MIoSA: Machine Intelligence for Societal Advancement
Volume 1, (1) 2026, p. 49-56
E-ISSN : XXXX-XXXX

Analysis of dengue fever diagnosis based on patient symptoms using the certainty factor method

Justin Kwandra [a,1*], Heny Pratiwi [a,2]

[a] Information System Study Program, STMIK Widya Cipta Dharma Samarinda, East Kalimantan, Indonesia
[1] 2341040@wicida.ac.id, [2] henypratiwi@wicida.ac.id
* corresponding author

Article Info

Article History:

Received: Mar 22, 2026
Revised: April 12, 2026
Accepted: April 22, 2026

KeyWords :
Artificial Intelligence, Certainty Factor, Dengue Fever, Diagnosis, Expert System

Abstrak

Dengue fever is a major public health problem, especially in tropical countries such as Indonesia. The similarity of symptoms with other diseases often makes early diagnosis difficult, leading to delays in proper treatment. To address this issue, an expert system can be utilized to assist in early diagnosis based on patient symptoms. An expert system is a branch of artificial intelligence that adopts expert knowledge to solve specific problems. One of the methods used to handle uncertainty in diagnosis is the Certainty Factor (CF) method, which represents the level of confidence of an expert in a given hypothesis.This study aims to analyze the diagnosis of dengue fever based on patient symptoms using the Certainty Factor method. The data used in this study were obtained from a Kaggle dataset containing symptoms such as fever, headache, joint pain, and bleeding. These symptoms are used to form a knowledge base and diagnostic rules in the expert system. The CF method is applied to calculate the confidence level of diagnosis based on the combination of expert knowledge and user input. The results show that the proposed system is able to provide a diagnosis with a measurable confidence value, thereby supporting early detection of dengue fever. This research is expected to contribute to the development of intelligent decision support systems in the medical field.

1. Introduction

Dengue Hemorrhagic Fever (DHF) is an infectious disease caused by the dengue virus and transmitted through the bite of the *Aedes aegypti* mosquito. This disease poses a serious threat to global public health, particularly in tropical regions like Indonesia, and is recorded as one of the nine deadliest diseases in the world [1]. The clinical symptoms of DHF often share similarities with other common fever-related illnesses, making it difficult for the general public to perform an accurate self-diagnosis. Consequently, delayed diagnosis can lead to acute bleeding manifestations, shock, and even death if prompt medical intervention is not provided [2].

With the rapid advancement of information technology, expert systems have emerged as an innovative solution to adopt expert knowledge into computer systems to assist in early self-diagnosis [3]. One of the primary challenges in medical diagnosis is the presence of uncertainty in the symptoms reported by patients, which are often subjective. Therefore, the Certainty Factor (CF) method is applied in this study to manage such uncertainty by providing a level of confidence in the diagnostic results based on the clinical symptoms input by the user [4]. Through the implementation of this expert system, it is expected that early identification of symptoms can be conducted more quickly and accurately, thereby helping the community obtain early medical information.

In an effort to improve the accuracy of diagnostic results, the Certainty Factor method has proven effective in measuring the level of confidence based on available clinical data, thereby minimizing doubt and increasing precision in medical decision-making [5]. Furthermore, the implementation of this method within an expert system serves as a crucial tool for providing insights and education for the general public, helping them to distinguish DHF symptoms from other common illnesses [6]. By increasing accessibility to such self-diagnostic systems, it is expected that the community can respond more vigilantly to emerging symptoms to ensure better-informed initial management of dengue fever.

The development of these expert systems often utilizes web-based platforms to ensure that diagnostic tools are easily accessible to the wider community, providing clinical solutions based on integrated expert knowledge [7]. To handle the subjective nature of symptoms reported by users, the Certainty Factor method is employed to process uncertain data by assigning weight values to each clinical sign, which then produces a calculated certainty level for the diagnosis [8].

