Matrik: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer
Vol. 22, No. 2, Maret 2023, pp. 405∼418
ISSN: 2476-9843, accredited by Kemenristekdikti, Decree No: 200/M/KPT/2020
DOI: 10.30812/matrik.v22i2.2777

Application of KNN Machine Learning and Fuzzy C-Means to Diagnose Diabetes

Anthony Anggrawan , Mayadi
Universitas Bumigora, Mataram, Indonesia

Article Info

Article history:

Received Janaury 17, 2023
Revised February 27, 2023
Accepted March 28, 2023

Keywords:

Diabetes
Machine Learning
Fuzzy C-means
K-Nearest Neighbor

ABSTRACT

The disease is a common thing in humans. Diseases that attack humans do not know anyone and do not know age. The disease experienced by a person starts from an ordinary level until it can be declared severe to the point of being at risk of death. In this study, the early diagnosis was carried out related to diabetes, where diabetes is a condition in which the sufferer's body has low sugar levels above normal. Symptoms experienced by sufferers include frequent thirst, frequent urination, frequent hunger, and weight loss. Based on these problems, a system is needed that can quickly find out the diagnosis experienced by a patient. This research aimed to diagnose diabetes early on based on early symptoms. The methods used are KNN and web-based fuzzy C-means. Creating a web-based system can represent medical personnel experts in a fast-diagnosing approach to diabetes. This system was a computer program embedded with the knowledge of the characteristics of diabetes. The results of testing the KNN and Fuzzy C-means applications and methods get an accuracy of 96% for the K-Nearest Neighbor method, while for the Fuzzy C-Means method with Confusion Matrix calculations, an accuracy of 96% is obtained, so it can be concluded that the Fuzzy C-means method Means better than the K-Nearest Neighbor method.

Corresponding Author:

Mayadi, +6287753166756
Faculty of engineering and Study Program computer science,
Univrsitas Bumigora, Mataram, Indonesia,
Email: mayadi.yadot@universitasbumigora.ac.id

How to Cite:A. Anggrawan and M. Mayadi, "Application of KNN Machine Learning and Fuzzy C-Means to Diagnose Diabetes", *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 22, no. 2, pp. 405-418, Mar. 2023.

1. INTRODUCTION

Illness is a common thing in humans in general. Usually, a person gets sick is caused of an unhealthy lifestyle [1, 2]. Disease in humans has often happened to anyone and knows no age. In this study, the disease studied was diabetes. This disease is one of the most common diseases suffered by humans around the world, and this disease is also a type of disease widely observed in many countries today. [3] Diabetes mellitus is also a disease with high complications, so it needs continuous medical treatment to reduce the impact of complications with glycemic checks. [4, 5] The number of people with diabetes from year to year is increasing. Approximately 14,200 to 18,900 global deaths are due to diabetes (type 1 and 2) [6]. Diabetes patients in Indonesia amounted to 10 million people in 2015 Diabetes patients in Indonesia amounted to 10 million people in 2015. Referring to data from the International Diabetes Federation, it is predicted that people with diabetes in Indonesia according to the WHO report, the number of DM sufferers in Indonesia will increase from 8.4 million in 2000 to around 12 million in 2030 and will increase to 16.2 million in 2040 [5, 7]. Another important thing that needs attention in medical services is the problem of the number of costs and the problem of inflexibility in medical services [8]. The process of medical examinations that are carried out manually often causes delays in the results of medical diagnoses [8]. In essence, by utilizing an electronic system, the treatment process can save or speed up treatment time or achieve better work efficiency, including saving costs [9]. The utilization of electronic systems results in fewer errors occurring. However, it also eliminates the possibility of failure in deciding the results of medical examinations or tests and failure in achieving the correct results. However, the success of curing diabetes is still limited; this includes a lack of success in detecting populations at risk of developing early diabetes early, resulting in increased rates of acute diabetes [10]. That is, by not quickly accepting or knowing the early symptoms of diabetes. You will be at high risk and make the people around you suffer the destructive effects of the disease [10]. Conversely, by being able to identify early and precisely, it can reduce the consequences of the more severe risks that occur [9, 10] In other words, it is important to develop various detection systems that can classify and identify someone susceptible to certain diseases [10]. There are two types of diabetes, namely type 1 diabetes which is insulin-dependent diabetes, in which the pancreas produces little insulin or does not produce insulin at all. Whereas in type 2 diabetes, the pancreas continues to produce insulin, but sometimes the levels are higher than normal where this occurs. will cause the body to develop immunity to its effects, resulting in a relative deficiency of insulin [4]. The initial symptoms of diabetes are usually preceded by three conditions, namely polyuria (increased bladder expenditure), polydipsia (excessive thirst), and polyphagia (increased hunger) [4, 11, 12]. People with diabetes who have experienced complications such as neuropathy (63.5%), retinopathy (42%), nephropathy (7.3%), macrovascular (16%), microvascular (6%), and diabetic foot wounds (15%) [13], while the mortality rate due to diabetic foot ulcers and gangrene reaches 17-23% and the amputation rate reaches 15-30%, besides that the mortality rate one year after amputation is 14.8% [3, 14, 15]. Given the above problems, a research system is needed to detect diabetes recipients early to reduce the increase in people with diabetes. This study aims to diagnose diabetes early by utilizing data mining, namely the KNN method and fuzzy c-means, then applying machine learning based into a WEB-based system so that it can be accessed anywhere and without expert assistance. In this research, the reason for using KNN and fuzzy c-means is that the Fuzzy -Means Cluster can introduce patterns that are more flexible and make it easier to solve calculations on diabetes data based on the formulated problems [14, 16], while the reason for using the KNN method is that it is widely used in research and is easy to process in the numerical form [15, 17, 18]. Meanwhile, Machine Learning in computer science is growing rapidly [10, 19]. Although in most scientific studies, machine learning is popular, and it is still very limited in health studies [18]. Machine learning can be used to assist in mining data so that it can be used to predict mining results accurately [20]. Machine learning is a technique that helps find case-based correlations to predict [21]. With the availability of big data, it is possible to develop various solutions using machine learning [22, 23]. With the development of advances in information and communication technology [24], it is very easy to obtain and collect the required big data. Predictive modeling is one solution that can be used in machine learning [25–27]. In addition, machine learning can uncover hidden patterns in big data so that patterns can be distinguished better and more accurately [11] so that machine learning can provide predictive results with high accuracy [10].

