Jurnal Minfo Polgan (JMP)
Volume 15, Nomor 1, Maret 2026
DOI : https://doi.org/10.33395/jmp.v15i1.16046
e-ISSN : 2797-3298
p-ISSN : 2089-9424

Web-Based Stroke Disease Classification System Using the Modified K-Nearest Neighbors Method

[1]Rizki Suwanda, [2]Bustami, [3]Muhammad Khardawi
[1*,2,3]Teknik Informatika, Universitas Malikussaleh, Lhokseumawe, Aceh, Indonesia
**Korespondensi*: rizkisuwanda@unimal.ac.id

Submitted : Feb 22, 2026 | **Accepted** : March 29, 2026 | **Published** : April 02, 2026

ABSTRACT

Classification is a systematic method of grouping data based on predefined analytical rules and principles. One of the classification methods employed in this study is the Modified K-Nearest Neighbors (MKNN) algorithm, which is recognized for its potential to achieve higher accuracy. MKNN is an extension of the traditional K-Nearest Neighbors (KNN) method, incorporating an additional ranking stage and a weighted voting mechanism using an alpha value of 0.5. The object of this study is stroke disease. In the medical context, stroke occurs due to a disruption of blood flow to the brain. Ischemic stroke is caused by the obstruction of blood vessels and is generally considered less severe, whereas hemorrhagic stroke results from the rupture of blood vessels and is categorized as a severe condition. Hospitals in Indonesia are required to provide prompt and accurate healthcare services, in accordance with Law Number 36 of 2009 concerning Health. Approximately 70% of stroke patients have a history of hypertension and heart disease, while around 87% experience psychological disorders such as anxiety and depression. Based on data obtained from Cut Meutia Regional General Hospital (RSUD Cut Meutia) in Lhokseumawe, the classification of stroke types is still performed manually through clinical observation. Therefore, this study proposes a stroke classification system based on the MKNN algorithm. The system utilizes 11 features and two diagnostic classes, namely ischemic stroke and hemorrhagic stroke, with a total of 100 medical record datasets divided into 80 training data and 20 testing data. Using a value of K = 5, the system achieved an average confidence accuracy of 81.19%, with a precision of 85.71%, recall of 80%, F1-score of 82.75%, and overall accuracy of 75%. The system was developed using the PHP programming language and a MySQL database.

Keywords*: Stroke Disease, Modified K-Nearest Neighbors, Web-Based System, Classification.*

ABSTRAK

Klasifikasi ialah metode pengelompokan data secara sistematis berdasarkan aturan dan prinsip analisis yang telah ditetapkan. Salah satu metode klasifikasi yang digunakan dalam penelitian ini ialah algoritma Modified K-Nearest Neighbors (MKNN), yang dikenal mampu menghasilkan tingkat akurasi yang lebih baik. MKNN merupakan pengembangan dari metode K-Nearest Neighbors (KNN) dengan penambahan tahapan perankingan dan mekanisme weighted voting menggunakan nilai alpha 0,5. Pada objek penelitian ini ialah penyakit stroke, dalam dunia medis penyakit ini terjadi akibat terganggunya aliran darah ke otak, di mana stroke iskemik disebabkan oleh penyumbatan pembuluh darah dan umumnya bersifat ringan, sedangkan stroke hemoragik terjadi akibat pecahnya pembuluh darah dan tergolong sebagai stroke berat. Setiap rumah sakit di Indonesia diwajibkan memberikan pelayanan kesehatan yang cepat dan akurat, hal ini sejalan dengan Undang - Undang Nomor 36 Tahun 2009 tentang kesehatan. Sekitar 70% penderita stroke memiliki riwayat hipertensi dan penyakit jantung, sementara sekitar 87% lainnya mengalami gangguan psikologis seperti kecemasan dan depresi. Berdasarkan data dari RSUD Cut Meutia Lhokseumawe, proses klasifikasi jenis stroke masih dilakukan secara manual melalui observasi klinis. Oleh karena itu, penelitian ini mengusulkan sistem klasifikasi penyakit stroke berbasis algoritma MKNN, sistem ini menggunakan 11 fitur dan 2 kelas diagnosis, yaitu stroke iskemik dan stroke hemoragik dengan total 100 data rekam medis yang sudah dibagi menjadi 80 data latih dan 20 data uji. Dengan nilai K=5, sistem menghasilkan tingkat akurasi confidence rata-rata sebesar 81,19%, serta nilai precision 85,71%, recall 80%, FI score 82,75% dan

@2026 The Author Published By Politeknik Ganesha Medan
This is an open access article under the CC BY licence (http://creativecommons.org/licenses/by/4.0/). which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properlycited.

accuracy 75%. Sistem ini dikembangkan menggunakan bahasa pemrograman PHP dan basis data MySQL.