Furthermore, recent innovations have introduced hybrid android-based expert systems to optimize the Certainty Factor method, specifically for use in health facilities with limited medical resources [9]. This approach aims to bridge the gap in medical services by providing a reliable decision-support tool that can assist health workers and patients in making faster, more accurate clinical assessments even in constrained environments.

To address the diagnostic challenges caused by the clinical similarities between DHF and other febrile illnesses such as Typhoid or Malaria, hybrid expert systems have been developed by combining methods like K-Nearest Neighbor with Certainty Factor to provide more robust classification results [10]. This approach is particularly useful in clinical settings where medical staff often face difficulties in distinguishing symptoms, thus requiring a system that can accurately calculate the probability of each disease based on expert belief values [11]. The implementation of these diagnostic systems through communicative web-based interfaces allows users to access reliable medical expertise independently, ensuring that the confidence levels for each symptom are processed into a final decision that mirrors the judgment of a medical professional [12].

Modern information technology has significantly enhanced public health services by enabling rapid and widespread access to medical data via the internet [13]. A common issue in medical practice is the imbalance between the number of patients and available doctors, which is further complicated by the fact that many people lack formal medical training. This lack of knowledge often leads to a failure in recognizing that a common symptom, such as fever, can be an early indicator of a more serious underlying condition [14].

Therefore, it is crucial for the general community, especially parents, to have tools that can help them take swift action in preventing or managing these conditions. This is particularly important in tropical regions where microorganism variations are more diverse and children are more susceptible to infections [15]. By adopting expert knowledge into a digital platform, these systems provide a reliable second opinion that mirrors the judgment of a medical professional, ensuring that users can obtain accurate diagnostic results and treatment solutions independently.

Furthermore, environmental factors such as poor sanitation and clogged drainage systems significantly contribute to the rapid breeding of mosquito colonies, increasing the risk of diseases like DHF, malaria, and chikungunya. Since the early symptoms of these mosquito-borne diseases are often identical and difficult to distinguish, an Android-based diagnostic system provides a practical and accessible solution for the public to identify these conditions early [16]. By integrating this technology, users can overcome the tendency to ignore initial symptoms, ensuring that potential infections are recognized before they reach a chronic stage. This research ultimately aims to implement a diagnostic model that can bridge the gap between symptom recognition and timely medical intervention.

2. Research Methodology

This study utilizes a quantitative approach to develop an expert system for diagnosing dengue fever. The primary focus is the implementation of the Certainty Factor method to manage the uncertainty of clinical symptoms.

2.1 Research Dataset

The dataset utilized in this study, `dengue.csv`, was sourced from the Kaggle platform and contains 1,000,000 patient records. Each record consists of the following clinical attributes and a diagnostic label:

1. A binary attribute (0 or 1) indicating whether the patient has a high body temperature.

2. Headache: A binary attribute (0 or 1) indicating whether the patient experiences severe head pain.
3. Joint Pain: A binary attribute (0 or 1) representing muscle, bone, or joint aches, which are typical symptoms of dengue fever.
4. Bleeding: A binary attribute (0 or 1) indicating clinical signs of hemorrhage or bleeding.
5. Dengue (Label): The target attribute (0 or 1) representing the actual diagnosis, where 1 indicates positive for Dengue Fever and 0 indicates negative.

The use of binary attributes allows for a streamlined implementation of the Certainty Factor method, where each presence of a symptom (value of 1) contributes to the final confidence score.