The following is a discussion of several studies related to this research carried out by previous researchers: T.K. Abbas, J.A.Hataf, and M.A. Abbas (2022) conducting research related to diabetes mellitus. This paper focused on five classification algorithms in machine learning, namely MLP, SVM, KNN, DT, and enhanced ANN algorithms. The difference from the current research is in terms of methods, namely using the classification method with KNN and clusters using Fuzzy C-means. For the final finalization, the previous research only analyzed without making applications, while in the current research, it was implemented into WEB-based applications [28]. Research conducted by S. Pungkas, S. Irfan, A. Tri (2017) entitled Comparison of the Performance of the CART and Nave Bayesian Algorithms for Diagnosing Diabetes Mellitus. This study used the CART and Nave Bayes methods using a dataset from the Pima Indian UCI database repository consisting of clinical data of patients who were detected positive and negative for diabetes mellitus. CART is 76.9337%, and Nave Bayes gets an accuracy value of 73.7569% [7]. Whereas the current research uses patient data from the hospital and uses the KNN and Cluster methods with Fuzzy C-means in the previous study only analyzed results. However, the present study implements it into a WEB-based system. R.M.Fadly, D.M.Ilham, and A.A.Dion (2017) discussed the performance of the Bayesian regularization neural network method with diabetes data related to the use of neurons in the hidden layer, which can affect the accuracy of the results of the classification process (the more, the more accurate), this is because by changing the number of neurons in the hidden layer, we can also change the network structure of the RBNN method (results can be optimal or not) [29]. While what is being done in the current research is to carry out the early diagnosis of diabetes with the KNN and Nave Bayes methods and then implement it into a WEB-based system. Based on the discussion from previous studies, this research is different. It has updated the method used and aims to increase the accuracy of predictions using a more dominant method known as data mining. In short, the research in this article differs from the prediction method used, but this research also builds Web-based applications that previous researchers did not do. With the existence of a built web application, it is useful for every layperson to predict diabetes based on an expert system that is built without the need to involve experts (medical doctors). This research has very useful implications for the benefit of the general public and also for the medical world, especially doctors and medical nurses, in predicting people with diabetes using a web-based electronic system (Android web).

The structure of writing this research is that the next section discusses the research methodology, namely a brief discussion of the methodology used in this study. Then, the third sub-section discusses Results and Discussion, which means explaining how to design a website-based application system interface and includes testing the website application system that was built and the results achieved. Finally, at the end of the manuscript, the conclusions from the research results are discussed and placed in the Conclusion sub-section.

2. RESEARCH METHOD

In this study, the process of research methodology uses several stages. The stages carried out in this study are shown in Figure 1:

