Comparison of Knowledge-Based Reasoning Methods to Measure the Effectiveness of Diagnostic Results

Puji Sari Ramadhan*, Marsono, Juniart Hutagalung, and Yohanni Syahra

Department of Informatics, STMIK Triguna Dharma, Medan, Indonesia

*pujisariramadhan@gmail.com

Abstract. This study describes a comparative analysis of methods in knowledge-based reasoning which aims to determine the best and optimal method in producing a diagnosis of a disease. The methods to be compared are Bayes’ Theorem, Certainty Factor, and Euclidean Probability. The three methods were chosen because they can generate hypotheses from several existing possibilities, this can be seen from the many uses of these methods in previous studies. With this research, it can be used as material for consideration or support in producing a diagnosis conclusion. The process of testing the comparison of these methods is carried out by selecting the highest diagnosis result value and performing a comparative analysis using exponential techniques. The results of this test indicate that the Euclidean Probability produces an accuracy value of 75%, then the Bayes Theorem obtains an accuracy value of 62% and the Certainty Factor obtains an accuracy value of 87%. In addition, the results of the comparison of methods using exponential techniques show that Euclidean Probability gets 84%, then Bayes’ Theorem gets 78% and the Certainty Factor gets 87%. With these results, it can be concluded that the Certainty Factor is better than Euclidean Probability and Bayes’ Theorem in diagnosing disease.

1. Introduction

Knowledge-based reasoning is a computer science concept that can analyze probability calculations to produce a conclusion [1]. Knowledge-based reasoning is also believed to be able to produce predictions for a diagnosis result[2]. This can be seen from several previous studies that used the concept of knowledge-based reasoning in predicting the results of disease diagnosis [3] and [4]. Besides, knowledge-based reasoning is also used in other fields to predict device damage, [5], [6], [7], and [8].

In knowledge-based reasoning, there are methods used to produce predictions for a case. For this reason, it is necessary to measure the effectiveness of the results of the diagnosis against the method of knowledge-based reasoning so that the best method for generating diagnostic conclusions can be found.

In this study, the knowledge-based reasoning method used in solving the case of diagnosis is Euclidean Probability, Bayes Theorem, and Certainty Factor. The three methods were chosen because they can generate hypotheses from several existing possibilities, this can be seen from the successful use of these methods in previous studies, such as the use of Bayes’ Theorem [9], [10], [11]. Then the use of Euclidean Probability in case of diagnosis [11], [12]. Besides, the Certainty Factor can generate a conclusion from several existing possibilities [13], [14].

Comparative testing of this method will use a case study of the diagnosis of virus dermatitis, which begins with gathering knowledge about virus dermatitis, then tracing symptoms using the Euclidean Probability method, Bayes Theorem, and Certainty Factor. Based on the results of the application of these methods, the diagnostic value of each method will be known to select the method that has the
greatest diagnostic value. Besides, to optimize testing in comparison to methods, comparisons will be carried out by applying exponential techniques in calculating the resulting diagnostic value so that the best method for diagnosis of virus dermatitis can be found. The exponential technique has been tested in comparing methods and can find priority alternatives from several existing alternatives so that the best alternative can be found among other alternatives [15], [16].

With this research, it can provide information about the most appropriate method of diagnosing the disease so that it can be used as a prediction result and material for initial diagnosis recommendations. Furthermore, this research will have an impact on knowledge and development in knowledge-based reasoning to solve problems related to diagnosis results.

2. Methodology

In conducting this research, a framework that contains stages or steps of work must be carried out. These stages consist of gathering knowledge, applying methods, and testing methods. The purpose of using this framework is to be able to solve the problems that occur regarding the comparison of knowledge-based reasoning methods in producing a diagnostic value for dermatitis virus disease. The following are the stages of the framework that has been prepared.

a. Data Collection of Knowledge

This stage is carried out to collect knowledge data and information relating to clinical symptoms data from symptomatic dermatitis virus disease. The technique of knowledge data in interviews. Based on the knowledge data that has been collected, a list of rules or rules can be formed into three, along with a list of ordering rules:

Rule-1: IF High Fever And Flu And Quickly feel tired And lethargic And rashes all over the body And joint pains And Headaches And peeling of the skin Then Disease = Varicella

Rule-2: IF High Fever And Rhinitis And Conjunctiva And Irritation And Cough And Body shivering And Body Chills Then Disease = Morbillivirus

Rule-3: IF High Fever And Skin burns And Reddish spots And Water bubbles on the skin And Itching Then Disease = Herpespidae.