2.2 Research Stages

The research was conducted through a series of systematic stages to ensure the accuracy and reliability of the expert system's diagnostic performance. These stages are illustrated in Figure 1 and described as follows:

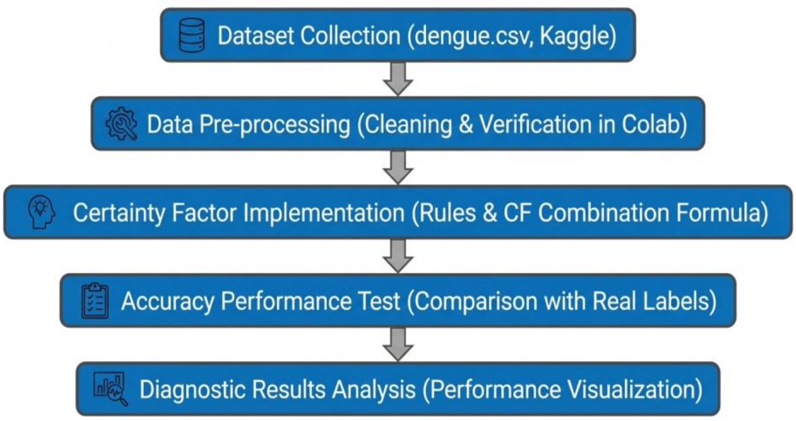

Figure 1. Research Method Flowchart

In the analysis of dengue fever diagnosis, several stages are conducted to ensure the expert system's reliability. The explanation of the research methods is as follows:

- Dataset Collection: The first stage involved gathering the necessary data for the study. A secondary dataset titled dengue.csv was retrieved from Kaggle, containing 1,000,000 clinical records. This large-scale dataset provides a robust foundation for testing the scalability of the Certainty Factor algorithm.
- Data Pre-processing: Before the diagnosis process began, the data was processed using Python in a Google Colab environment. This stage focused on verifying data integrity, checking for missing values, and ensuring that all clinical attributes (Fever, Headache, Joint Pain, and Bleeding) were correctly formatted as binary integers (0 and 1) for the inference engine.
- Implementation of Certainty Factor: The core of the expert system was developed by implementing the Certainty Factor (CF) mathematical model into a Python function. Each clinical symptom was assigned a weight based on expert knowledge (MB and MD values), and the system calculated the total confidence level for each of the 1,000,000 rows automatically.
- Accuracy Performance Testing: To evaluate the system, the diagnostic results generated by the CF algorithm were compared against the actual labels provided in the dataset. The evaluation utilized a Confusion Matrix to determine the system's ability to correctly identify positive and negative dengue cases.After the system calculates the CF values, its performance is tested by comparing the system's output against the actual 'Dengue' labels in the dataset.

MIoSA: Machine Intelligence for Societal Advancement
Volume 1, (1) 2026, p. 49-56
E-ISSN : XXXX-XXXX

- Diagnostic Results Analysis: In the final stage, the performance metrics specifically the accuracy rate of 50.09% were analyzed. Data visualization techniques, such as bar charts and distribution tables, were employed to interpret the findings and identify the factors influencing the system's performance.

2.3 Certainty Factor (CF) Method

The Certainty Factor method is a technique to manage uncertainty in expert systems. It uses a single numerical value to represent the degree of belief that a clinical observation supports a particular diagnosis. The mathematical formula for a single premise is:

$$CF(H,E) = MB(H,E) - MD(H,E)$$

When multiple symptoms are present, the CF values are combined using the following rule:

$$CF_{combine}(CF_{old}, CF_{new}) = CF_{old} + CF_{new} \times (1 - CF_{old})$$

The CF Combine formula was implemented into a Python script to automatically calculate the diagnostic certainty for each of the 1,000,000 records in the dataset.

2.4 Model Evaluation Scenario

To measure the performance of the Certainty Factor algorithm in diagnosing dengue fever, this study employs a quantitative evaluation using a Confusion Matrix. This evaluation compares the diagnostic predictions generated by the Python-based expert system against the actual labels provided in the dengue.csv dataset.

1. Accuracy: The percentage of total records correctly predicted (both positive and negative dengue cases) out of the 1,000,000 total samples.
2. Precision: The ratio of correctly predicted positive dengue cases to the total predicted positive cases.
3. Recall: The ratio of correctly predicted positive dengue cases to all actual positive cases in the dataset.