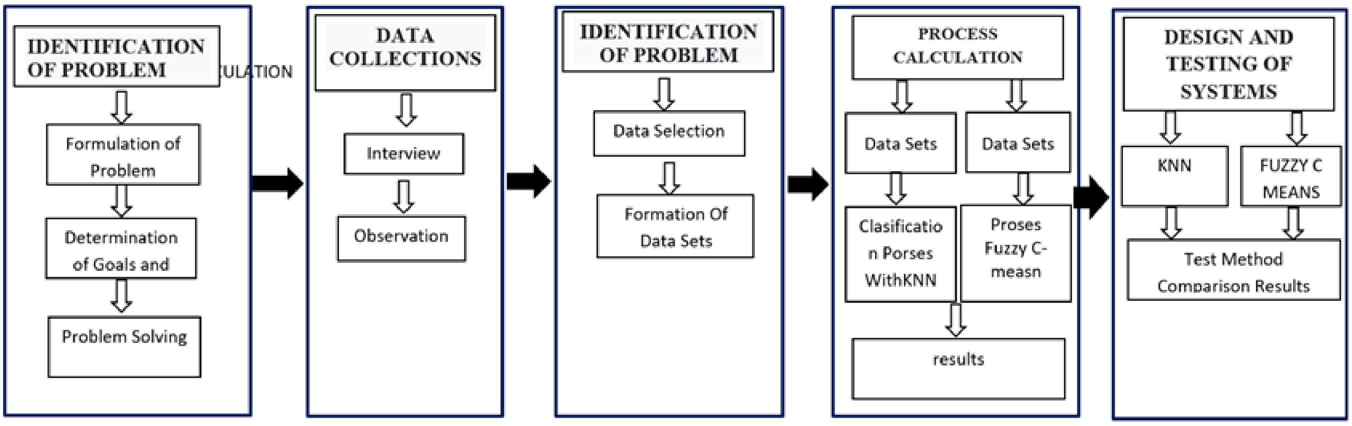

Figure 1. The research methodology process

2.1. Identification of problems

Identifying problems in this study is a step to finding out problems related to diabetes. Therefore, this process is carried out to collect information related to diabetes problems. Furthermore, this stage is carried out to obtain symptoms from patients and diabetes specialist doctors, where this information is useful for building expert systems with machine learning.

2.2. Data collection

This data collection stage is carried out through interviews and observations, and the process is carried out to related parties, for symptoms to experts, and diagnoses to patients. Interviews and observations were conducted to obtain information about the data used in the diagnosis. The interview process was conducted with experts (**parties at the Tanjung Health Center**) and obtained several symptoms, namely fatigue, difficult-to-heal wounds, blurred vision, frequent hunger and thirst, and a history of heredity [30].

2.3. Data Preprocessing

Preprocessing is one of the stages in the data mining process. This stage is also the process of converting raw data into a form that is easy to understand. In this study, the authors divided the preprocessing stage into several processes, including data weighting and dataset formation.

2.4. Expert System Design and Testing

Tahapan The design and testing phase is the design and performance process of the system being built. Testing systems built with machine learning must produce accurate performance, both system performance processes and process results from methods in the form of accurate resultsthe results of this accuracy test aim to determine how accurate the system is in diagnosing diabetes. The amount of data used to diagnose diabetes is 120 patients data with five symptoms.

3. RESULT AND ANALYSIS

3.1. Problem Identification

From helping to get problems related to knowledge, this process or stage aims to obtain the required data knowledge. Programming knowledge obtained to obtain useful data in solving the logic of diagnosing diabetes.

3.2. Data collection

In collecting data, there are two types of attributes and the process of determining the data: interviews and lab test results. At the time of the interview, information was obtained about how a person could get diabetes and how the examination process was in the lab. Lab tests were used to test whether diabetes was positive and to collect information on what factors led to diabetes and what data was obtained. In the interview process, five questions related to diabetes symptoms were obtained with 120 data. Table 1 shows the symptom survey instrument experienced by the patient, while Table 2 contains symptom data obtained from the survey.

Table 1. Symptom Question

No.	Question Symptom
1	Often Feeling Tired
2	Wounds are hard to heal
3	Blurred vision
4	Frequently Feeling Hungry (Pholyphagia)
5	History of Descendants

Table 2. Symptom Data from Interviews

No.	Often Feeling Tired	Hard To Heal Wounds	Blurred vision	Often Feeling Hungry	History of Descendants
1	Often Feeling Tired	Very often	often	Very often	Yes
2	Very often	No	Seldom	Seldom	No
3	No	Seldom	Seldom	Very often	No
4	Very often	Very often	often	Very often	Yes
5	Very often	often	Very often	Very often	Yes
6	Often Feeling Tired	Seldom	Very often	Very often	Yes
7	seldom	Seldom	Seldom	No	No
8	often	Very often	often	Very often	Yes
9	NO	No	often	No	No
10	seldom	No	Seldom	No	No
11	Very often	often	often	Very often	Yes
18	...	...	...	...	...
19	...	...	...	...	...
119	seldom	Very often	often	often	Yes
120	Very often	often	Seldom	Very often	Yes

3.3. Data Preprocessing

1. Data Weighting

Stages of data weighting aim to give value to each attribute. The weighting results will be used for the calculation process in the system. The attributes used are: often feeling tired, difficult to heal wounds, blurred vision, often feeling hungry (polyphagia), and history of heredity, while the weighting levels are in Table 3.

Table 3. Attribute Weight Value

No	Information	Weight
1	No	0
2	Seldom	1
3	Often	2
4	Very often	3

So that in Table 3, on the symptom weighting score given by the expert on the certainty factor, a score of 0 indicates that the user does not experience these symptoms. If the answer to the question is rare, then the patient rarely experiences symptoms with a score of 1, for often, the patient experiences the symptoms asked with a score of 2, and the answer is very often a score of 3.

2. Formation of datasets.

The formation of datasets in this study is the process of changing data into sentences and then converting them into weights, where the goal is to be processed using the KNN and Fuzzy C-means methods. The weighting results are shown in Table 4.

