

Clustering Algorithm

Implementation of Fuzzy K-Nearest Neighbor Method in Dengue Disease Classification

Aulia Jannah¹, Abdillah Husaini¹, Aulia Ichsan², Mulkan Azhari³

¹ Department of Information Technology, Universitas Muhammadiyah Sumatera Utara, Medan, 20238, North Sumatera, Indonesia

² Department of Information System, Universitas Deli Sumatera, Medan, 20143, North Sumatera, Indonesia

³ Department of Science Data, Universitas Muhammadiyah Sumatera Utara, Medan, 20238, North Sumatera, Indonesia

ARTICLE INFORMATION

Received: January 07, 2024

Revised: February 13, 2024

Available Online: February 15, 2024

KEYWORDS

Dengue Fever; Classification; FKNN; K-Fold Cross Validatio; Confusion Matrixclustering

CORRESPONDENCE

Phone: +62 822-4707-8892

E-mail: aj08062080@gmail.com

A B S T R A K

Dengue hemorrhagic fever (DHF) is a condition brought on by infection with the dengue virus. DHF is a severe illness with hemorrhagic clinical signs that can result in shock and death. One of the four viral serotypes of the genus *Flavivirus* is responsible for DHF. DHF symptoms include fever, joint pain, red skin patches, and others that are similar to those of other illnesses. So that there are no errors in illness prediction, strong accuracy and accuracy are required when classifying DHF patients or not. The Fuzzy K-Nearest Neighbor (FKNN) method is used in this study to classify dengue sickness in order to obtain the best classification outcomes. In this investigation, k was searched for eight times, with values of 3,5,7,9,11,13,15, and 20. Each K's accuracy statistics are 75.15, 75.16, 77.58%, 79.51%, 85.01%, 78.14%, and 75.20 percent. K = 13, which has an accuracy score of 85.01%, yields the highest accuracy.

INTRODUCTION

Dengue hemorrhagic fever (DHF) is a condition brought on by infection with the dengue virus. DHF is a severe illness with bleeding as one of its clinical symptoms, which can result in shock and death. One of the four viral serotypes of the family *Flaviviridae*'s genus *Flavivirus* is responsible for DHF [1]. DHF symptoms include fever, joint pain, red skin patches, and others that are similar to those of other illnesses. Therefore, in order to avoid mistakes in disease prediction, good precision is required for classifying DHF patients.

Before a medical examination is carried out, classification can be used as a first point of reference to determine whether or not a person is a member of the dengue class. Supervised learning is classification. The class of an object whose class is unknown can be predicted via classification. K-Nearest Neighbor, Fuzzy K-Nearest Neighbor (FKNN), Naives Bayes, Neural Network, Support Vector Machines (SVM), and Decision Tree are examples of frequently used categorization techniques. [2].

The Fuzzy K-Nearest Neighbor (FKNN) artificial intelligence (AI) method is the one with the best performance. The Fuzzy K-NN approach classifies data by looking for values that belong to each class. The FKNN method has been widely used by other researchers in identifying various studies such as, research conducted in 2019 regarding "Diagnosis of Chili Diseases Using the Fuzzy K-Nearest Neighbor (FKNN) Method", by obtaining the results of diagnosing diseases that attack chili plants, using testing of the K value 4 times, namely the value of K = 5, K = 10, K = 15 and K = 20, with the test results obtained, namely optimal accuracy of 92% at the value of K = 5 [3]. Another study on the FKNN method's application was carried out in 2019 and was titled "Implementation of the FKNN (Fuzzy K-Nearest Neighbor) Method for Diagnosing Potato Plant Diseases". Results were obtained in the form of results for diagnosing potato plant diseases

<https://doi.org/10.56211/hanif.v1i2.14>

[Attribution-ShareAlike 4.0 International](https://creativecommons.org/licenses/by-sa/4.0/) Some rights reserved

by testing four times at the values of $K = 5$, $K = 10$, $K = 15$, and $K = 20$, with the best accuracy results of 91.4% at the value of $K = 15$ [4]. Another study, "Classification of Chili Plant Quality Using the Fuzzy K-Nearest Neighbor (FKNN) Method," was also carried out in 2017 and produced results with an optimal accuracy of 96.67% of sample data up to 100 data with the division of training data and testing data of 70 for training data and 30 for testing data [5]. In order to diagnose this dengue disease using the Fuzzy K-Nearest Neighbor (FKNN) method and obtain good classification results, the FKNN method has been used in a number of research. By recognizing this DHF sickness, we can aid medical professionals in accurately diagnosing DHF disease and reduce the rising death toll brought on by DHF disease. It is anticipated that the Fuzzy K-Nearest Neighbor (FKNN) approach will produce more precise, accurate classification results when used to identify dengue sickness.