Table 1. Data Collection of Knowledge

| No | Symptom Code | Symptom Name       | The Weight Value of Disease Types |
|----|--------------|--------------------|-----------------------------------|
|    |              |                    | Varicela | Morbillivirus | Herpespidae |
| 1  | G01          | High Fever         | 0,6      | 0,4           | 0,2         |
| 2  | G02          | Flu                | 0,4      |               |             |
| 3  | G03          | Quickly feel tired | 0,2      |               |             |
| 4  | G04          | lethargic          | 0,4      |               |             |
| 5  | G05          | Rashes all over the body | 0,4 |               |             |
| 6  | G06          | Joint pains        | 0,4      |               |             |
| 7  | G07          | Headaches          | 0,2      |               |             |
| 8  | G08          | Peeling of the skin| 0,2      |               |             |
| 9  | G09          | Rhinitis           | 0,4      |               |             |
| 10 | G10          | Conjunctiva        | 0,6      |               |             |
| 11 | G11          | Irritation         | 0,4      |               |             |
| 12 | G12          | Cough              | 0,2      |               |             |
| 13 | G13          | Body shivering     | 0,2      | 0,2           |             |
b. Implementation of the Euclidean Probability

The next step is to calculate using Euclidean Probability. In [12] explains Euclidean Probability is used in generating diagnosis conclusions with knowledge data that has been included in the database of expertise in the form of Equation 1:

\[
EP = \sqrt{(E_1 \times NBE_1)^2 + (E_2 \times NBE_2)^2 + \ldots + (E_n \times NBE_n)^2}
\]  

(1)

Where EP is the result of Euclidean Probability, E is the condition value on evidence, and NBE is the weight value of the evidence.

c. Implementation of Bayes Theorema

In the next section, the calculation of the diagnosis value will be carried out using the Bayes Theorem. In [17] suggests that the Bayes Theorem is a method that combines rules with probability or Bayes values in generating conclusions based on the problems that occur with the following steps:

The first stage is to calculate the symptom weight value in the form of Equation 2,

\[
P(H) = \frac{P(H1)}{n} \sum_{i=1}^{n} \frac{P(H | Hi) \times P(E | Hi)}{P(E | Hi) \times P(Hi)}
\]  

(2)

Where P (H) is the result of the hypothesized probability, P (H1) is the condition value for each evidence.

Furthermore, the second stage is to calculate the value of the universe without looking at the evidence in the form of Equation 3,

\[
\sum_{i=1}^{n} P(Hi) \times P(E | Hi) = n
\]  

(3)

Where P (Hi) is the condition value for each evidence, P (E | Hi) is the condition value against the hypothesis value.

Then the third stage is to calculate the probability value with the value of influential evidence in the form of Equation 4,

\[
P(Hi) = \frac{P(E | Hi) \times P(Hi)}{\sum_{i=1}^{n} P(E | Hi) \times P(Hi)}
\]  

(4)

Where P (Hi) is the condition value for each evidence, P (E | Hi) is the condition value against the hypothesis value.

Then the fourth stage is to calculate the hypothetical value of total Bayes in the form of Equation 5,

\[
\sum_{i=1}^{n} P(Ei | Hi) = Bayes1 + Bayes2 + \ldots + Bayesn
\]  

(5)

Where Bayes is the value of the Bayes calculation results that have been done.

d. Implementation of Certainty Factor

The next work step is to perform calculations using Certainty Factors, In [18] it explains that Certainty Factor is a method that can generate possible values of complex problems based on conditions that have uncertainty in the form of Equation 6,

\[
CF(cf_1, cf_2) = \begin{cases} 
  cf_1 + cf_2(1 - cf_1), & \text{jika } cf_1 \text{ dan } cf_2 > 0 \\
  cf_2 + cf_1(1 - cf_2), & \text{jika } cf_1 \text{ atau } cf_2 < 0 \\
  \frac{1 - \min(cf_1, cf_2)}{cf_1 + cf_2(1 - cf_1)}, & \text{jika } cf_1 \text{ dan } cf_2 < 0
\end{cases}
\]  