Table 4. Attribute Weight Values

Patient	Often Feel tired	Wound Hard to Cure	Vision Blur	Feel Often Hungry (Polifagia)	History of Descendants	Status
1	2	3	2	3	1	Diabetes
2	3	0	1	1	0	Negatives
3	0	1	1	0	0	Negatives
4	3	3	2	3	1	Diabetes
5	3	2	3	3	1	Diabetes
6	2	1	3	3	1	Diabetes
7	1	1	1	0	0	Negatives
8	2	3	2	3	1	Diabetes
9	0	0	2	0	0	Negatives
10	1	0	1	0	0	Negatives
11	3	2	2	3	1	Diabetes
12	2	3	3	3	1	Diabetes
13	0	1	1	1	0	Negatives
14	2	3	3	3	1	Diabetes
...	...	...	...	...	...	...
120	3	2	1	3	1	Diabetes

Table 4 is the result of weighting the data. The weighted value entered is the weight of the questions given to the patient. The goal is to be able to perform calculations using KNN and fuzzy c-means. The results of each weighting are then divided into two classes, namely negative and diabetic.

3.4. Testing the Machine Learning Method with KNN

Testing the performance of the machine learning method is carried out with a sample of diabetes patient case data based on Table 4 in the process of testing the machine learning method using KNN. For example, in one case, medication The patient has symptoms: Often feeling tired: very often, Wound does not heal: often, Blurred vision: rarely, and Often feeling hungry (Polyphagia): very often. Family history: yes Status: ?
The first step is

1. Determine the number of neighbors (k) that will be used as a consideration for class determination. For example, suppose K=3. Then, calculate the distance between the test data and the training data.
2. Perform calculations with the Euclidean Distance method. The training data consists of 30 data, which will be calculated individually with data testing to determine the distance for each data.
Data 1
$dis = \sqrt{(2-3)2+(3-2)2+(2-1)2+(3-3)2+(1-1)2} = 1,732$
Data 2
$dis = \sqrt{(3-3)2+(0-2)2+(1-1)2+(1-3)2+(1-1)2} = 3$
Data 3
$dis = \sqrt{(0-3)2+(1-2)2+(1-1)2+(0-3)2+(0-1)2} = 4,472$
Data 4
$dis = \sqrt{(3-3)2+(3-2)2+(2-1)2+(3-3)2+(1-1)2} = 1,414$
...
...
$dis = \sqrt{(1-3)2+(0-2)2+(1-1)2+(0-3)2+(0-1)2} = 4,24$
Data 26
$dis = \sqrt{(0-3)2+(0-2)2+(1-1)2+(0-3)2+(0-1)2} = 4,79$
Data 27
$dis = \sqrt{(1-3)2+(1-2)2+(1-1)2+(0-3)2+(0-1)2} = 3,87$
Data 28
$dis = \sqrt{(3-3)2+(3-2)2+(1-1)2+(3-3)2+(1-1)2} = 1$
...
Data 120
$dis = \sqrt{(1-3)2+(1-2)2+(0-1)2+(0-3)2+(0-1)2} = 4$
3. Retrieve data with the shortest distance.
Based on the test data, after calculating with the Euclidean Distance formula above, sort them according to the smallest value based on the neighbor value (k), namely k = 3. The results obtained are k (closest neighbors) data 11, 19, and 28. They are shown in Table 5.

Table 5. Data with the Shortest Distance

Patient	Often Feeling Tired	Wounds Hard to Heal	Blurred vision	Often Feeling Hungry (Polifagia)	History Descendants	Status
11	3	2	2	3	1	Diabetes
19	3	3	1	3	1	Diabetes
28	3	3	1	3	1	Diabetes

4. Define Class. From the data calculated in Table 5, it can be concluded that the patient in the previous test data had diabetes. This is evidenced by the 3 data with the closest distance whose status also has a history of diabetes

3.5. Trial of the Fuzzy C-means Machine Learning Method

In the calculation process with fuzzy c-means with data in Table 4, The initial steps in calculating fuzzy-means are [31, 32]: Determine the number of clusters (c), rank (w), maximum iteration (MaxIter), smallest error expected (), the initial objective function (P0) and initial iteration. The values used are as in Table 6.

Table 6. Initial Terms of the Fuzzy C-Means Process

Number of Clusters	c	2
Rank	w	2
Maximum Iteration	MaxIter	100
Smallest Error	e	0,01
F. Initial Purpose	P0	0
Initial Iteration	t	1

Determine the membership of the cluster randomly, which, if added up = 1. After that, the 1st iteration is carried out. The trial data used is diabetes patient data in Table 7, with a total of 120, and comes from patient data at the health center with predetermined parameters.