METHOD

Fuzzy K-Nearest Neighbor (FKNN)

Dengue hemorrhagic fever (DHF) is a condition brought on by infection with the dengue virus. DHF is a severe illness with hemorrhagic clinical signs that can result in shock and death. One of the four viral serotypes of the family Flaviviridae's genus Flavivirus is responsible for DHF. Because there is no cross-protection and any serotype can produce an outbreak (hyperendemicity), this is possible. *Aedes aegypti* and *Aedes albopictus* mosquitoes can transmit the virus to humans. Humans, viruses, and intermediate vectors are the three main contributors to the spread of dengue virus infection. Humans contract the dengue virus from the *Aedes aegypti* mosquito. Although they are less significant vectors, *Aedes albopictus*, *Aedes polynesiensis*, and numerous other species can also transmit the virus. When biting people who have viremia, these *Aedes* carry dengue virus. High temperature that reaches 40 degrees Celsius, a severe headache, pain in the muscles, bones, and joints are all signs of dbd. discomfort behind the eyes, reduced appetite vomiting and nauseous, enlarged lymph nodes 2–5 days after a fever, there is a reddish rash [1].

The K-Nearest Neighbor classifier is combined with fuzzy approaches in the fuzzy K-Nearest Neighbor (FKNN) classification method. The membership value of the data in each class is defined by fuzzy set theory, which generalizes the traditional K-NN theory. The FKNN technique determines the value of the test data's membership degree for each class, then takes the largest membership degree value generated. The result class of the classification will be the one with the highest membership degree value.

The training data and test data that will be utilized in the FKNN classification process are input first, followed by a determination of the number of nearest neighbor K values that will be employed in the calculation. Then, on the test data and training data, the distance computation is performed using the Euclidean distance function. The Euclidean distance computation formula [6] is shown in equation (1).

$$d(x_{i1,2}) = \sqrt{\sum_{n=1}^p (x_{i1n} - x_{i2n})^2}$$

There are various symbols used in the computation of the closest distance, including d_i , which stands for proximity distance (also known as Euclidean distance), p , which stands for the number of characteristics, and x_1 , which stands for training data, and x_2 , which stands for test data. The training data are calculated on in the FKNN calculation to determine the data class. The formula for class determination in training data is contained in the equation that can be seen in equation [7].

A training data is identified by the symbols n_j , which stands for the number of individuals in class j in the training data, N , which stands for the number of training data utilized in the study, and J , which stands for the class in the data. The FKNN method's final computation is to ascertain each class's membership. Calculating the membership of each class is done in the equation that can be seen in equation (3) [8].

The membership value taken is the biggest membership value after receiving the results of the membership computation for each class. The classification of the dengue virus using the FKNN algorithm yields the largest membership value.

Cross-validation or often referred to as rotation estimation is a model validation technique to assess the optimality of the analysis results, besides that crossvalidation is also a composition technique in determining the amount of training data

and testing data to be used. There are several methods in cross-validation including the first k-fold method. In the k-fold method, data is randomly segmented into k equal-sized partitions. During the process, one of the partitions is selected to be the testing data, while the rest is used for training data. The second cross-validation method is the holdout method. In the holdout method, the initial labeled data is partitioned into two sets randomly called training data and testing data. The proportion of data reserved for training data and testing data depends on the analysis such as 70%- 30% or 2/3 for training and 1/3 for testing, but according to [9] and [10] in general the ratio used is 2:1 for training data versus testing data.

Utilizing a confusion matrix, the categorization findings of DHF sickness are evaluated. This approach uses a matrix, as shown in Table 1, to display the classification results.

