value of 75 and *r* values of three or four consistently produces the highest levels of accuracy, recall, and precision. In a system designed to detect high-risk diseases such as stroke, recall is particularly crucial, as a low recall value implies misdiagnosing some patients with stroke as healthy, leading to serious risks. Therefore, the optimal configuration for the expert system for stroke diagnosis is the Minkowski Distance Similarity method with a threshold value of 75 at r = 3 or r = 4, achieving a system accuracy rate of 88.89%, a recall of 88%, and a precision of 100%.

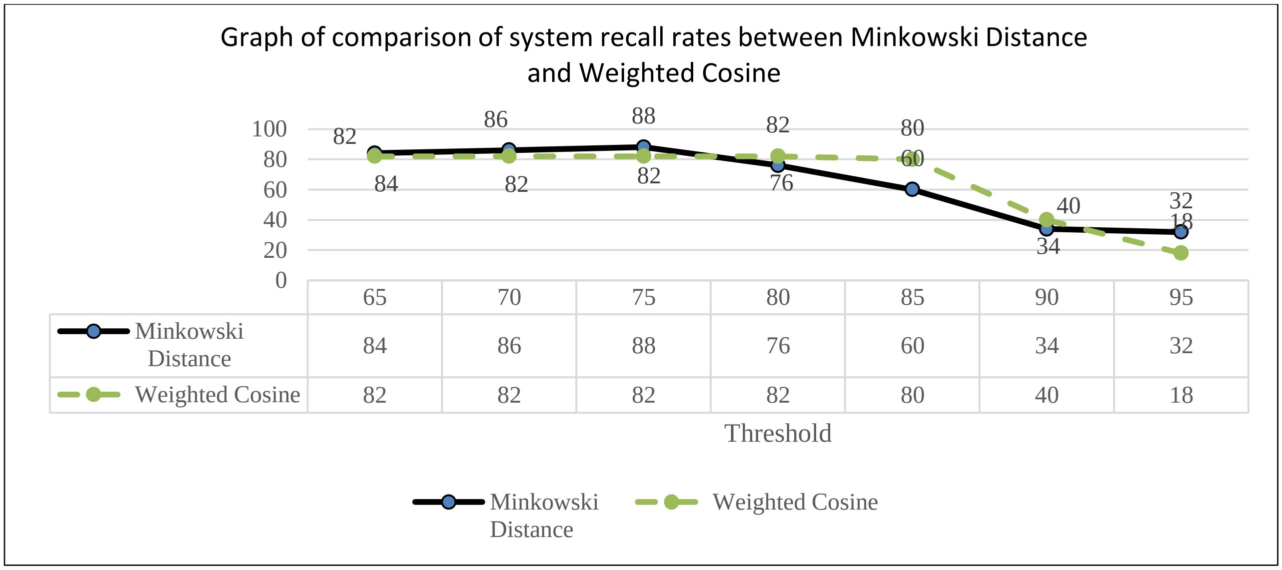

Fig. 6. Graph of comparison of system recall rates between Minkowski Distance, Weighted Cosine and Jaccard Coefficient.