(6)

\[14\text{ G14 Skin burns 0,4}\]
\[15\text{ G15 Reddish spots 0,6}\]
\[16\text{ G16 Water bubbles on the skin 0,2}\]
\[17\text{ G17 Itching 0,2}\]
Where \( Cf \) is the result of the Certainty Factor, \( Cf_{1,2} \) is the condition value on the evidence.

f. Testing Method

After the implementation of the Euclidean Probability method, Bayes Theorem and Certainty Factor has been carried out, then the method comparison will be carried out by ranking the results of the diagnostic value, accuracy and comparing the diagnostic values of the three methods using exponential techniques to produce the highest priority alternative conclusions. In [4] argued that exponential techniques can find priority alternatives from several existing alternatives so that the best alternative can be found among other alternatives.

\[
\text{Accuracy} = \left( \frac{\sum \text{appro}}{\sum \text{Data}} \right) \times 100\%
\]

(7)

Where Accuracy is the result of the accuracy of the available alternatives, \( \sum \text{appro} \) is the amount of data that has a match and \( \sum \text{Data} \) is the total number of data.

\[
E(H_n) = \sum_{i=1}^{n} (W_i)^j
\]

(8)

Where \( E(H_n) \) is the exponential value of the available alternatives, \( W_i \) is the alternative value and \( j \) is the alternative weight value.

3. Result and Discussion

In this study, 8 diagnostic sample data were taken to test the diagnostic value of the Euclidean Probability method, Bayes' Theorem, and Certainty Factor. The following is a sample data of cases of dermatitis virus disease:

| Table 2. Case Data |
|-------------------|
| No | Case Data |
|----|-----------|
| K1 | G01, G05, G08 |
| K2 | G01, G02, G03 |
| K3 | G10, G11 |
| K4 | G04, G05, G08 |
| K5 | G01, G16, G17 |
| K6 | G11, G12, G13 |
| K7 | G10, G12, G01 |
| K8 | G02, G13, G17 |

From the case data, then the calculation will be carried out using the Euclidean Probability method, Bayes Theorem, and Certainty Factor based on predetermined rule bases.

a. Euclidean Probability Calculation

At this stage, calculations will be carried out using the Euclidean Probability method, through equation (1). From these calculations, the following results are obtained:

| Table 3. Euclidean Probability Results |
|---------------------------------------|
| No | Case Data | Weight Value | Diagnosis Results |
|----|-----------|--------------|-------------------|
| K1 | G01, G05, G08 | 0.75 | Varicella |
| K2 | G01, G02, G03 | 0.75 | Varicella |
| K3 | G10, G11 | 0.72 | Morbillivius |
| K4 | G04, G05, G08 | 0.60 | Varicella |

| No | Case Data | Weight Value | Diagnosis Results |
|----|-----------|--------------|-------------------|
| K5 | G01, G16, G17 | 0.35 | Herpespidae |
b. Theorema Bayes Calculation

At this stage, calculations will be carried out using the Bayes Theorem method, starting by calculating the symptom weight value in the form of Equation (2). From these calculations, the following results are obtained:

| No | Case Data       | Varicella Value | Morbillivius Value | Herpespidae Value |
|----|----------------|-----------------|--------------------|------------------|
| K1 | G01, G05, G08  | 1.20            | 0.4                | 0.2              |
| K2 | G01, G02, G03  | 1.20            | 0                  | 0                |
| K3 | G10, G11       | 0.00            | 1.00               | 0.00             |
| K4 | G04, G05, G08  | 1.00            | 0                  | 0                |
| K5 | G01, G16, G17  | 0.60            | 0.4                | 0.60             |
| K6 | G11, G12, G13  | 0.00            | 0.80               | 0.20             |
| K7 | G10, G12, G01  | 0.60            | 1.20               | 0.2              |
| K8 | G02, G13, G17  | 0.40            | 0.2                | 0.4              |

Then calculate the value of the universe without looking at the evidence in the form of Equation (3). From these calculations, the following results are obtained:

| No | Case Data       | Varicella Value | Morbillivius Value | Herpespidae Value |
|----|----------------|-----------------|--------------------|------------------|
| K1 | G01, G05, G08  | 0.47            | 0.4                | 0.2              |
| K2 | G01, G02, G03  | 0.47            | 0                  | 0                |
| K3 | G10, G11       | 0               | 0.52               | 0.00             |
| K4 | G04, G05, G08  | 0.36            | 0                  | 0                |
| K5 | G01, G16, G17  | 0.6             | 0.4                | 0.20             |
| K6 | G11, G12, G13  | 0               | 0.30               | 0.20             |
| K7 | G10, G12, G01  | 0.6             | 0.47               | 0.2              |
| K8 | G02, G13, G17  | 0.4             | 0.2                | 0.20             |