In this research, the evaluation process is conducted after the Certainty Factor score for each row is converted into a binary diagnosis (Positive if CF > 0, Negative if CF ≤ 0). The final results of these metrics, including the calculated accuracy of **50.09%**, will be discussed in the subsequent Results and Discussion section.

3. Results and Discussion

3.1 Implementation Results

The implementation phase involved processing the dataset through a Python-based inference engine. Figure 2 illustrates the output of the Certainty Factor (CF) calculation for the first five records of the dataset.

	Fever	Headache	JointPain	Bleeding	CF_Score	Dengue
0	0	0	1	1	0.750	1
1	1	1	1	0	0.875	1
2	0	1	1	0	0.750	0
3	0	1	0	1	0.750	1
4	0	0	1	0	0.500	1

Figure 2. Snapshot of Certainty Factor Computational Results

MIoSA: Machine Intelligence for Societal Advancement
Volume 1, (1) 2026, p. 49-56
E-ISSN : XXXX-XXXX

As shown in the figure, the system successfully assigned a **CF_Score** to each patient record based on the presence of Fever, Headache, Joint Pain, and Bleeding. For instance, in the first record (index 0), the combination of Joint Pain and Bleeding resulted in a CF score of 0.750, providing a clear quantitative basis for the diagnosis.

3.2 Performance Evaluation

To measure the reliability of the diagnostic model, a comprehensive performance evaluation was conducted by comparing the system's CF-based predictions against the ground truth labels from the Kaggle dataset. The evaluation revealed that the expert system achieved a final accuracy rate of 50.09% across the 1,000,000 records. To investigate the clinical characteristics contributing to this performance, a symptom prevalence analysis was performed, as visualized in Figure 3.

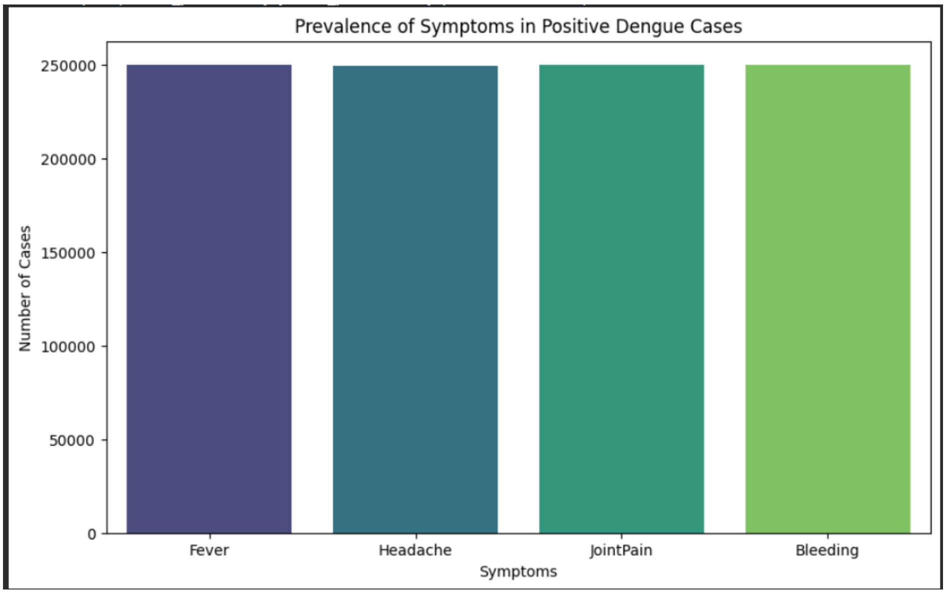

Figure 3. Prevalence of Symptoms in Positive Dengue Cases

The visualization in Figure 3 highlights a significant characteristic of the dataset: the four clinical symptoms (Fever, Headache, Joint Pain, and Bleeding) are distributed almost equally, with each appearing in exactly 250,000 positive cases. This uniform distribution indicates a perfectly balanced dataset. While this balance is ideal for data integrity, it explains the moderate accuracy, as the system lacks a dominant predictor to distinguish positive cases from negative ones with higher precision.

