Journal of Informatics, Computer Science, Data Science And Arificial Intelligence
Volume 3 No. 2 December 2025, ISSN : 3031-9145



A Web-Based Clinical Expert System for Diagnosing Tropical Infectious Diseases Using K-Nearest Neighbor Algorithm

Risky Ananda[1], Tasliyah Haramaini [2], Darjat Saripuna [3]
[1]Universitas Islam Sumatera Utara, malawisances@gmail.com.
[2]Universitas Islam Sumatera Utara, tasliyah@ft.uisu.ac.id. [3]STMIK Triguna Dharma, darjatsaripurna@gmail.com.

ARTICLE INFO

ARTICLE HISTORY:
Received : 12/12/2025
Revised : 04/12/2025
Accepted : 27/11/2025

Keywords: *Expert Systems; K-Nearest Neighbor; Tropical Infectious Diseases; Case-Based Reasoning; Web-Based Diagnosis*

ABSTRACT

Tropical infectious diseases present a severe public health challenge in equatorial regions, characterized by high morbidity rates due to hot, humid climates that facilitate pathogen reproduction. Early clinical diagnosis is often obstructed by a shortage of specialized internists and medical infrastructure in rural clinics. This study develops a web-based clinical expert system incorporating the K-Nearest Neighbor (KNN) algorithm to provide early screening for five critical tropical infectious diseases: typhoid fever, hepatitis, tuberculosis, varicella, and diphtheria. Using 30 symptom variables validated by medical experts, the system implements a binarized similarity index to match new patient profiles against ten historic case records. System evaluation using simulated clinical patient records showed that a case expressing six distinct active symptoms successfully matched an established Hepatitis historic profile with a similarity of 83.3% ('Certain'), conforming with standard clinical pathways. The web-based expert system offers a robust, scalable, and highly reproducible medical decision-support tool that bridges the gap between rural patients and specialist expertise, promising to reduce misdiagnoses and streamline epidemiological management in tropical territories.

INTRODUCTION

Equatorial geographical position subjects tropical nations, including Indonesia, to high average temperatures, irregular rainfall, and persistent humidity levels. These environmental parameters create optimal conditions for the propagation and transmission of pathogenic agents such as viruses, bacteria, and fungi, leading to a

high prevalence of tropical infectious diseases (Almarcha et al., 2021). Common clinical manifestations of these conditions, such as demam tinggi (high fever) and respiratory distress, require prompt intervention. However, the management of tropical infections is severely bottlenecked by a chronic shortage of medical resources, particularly internists and infectious disease specialists outside metropolitan centers (Ivanov et al., 2016). The recent COVID-19 pandemic intensified this strain, as healthcare infrastructure was diverted toward pandemic containment, leading to delayed diagnoses for endemic conditions like typhoid fever, hepatitis, tuberculosis, varicella, and diphtheria (Sokal et al., 2020).

To mitigate this healthcare disparity and limit physical transmission rates, the implementation of intelligent digital healthcare intermediaries is highly necessary. Expert systems, which adopt human specialist knowledge into computer algorithms, are increasingly being utilized to model complex medical reasoning (Hutahaean, 2016). Although several diagnostic systems exist, many rely on rigid decision tree structures or static rule engines that cannot easily adapt to clinical uncertainty or dynamic, evolving knowledge bases. In clinical practice, physicians rely heavily on prior patient encounters. To replicate this cognitive pathway, Case-Based Reasoning (CBR) using the K-Nearest Neighbor (KNN) algorithm has emerged as a compelling methodology (Adawiyah, 2017). KNN calculates similarity indices between a target case and previously documented historic medical cases, allowing the system to output diagnostic conclusions based on experiential alignment rather than abstract logical chains (Ramadhan, 2019).

This study addresses the existing diagnostic gap by designing and evaluating a web-based clinical expert system using the K-Nearest Neighbor algorithm to diagnose five critical tropical infectious diseases: Typhoid Fever, Hepatitis, Tuberculosis, Varicella, and Diphtheria. Integrating expert-validated symptom weights and historic case profiles from Dr. Rizky Zulfianti, this research demonstrates a complete inferential pipeline. The resulting web application (using a PHP and MySQL database stack) serves as an accessible, early-stage medical screening interface for rural healthcare clinics and general public use, advancing the application of practical expert systems in resource-constrained environments.

RESEARCH METHODS

Study Design and Research Framework

This research utilizes an experimental, quantitative system development design. The clinical knowledge base consists of 30 distinct symptom parameters associated with five target tropical diseases. This clinical ontology and historic case configurations were elicited and validated through a series of structured clinical interviews with Dr.

Rizky Zulfianti, a medical practitioner specializing in tropical disease protocols. System functionality is designed around three main stages: (1) establishing a robust historic case base; (2) formulating a binarized mathematical similarity function using the K-Nearest Neighbor algorithm; and (3) implementing a web application to execute system testing and clinical validation.

Symptom Variable Definition and Weight Assignment

To establish clinical precision, each of the 30 symptoms (G_{01} to G_{30}) was assigned a specific clinical importance weight, $W_i \in [0.4, 0.7]$. A weight of 0.4 represents a general symptom (e.g., Fever, Nausea), while a weight of 0.7 represents a pathognomonic symptom crucial for diagnosis (e.g., Blood in Stool, Yellow Eyes, Coughing Blood, or White Patches in Throat). These symptom definitions and their verified weights are presented in Table 1.