Kata Kunci: Penyakit Stroke, Modified K-Nearest Neighbor, Klasifikasi.

INTRODUCTION

In the context of healthcare service regulations, every hospital in Indonesia is mandated to provide accurate and prompt medical services, particularly in handling emergency cases such as stroke. This requirement is in accordance with Law Number 36 of 2009 concerning Health. Theoretically, stroke is one of the most prevalent health problems, especially in developing countries and among adults. Approximately 70% of stroke patients in developing countries have a history of hypertension and heart disease, while 87% experience psychological disorders such as anxiety and depression. Globally, stroke remains one of the leading causes of mortality and long-term disability. To determine whether a stroke is caused by vascular obstruction (ischemic stroke) or intracerebral hemorrhage (hemorrhagic stroke), physicians generally rely on clinical symptoms and brain imaging examinations, such as computed tomography (CT) scans (Setiawan et al., 2024).

According to global statistical data, approximately 15 million people worldwide experience stroke each year, meaning that one in six individuals is at risk of developing the disease. In Indonesia, stroke cases are most prevalent among the elderly aged 75 years and above, accounting for 43.1%, whereas in the younger age group of 15 to 24 years, the prevalence is only around 0.2% (Akmal et al., 2023). If not managed appropriately, stroke can significantly increase the risk of mortality. The high incidence of stroke is attributed to the substantial prevalence of risk factors and the lack of optimal early detection. Timely and accurate intervention is crucial to prevent severe consequences, including long-term disability or death.

Therefore, Cut Meutia Regional General Hospital (RSUD Cut Meutia) in Lhokseumawe requires clear and reliable information regarding the type of stroke experienced by patients. To support this need, a system capable of identifying and classifying stroke types both ischemic and hemorrhagic is essential. Such a system would not only assist the community in recognizing stroke types at an earlier stage but also support healthcare professionals in establishing more accurate and efficient diagnoses.

This study applies a classification technique, which is an approach used to assign each object into a specific group or class based on predefined characteristics. One of the algorithms frequently employed in classification processes is the K-Nearest Neighbors (KNN) algorithm and its development, known as the Modified K-Nearest Neighbors (MKNN). The KNN method determines the class of new data based on a number of its nearest neighbors (the K value). Although this method is popular due to its simplicity, KNN has a limitation in determining the optimal value of K, which may reduce the accuracy of the classification results.

To address this limitation, the Modified K-Nearest Neighbors (MKNN) algorithm was developed, incorporating a data ranking process and a weighting mechanism to improve classification accuracy. However, under certain conditions, MKNN may produce lower accuracy than KNN at specific values of K. A comparative analysis between these two algorithms is therefore conducted to obtain a clearer understanding of which method is more appropriate for classifying stroke disease. This comparison is intended to assist researchers in selecting the algorithm that yields the most optimal performance (Pratama et al., 2025).

A previous study conducted by Alwie et al. (2020), entitled *Comparison of the K-Nearest Neighbor and Modified K-Nearest Neighbor Methods for Classifying Families at Risk of Stunting*, demonstrated that both KNN and MKNN were capable of effectively classifying families at risk of stunting. The testing results indicated that MKNN consistently outperformed KNN in maintaining high accuracy, exhibiting stable performance across different data split ratios (90:10, 80:20, and 70:30) and various values of *k*. The findings revealed that KNN showed performance fluctuations, despite achieving the highest accuracy at certain *k* values. In addition to accuracy, MKNN demonstrated superiority in maintaining balanced evaluation metrics even as the proportion of test data increased. In contrast, KNN experienced performance declines as the value of *k* varied and the test dataset expanded. The ranking process and weighting mechanism incorporated in MKNN highlighted its advantages in performing not only accurate but also reliable classification in complex scenarios. Consequently, MKNN was identified as a more effective method for detecting families at risk of stunting.