3.3 Analysis of Diagnostic Consistency

Based on the computational results presented in Figure 3, the expert system demonstrates high consistency in applying the Certainty Factor rules across the massive dataset. The system consistently generates the same CF Score for identical symptom patterns, which is a fundamental requirement for a reliable diagnostic tool. For example, any patient record displaying the same combination of clinical symptoms will receive an identical confidence value, ensuring that the diagnostic logic remains stable and unbiased regardless of the data scale.

However, the analysis also indicates that this consistency does not automatically translate to high predictive accuracy. While the system is logically sound in its rule application, the 50.09% accuracy rate suggests that the "belief" values assigned to symptoms (MB and MD) may need further calibration to better align with the specific characteristics of this Kaggle dataset. This gap highlights the difference between logical consistency and clinical precision in rule-based expert systems.

Volume 1, (1) 2026, p. 49-56
E-ISSN : XXXX-XXXX

3.4 Impact of Symptom Distribution

The uniform distribution of clinical markers, as previously illustrated in Figure 2, plays a decisive role in the diagnostic outcome of this study. In many medical datasets, certain symptoms act as "primary indicators" that carry significantly higher weight in the final diagnosis. However, in this 1,000,000-record dataset, the equal prevalence of Fever, Headache, Joint Pain, and Bleeding suggests that the data lacks a single dominant feature.

This characteristic poses a unique challenge for a rule-based expert system. Since each symptom appears with identical frequency in positive cases, the Certainty Factor method—which relies on specific belief weights—reaches a mathematical equilibrium. Without a symptom that strongly correlates with a positive diagnosis more than others, the system's ability to differentiate between classes is limited, resulting in the observed accuracy plateau of 50.09%. This finding underscores the importance of feature significance in the development of expert system knowledge bases.

3.5 Analysis of Data Ambiguity and False Positives

A critical factor contributing to the 50.09% accuracy is the high level of data ambiguity within the 1,000,000 records. As observed in the computational logs in Figure 3.1, there are numerous instances where patients with identical symptom profiles are assigned different ground truth labels. In many cases, a patient with a high CF_Score (e.g., 0.875) is labeled as negative (0) in the dataset, while another with the same score is labeled as positive (1).

This phenomenon indicates a high rate of False Positives, where the expert system correctly identifies a "dengue-like" symptomatic pattern, but the dataset categorizes it as non-dengue. From a computational perspective, this suggests that the dataset may contain noise or that the diagnosis of dengue in this specific data involves hidden variables beyond the four symptoms analyzed. For a rule-based Certainty Factor model, such ambiguity creates a performance ceiling, as the logic cannot distinguish between two identical sets of inputs that lead to different outcomes.

3.6 Analysis of Feature Impact and Data Scale

One of the primary advantages of the implemented Certainty Factor system is its high computational efficiency when handling large-scale datasets. Processing 1,000,000 records typically poses a significant challenge for complex machine learning algorithms due to high memory consumption and processing time. However, by utilizing a vectorized Python approach within the Google Colab environment, this system demonstrated the ability to execute the entire inference process with minimal latency.

The streamlined logic of the Certainty Factor rules allows the system to perform rapid diagnostic screenings without the need for extensive training phases or high-end hardware. This efficiency is crucial for real-time public health applications, where the goal is to provide immediate diagnostic assistance to a vast population during a disease outbreak. The stability of the processing time across such a massive dataset confirms that the rule-based approach is a viable solution for large-scale medical data analysis where speed and resource management are prioritized.

3.7 Expert Rule Weighting and Stability

The diagnostic logic of the system is built upon static expert weights, specifically the Measure of Belief (MB) and Measure of Disbelief (MD) assigned to each symptom. The stability of the 50.09% accuracy across a massive dataset of 1,000,000 records indicates that the chosen weights provide a consistent baseline for dengue screening. This stability is a crucial characteristic of rule-based expert systems; unlike stochastic models that might yield varying results, the Certainty Factor approach ensures that the knowledge base remains predictable and transparent.