Table 1. Confusion Matrix

Predicted Value	TP	FP
	FN	TN

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$$

General Architecture

This study was conducted in stages, commencing with the process of studying the literature, data gathering, classification utilizing fuzzy K-nearest Neighbor (FKNN) technology, and testing. Literature reviews are used to support the issues brought up in various sources that are connected to the issues raised. Then continue gathering data. The secondary data used in this study was acquired from UCI machine learning. There are 300 data in the set. The data is then standardized using the data pretreatment and normalization processes, ensuring that no one dominates the data throughout testing. The data are normalized using the Min-Max Normalization method, which limits the range of the value scale in all data to 0 to 1. Equation 3 shows the min-max normalization calculation formula.

$$V' = \frac{v - \text{min}A}{\text{Max} A - \text{Min} A}$$

At the data preprocessing stage, the v formula is employed for the min-max normalization method of data normalization. The next step in the normalization process is data division using k-fold cross validation on the DBD dataset after the data has successfully passed the data preprocessing stage. Because all datasets utilized in the study will be used as training data and testing data when testing is conducted, k-fold cross validation is employed in data sharing because the test findings will be more ideal and good [11][12]. The Fuzzy K-Nearest Neighbor (FKNN) approach is then used to classify the data once it has been divided. When utilizing the FKNN method for classification, k selection is used, with values of K = 3, K = 5, K = 7, K = 9, K = 11, K = 13, and K = 20. A confusion matrix is used in testing to determine the accuracy level of classifying dengue disease based on the data that has been used after classification using the Fuzzy K-Nearest Neighbor (FKNN) approach. The FKNN classification flowchart using dengue data is shown in the picture in Figure 1.

