Comparison of Case-Based Reasoning and Certainty Factor Methods for Dengue Diagnosis

Hamdani Hamdani
dept. Informatics
Faculty of Engineering
Mulawarman University
Samarinda, Indonesia
hamdani@unmul.ac.id

Della Aswintha Asdedi
dept. Informatics
Faculty of Engineering
Mulawarman University
Samarinda, Indonesia
dellaaswinthaa13@gmail.com

Heliza Rahmania Hatta
dept. Informatics
Faculty of Engineering
Mulawarman University
Samarinda, Indonesia
heliza_rahmania@yahoo.com

Anindita Septiarini
dept. Informatics
Faculty of Engineering
Mulawarman University
Samarinda, Indonesia
anindita@unmul.ac.id

Novianti Puspitasari
dept. Informatics
Faculty of Engineering
Mulawarman University
Samarinda, Indonesia
novipuspitasari@unmul.ac.id

Widyastuti Andriyani
Master of Information Technology
STMIK Akakom
Yogyakarta, Indonesia
widy.ugm@gmail.com

2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI) | 978-1-6654-0151-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISRITI54043.2021.9702795

Abstract*— Dengue Hemorrhagic Fever (DHF) is a disease caused by a dengue virus infection that is transmitted through the bite of Aedes Aegypti and Aedes Albopictus mosquitoes. People with dengue fever who are diagnosed based on symptoms that appear without any medical considerations can lead to mishandling. Therefore, a computer-based system is needed to overcome these problems. This study aims to compare two methods: Case-Based Reasoning (CBR) and Certainty Factor (CF) for diagnosing dengue. The CBR method computed the similarity value of the new case disease with the previous case to justify the disease. Meanwhile, the CF method used the certainty value of the disease. The dataset consists of 110 patient data divided into three categories of dengue fever, namely, dengue fever, dengue hemorrhagic fever, and dengue shock syndrome. There are 19 symptoms of the patient recorded. The expert system presents the result of DHF diagnosis, the appropriate treatment solutions, and the comparison results of the diagnosis from the CBR and CF methods. Based on the evaluation result, it shows that the CBR method achieves the value of accuracy, precision, and recall is 84.455%, 84.455%, and 100%, while the CF method obtain 80.909%, 80.909%, and 100%, respectively. Based on the test results, it shows that there are differences in values, so that the conclusion is that the CBR method is more accurate in diagnosing DHF.

Keywords*— Expert system, dengue disease, dengue virus, Case-Based reasoning, certainty factor.

I. INTRODUCTION

The challenge of health development in Indonesia is the triple burden of health problems, namely the increase in non-communicable diseases, the high prevalence of infectious diseases, and diseases that should be able to be overcome reappear. In addition, weather and environmental factors also affect the emergence of diseases, one of which is Dengue Hemorrhagic Fever (DHF) [1], [2]. DHF is a dengue virus infection transmitted through the bite of Aedes Aegypti and Aedes Albopictus mosquitoes. Usually, DHF is found in tropical or subtropical areas [3].

People diagnose the symptoms of DHF based on known characteristics without medical considerations. Therefore, causing the wrong handling and not closing the possibility can be dangerous if not treated immediately [4]. In this case, to minimize the occurrence of this, it is necessary to apply technology that can be used to diagnose dengue. One of the uses of this technology is to use an expert system. This system adopts human knowledge, such as health experts, agricultural experts, and so on, to solve a problem [5]–[7]. In the expert system, some methods can solve problems, especially diagnosing diseases, such as Case-Based Reasoning (CBR) and Certainty Factor (CF).

Related research in the application of the CBR and CF methods has been carried out, namely, the study of Vedayoko, who applied the CBR and K-NN methods to diagnose intestinal diseases. Based on the level of system accuracy of 20 test data obtained a percentage of 95% [8]. Meanwhile, Elkader's research uses 400 case data with 25 features that utilize a series of data mining algorithms. These data are used to help diagnose chronic kidney disease in CBR. Based on the highest classification accuracy 99%, precision 99%, recall 99% F-Measure [9].

The CF method is used to perform early detection of Hand, Foot, and Mouth Disease (HFMD). This method's accuracy, precision, and recall are 100% based on the test results obtained [10]. Meanwhile, a similar study conducted by Setiabudi for the detection of dental disease, looking for the system's accuracy from 20 patients, 19 cases were suitable and 1 case that was not suitable. The results of the system testing carried out resulted in an accuracy rate of 95% [11].