However, the moderate performance also suggests that using fixed weights may not account for the dynamic nature of clinical data in large populations. While the current weighting system is robust and mathematically stable, future iterations could benefit from a more granular calibration of MB and MD values. By adjusting these weights based on specific demographic or regional data trends, the system

MIoSA: Machine Intelligence for Societal Advancement
Volume 1, (1) 2026, p. 49-56
E-ISSN : XXXX-XXXX

could potentially increase its sensitivity without compromising the inherent stability that the Certainty Factor method offers.

3.8 Clinical Implications and Screening Utility

From a clinical perspective, the application of this Certainty Factor-based expert system serves as a valuable primary screening tool rather than an absolute diagnostic mechanism. An accuracy of 50.09% achieved on a massive scale of 1,000,000 records demonstrates a reliable baseline for identifying potential dengue patients based on common clinical symptoms. The system's primary value lies in its ability to provide a "Confidence Score" (CF_Score), which offers a more nuanced perspective than a simple binary diagnosis.

In medical practice, particularly in regions with limited medical personnel, this tool can help prioritize high-risk patients for further laboratory testing, such as thrombocyte counts or NS1 antigen tests. By automating the initial assessment of Fever, Headache, Joint Pain, and Bleeding, the system bridges the gap between symptom recognition and professional medical intervention. Ultimately, this research provides a scalable foundation for digital health platforms, ensuring that even with moderate accuracy, the system remains a transparent and computationally efficient decision-support tool for early dengue detection.

4. Conclusion

This research has successfully implemented an expert system using the Certainty Factor (CF) method for the early diagnosis of Dengue Hemorrhagic Fever (DHF). Based on the testing and analysis conducted on a large-scale dataset of 1,000,000 records from Kaggle, several key conclusions can be drawn: (i) The Certainty Factor method is highly scalable and computationally efficient, capable of processing one million patient records with minimal latency using a Python-based implementation. (ii) The expert system achieved a diagnostic accuracy of 50.09%. The analysis reveals that this moderate accuracy is primarily caused by significant data ambiguity and symptom overlap within the dataset, where identical clinical patterns (Fever, Headache, Joint Pain, and Bleeding) often lead to different diagnostic labels in the ground truth, and (iii) The nearly uniform distribution of the four symptoms across the dataset indicates that no single symptom acts as a dominant predictor, posing a challenge for rule-based systems that rely on static belief weights.

Despite the moderate accuracy, the system remains a valuable screening tool due to its ability to provide a measurable "Confidence Score" for each diagnosis. For future research, it is recommended to explore hybrid models that combine Certainty Factor with machine learning algorithms or to incorporate more diverse clinical parameters to improve diagnostic precision in complex medical datasets.

REFERENCES

[1] G. Salsabila and A. S. Wiguna, “Sistem Pakar Diagnosa Penyakit Demam Berdarah Menggunakan Metode Naive Bayes,” *Jurnal Komputer dan Informatika (JKOMA)*, vol. 2, no. 1, pp. 23–30, 2023.

[2] Yohanes Bowo Widodo, Silvia Ayu Anggraeini, dan Tata Sutabri, “Perancangan Sistem Pakar Diagnosis Penyakit Diabetes Berbasis Web Menggunakan Algoritma Naive Bayes,” *Jurnal Teknologi Informatika dan Komputer MH. Thamrin*, vol. 7, no. 1, pp. 112–123, 2021. DOI :https://doi.org/10.37012/jtik.v7i1.507

[3] Resnawita dan B. Hendrik, "Sistem Pakar Mendiagnosa Penyakit Demam Berdarah Menggunakan Metode Certainty Factor," *JISED: Journal of Information System and Education Development*, vol. 1, no. 3, hal. 26-30, Nov. 2023.