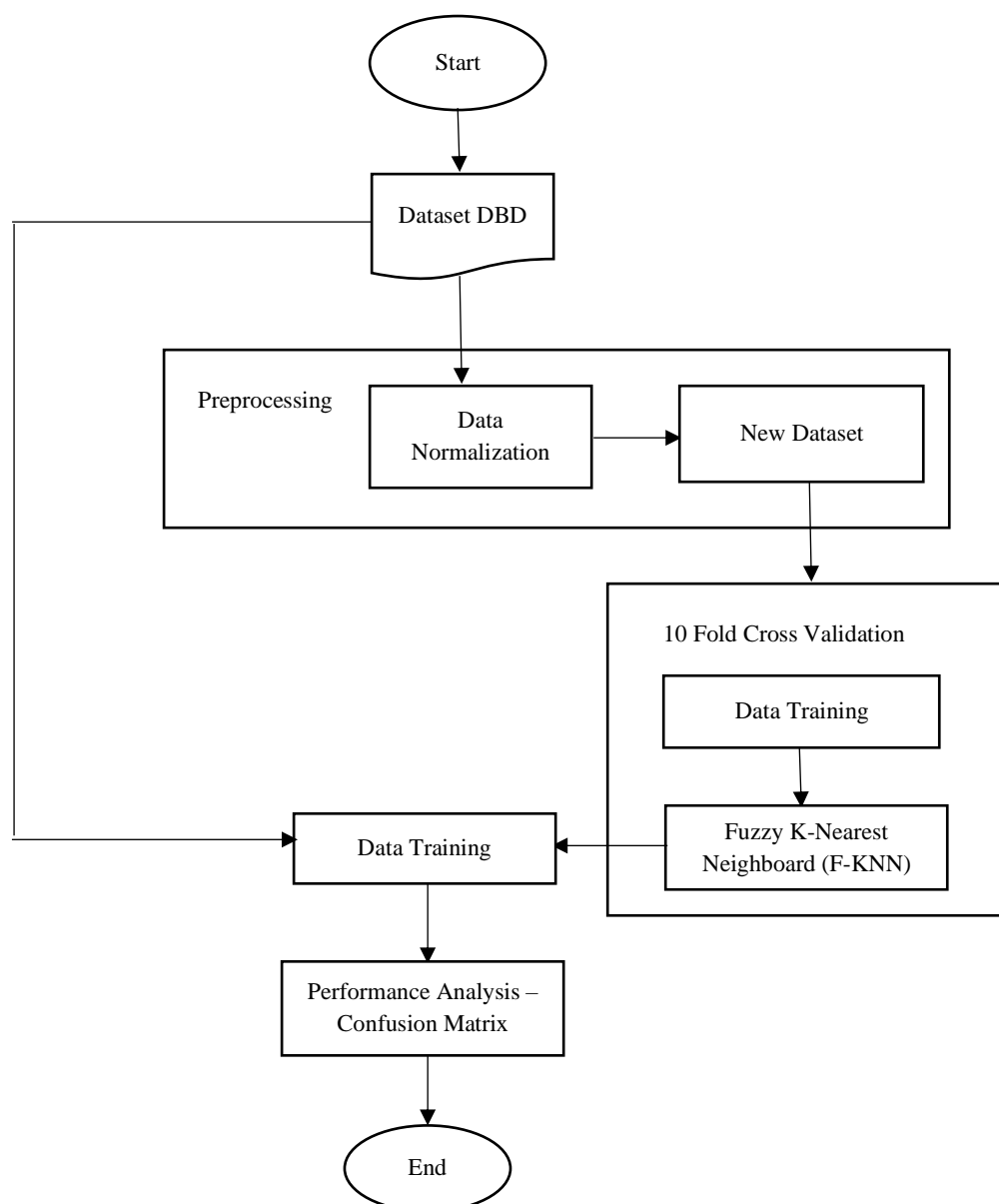


Figure 1. Flowchart of FKNN Classification on DHF Data

The stages in the flowchart above are:

1. Dataset DBD
2. After determining the BDB dataset, it is processed into data normalization
3. Then it is validated into training data
4. And the training data was processed using the Fuzzy K-Nearest Neighborboard method
5. The final stage is Performance Analysis – Confusion Matrix

RESULTS AND DISCUSSION

The Fuzzy K-Nearest Neighbor (FKNN) approach is used in this work to categorize the dengue virus. Secondary data obtained from UCI machine learning was utilised in this investigation. There are two classes, 13 attributes, and 300 data in the set. K-fold cross validation is used to divide the dataset into training and testing data in the study on the categorization of DHF disorders using the Fuzzy K-Nearest Neighbor (FKNN) approach. Data processing for FKNN computations is split by 70% training data and 30% testing data. 240 training data and 60 testing data are present. The 10-fold cross validation process yields data division. Searching for the optimal K value in the initial stage. Testing for the K value employed in this study is done at the following K values: 3, 5, 7, 9, 11, 13, and 20. Table 2 below shows the results of the accuracy value for various K.

Table 2. Accuracy With Different K

	Accuracy
K = 3	75,15 %
K = 5	75,16 %
K = 7	77,76 %
K = 9	79,80 %
K = 11	83,12 %
K = 13	85,01 %
K = 15	78,14 %
K = 17	75,20 %

The study's accuracy findings were calculated utilizing the confusion matrix calculating method. According to some of the accuracy statistics, the Fuzzy KNearest Neighbor (FKNN) technique performs well in classifying DHF illnesses. As can be seen from the data above, K = 13 produces the best accuracy, with an accuracy value of 85.01%.

CONCLUSIONS

Based on the research that has been done, it can be said that the Fuzzy K-Nearest Neighbor (FKNN) method is most accurate at K = 13, with an accuracy of 85.01% in testing, when used to classify dengue diseases, using a dataset division of 70% training data and 30% testing data on 300 datasets.