Next, calculate the probability value with the value of influential evidence in the form of Equation (4). From these calculations, the following results are obtained:

| No | Case Data       | Varicella Value | Morbillivius Value | Herpespidae Value |
|----|----------------|-----------------|--------------------|------------------|
| K1 | G01, G05, G08  | 0.64            | 0.4                | 0.2              |
| K2 | G01, G02, G03  | 0.64            | 0                  | 0                |
| K3 | G10, G11       | 0               | 0.69               | 0                |
| K4 | G04, G05, G08  | 0.44            | 0                  | 0                |
| K5 | G01, G16, G17  | 0.6             | 0.4                | 0.33             |
| K6 | G11, G12, G13  | 0               | 0.67               | 0.2              |
| K7 | G10, G12, G01  | 0.6             | 0.64               | 0.2              |
| K8 | G02, G13, G17  | 0.4             | 0.2                | 0.5              |
After that, calculate the hypothetical value of total Bayes in the form of Equation (5). From these calculations, the following results are obtained:

### Table 7. Bayes Value Results

| No | Case Data     | Bayes Value | Diagnosis Results |
|----|---------------|-------------|-------------------|
| K1 | G01, G05, G08 | 0.51        | Varicella         |
| K2 | G01, G02, G03 | 0.51        | Varicella         |
| K3 | G10, G11     | 0.56        | Morbillivirus     |
| K4 | G04, G05, G08 | 0.31        | Varicella         |
| K5 | G01, G16, G17| 0.6         | Varicella         |
| K6 | G11, G12, G13| 0.54        | Morbillivirus     |
| K7 | G10, G12, G01| 0.51        | Morbillivirus     |
| K8 | G02, G13, G17| 0.34        | Herpespidae       |

### c. Certainty Factor Calculation

At this stage, calculations will be carried out using the Certainty Factor method, through equation (6). From these calculations, the following results are obtained:

### Table 8. Certainty Factor Results

| No | Case Data     | CF Value | Diagnosis Results |
|----|---------------|----------|-------------------|
| K1 | G01, G05, G08 | 0.8      | Varicella         |
| K2 | G01, G02, G03 | 0.8      | Varicella         |
| K3 | G10, G11     | 0.76     | Morbillivirus     |
| K4 | G04, G05, G08 | 0.71     | Varicella         |
| K5 | G01, G16, G17| 0.48     | Herpespidae       |
| K6 | G11, G12, G13| 0.61     | Morbillivirus     |
| K7 | G10, G12, G01| 0.8      | Morbillivirus     |
| K8 | G02, G13, G17| 0.4      | Varicella         |

### d. Method Comparison Results

After doing calculations with Euclidean Probability, Bayes Theorem, and Certainty Factor, the next process is to rank the diagnostic results of the method. The following is a ranking of the diagnostic values:

From the ranking results that have been displayed, it can be seen that the Certainty Factor method has a higher diagnostic value than the Euclidean Probability and Bayes Theorem. Furthermore, the accuracy value of each method will be calculated according to equation (7). The following is the result of the accuracy that has been obtained:

![Figure 1. Ranking Results](image-url)
From the results of the accuracy that has been shown, it can be seen that the Certainty Factor method (87%) has a higher accuracy value than the Euclidean Probability (75%) and Bayes' Theorem (62%). To provide accurate method comparison results, it is necessary to apply the exponential technique through equation (8). The following is the result of the exponential technologies that have been obtained:

From the results of the calculation of 8 case samples using the exponential technique, it is known that the Certainty Factor method has an average value of 0.87 while the Euclidean Probability produces an average value of 0.84 and Bayes' Theorem has an average value of 0.78. Based on these exponential results, it can be concluded that the Certainty Factor method is better than the Euclidean Probability and Bayes Theorem.

4. Conclusion
Based on the results of research that has been done regarding the comparison of knowledge-based reasoning methods using a comparison of accuracy values and exponential techniques, it can be obtained the results of the comparison of the accuracy value of the Euclidean Probability of 75%, then the Bayes Theorem produces an accuracy value of 62%, and Certainty Factor produces 87%. Besides, the Euclidean Probability exponential technique gets a priority value of 0.84, then the Bayes Theorem gets a priority value of 0.78, and the Certainty Factor gets a priority value of 0.87. With the results of the comparison of the methods carried out, it can be concluded that the Certainty Factor method is better than the Euclidean Probability and Bayes Theorem in making a diagnosis of disease.