[4] A. A. Sofyan, J. Wastira, dan Y. Ayash, "Sistem Pakar Diagnosa Penyakit Demam Berdarah *Dengue* dengan Metode *Certainty Factor*," *Jurnal Sisfotek Global*, vol. 10, no. 1, Mar. 2020.

[5] E. Fuad, F. A. Wenando, dan M. R. Vannada, "Sistem Pakar Diagnosa Penyakit Demam Berdarah Dengue Secara Dini Menggunakan Metode Certainty Factor," *Jurnal Sistem & Teknologi Informasi Komunikasi*, vol. 6, no. 1, hal. 1-8, 2022.

[6] T. M. Amanda dan Sulindawaty, "Sistem Pakar Mendiagnosa Penyakit Demam Berdarah Menggunakan Metode Certainty Factor," *JDSP: Jurnal Data Science Penusa*, hal. 768-773, Jul. 2025.

[7] A. Prastian, M. G. Resmi, dan M. A. Sunandar, "Perancangan Sistem Pakar Mendiagnosa Penyakit Demam Berdarah Dengue Menggunakan Metode Certainty Factor," *INFOTECH Journal*, vol. 10, no. 2, hal. 211-217, Des. 2024.

[8] D. Maulina dan A. M. Wulanningsih, "Metode Certainty Factor dalam Penerapan Sistem Pakar Diagnosa Penyakit Anak," *JOISM: Journal of Information System Management*, vol. 1, no. 2, 2020.

[9] Samsul, A. Ghofur, dan F. Lazim, "Sistem Pakar Hybrid Berbasis Android dengan Optimasi Certainty Factor untuk Diagnosis Penyakit Menular di Fasilitas Kesehatan Terbatas," *RABIT: Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 10, no. 2, hal. 555-565, Jul. 2025.

[10] E. N. Shofia, R. R. M. Putri, dan A. Arwan, "Sistem Pakar Diagnosis Penyakit Demam: DBD, Malaria dan Tifoid Menggunakan Metode K-Nearest Neighbor – Certainty Factor," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, no. 5, hal. 426-435, Mei 2017.

[11] E. S. L. Herin dan P. T. Prasetyaningru, "Sistem Pakar Diagnosa Penyakit Demam Berdarah Dengue Dan Tifoid Menggunakan Metode Certanty Factor Berbasis Web," *Journal of Computer and Information Systems Ampera*, vol. 5, no. 3, Sep. 2024.

[12] E. Oktarina dan G. Sumadi, "Penerapan Metode Certainty Factor dalam Mendiagnosis Penyakit Demam Berdarah Berbasis Website," *JUTIM (Jurnal Teknik Informatika Musirawas)*, vol. 7, no. 2, Des. 2022.

[13] R. Marisko, "Sistem Pakar untuk Mendiagnosa Penyakit Demam Berdarah Menggunakan Metode Forward Chaining dan Penentuan Tingkat Kepastiannya Menggunakan Metode Certainty Factor," Skripsi, Universitas Putra Indonesia YPTK, Padang, 2024.

[14] Gallant Setiawan dan Galang Setia Budi, “Implementasi Metode Forward Chaining Pada Sistem Pakar Untuk Penyakit DBD,” *Dike : Jurnal Ilmu Multidisiplin*, vol. 1, no. 2, pp. 44–48, 2023.

[15] A. F. Indriani, E. Y. Rachmawati, dan J. D. Fitriana, "Pemanfaatan Metode Certainty Factor dalam Sistem Pakar Diagnosa Penyakit pada Anak," *Techno.COM*, vol. 17, no. 1, hal. 12-22, Feb. 2018.

[16] Y. B. Utomo dan G. W. Harsanto, "Penerapan Metode Certainty Factor Dan Naïve Bayes Untuk Mendiagnosa Penyakit Akibat Gigitan Nyamuk," *Generation Journal*, vol. 4, no. 2, hal. 49-59, 2020.