REFERENCES

- [1] Sari., I.P, Al-Khowarizmi., A, & Batubara., I.H. (2021). Cluster Analysis Using K-Means Algorithm and Fuzzy C-Means Clustering For Grouping Students' Abilities In Online Learning Process. *Journal of Computer Science, Information Technology and Telecommunication Engineering* 2 (1), 139-144
- [2] Sari., I.P, Fahroza., M.F, Mufit., M.I, & Qathrunad., I.F. (2021). Implementation of Dijkstra's Algorithm to Determine the Shortest Route in a City. *Journal of Computer Science, Information Technology and Telecommunication Engineering* 2 (1), 134-138
- [3] Sari., I.P, Batubara., I.H, & Al-Khowarizmi., A. (2021). Sensitivity Of Obtaining Errors In The Combination Of Fuzzy And Neural Networks For Conducting Student Assessment On E-Learning. *International Journal of Economic, Technology and Social Sciences (Injects)* 2 (1), 331-338
- [4] Ramadhani., F, Satria., A, & Sari., I.P. (2023). Implementasi Metode Fuzzy K-Nearest Neighbor dalam Klasifikasi Penyakit Demam Berdarah. *Hello World Jurnal Ilmu Komputer* 2 (2), 58-62
- [5] Sari., I.P, Al-Khowarizmi., A, Ramadhani., F, & Sulaiman., O.K. (2023). Implementation of the Selection Sort Algorithm to Sort Data in PHP Programming Language. *Journal of Computer Science, Information Technology and Telecommunication Engineering* 4 (1)
- [6] Sari., I.P & Batubara., I.H. (2021). Optimization of the FP-Growth Algorithm in Data Mining Techniques to Get the Electric Power Theft Pattern for the Development of Smart City. *2021 4th International Conference of Computer and Informatics Engineering (IC2IE)*, 293-298
- [7] Batubara., I.H, & Sari., I.P. (2021). Combination of Analytic Hierarchy Process (AHP) Method and Profile Matching Method with Matrix Decomposition in Determining Olympiad Candidates. *International Journal of Economic, Technology and Social Sciences* 2, 470-477
- [8] Sari., I.P, Batubara., I.P, Al-Khowarizmi., A, & PP Hariani. (2022). Perancangan Sistem Informasi Pengelolaan Arsip Digital Berbasis Web untuk Mengatur Sistem Kearsipan di SMK Tri Karya. *Wahana Jurnal Pengabdian kepada Masyarakat* 1 (1), 18-24
- [9] Sitompul., D.N, Rahmatika., A, & Sari., I.P. (2023). Application of The Sales and Purchase Program Using The Rapid Application Development Model. *Al'adzkiya International of Computer Science and Information Technology (AloCSIT) Journal* 4 (1), 6-16
- [10] Sari., I.P, A Syahputra, N Zaky, RU Sibuea, & Z Zakhir. (2022). Perancangan sistem aplikasi penjualan dan layanan jasa laundry sepatu berbasis website. *Blend sains jurnal teknik* 1 (1), 31-37