Dengue is a dengue virus infection that is transmitted through the bite of the Aedes Aegypti and Aedes Albopictus mosquitoes [12], Symptoms usually begin 3-14 days after infection. Based on the type of disease, DHF is divided into 3 phases, namely Dengue Fever (DD), Dengue Hemorrhagic Fever (DHF), and Dengue Shock Syndrome (DSS). The research conducted by Aysha aims to assist doctors and patients in diagnosing dengue disease and provide them with information on how to prevent dengue disease and understand the signs and symptoms of dengue disease. This study applies a knowledge-based system using the SL5 object language. The results obtained that this system is very user-friendly and easy to use [12].

Based on these problems, this study created a system comparing CBR and CF methods on the problem of diagnosing dengue hemorrhagic fever, which in turn is expected to help minimize misdiagnosis and make it easier to provide information about dengue fever in the community.

II. PROPOSED METHOD

A. Knowledge Acquisition

The data used comes from the medical records of patients with dengue fever at the Dirgahayu Hospital. In addition, data on dengue fever and clinical symptoms of DHF were obtained from the knowledge of doctors (experts/experts) in internal medicine and general practitioners directly, as well as references to books and journals provided by experts.

B. Case Base Reasoning

Each case source has two parts, namely, the problem and the solution. The CBR process usually operates in four sequential phases: retrieve, reuse, revise and retain [13], [14]. The stages of the CBR process are shown in Fig. 1.

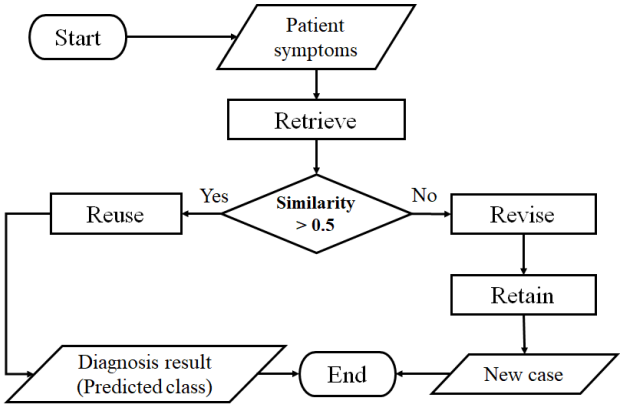

Fig. 1 Process stages in the CBR method

The process detail in Fig. 1 describe as follow:

1. Retrieve

Searches from past cases on a case base that is most similar to the current case. In this process, it will be searched using the K-NN algorithm. The case with the greatest similarity value is considered the most "similar" case. The equation to find similarity is in the equation (1).

$$Similarity\ (T,S) = \frac{\sum_{i=1}^{n} f(T_i S_i) * w_i}{\sum_{i=1}^{n} w_i} \quad (1)$$

where,

S_i: The feature i that is in the source case

T_i: the feature i that is in the target case

w_i: feature weight to i

n: number of features

$f(T_i S_i)$: The similarity function between case T and case T, if there is a similarity case that will be worth 1, no similarity will be worth 0. Between case T and case T, there is a similarity case that will be worth 1, and no similarity will be worth 0 [8].

2. *Reuse*

Reuse the previous case solutions as a reference for solving new case problems.

3. *Revise*

Improvement of solutions by testing simulations of confirmed cases to improve solutions.

4. *Retain*

New solutions are saved to the case base for future case solving [14].

C. Certainty Factor

The CF method is used to deal with a problem whose answer is uncertain [15]–[17]. The stages of the CF process are depicted in Fig. 2.

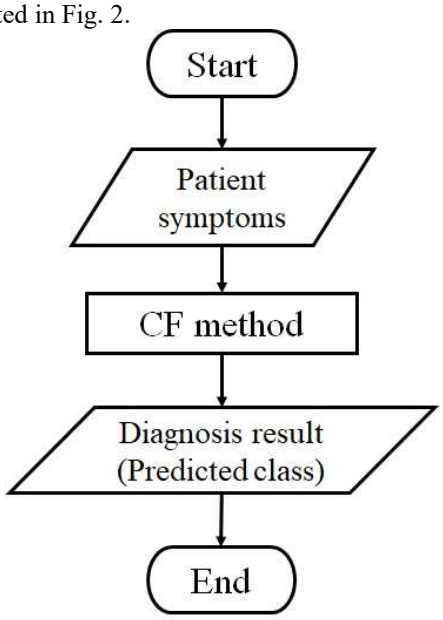

Fig. 2 Process stages in the CF method

CF introduces the concept of uncertainty and certainty, which is then as in Equation (2).

$$CF[h,e] = MB\ [h,e] - MD\ [h,e] \quad (2)$$

where,

$CF[h,e] =$: CF from hypothesis H, which is influenced by e symptoms (evidence). The magnitude is between -1 to 1. A value of -1 indicates absolute distrust, while a value of 1 indicates absolute confidence.