Through the implementation of this method, the study is expected to provide an innovative solution for improving the management of stroke disease classification. The resulting system is designed as a web-based classification platform that facilitates both patients and healthcare services at RSUD Cut Meutia Lhokseumawe by delivering diagnostic support that is fast, precise, and accurate.

RESEARCH METHOD

Research Workflow

The data utilized in this study consisted of medical records of patients diagnosed with stroke. The research was conducted from April 2025 to October 2025 and followed a series of systematic steps, as outlined below:

1. Problem Identification: Identifying and defining the research problem related to stroke disease.
2. Objective Determination: Establishing the research objective, namely implementing the Modified K-Nearest Neighbors (MKNN) method to address the problem of stroke disease classification.
3. Data Collection: Collecting data required for the study, including training data and sample datasets obtained from RSUD Cut Meutia Lhokseumawe, as well as conducting direct interviews with relevant stakeholders.
4. System Design: Designing the system based on the predefined workflow in accordance with the research implementation plan.
5. System Implementation: Implementing the system by applying the MKNN method that had been designed and developed.
6. System Testing: Conducting system testing, including program debugging stages, to ensure that the system operates properly and is capable of generating accurate classification results as intended.

Stroke Disease Classification System Scheme

The classification of stroke disease using the Modified K-Nearest Neighbors (MKNN) method consists of several stages. The first stage involves inputting a dataset of stroke-related clinical features that have been normalized. These features include gender, age, hypertension, history of heart disease, CT scan results, slurred speech, decreased level of consciousness, right-sided limb weakness, left-sided limb weakness, smoking status, and headache. Next, the Euclidean distance is calculated between each training dataset instance and the testing dataset instance (new data). Based on the computed Euclidean distances, a predetermined number of nearest neighbors (K) is selected. In this study, the value of K was set to 5. Subsequently, a ranking value is calculated based on the previously obtained results. The final stage involves computing the weighted voting to determine the final classification of the testing data. Afterward, the average confidence value is calculated from the overall results. For clarity, the flowchart of the Modified K-Nearest Neighbors (MKNN) classification process illustrates the overall workflow of the stroke disease classification system.

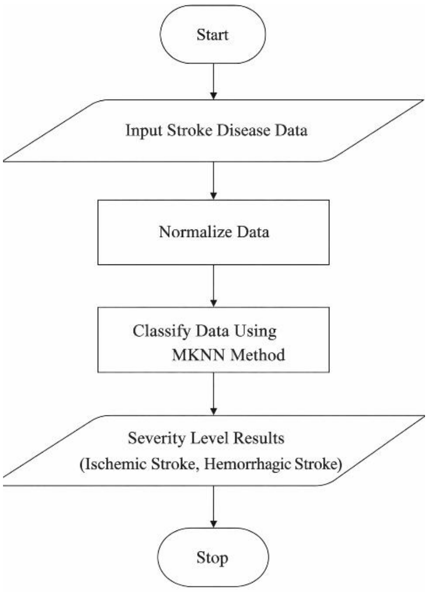

Figure 1. Overall System Architecture

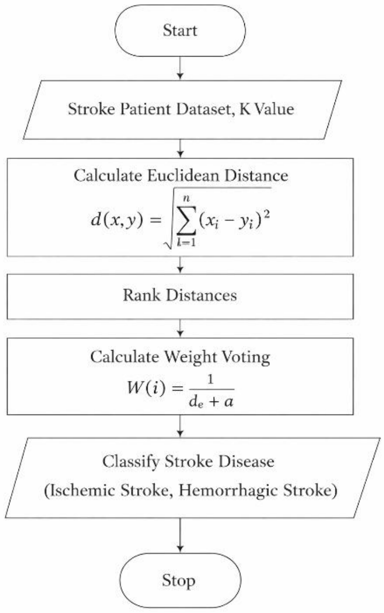

Figure 2. MKNN System Workflow Diagram

The system workflow is described sequentially from the initial stage to the final stage, with each step explained as follows:

1. Start: The system execution begins.
2. Input Stroke Data: The user enters patient data related to stroke, including attributes such as gender, age, hypertension, history of heart disease, CT scan results, slurred speech, decreased level of consciousness, right-sided limb weakness, left-sided limb weakness, smoking status, and headache.
3. Classification Process: The user selects the classification process option, after which the system performs the Modified K-Nearest Neighbors (MKNN) algorithm computation.
4. Display of Results: The system presents the final classification outcome, categorizing the stroke type as either ischemic stroke or hemorrhagic stroke.
5. End: The system process is completed and terminated.