References

[1] Y. Purnama, I. Ismail, D. Noviandri, Y. Hendriyani, P. T. Nguyen, And I. P. A. Darmawan, “Expert System In Detecting Children’s Intelligence Using Certainty Factor,” J. Crit. Rev.,
[2] F. Riana, T. Informatika, F. Satrya, F. Kusumah, And T. Informatika, “Expert System Of Diagnosing Disease In Banana Plants Using Dempster-Shafer And Certainty Factor,” Vol. 410, No. Imcete 2019, Pp. 100–103, 2020.

[3] M. M. Abu-Saqer And S. S. Abu-Naser, “Knowledge Based System For Uveitis Disease Diagnosis,” Vol. 3, No. 5, Pp. 18–25, 2019.

[4] A. F. Ahmed And S. S. A. Naser, “Anemia Expert System Diagnosis Using SL5 Object,” Vol. 3, No. 5, Pp. 9–17, 2019.

[5] F. M. Hadini, “Detection System Milkfish Formalin Android-Based Method Based On Image Eye Using Naive Bayes Classifier,” Vol. 9, No. 1, Pp. 2–5, 2018.

[6] A. Widjaja And A. B. Susilo, “Expert System To Identify Damage Cisco As5300 Device With The Method Of Forward Chaining-Based Client-,” Vol. 9, Pp. 787–805, 2018.

[7] M. Zuhaer And M. H. M. Alhabib, “Face Recognition System Based On Kernel Discriminant Analysis , K-Nearest Neighbor And Support Vector Machine,” Vol. 5, No. 3, Pp. 335–338, 2018.

[8] Ö. F. Arar And K. Ayan, “A Feature Dependent Naive Bayes Approach And Its Application To The Software Defect Prediction Problem,” Appl. Soft Comput. J., 2017.

[9] A. Joshi And M. Ashish, “Analysis Of K-Nearest Neighbor Technique For Breast Cancer Disease Classification,” Int. J. Recent Sci. Res., Vol. 8, No. 8, Pp. 1005–19008, 2017.

[10] S. Qwaider And S. S. A. Naser, “Expert System For Diagnosing Ankle Diseases,” No. August, 2017.

[11] B. F. Yanto, I. Werdiningsih, And E. Purwanti, “Aplikasi Sistem Pakar Diagnosa Penyakit Pada Anak Bawah Lima Tahun Menggunakan Metode Forward Chaining,” J. Inf. Syst. Eng. Bus. Intell., Vol. 3, No. 1, Pp. 61–67, 2017.

[12] P. S. Ramadhan, “Penerapan Euclidean Probability Dalam Pendeteksian Penyakit Impetigo,” Vol. 4, No. 1, Pp. 11–16, 2019.

[13] L. S. Trofimova, “Application Of The Bayes Theory Prinicples In Planning A Motor Transport Company Work,” Vol. 113, No. Fred 2019, Pp. 156–160, 2020.

[14] J. Wu, S. Pan, X. Zhu, P. Zhang, And C. Zhang, “Sode: Self-Adaptive One-Dependence Estimators For Classification,” Pattern Recognit., Vol. 51, Pp. 358–377, 2016.

[15] Y. Sonatha, C. Prabowo, H. Amnur, And R. Afyenni, “Eksponsensial Of Decision Support System,” Indones. J. Electr. Engineering Informatics, Vol. 6, No. 3, Pp. 323–332, 2018.

[16] F. Masya, H. Prastiantawan, And S. Mubaroq, “Application Design To Diagnosis Of Bone Fracture ( Traditional ) Using Forward Chaining Methods,” Int. Res. J. Comput. Sci., Vol. 3, No. 09, Pp. 23–30, 2016.

[17] M. S. Hossain, F. Ahmed, Fatema-Tuj-Johora, And K. Andersson, “A Belief Rule Based Expert System To Assess Tuberculosis Under Uncertainty,” J. Med. Syst., Vol. 41, No. 3, 2018.

[18] F. M. Salman And S. S. Abu-Naser, “Expert System For Castor Diseases And Diagnosis,” Vol. 3, No. 3, Pp. 1–10, 2019.