$MB\ [h,e]$: The measure of Belief against hypothesis h, if given evidence e (between 0 and 1).

$MD\ [h,e]$: The measure of Disbelief against hypothesis h, if given evidence e (between 0 and 1).

h : Hypothesis

e : Events or facts (evidence) [18][19]

Some combined evidence in determining the CF of a hypothesis. If e1 and e2 are observations as in equations (3) and (4).

MB(h,e1^e2) = MB[h,e1]+(MB[h,e2]*(1-MB[h,e1])) (3)

MD(h,e1^e2) = MD[h,e1]+(MD[h,e2]*(1-MD[h,e1])) (4)

D. Testing

The confusion matrix is commonly used in the field of machine learning and statistical classification problems. The confusion matrix is also known as an error matrix [20]. The confusion matrix is described as a table that states the test data classified as true and false [21]. The following stages of testing using the confusion matrix model are in Table I.

TABLE I. CONFUSION MATRIX

		Prediction Class	
		(+)	**(-)**
Class	**(+)**	*True Positive*	*False Negative*
	(-)	*False Positive*	*True Negative*

where,

a. True Positives (TP) is the number of data records with true positive and positive predictive values,
b. False Positives (FP) is the number of data records with true negative values that are incorrectly predicted as positive values,
c. False Negatives (FN) is the number of data records with true positive values that are falsely predicted as negative predicted values,
d. True Negatives (TN) is the number of data records with true negative values that are correctly predicted as negative values.

The value generated from the confusion matrix method is in the form of evaluation, namely accuracy, precision, and recall are contained in the equation (5), (6), and (7) [22].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \quad (5)$$

$$Precision = \frac{TP}{FP+TP} * 100\% \quad (6)$$

$$Recall = \frac{TP}{TP+FN} * 100\%. \quad (7)$$

III. Result and Discussion

A. System Description

This expert system compares the CBR and CF methods in diagnosing Dengue Hemorrhagic Fever (DHF). It is hoped that the system built can minimize misdiagnosis and make it easier to provide information about DHF to the community. The system's output is the diagnosis results in disease categories and treatment solutions and a comparison of the diagnosis results from the two methods.

B. Knowledge Base

The knowledge base contains data on the category of dengue virus infection, presented in Table II, and the classification of symptoms by disease category is shown in Table III. The weight was provided by the expert in Table III.

TABLE II. Dengue virus infection category

Disease Code	Disease Name
P001	Dengue Fever (DD)
P002	Dengue Hemorrhagic Fever (DHF)
P003	Dengue Shock Syndrome (DSS)

TABLE III. Classification of symptoms by disease category

Disease Code	Symptom	DD	DHF	DSS	Weight
G001	Fever with a temperature between 39 - 40 and the fever is biphasic (The fever will go down on the 3rd day, but the fever will rise again on the 4th or 7th day)	✓	✓	✓	0,90
G002	Headache	✓	✓	✓	0,44
G003	Pain behind the eyes	✓	✓	✓	0,80
G004	Joint or bone pain	✓	✓	✓	0,80
G005	Muscle ache	✓	✓	✓	0,80
G006	Maculopapular skin rash (rash on the skin surface in the first 2 to 3 days)	✓	✓	✓	0,80
G007	Petechiae (red spots on the skin that can be seen on the back, legs, hands, and arms)	-	✓	✓	0,90
G008	Bleeding manifestations (bleeding bowel movements, bleeding in the gums, nose, and other places)	-	✓	✓	0,90
G009	Shock	-	-	✓	0,90
G010	Nervous	-	-	✓	0,90
G011	Throws up	✓	✓	✓	0,58
G012	Constipation	-	✓	✓	0,53
G013	Diarrhea	-	✓	✓	0,16
G014	Upper abdominal pain or heartburn	-	✓	✓	0,50
G015	Red eyes	✓	✓	✓	0,34
G016	Pain in the lower jaw	-	✓	✓	0,40
G017	Cough	✓	✓	✓	0,21
G018	Laryngitis	✓	✓	✓	0,21
G019	Inflammation of the nasal cavity	-	✓	✓	0,12

C. System Implementation

The implementation of this web-based system is used to find out the results of the comparison of the CBR and CF methods, where doctors directly fill in diagnosis data and disease data, which contains information related to DHF case data at Dirgahayu Hospital Samarinda. Symptom data in 19 general clinical symptoms consist of DHF, DD, and DSS diseases. This system interface consists of consultations, calculations using the CBR and CF methods based on comparing the results of the diagnosis so that the

system knows the results of the comparison of these methods.