RESULT AND DISCUSSION

The Modified K-Nearest Neighbors (MKNN) algorithm was applied to classify stroke disease based on 11 attributes: gender, age, hypertension, history of heart disease, CT scan results, slurred speech, decreased level of consciousness, right-sided limb weakness, left-sided limb weakness, smoking status, and headache. These attributes were processed using the MKNN algorithm to categorize the data into two classes: ischemic stroke and hemorrhagic stroke. The dataset was obtained from Cut Meutia Regional General Hospital (RSUD Cut Meutia) Lhokseumawe and consisted of 100 medical records of patients diagnosed with either ischemic or hemorrhagic stroke. For experimental evaluation, the dataset was divided into 80 training instances and 20 testing instances. The testing results showed that the proposed MKNN-based classification system achieved an overall accuracy of 75% and an average confidence value of 81.19%. Furthermore, evaluation using a confusion matrix produced a precision value of 85.71%, a recall value of 80%, and an F1-score of 82.75%. These results indicate that the MKNN algorithm demonstrates satisfactory performance in classifying stroke types based on the selected clinical attributes.

The following outlines the steps for calculating the Euclidean Distance using the Modified K-Nearest Neighbors (MKNN) method. In this section, two examples of manual calculations are provided to illustrate each stage of the process:

1. Determining the Value of K
   In this study, the value of K was set to 5 for the classification of stroke disease.
2. Calculating the Euclidean Distance Between Training and Testing Data
   The Euclidean distance was computed between the 80 training data instances and each of the 20 testing data instances. Each training object was individually compared with every

testing object. The following section presents ten sample calculations of Euclidean distance between selected training and testing data instances to demonstrate the computation process.

$$d_{(1,1)} = \sqrt{\begin{matrix}(0{,}700-0{,}171)^2+(1-0)^2+(1-0)^2+(1-1)^2+\\(0-0)^2(1-0)^2(0-0)^2(0-1)^2(0-0)^2(0-1)^2\end{matrix}}$$
$$= \sqrt{\begin{matrix}(0{,}529)^2+(1)^2+(1)^2+(0)^2+(0)^2+(1)^2+\\(0)^2+(-1)^2+(0)^2+(-1)^2\end{matrix}}$$
$$= \sqrt{0{,}279841+1+1+0+0+1+0+1+0+1}$$
$$= \sqrt{5{,}279841}$$
$$=2{,}29779$$

$$d_{(1,2)} = \sqrt{\begin{matrix}(0{,}700-0{,}300)^2+(1-1)^2+(1-1)^2+(1-0)^2+\\(0-0)^2(1-1)^2(0-0)^2(0-0)^2(0-1)^2(0-0)^2\end{matrix}}$$
$$= \sqrt{\begin{matrix}(0{,}400)^2+(0)^2+(0)^2+(1)^2+(0)^2+(0)^2+\\(0)^2+(0)^2+(-1)^2+(0)^2\end{matrix}}$$
$$= \sqrt{0{,}16+0+0+1+0+0+0+0+1+0}$$
$$= \sqrt{2{,}16}$$
$$=1{,}46969$$

Weighted Voting Calculation

The weighted voting stage is performed by applying a formula that divides one by the Euclidean distance between the training data and the testing data, incorporating an alpha (α) value of 0.5. This weighting mechanism assigns greater influence to neighbors with shorter distances, thereby strengthening their contribution to the final classification decision. The following section presents ten examples of manual weighted voting calculations to illustrate the computational procedure.