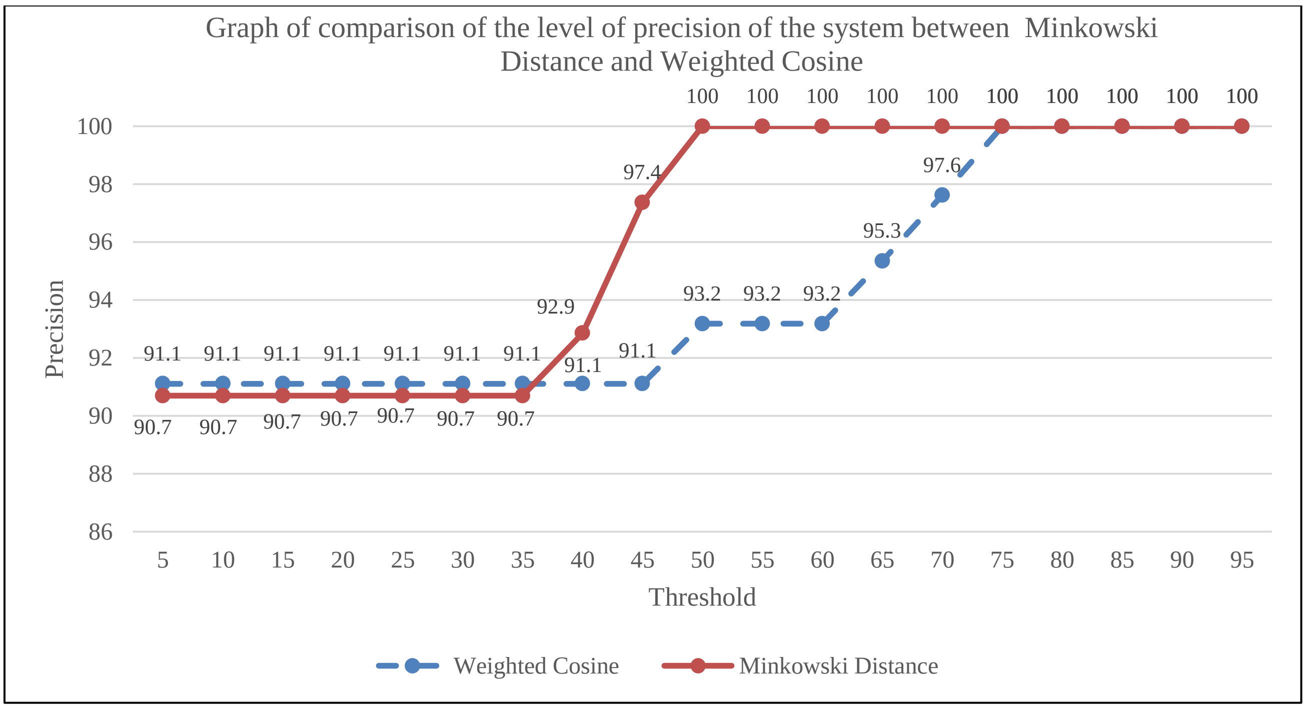

Fig. 7. Graph of comparison of the level of precision of the system between Minkowski Distance and Weighted Cosine.

A. *Limitation*

This research is a follow-up research that uses the same dataset from the work of Nelson and the team [26]. This study has not added a dataset with the latest cases for stroke diagnosis. Further research can be carried out by collecting and adding a dataset of cases in the last five years.

V. Conclusion

This research is a follow-up study that uses a dataset of stroke cases. Research focuses on the effectiveness of algorithms for stroke diagnosis. The system developed is an expert system with a CBR approach. The study focuses on finding the most effective algorithm for diagnosing stroke. In previous research, the Jaccard Coefficient algorithm was applied with an accuracy level of 81.67%, and Minkowski Distance Similarity was applied with an accuracy of 88.89%. In this research, the Weighted Cosine algorithm was applied, resulting in an accuracy of 83.33%. Through the comparison of the applications of the three algorithms, it becomes apparent that the Minkowski Distance Similarity algorithm exhibits a superior level of accuracy and sensitivity (or recall) in contrast to systems utilizing the Jaccard Coefficient method and the Weighted Cosine method. When the threshold is set at 75 or above, the system attains an accuracy rate of 88.89%, along with a recall of 88%. In comparison to the Weighted Cosine method alone, the precision level is 100%.

Further research is being carried out to develop a rule generator to automate the formation of a knowledge base in a rule-based system format. The anticipated outcome of this research is an enhancement in the efficiency of tracking case times for decision-making. Apart from this, updates to the stroke case dataset must also be carried out to enrich the case data

ACKNOWLEDGMENT

We express our gratitude to the Faculty of Information Technology and the Institute for Research and Community Service (LPPM) at Duta Wacana Christian University for providing the funding for both the research and its publication.

REFERENCES

[1] "Apa itu Stroke," http://p2ptm.kemkes.go.id/infographic-p2ptm/stroke/apa-itu-stroke. 2018.