The consultation system is carried out by filling in the data first. Next, choose the symptoms felt by the patient. This application then performs a comparison of CBR and CF, which displays the results of consultations that patients have carried out. The results consist of the name of the disease and the calculation results of each method. System performance is determined based on how well the system classifies data.

The next result test uses an uncertainty matrix for the accuracy of the CBR and the CF method on the input data for DHF diagnosis. The following is a test using the confusion matrix in the CBR method presented in Table IV and the CF method in Table V.

Based on the results of the test data from Table IV using the CBR method, accuracy, precision, and recall are obtained based on equations (5), (6), and (7), then the results are as follows:

$$Accuracy = \frac{94+0}{94+0+16+0} * 100\% = 84{,}455\%$$

$$Precision = \frac{94}{94+16} * 100\% = 84{,}455\%$$

$$Recall = \frac{94}{94+0} * 100\% = 100\%$$

TABLE IV. CONFUSION MATRIX CBR METHOD

Identification Type	Test data	TP	TN	FP	FN
Dengue fever	110	94	0	16	0

TABLE V. CONFUSION MATRIX CF METHOD

Identification Type	Test data	TP	TN	FP	FN
Dengue fever	110	89	0	21	0

The results are based on the test data from Table V using the CF method, then the accuracy, precision, and recall are obtained based on equations (5), (6), and (7), then the results are as follows:

$$Accuracy = \frac{89+0}{89+0+21+0} \times 100\% = 80{,}909\%$$

$$Precision = \frac{89}{89+21} \times 100\% = 80{,}909\%$$

$$Recall = \frac{89}{89+0} \times 100\% = 100\%$$

Meanwhile, the comparison results using the confusion matrix based on the CBR and CF methods are in Table VI.

TABLE VI. THE RESULTS OF THE COMPARISON OF THE CONFUSION MATRIX CBR AND CF METHODS

Method	Accuracy	Precision	Recall
CBR	84,455%	84,455%	100%
CF	80,909%	80,909%	100%

Based on system testing using confusion matrix, the resulting comparison for the CBR method obtained an accuracy value of 84.455% %, precision 84.455%, and recall 100%, while the CF method obtained an accuracy value of 80.909%, precision 80.909%, and recall 100%.

IV. CONCLUSION

Based on the research results, it has been done that the results of the expert system for diagnosing dengue use two methods of CBR and CF. The performance comparison from the CBR method resulted in the value of accuracy, precision, and recall of 84.455%, 84.455%, and 100%, respectively; meanwhile, the CF method produced 80.909%, 80.909%, and 100%. The test results show that the CBR method has a higher accuracy rate than the CF method in the problem of diagnosing DHF.

ACKNOWLEDGMENT

The author would like to thank the Faculty of Engineering, Mulawarman University for the support of research funds in 2021.

REFERENCES

[1] R. A. Werdhani, "Medical problem in Asia pacific and ways to solve it : The roles of primary care / family physician (Indonesia Xperience)," pp. 1523–1527, 2019, doi: 10.4103/jfmpc.jfmpc.

[2] A. F. Arham, M. R. Razman, L. Amin, and Z. Mahadi, "Dengue Review: Issues, Challenges and Public Attitudes," *Int. J. Acad. Res. Bus. Soc. Sci.*, vol. 8, no. 4, pp. 980–997, 2018, doi: 10.6007/IJARBSS/v8-i4/4125.

[3] A. Wilder-Smith, E.-E. Ooi, O. Horstick, and B. Wills, "Dengue," *Lancet*, vol. 393, no. 01, pp. 350–363, 2019, doi: 10.1016/S0140-6736(18)32560-1.

[4] S. Nugroho, S. H. Waluyo, N. Agustina, R. Faslah, and R. M. Ali, "Rapid Application Development of Decision Support System Using Naive Bayes Classification to Classify Dengue Hemorrhagic Diseases," vol. 2, no. 1, pp. 1–5, 2016.

[5] H. R. Hatta, F. Ulfah, D. M. Khairina, H. Hamdani, and S. Maharani, "Web-expert system for the detection of early symptoms of the disorder of pregnancy using a forward chaining and bayesian method," *J. Theor. Appl. Inf. Technol.*, vol. 95, no. 11, pp. 2589–2599, 2017.