Table 7. Random Membership Results

	C1	**C2**	**Amount**
1	0,3	0,7	1
2	0,4	0,6	1
3	0,9	0,1	1
4	0,8	0,2	1
…	…	…	…
24	0,1	0,9	1
25	0,6	0,4	1
26	0,2	0,8	1
…	…	…	…
120	1	0	1

From the calculation results, the iteration stops for up to 5 iterations, and the objective function fulfills = 0.001. Then the iteration process can be stopped. From the final results, the results of cluster 1 were 54 and cluster 2 were 64 in Table 8 as follows:

Table 8. Final Process Cluster Results

ID	Nama	C1	C2	Cluster
A001	pasien 1	0.035	0.965	C2
A002	pasien 2	0.677	0.323	C1
A003	Pasien 3	0.921	0.079	C1
A004	Pasien 4	0.033	0.967	C2
…	…	…	…	…
A107	Pasien 107	0.332	0.668	C2
A108	Pasien 108	0.937	0.063	C1
A109	Pasien 109	0.076	0.924	C2
…	…	…	…	…
A119	Pasien 119	0.205	0.795	C2
A120	Pasien 120	0.094	0.906	C2

3.6. System Development and Expert System Testing

The expert system design stage in this study is to make an initial design for modeling, and the model is built based on the data that has been collected. System design or development is planning use case design diagrams, data flow diagram designs (DFD), database designs, and flowcharts built on application programs. After passing through the design stage, then programming development is carried out. The programming stage is implementing the system design into a computer programming language. The programming language used to build applications is PHP, and database creation with MySQL. The expert system application built is stored on a server computer which aims to be accessible anywhere [33]. Untuk mendapatkan hasil yang diharapkan tahap berikutnya melakukan akuisisi pengetahuan untuk memperoleh data pengetahuan yang dibutuhkan. Pengetahuan yang didapat akuisisi data berguna dalam memecahkan logika pemrograman mendiagnosis penyakit diab To get the expected results, the next step is to acquire knowledge to obtain the required knowledge data. The knowledge gained from data acquisition is useful in solving programming logic for diagnosing diabetes.

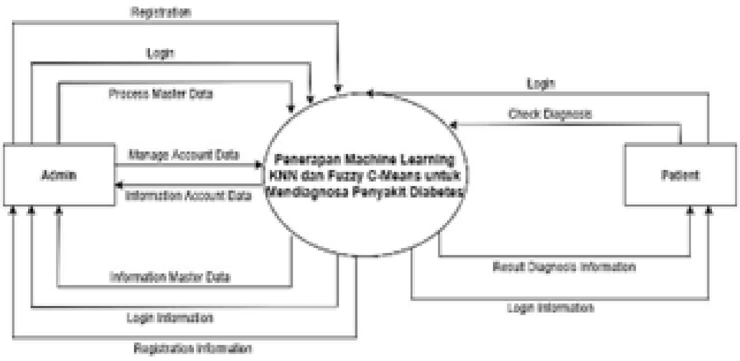

Figure 2. Data flow context diagram in machine learning expert systems

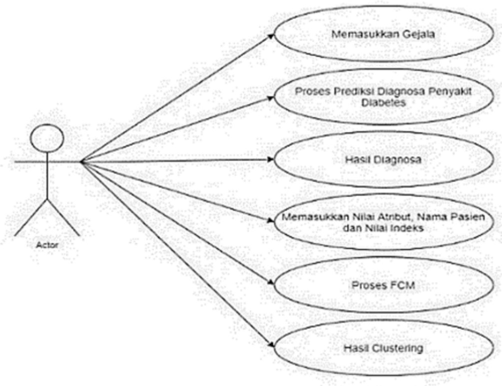

Figure 3. Use case on expert system machine Learning

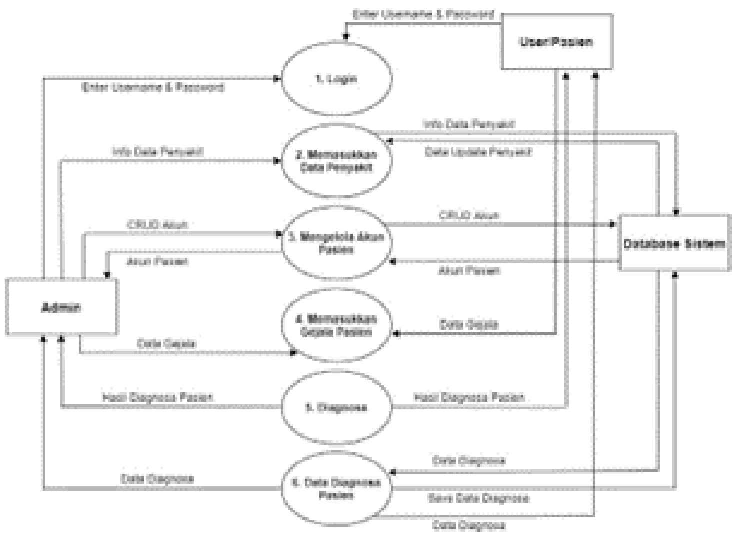

Figure 4. Overview of data flow diagrams in Machine Learning

Figure 2 and Figure 4 illustrate the Data Flow Diagram (DFD) or data flow originating from and where the data is processed in the expert system that was built. The context diagram in Figure 3 shows the data flow of the global system.