[2] P. J. F. Lucas and L. C. Van Der Gaag, Principles of expert systems. Singapore: Addison-Wesley Publishing, 1991. [Online]. Available: https://www.researchgate.net/publication/220694050

[3] A. Abraham, "Rule-based Expert Systems," in Handbook of Measuring System Design, P. H. Sydenham and R. Thorn, Eds. John Wiley & Sons, Ltd., 2005, pp. 909–919.

[4] I. Gunawan and Y. Fernando, "Sistem Pakar Diagnosa Penyakit Kulit pada Kucing Menggunakan Metode Naive Bayes Berbasis Web," J. Inform. dan Rekayasa Perangkat Lunak, vol. 2, no. 2, pp. 239–247, 2021, [Online]. Available: http://jim.teknokrat.ac.id/index.php/informatika/article/view/927/380

[5] F. Anjara and A. A. Jaharadak, "Expert System for Diseases Diagnosis in Living Things: A Narrative Review," J. Phys. Conf. Ser., vol. 1167, no. 1, 2019, doi: 10.1088/1742-6596/1167/1/012070.

[6] H. Henderi, F. Al Khudhorie, G. Maulani, S. Millah, and V. T. Devana, "A Proposed Model Expert System for Disease Diagnosis in Children to Make Decisions in First Aid," INTENSIF J. Ilm. Penelit. dan Penerapan Teknol. Sist. Inf., vol. 6, no. 2, pp. 139–149, 2022, doi: 10.29407/intensif.v6i2.16912.

[7] A. Andriani, A. Meyliana, Sardiarinto, W. E. Susanto, and Supriyanta, "Certainty Factors in Expert System to Diagnose Disease of Chili Plants," 2018 6th Int. Conf. Cyber IT Serv. Manag. CITSM 2018, no. Citsm, 2019, doi: 10.1109/CITSM.2018.8674264.

[8] B. Basiroh and S. W. Kareem, "Analysis of Expert System for Early Diagnosis of Disorders During Pregnancy Using the Forward Chaining Method," Int. J. Artif. Intell. Res., vol. 5, no. 1, pp. 44–52, 2021, doi: 10.29099/ijair.v5i1.203.

[9] B. A. Wijaya and J. P. Tanjung, "An Expert System For Diagnosis Eye Diseases On Human Using Certainty Factor Method Based Web," SinkrOn, vol. 5, no. 1, pp. 78–83, 2020, doi: 10.33395/sinkron.v5i1.10579.

[10] X. Huang et al., "A Generic Knowledge Based Medical Diagnosis Expert System," ACM Int. Conf. Proceeding Ser., vol. 1, no. 1, pp. 462–466, 2021, doi: 10.1145/3487664.3487728.

[11] I. Setiawan and M. Batara, "Expert System Design to Diagnose Pests and Diseases on Local Red Onion Palu Using Bayesian Method," BAREKENG J. Math. Its Appl., vol. 17, no. 1, pp. 371–382, 2023.

[12] M. M. Richter and R. O. Weber, Case-Based Reasoning. New York: Springer Berlin Heidelberg, 2013. doi: 10.1007/978-3-642-40167-1.

[13] I. Nugraha and M. Siddik, "Penerapan Metode Case Based Reasoning (CBR) Dalam Sistem Pakar Untuk Menentukan Diagnosa Penyakit Pada Tanaman Hidroponik," J. Mhs. Apl. Teknol. Komput. dan Inf., vol. 2, no. 2, pp. 91–96, 2020, [Online]. Available: https://www.ejournal.pelitaindonesia.ac.id/JMApTeKsi/index.php/JOM/article/view/575/387

[14] I. Y. Subbotin and M. G. Voskoglou, "Applications of fuzzy logic to Case-Based Reasoning," vol. 11, pp. 7–18, 2012, [Online]. Available: https://www.researchgate.net/publication/223129950

[15] A. S. Soroto, A. Fuad, S. Lutfi, J. J. Metro, and K. T. Selatan, "Penerapan Metode Case Based Reasoning (CBR) untuk Sistem Penentuan Status Gunung Gamalama," J. Inform. dan Komput., vol. 02, no. 2, pp. 70–75, 2018.