[6] N. Puspitasari, J. A. Widians, E. Budiman, M. Wati, and A. E. Ramadhan, "Dayak Onion (Eleutherine palmifolia (L) Merr) as An Alternative Treatment in Early Detection of Dental Caries using Certainty Factor," 2020, doi: 10.1109/ISRITI51436.2020.9315469.

[7] A. Subekti and E. A. E. Ningtyas, "Dentist Expert System Software in Dental Caries Detection," *Adv. Sci. Lett.*, vol. 23, no. 4, pp. 3288–3290, 2017.

[8] L. G. Vedayoko, E. Sugiharti, and M. A. Muslim, "Expert System Diagnosis of Bowel Disease Using Case Based Reasoning with Nearest Neighbor Algorithm," *Sci. J. Informatics*, vol. 4, no. 2, pp. 7–10, 2017.

[9] S. A. Elkader, M. Elmogy, S. El-sappagh, and A. N. H. Zaied, "A framework for chronic kidney disease diagnosis based on case based reasoning," *Int. J. Adv. Comput. Res.*, vol. 8, no. 35, pp. 59–71, 2018.

[10] D. Novitasari, B. Irawan, and A. L. Prasasti, "Early Detection of Hand , Foot , and Mouth Disease based on Palmprint using Certainty Factor as Expert System Method based on Android Early Detection of Hand , Foot , and Mouth Disease based on Palmprint using Certainty Factor as Expert System Method base," *J. Phys. Conf. Ser. Pap.*, vol. 1201, no. 1, pp. 1–10, 2019, doi: 10.1088/1742-6596/1201/1/012055.

[11] W. U. Setiabudi, E. Sugiharti, and F. Y. Arini, "Expert System Diagnosis Dental Disease Using Certainty Factor Method," *Sci. J. Informatics*, vol. 4, no. 1, pp. 43–50, 2017, doi: 10.15294/sji.v4i1.8463.

[12] A. I. Mansour and S. S. Abu-naser, "Knowledge Based System for the Diagnosis of Dengue Disease," vol. 3, no. 4, pp. 12–19, 2019.

[13] N. Choudhury, "A Survey on Case-based Reasoning in Medicine," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 8, pp. 136–144, 2016.

[14] M. I. Insani and A. T. Putra, "Implementation of Expert System for Diabetes Diseases using Naïve Bayes and Certainty Factor Methods," *Sci. J. Informatics*, vol. 5, no. 2, pp. 185–193, 2018.
[15] Y. Findawati and A. I. Afrina, "Expert system diagnose disease dermatitis using web based certainty factor," 2018, doi: 10.1088/1757-899X/403/1/012068.
[16] J. A. Widians, N. Puspitasari, and A. Febriansyah, "Disease Diagnosis System Using Certainty Factor," *ICEEIE 2019 - Int. Conf. Electr. Electron. Inf. Eng. Emerg. Innov. Technol. Sustain. Futur.*, vol. 6, pp. 303–308, 2019, doi: 10.1109/ICEEIE47180.2019.8981421.
[17] I. Sumatorno, D. Arisandi, A. P. U. Siahaan, and M. Mesran, "Expert System of Catfish Disease Determinant Using Certainty Factor Method," *Int. J. Recent Trends Eng. Res.*, vol. 3, 2017.
[18] D. T. Yuwono and A. Fadlil, "Comparative Analysis of Dempster-Shafer Method and Certainty Factor Method On Personality Disorders Expert Systems," *Sci. J. Informatics*, vol. 6, no. 1, pp. 12–22, 2019.
[19] K. Muludi, "Implementation of Forward Chaining and Certainty Factor Method on Android-Based Expert System of Tomato Diseases Identification," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 9, pp. 451–456, 2018.
[20] R. S. Sambyal, T. Javid, and A. Bansal, "Performance Analysis of Data Mining Classification Algorithms to Predict Diabetes," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 4, no. 1, pp. 56–63, 2018.
[21] F.J. Ariza-Lopez, J. Rodriguez-Avi, and M.V. Alba-Fernandez, "Complete control of an observed confusion matrix," *IGARSS-IEEE Int. Geosci. Remote Sens.*, vol. 2, no. 11, pp. 1222–1225, 2018.
[22] A. Septiarini, Hamdani, and D. M. Khairina, "The contour extraction of cup in fundus images for glaucoma detection," International Journal of Electrical and Computer Engineering, vol. 6, pp. 2797–2804, 2016.