- [11] Batubara., I.H. & Sari., I.P. (2021). Improving Critical Thinkingability Through Guided Discovery Methods Assisted By Cabri 3d Software. *International Journal of Economic, Technology and Social Sciences (Injects)* 2 (1), 325-330
- [12] Apdilah., D, Sulaiman., O.K, & Sari., I.P. (2021). Optimization Of The Fuzzy C-Means Cluster Center For Credit Data Grouping Using Genetic Algorithms. *Al'adzkiya International of Computer Science and Information Technology (AIOCSIT) Journal* 2 (2), 156-163
- [13] Hariani.,P.P, Sari.,I.P, & Batubara., I.H. (2021). Implementasi e-Financial Report BUMDes. *IHSAN: JURNAL PENGABDIAN MASYARAKAT* 3 (2), 169-177
- [14] Sari., I.P, A Azzahrah, FQ Isnaini, L Nurkumala, & A Thamita. (2022). Perancangan sistem absensi pegawai kantor secara online pada website berbasis HTML dan CSS. *Blend sains jurnal teknik* 1 (1), 8-15
- [15] Sulaiman., O.K, Sari., I.P, & Satria., A. (2021). Implementation Data Mining For Level Analysis Traffic Violation By Algorithm Association Rule. *Al'adzkiya International of Computer Science and Information Technology (AIOCSIT) Journal* 2 (2), 128-135
- [16] Sari.,I.P, & Ramadhani., F. (2021). Pengaruh Teknologi Informasi Terhadap Kewirausahaan Pada Aplikasi Perancangan Jual Beli Jamu Berbasis WEB. *Prosiding Seminar Nasional Kewirausahaan* 2 (1), 874-878.
- [17] Sari., I.P, A Jannah, AM Meuraxa, A Syahfitri, & R Omar. (2022). Perancangan Sistem Informasi Penginputan Database Mahasiswa Berbasis Web. *Hello World Jurnal Ilmu Komputer* 1 (2), 106-110.
- [18] Sari., I.P, Ramadhani., F, Satria., A, & Apdilah., D. (2023). Implementasi Pengolahan Citra Digital dalam Pengenalan Wajah menggunakan Algoritma PCA dan Viola Jones. *Hello World Jurnal Ilmu Komputer* 2 (3), 146-157
- [19] Sari., I.P, & Batubara., I.H. (2021). Perancangan Sistem Informasi Laporan Keuangan Pada Apotek Menggunakan Algoritma K-NN. *Seminar Nasional Teknologi Edukasi dan Humaniora (SiNTESa)* 1 (2021 - ke 1
- [20] Ramadhani., F, A Satria, & Sari., I.P. (2022). Aplikasi Internet Berbasis Website sebagai E-Commerce Penjualan Komponen Sport Car. *Blend Sains Jurnal Teknik* 1 (2), 69-75
- [21] Sari., I.P, & Batubara., I.H. (2021). User Interface Information System for Using Account Services (Joint Account) WEB-Based. *International Journal of Economic, Technology and Social Sciences (Injects)*, 462-469
- [22] Batubara., I.H, Sari., I.P, EFS Siregar, & BS Lubis. (2021). Meningkatkan Kemampuan Penalaran Matematika Melalui Metode Penemuan Terpandu Berbantuan Software Autograph. *Seminar Nasional Teknologi Edukasi Sosial dan Humaniora* 1 (1), 699-705
- [23] Sari., I.P, Al-Khowarizmi., A, & Batubara., I.H. (2021). Implementasi Aplikasi Mobile Learning Sistem Manajemen Soal dan Ujian Berbasis Web Pada Platform Android. *IHSAN: JURNAL PENGABDIAN MASYARAKAT* 3 (2), 178-183
- [24] Ramadhani., F, & Sari., I.P. (2021). Pemanfaatan Aplikasi Online dalam Digitalisasi Pasar Tradisional di Medan. *Prosiding Seminar Nasional Kewirausahaan* 2 (1), 806-811
- [25] Sari., I.P, Batubara., I.H, & M Basri. (2022). Implementasi Internet of Things Berbasis Website dalam Pemesanan Jasa Rumah Service Teknisi Komputer dan Jaringan Komputer. *Blend Sains Jurnal Teknik* 1 (2), 157-163
- [26] PP Hariani, Sari., I.P, & Batubara., I.H. (2021). Android-Based Financial Statement Presentation Model. *JURNAL TARBIYAH* 28 (2), 1-16
- [27] Ramadhani., F, Sari., I.P, & Satria., A. (2024). Perancangan UI/UX Surat Keterangan Waris dalam Pengembalian Dana Haji Berbasis Web. *Blend Sains Jurnal Teknik* 2 (3), 198-203.
- [28] Sari., I.P, Sulaiman., O.K, Ramadhani., F, & Satria., A. (2023). Perancangan Sistem Manajemen Surat Berbasis Web Pada Kantor Camat Tano Tombangan Angkola. *INCODING: Journal of Informatics and Computer Science Engineering* 3 (2), 61-76.
- [29] Sari., I.P, Al-Khowarizmi., A, , Jannah., A, Meuraxa., A.M, & Tanjung., M.I. (2023). Web-Based Offline GamSuit Design: A Model Overview. *Journal of Computer Science, Information Technology and Telecommunication Engineering* 4 (2), 389-394.
- [30] Sari., I.P, Sulaiman., O.K, Al-Khowarizmi., A, & Azhari., M. (2023). Perancangan Sistem Informasi Pelayanan Masyarakat pada Kelurahan Sipagimbar dengan Metode Prototype Berbasis Web. *Blend Sains Jurnal Teknik* 2 (2), 125-134.
- [31] Sari., I.P, Al-Khowarizmi, A., Sulaiman., O.K, & Apdilah., D. (2023). Implementation of Data Classification Using K-Means Algorithm in Clustering Stunting Cases. *Journal of Computer Science, Information Technology and Telecommunication Engineering* 4 (2), 402-412