$$W(1,1) = \left(\frac{1}{2{,}29779+0{,}5}\right) = 0{,}357424968$$
$$W(1,2) = \left(\frac{1}{1{,}46969+0{,}5}\right) = 0{,}507694104$$
$$W(1,3) = \left(\frac{1}{2{,}47972+0{,}5}\right) = 0{,}335602003$$
$$W(1,80) = \left(\frac{1}{2{,}01828+0{,}5}\right) = 0{,}397096431$$
$$W(2,2) = \left(\frac{1}{1{,}75103+0{,}5}\right) = 0{,}444241081$$
$$W(2,3) = \left(\frac{1}{1{,}74901+0{,}5}\right) = 0{,}444640086$$
$$W(2,4) = \left(\frac{1}{1{,}02566+0{,}5}\right) = 0{,}655454033$$
$$W(2,80) = \left(\frac{1}{1{,}00816+0{,}5}\right) = 0{,}663059622$$
$$W(3,3) = \left(\frac{1}{1{,}06181+0{,}5}\right) = 0{,}640282749$$
$$W(3,4) = \left(\frac{1}{2{,}26208+0{,}5}\right) = 0{,}362045994$$

Determination Of The Testing Data Class

The final step is to determine the class of each testing instance based on the predetermined value of *K = 5*. The classification decision is made by selecting the class with the highest total weighted voting value among the five nearest neighbors. In other words, the testing data are assigned to the class whose cumulative weight is the largest.

The following table presents the ordered results of the weighted voting calculations used to determine the final class for each testing instance.

Table 1. Classification Results

NO	*d*	*w*	CLASS	RESULTS
1	1.53	1,458333333	SH	SH
	1.36	0,666666667	SH	
	1.18	0,52238806	SH	
	1.21	0,52238806	SH	
	1.22	0,520043143	SI	
2	2.55	1,489361702	SI	SI
	2.74	0,66625894	SI	
	2.68	0,665040975	SH	
	2.71	0,664457319	SI	
	2.70	0,663028221	SI	
3	3.65	0,666621321	SI	SI

NO	*d*	*w*	CLASS	RESULTS
	3.66	0,66625894	SI	
	3.9	0,665536242	SI	
	3.69	0,664457319	SI	
	3.31	0,663786089	SI	
4	4.17	0,666485347	SI	SI
	4.60	0,665942422	SI	
	4.25	0,663786089	SI	
	4.38	0,660245785	SH	
	4.33	0,655398235	SI	
5	5.53	1,75	SH	SH
	5.36	0,663786089	SH	
	5.22	0,522230627	SI	
	5.21	0,521445845	SH	
	5.18	0,520820858	SH	

Accuracy Calculation

To determine the accuracy value, a comparison was conducted between the predicted classification results (total weighted votes) and the actual class labels. Accuracy reflects the proportion of correctly classified instances out of the total number of testing data. The following table presents the results of the accuracy calculation based on the comparison between the predicted classes and the actual classes.

Table 2. Prediction Accuracy Value Results

NO	CORRECT STATUS	TOTAL VOTE VALUE	ACCURACY VALUE
1	SH	3,1698	85,91
		0,5200	
2	SI	3,4831	83,97
		0,665	
3	SI	3,3267	100
		0	
4	SI	2,6516	80,06
		0,6602	
5	SH	3,4561	86,87
		0,5222	
6	SI	3,3217	100
		0	
7	SH	1,5626	57,71
		1,1449	
8	SI	1,8191	60,90
		1,1678	
9	SI	2,6492	79,96
		0,6638	
10	SI	1,2434	54,48
		1,0389	
11	SI	2,652	66,33
		1,3462	
12	SI	2,7473	100
		0	
13	SH	1,994	62,99
		1,1718	
14	SH	1,5562	59,85
		1,0441	
15	SI	2,8909	100
		0	
16	SI	6,2635	100
		0	
17	SH	3,2293	100

NO	CORRECT STATUS	TOTAL VOTE VALUE	ACCURACY VALUE
		0	
18	SI	3,1826	62,72
		1,8919	
19	SI	2,3933	82,09
		0,5223	
20	SI	7,1489	100
		0	

Confusion Matrix Analysis

The performance evaluation of the MKNN classification model was conducted using a confusion matrix based on 20 testing instances. The results of the actual and predicted classifications are presented as follows:

Table 3. Confusion Matrix (Actual vs. Predicted)