In contrast, the use case description in Fig. 2 shows a more detailed data flow that the system performs and engages with external data. The flow chart in Figure 5 shows a series of flow relationships in the expert system built in this study or a demonstration of the overall process sequence in building the expert system in this lesson. The flowchart contains a more detailed description of how each step of the procedure is carried out, building an expert system on machine learning that can diagnose the user and the type of drug used by the user.

Figure 5. Disease diagnosis page in Machine Learning.

Figure 5 shows the page for making a diagnosis. This diagnosis page consists of the patient's name, often feeling tired, wounds that are difficult to heal, blurred vision, frequent hunger, and family history. In filling in the symptoms, first, enter the value of K. Next, the value of K finds the closeness value to determine the final value or class prediction results. Whereas in the Fuzzy C-means process, it is not necessary to determine the value of K, as shown in Figure 7.

Figure 6. The disease diagnosis results in a page in Machine Learning

Figure 6 is the process of the diagnosis results. The results of the diagnosis explain the results of diagnosing the disease. The prediction results consist of the patient's name, the results of the diagnosis, and the accuracy of the diagnosis.

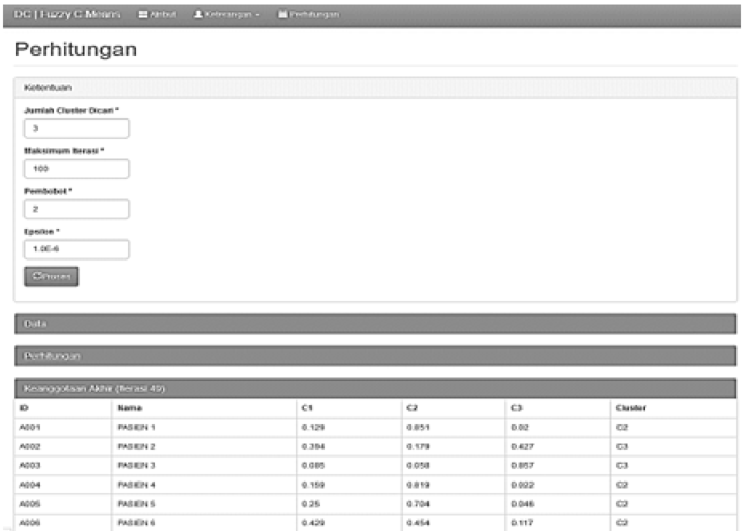

Figure 7. Process page and fuzzy c-means results

Figure 7 is a page for diagnosing calculations using fuzzy c-means. The process on the diagnostic page with fuzzy c-means consists of filling in the number of clusters, the number of iterations, the number of weights, and epsilons. For each condition, it is mandatory to fill in because in this study, looking for 3 clusters must fill in 3 clusters. A maximum iteration process of 100 can be less or more, epsilon (smallest error) 0.01, and weighing 2. From this process, the calculation results for cluster 1 are obtained; cluster 1 is as many as 54, and cluster 2 is as many as 64.

3.7. Method Accuracy Testing

From the results of calculating accuracy with the Confusion Matrix, the results of an accuracy comparison are obtained where the Fuzzy C-Means method is 96% better than the K-Nearest Neighbor method of 86.33% by showing the comparison of accuracy values in Table 9 as follows:

Table 9. Accuracy Comparison

K-Nearest Neighbor	Fuzzy C-Means
83,33%	96%

Based on the comparison shown in Table 9, fuzzy C-means gets the largest value compared to K-Nearest Neighbor and, as a comparison with the previous fuzzy c-means process, which has the highest accuracy.

Table 10. Comparison of the Work of This Article with Several Previous Related Works

Research By	Metode	Attribute	Building Applications		Accuracy		Data source	
			Ya	Tidak	90%	90%	Lab	Question and answer
Putu I & Lesmana D	Decision Tree J48	9	-	✓	✓	-	✓	-
Hana M	Decision Tree C4.5	17	-	✓	✓	-	✓	-
Aris F & Benyamin	C4.5	8	-		-	-	✓	-
Andanika S & Fitriyani	C4.5	15	-	✓	-	✓	✓	-
Islam S, Banik P, Rahman M et al	-			-	-	-	-	
Maulidah N, Abdilah A, Nurlelah E et al	Nave Bayes	8	-	✓	✓	-	✓	-
Dennedy M, Rizza R, & Dinneen S		-	-	-	-	-	✓	✓
Ramadhan P	Sistem Pakar	-	-	-	-	-	-	✓
Achenbach P, Hippich M, Zapardiel-Gonzalo J et al	Decision Tree CART	-	-	-	-	-	✓	✓
Penelitian Sekarang	Fuzzy C-means dan KNN	6	✓	-		✓	✓	✓

Based on Table 10, several differences have never been made in research. New things that have never been done by other researchers before. Table 10 shows some comparisons of the differences between the work of this article and several related previous works.