[16] A. Yuli Vandika and A. Cucus, "Sistem Deteksi Awal Penyakit TBC dengan Metode CBR," Pros. Semin. Nas. Darmajaya, 2017.

[17] J. Dou et al., "Automatic Case-Based Reasoning Approach for Landslide Detection: Integration of Object-Oriented Image Analysis and a Genetic Algorithm," Remote Sens., vol. 7, no. 4, pp. 4318–4342, 2015, doi: 10.3390/rs70404318.

[18] F. Tempola, A. Musdholifah, and S. Hartati, "Case Based Reasoning Untuk Penentuan Kelayakan Mahasiswa Penerima Beasiswa," 2015.

[19] M. Nishom, "Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square," J. Inform. J. Pengemb. IT, vol. 4, no. 1, pp. 20–24, 2019, doi: 10.30591/jpit.v4i1.1253.

[20] J. Firdaus, I. Y. Purbasari, and H. P. Swari, "Implementasi Case-Based Reasoning pada Sistem Prediksi Masa Studi dan Predikat Kelulusan Mahasiswa (Studi Kasus : Fakultas Ilmu Komputer, UPN 'Veteran' Jawa Timur)," 2020.

[21] R. Adawiyah, "Implementasi Metode Minkowsky Distance untuk Deteksi Kelahiran Bayi Prematur Berbasis Case-Based Reasoning," J. Inform. dan Komputer) Akreditasi KEMENRISTEKDIKTI, vol. 3, no. 1, 2020, doi: 10.33387/jiko.

[22] A. Mubarak, M. Salmin, A. Fuad, and S. Do Abdullah, "Penalaran Berbasis Kasus Untuk Diagnosis Penyakit Malaria Dengan Menggunakan Metode Minkowsky Distance," J. Ilm. Ilk. - Ilmu Komput. Inform., vol. 5, no. 1, 2022, doi: 10.47324/ilkominfo.v4i3.136.

[23] M. E. Soinbala, D. Rony Sina, and M. Boru, "Case Based Reasoning untuk mendiagnosis Gizi Buruk pada Anak Usia 0-5 Tahun Menggunakan Metode Cosine Similarity," J-ICON, vol. 7, no. 1, pp. 67–71, 2019.

[24] M. Zainuddin, K. Hidjah, and W. Tunjung, "Penarapan Case-Based Reasoining (CBR) untuk Mendiagnosis Penyakit Stroke Menggunakan Algoritma K-Nearest Neighbor," CITISEE, 2016.

[25] I. Warman, "Sistem Pakar Identifikasi Penyakit Tanaman Padi Menggunakan Case-Based Reasoning," 2017.

[26] R. Nelson, A. Harjoko, and A. Musdholifah, "Case-Based Reasoning for Stroke Disease Diagnosis," IJCCS (Indonesian J. Comput. Cybern. Syst., vol. 12, no. 1, p. 33, 2018, doi: 10.22146/ijccs.26331.

[27] A. Rumuy, R. Delima, K. P. Saputra, and J. Purwadi, "Application of the Minkowski Distance Similarity Method in Case-Based Reasoning for Stroke Diagnosis," JUITA J. Inform., vol. 11, no. 2, pp. 323–332, 2023, [Online]. Available: https://jurnalnasional.ump.ac.id/index.php/JUITA/article/view/18582/pdf

[28] M. K. Jha, D. Pakhira, and B. Chakraborty, "Diabetes Detection and Care Applying CBR Techniques ," IJSCE, vol. 2, no. 6, 2013.

[29] E. Faizal and H. Hamdani, "Weighted Minkowski Similarity Method with CBR for Diagnosing Cardiovascular Disease," Int. J. Adv. Comput. Sci. Appl., vol. 9, no. 12, 2018, doi: 10.14569/IJACSA.2018.091244.

[30] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," 2021.

[31] R. Nelson, A. Harjoko, and A. Musdholifah, "Rekam Medik Stroke.xlsx." 2018.