ACTUAL	PREDICTED	
	SI	SH
SI	12	3
SH	2	3

From the confusion matrix above, the following evaluation metrics were calculated:

Precision:

$$Precision = \frac{TP}{TP+FP}$$

$$Precision_{SI} = \frac{12}{12+2} \times 100\% = \frac{12}{14} \times 100\% = 85{,}71\%$$

$$Precision_{SH} = \frac{3}{3+3} \times 100\% = \frac{3}{6} \times 100\% = 50\%$$

Recall:

$$Recall = \frac{TP}{TP+FN}$$

$$Recall_{SI} = \frac{12}{12+3} \times 100\% = \frac{12}{15} \times 100\% = 80\%$$

$$Recall_{SH} = \frac{3}{3+2} \times 100\% = \frac{3}{5} \times 100\% = 60\%$$

Accuracy:

$$Accuracy = \frac{TP+TN}{\text{Jumlah Data}}$$

$$Accuracy = \frac{12+3}{20} \times 100\% = \frac{15}{20} \times 100\% = 75\%$$

F1-Score:

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$$F1\text{-}Score_{SI} = 2 \times \frac{85{,}71\% \times 80\%}{85{,}71\% + 80\%} = 2 \times \frac{6.856{,}8}{165{,}71}$$

$$F1\text{-}Score_{SI} = \frac{13.713{,}6}{165{,}71} = 82{,}75\%$$

$$F1\text{-}Score_{SH} = 2 \times \frac{50\% \times 60\%}{50\% + 60\%} = 2 \times \frac{3.000}{110}$$

$$F1\text{-}Score_{SH} = \frac{6.000}{110} = 54{,}54\ \%$$

Table 4. Summary of Evaluation Results / Confusion Matrix

CLASS	PRECISION	RECALL	ACCURACY
SI	85,71%	80%	75%
SH	50%	60%	

Based on the results presented in Table 5, the classification of stroke disease using the Modified K-Nearest Neighbors (MKNN) method achieved a precision of 85.71%, recall of 80%, and overall accuracy of 75%. These findings indicate that the model performs well in identifying ischemic stroke (SI), although its performance in detecting hemorrhagic stroke (SH) remains comparatively lower.

System Testing Results

The system testing phase was conducted to evaluate whether the developed system operates in accordance with the previously established design specifications. This testing process aims to verify the functional validity of the system and to ensure that each component performs as intended. The following table presents the system testing results, which indicate whether each functional module of the system is valid and operates correctly.

Table 5. System Testing Results

No	Feature	Test Case	Expected Result	Test Result	Status
1	Login Page	Enter valid username and password	Display login page	Successfully displayed	Valid
2	Dashboard Page	Access dashboard interface	Display dashboard page	Successfully displayed	Valid
3	Training Data Page	View training data	Display training data table	Successfully displayed	Valid
4	Testing Data Page	View testing data	Display testing data table	Successfully displayed	Valid
5	MKNN Classification Page	Access classification feature	Display classification page	Successfully displayed	Valid
6	Classification Results Page	View classification statistics	Display classification results table	Successfully displayed	Valid
7	User Management Page	Add new user account	Display total number of users	Successfully displayed	Valid
8	User Dashboard	Access user dashboard	Display user dashboard page	Successfully displayed	Valid
9	User Classification Page	Access user classification feature	Display user classification page	Successfully displayed	Valid
10	User Classification Results Page	View user classification statistics	Display classification results table	Successfully displayed	Valid

Based on the system testing results, all functional features operated as expected and were validated successfully. This indicates that the web-based MKNN classification system is functioning properly in accordance with the designed specifications.

CONCLUSION

The classification system was developed as a web-based application to distinguish between ischemic stroke (SI) and hemorrhagic stroke (SH) using 11 clinical attributes. The symptom data were represented in binary form (0/1) and normalized prior to processing. The method implemented was the Modified K-Nearest Neighbors (MKNN) algorithm with $K = 5$, Euclidean distance measurement, and a weighted voting parameter (α) of 0.5. The system was designed using a context diagram, activity diagram, Data Flow Diagram (DFD), and Entity Relationship Diagram (ERD), and provides two user roles: administrator and user.