4. CONCLUSION

For the results of research on diagnosing diabetes using the K-Nearest Neighbor and Fuzzy C-Means methods, the following conclusions are obtained. After testing with the K-Nearest Neighbor method, the highest accuracy was obtained at 83.33%, while the Fuzzy C-Means method obtained an accuracy of 96%. Therefore, based on the results of comparative testing of accuracy and working methods, it can be concluded that the Fuzzy C-Means method is better than the K-Nearest Neighbor because, in this study, Fuzzy C-means has the highest value, application development uses Fuzzy C-means.

The novelty of this research is that this research article is not the same as previous research in terms of the research method used. Another novelty of this research is to build a web-based application that facilitates the work of medical experts and can be used by ordinary people to predict whether someone has diabetes accurately. The results of this study increase the accuracy (has results of accuracy) of diabetes predictions that have been carried out by previous researchers, including research conducted by: T.K.Abbas, J.A.Hataf, M.A.Abbas (2022), S. Pungkas, S. Irfan, A.Tri (2017) and R.M. Fadly, D.M. Ilham, A.A.Dion (2017).

This study has limitations in predicting diabetes using KNN and Fuzzy C-means, although many other data mining methods exist. Therefore, it is suggested for further research to conduct research with other methods such as random forest, SVM, ANN, C4.5, and other methods, as well as other types of diseases.

5. ACKNOWLEDGEMENTS

alhamdulillah Praise and gratitude the author goes to the presence of God Almighty for all His grace and gifts so that this journal can be completed. Thank God, the results of this journal are very good, as expected both from the process method and the system design that was built, hopefully it can be utilized and help ease the burden on hospitals, especially in treating patients. and for future research it is hoped that it can increase the accuracy of adding symptoms, methods and adding data.

6. DECLARATIONS

AUTHOR CONTRIBUTION

In this study compiled by the authors who are divided into their respective tasks. Anthony Anggrawan Develop, analyze and test the system, find references and interpret data. Mayadi collects data and analyzes the methods used, designs tables and makes system designs

FUNDING STATEMENT
in this research did not receive grants from anywhere including the public, commercial, or non-profit sectors.

COMPETING INTEREST
In this study, there are no reserves related to competing financial, public and institutional interests.

REFERENCES

[1] P. W. Ramadhan, "Sistem Pakar Diagnosa Penyakit Umum Menggunakan Metode Forward Chaining Berbasis Web," *Jurnal Informatika*, vol. 6, pp. 1–14, 2019.

[2] F. Handayanna, "Diabetes Mellitus Dengan Metode Support Vector Machine Berbasis Particle Swarm," *Jurnal Teknik Informatika STMIK Antar Bangsa*, vol. II, no. 1, pp. 30–37, 2016.

[3] R. A. Siallagan and Fitriyani, "Prediksi Penyakit Diabetes Mellitus Menggunakan Algoritma C4.5," *Jurnal Responsif : Riset Sains dan Informatika*, vol. 3, no. 1, pp. 44–52, 2021.

[4] M. Dennedy, R. Rizza, and S. Dinneen, *Classification and Diagnosis of Diabetes Mellitus. Endocrinology: Adult and Pediatric*. Elsevier Inc., 2015.

[5] F. M. Hana, "Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.5," *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 32–39, 2020.

[6] E. Cousin *et al.*, "Diabetes mortality and trends before 25 years of age: an analysis of the Global Burden of Disease Study 2019," *The Lancet Diabetes and Endocrinology*, vol. 10, no. 3, pp. 177–192, 2022.

[7] P. Subarkah, I. Santiko, and A. Tri, "Perbandingan Kinerja Algoritma Cart dan Naive Bayesian untuk Mendiagnosa Penyakit Diabetes Melitus," *Conference on Information Technology, Information System and Electrical Engineering*, p. 17, 2017.

[8] A. Bevan and N. Patel, "An Electronic Prescription Alerting System-Improving the Discharge Medicines Process," *Archives of disease in childhood*, vol. 101, no. 9, p. e2, sep 2016.

[9] A. Tsyben, N. Gooding, and W. Kelsall, "Assessing the Impact of a Newly Introduced Electronic Prescribing System Across a Paediatric Department Lessons Learned," *Archives of Disease in Childhood*, vol. 101, no. 9, pp. e2—-e2, 2016. [Online]. Available: https://adc.bmj.com/content/101/9/e2.13

[10] A. Anggrawan *et al.*, "Machine Learning for Diagnosing Drug Users and Types of Drugs Used," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 11, pp. 111–118, 2021.

[11] F. Aris and Benyamin, "Penerapan Data Mining untuk Identifikasi Penyakit Diabetes Melitus dengan Menggunakan Metode Klasifikasi," *Router Research*, vol. 1, no. 1, pp. 1–6, 2019.

[12] I. P. D. Lesmana, "Pengembangan Decision Tree J48 Untuk Diagnosis Penyakit Diabetes Mellitus," *Seminar Nasional Teknologi Informasi & Komunikasi Terapan*, vol. 2, no. 1, p. 23, 2012.