The dataset consisted of 100 patient records obtained from RSUD Cut Meutia, divided into 80 training instances and 20 testing instances. The testing results indicated that 15 instances were correctly classified, while 5 instances were misclassified.

The model achieved a precision of 85.71%, recall of 80%, overall accuracy of 75%, and an

F1-score of 82.75%, with an average confidence value of 81.19%. These results demonstrate that the proposed MKNN-based classification system exhibits good performance in classifying stroke types.

REFERENCES

Akmal, K., Faqih, A., & Dikananda, F. (2023). Perbandingan metode algoritma Naïve Bayes dan K-Nearest Neighbors untuk klasifikasi penyakit stroke. *Jurnal Mahasiswa Teknik Informatika, 7*(1). https://doi.org/10.52436/1.jutif.2024.5.4.1951

Ambarita, R. S., & Witarsyah, D., & Hamami, F. (2023). Implementasi metode asosiasi untuk analisis penempatan produk retail. *E-Proceedings, 10*(2), 1519–1527. http://dx.doi.org/10.30813/j-alu.v7i1.5991

Apriani, S., Kurniawan, D., & Rahman, A. (2025). Optimalisasi website pemerintahan desa menggunakan Laravel dan MySQL. *Jurnal Inovasi Komputer (INOKOM), 1*(2), 96–108. https://doi.org/10.71200/inokom.v1i2.76

Aulia, Y., Andriyansyah, A., Suharjito, S., & Nensi, S. W. (2024). Analisis prediksi stroke dengan membandingkan tiga metode klasifikasi decision tree, Naïve Bayes, dan random forest. *Jurnal Ilmu Komputer dan Informatika, 3*(2), 89–98. https://doi.org/10.54082/jiki.90

Darnila, E., Yunizar, Z., & Alinda, D. G. (2021). Internet network classification in Malikussaleh University using Naïve Bayes method. *METHOMIKA: Jurnal Manajemen Informatika dan Komputerisasi Akuntansi, 5*(1), 48–53. https://doi.org/10.46880/jmika.vol5no1.pp48-53

Idris, F. J., Ramadhani, R., & Mutoffar, M. M. (2024). Klasifikasi penyakit kanker paru menggunakan perbandingan algoritma machine learning. *Jurnal Media Akademik (JMA), 2*(2). https://doi.org/10.62281/v2i2.145

Maskuri, N. M. H., K., S., & Bhakti, H. R. M. (2022). Penerapan algoritma K-Nearest Neighbor (KNN) untuk memprediksi penyakit stroke. *Jurnal Ilmiah Intech: Information Technology Journal of UMUS, 4*(1), 130–140. https://doi.org/10.46772/intech.v4i01.751

Naldy, E. T., & Andri, A. (2021). Penerapan data mining untuk analisis daftar pembelian konsumen dengan menggunakan algoritma Apriori pada transaksi penjualan toko bangunan MDN. *Jurnal Nasional Ilmu Komputer, 2*(2), 89–101. https://doi.org/10.47747/jurnalnik.v2i2.525

Novitasari, F., Haerani, E., Nazir, A., Jasril, & Insani, F. (2023). Sistem klasifikasi penyakit jantung menggunakan teknik pendekatan SMOTE pada algoritma Modified K-Nearest Neighbor. *BITS, 5*(1), 274–284. https://doi.org/10.47065/bits.v5i1.3610

Paramitha, Y. N., Nuryaman, A., Faisol, A., Setiawan, E., & Nurvazly, D. E. (2023). Klasifikasi penyakit stroke menggunakan metode Naïve Bayes. *Jurnal Siger Matematika, 4*(1), 11–16. https://doi.org/10.24198/jmi.v16.n1.27804.61-73

Pranajaya, R., & Suwanda, R. (2021). Sistem informasi inventaris perangkat IT menggunakan QR code berbasis website pada Politeknik LP3I Medan. In *Seminar Nasional Sains dan Teknologi Informasi (SENSASI)* (Vol. 3, No. 1, pp. 182–186).

Pratama, D., Insani, F., Yanto, F., & Afrianty, L. (2025). Perbandingan metode K-Nearest Neighbor dan Modified K-Nearest Neighbor untuk klasifikasi keluarga beresiko stunting. *Jurnal Ekonomi, 7*(1), 37–48. https://doi.org/10.31849/zn.v7i1.24698

Ramdani, A., Sofyan, D. C., & Ramdani, F. (2022). Algoritma klasifikasi data mining untuk memprediksi masyarakat dalam menerima bantuan sosial. *Jurnal Ilmiah Sistem Informasi, 1*(2), 39–47. https://doi.org/10.51903/juisi.v1i2.363

Retno, S., Hasdyna, N., & Yafis, B. (2024). K-NN with purity algorithm to enhance the classification of the air quality dataset. *Journal of Advanced Computer Knowledge and Algorithms, 1*(2), 42. https://doi.org/10.29103/jacka.v1i2.15890

Rezky, S. W., Mansyuri, U., Dwiki, G., & Aryono, P. (2025). Perancangan sistem informasi pemesanan sayur dan sembako secara online berbasis web di Mitra Sayur Uni Leni Pasar Kranggot Cilegon. *JATILIMA, 7*(2), 316–329. https://doi.org/10.54209/jatilima.v7i02.1539

Riany, A. F., & Testiana, G. (2023). Penerapan data mining untuk klasifikasi penyakit stroke menggunakan algoritma Naïve Bayes. *Jurnal SAINTEKOM, 13*(1), 42–54. https://doi.org/10.33020/saintekom.v13i1.352

Ricardo, H., & Fajrin, A. A. (2022). Aplikasi sistem pakar mendeteksi kerusakan pada mesin Toyota 4A-FE berbasis web. *Jurnal Comasie, 6*(2), 11–21. https://doi.org/10.37676/jmi.v17i1.1317

Setiawan, A., Waleska, F., Muhammad, A. R., P., & Efrizoni, L. (2024). Pengaruh penyuluhan kesehatan tentang penyakit stroke terhadap tingkat pengetahuan keluarga. (1). https://doi.org/10.56359/kolaborasi.v1i1.5

Shaumi, A. R., Ali, M. F., & Simbolon, M. T. A. M. (2022). Penerapan data mining menggunakan metode teknik classification untuk melihat potensi kepatuhan wajib pajak bumi dan bangunan. *JUKI: Jurnal Komputer dan Informatika, 4*(2), 171–182. https://doi.org/10.53842/juki.v4i2.131

Solikhin, I., Sobri, M., & Saputra, R. (2018). Sistem informasi pendataan pengunjung perpustakaan (Studi kasus: SMKN 1 Palembang). *Jurnal Ilmiah Betrik, 9*(3), 140–151. https://doi.org/10.36050/betrik.v9i03.40

Suwanda, R., Anshari, S. F., Fhonna, R. P., & Setiawan, T. (2025). Penerapan metode Naïve Bayes dalam sistem rekomendasi pemilihan program studi pendidikan tinggi berbasis website. *Jurnal Minfo Polgan, 14*(2), 2361–2369. https://doi.org/10.33395/jmp.v14i2.15858

Suwanda, R., Syahputra, Z., & Zamzami, E. M. (2020). Analysis of Euclidean distance and Manhattan distance in the K-means algorithm for variations number of centroid K. *Journal of Physics: Conference Series, 1566*(1), 012058. IOP Publishing.

Suwanda, R., Yunizar, Z., & Mauliza, N. (2023). Sistem pengaduan pelanggan berbasis website pada PT PLN (Persero) ULP Krueng Geukueh Aceh Utara. *Jurnal Minfo Polgan, 12*(2), 461–467. https://doi.org/10.33395/jmp.v12i2.12432

Tjut Adek, R., Fikry, M., & Khalil, U. (2021). News opinion classification application with Support Vector Machine algorithm using CodeIgniter framework. *Journal of Informatics and Telecommunication Engineering, 5*(1), 160–166. https://doi.org/10.31289/jite.v5i1.5189

Zharif, D. Z., Reza, M., Prasetyo, P., Abyan, M. A., Lathif, T., & Suryanto, M. (2025). Etika dan privasi dalam era kecerdasan buatan di rumah sakit: Systematic literature review. *JATI, 9*(1), 1132–1138. https://doi.org/10.36040/jati.v9i1.12527