[13] P. Achenbach *et al.*, "A Classification and Regression Tree Analysis Identifies Subgroups of Childhood Type 1 Diabetes," *eBioMedicine*, vol. 82, p. 104118, 2022. [Online]. Available: https://doi.org/10.1016/j.ebiom.2022.104118

[14] N. Dwitiyanti, N. Selvia, and F. R. Andrari, "Penerapan Fuzzy C-Means Cluster dalam Pengelompokkan Provinsi Indonesia Menurut Indikator Kesejahteraan Rakyat," *Faktor Exacta*, vol. 12, no. 3, p. 201, 2019.

[15] Y. Ambar, Kusrini, and Henderi, "Penerapan Algoritma K-Nearest Neighbour Dalam Menentukan Pembinaan Koperasi Kabupaten Kotawaringin Timur," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, no. 5, pp. 1–8, 2019.

[16] N. G. A. Dasriani, M. Mayadi, and A. Anggrawan, "Klasterisasi Lokasi Promosi PMB Dengan Fuzzy C-means Masa Pandemi Covid 19," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 2, pp. 327–336, 2022.

[17] Wahyono, "Peningkatan Kecepatan Algoritma K-NN Untuk Sistem Pengklasifikasian Kendaraan Bermotor," *Techno.Com*, vol. 19, no. 2, pp. 190–196, 2020.

[18] Z. Zhang, "Introduction to machine learning : k-nearest neighbors," *Annals of Translational Medicine*, vol. 4, no. 11, pp. 1–7, 2016.

[19] D. Prasad *et al.*, "System Model for Prediction Analytics Using K-Nearest Neighbors Algorithm," *Journal of Computational and Theoretical Nanoscience*, vol. 16, no. 10, pp. 4425–4430, 2019.

[20] P. Guleria, S. Ahmed, A. Alhumam, and P. N. Srinivasu, "Empirical Study on Classifiers for Earlier Prediction of COVID-19 Infection Cure and Death Rate in the Indian States," *Healthcare (Switzerland)*, vol. 10, no. 1, 2022.

[21] A. Yosipof, R. C. Guedes, and A. T. García-sosa, "Data Mining and Machine Learning Models for Predicting Drug Likeness and Their Disease or Organ Category," *Frontiers in Chemistry*, vol. 6, no. May, pp. 1–11, 2018.

[22] J. Bullock, A. Luccioni, K. H. Pham, and M. Luengo-oroz, "Mapping the Landscape of Artificial Intelligence Applications against COVID-19," *Journal of Artificial Intelligence Research*, vol. 69, no. 1, pp. 807–845, 2020.

[23] L. Wynants *et al.*, "Prediction models for diagnosis and prognosis of covid-19 : systematic review and critical appraisal," *The BMJ*, vol. 369, pp. 1–22, 2020.

[24] A. Anggrawan, "Interaction between Learning Preferences and Methods in Face-to-Face and Online Learning," *ICIC Express Letters*, vol. 15, no. 4, pp. 319–326, 2021.

[25] J. Gong *et al.*, "A Tool to Early Predict Severe Corona Virus Disease 2019 (COVID-19) : A Multicenter Study using the Risk Nomogram in Wuhan and Guangdong, China," *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 0813, no. April, pp. 1–11, 2018.

[26] X. Jiang *et al.*, "Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity," *Computers, Materials and Continua*, vol. 63, no. 1, pp. 537–551, 2020.

[27] J. Liu *et al.*, "Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage," *Journal of Translational Medicine*, vol. 18, no. 1, pp. 1–12, 2020. [Online]. Available: https://doi.org/10.1186/s12967-020-02374-0

[28] K. A. Tikki, H. H. Jaber, and A. A. Mahdi, "Estimation of Serum Uric Acid in Type II Diabetic Patients and Relationship with Anemia," *HIV Nursing*, vol. 22, no. 2, pp. 1523–1526, 2022.

[29] M. Rahman, D. Alamsah, M. Darmawidjadja, and I. Nurma, "Klasifikasi Untuk Diagnosa Diabetes Menggunakan Metode Bayesian Regularization Neural Network (RBNN)," *Jurnal Informatika*, vol. 11, no. 1, p. 36, 2017.

[30] E. Purwaningsih and E. Nurelasari, "Penerapan K-Nearest Neighbor Untuk Klasifikasi Tingkat Kelulusan Pada Siswa," *Syntax : Jurnal Informatika*, vol. 10, no. 01, pp. 46–56, 2021.

[31] A. W. Fadillah, A. Tejawati, and N. Puspitasari, "Penerapan Fuzzy C-Means Pada Curah Hujan Di Kalimantan Timur," *Jurnal Rekayasa Teknologi Informasi (JURTI)*, vol. 2, no. 1, p. 82, 2018.

[32] R. J. Sarjanako, "Penerapan Fuzzy C-Means Clustering Untuk Mengoptimalkan Penentuan Media Promosi," *Teknois : Jurnal Ilmiah Teknologi Informasi dan Sains*, vol. 6, no. 1, pp. 29–37, 2019.

[33] A. Anggrawan, N. Ibrahim, S. Muslim, and C. Satria, "Interaction between learning style and gender in mixed learning with 40and 60Science and Applications, vol. 10, no. 5, pp. 407–413, 2019.

[This page intentionally left blank.]