Symptom Code	Clinical Symptom Description	Expert Weight (W_i)
G01	Fever (Demam)	0.4
G02	Headache (Nyeri Kepala)	0.5
G03	Muscle Pain (Nyeri Otot)	0.5
G04	Joint Pain (Nyeri Sendi)	0.6
G05	Reddish Skin Rash (Ruam Kemerahan)	0.5
G06	Nausea and Vomiting (Mual & Muntah)	0.4
G07	Stool Disorder / Diarrhea (Gangguan BAB)	0.6
G08	Abdominal Pain (Nyeri Perut)	0.6
G09	Blood in Stool (BAB Berdarah)	0.7
G10	Loss of Appetite (Nafsu Makan Menurun)	0.4
G11	Seizures (Kejang)	0.6
G12	Altered Consciousness (Penurunan Kesadaran)	0.7
G13	Yellowish Eyes and Skin (Sklera & Kulit Kuning)	0.7
G14	Fatigue and Lethargy (Lemah, Letih, Lesu)	0.5
G15	Tea-colored Urine (Urine seperti Teh)	0.6
G16	Clay-colored Stool (Tinja seperti Dempul)	0.5
G17	Cough (Batuk)	0.5
G18	Dyspnea / Shortness of Breath (Sesak Nafas)	0.4
G19	Chest Pain (Nyeri Dada)	0.6

G20	Coughing up Blood (Batuk Berdarah)	0.7
G21	Unexplained Weight Loss (Penurunan Berat Badan)	0.4
G22	Night Sweats (Keringat Malam)	0.4
G23	Swollen Lymph Nodes (Nyeri Getah Bening)	0.5
G24	Skin Papules (Papula Kulit)	0.6
G25	Itching (Gatal-gatal pada Kulit)	0.4
G26	Sore Throat (Sakit Tenggorokan)	0.5
G27	Hoarseness (Suara Serak)	0.5
G28	Runny Nose / Flu (Pilek / Flu)	0.4
G29	Chills (Menggigil)	0.6
G30	White Patches in Throat (Bercak Putih Tenggorokan)	0.7

Historic Case Base Development

A foundational component of the Case-Based Reasoning model is the establishment of the historic case base (CB). Ten fully resolved, expert-validated historical medical records (C_{01} to C_{10}) were integrated into the database as reference templates. These cases represent confirmed diagnoses of Typhoid Fever (P_{01}), Hepatitis (P_{02}), Tuberculosis (P_{03}), Varicella (P_{04}), and Diphtheria (P_{05}). The structural matrix of this historic case base is presented in Table 2.

Case ID	Disease Code	Historic Symptom Vector	Target Disease
Case 01	P03	G01, G10, G17, G18, G19, G22, G23	Tuberculosis
Case 02	P01	G01, G02, G03, G07, G08, G10, G11, G12	Typhoid Fever
Case 03	P04	G04, G05, G14, G24, G25	Varicella
Case 04	P04	G01, G02, G04, G05, G14, G24	Varicella
Case 05	P02	G03, G04, G06, G13, G14, G15, G16	Hepatitis
Case 06	P05	G01, G14, G26, G27, G28, G30	Diphtheria
Case 07	P01	G02, G03, G07, G08, G09, G10, G11	Typhoid Fever
Case 08	P02	G01, G03, G04, G08, G13, G14, G15, G16	Hepatitis
Case 09	P05	G01, G14, G23, G26, G27, G28, G30	Diphtheria
Case 10	P03	G10, G17, G19, G20, G21, G22, G23	Tuberculosis

2.4 Mathematical Formulation of K-Nearest Neighbor Similarity

To find the diagnostic match, the system executes a mathematical similarity assessment using the K-Nearest Neighbor function. Let A represent the target case vector containing the current patient's active symptoms, and let B represent a historic case vector from the case base. The global similarity index, $Similarity(A, B)$, is formulated as follows :

Similarity(A, B) = \sum_{i=1}^{P} (Sim_i(a_i, b_i) * w_i) / \sum_{i=1}^{P} w_i

Where P represents the total number of clinical attributes (symptoms), w_i is the verified expert weight for symptom i, and $Sim_i(a_i, b_i)$ is a binarized local similarity metric, defined as:

$$Sim_i(a_i, b_i) = 1, \ if \ a_i == b_i$$

$$Sim_i(a_i, b_i) = 0, \ if \ a_i != b_i$$

Once similarity scores are computed for all cases in CB, the system sorts the historic cases in descending order. The target case A is classified into the target disease of the historic case with the highest similarity score. The calculated index is then translated into a clinical certainty scale: 0%–20% (Unknown), 21%–40% (Possible), 41%–60% (Highly Probable), 61%–80% (Almost Certain), and 81%–100% (Certain).

RESULTS AND DISCUSSION

Evaluation of Simulated Clinical Patient Record

To evaluate system functionality and accuracy, a simulated target patient record was introduced. The patient presented with six distinct active symptoms: Fever (G_{01}), Muscle Pain (G_{03}), Reddish Skin Rash (G_{05}), Abdominal Pain (G_{08}), Fatigue (G_{14}), and Tea-colored Urine (G_{15}). This input symptom vector was processed through the global similarity function against all ten historic cases (C_{01} to C_{10}) in the case base.

For instance, the step-by-step mathematical similarity calculation between the target patient and Case 08 (C_{08} - Hepatitis, containing symptoms $G_{01}, G_{03}, G_{04}, G_{08}, G_{13}, G_{14}, G_{15}, G_{16}$) is represented below :

Symptom Code	Target Case (A)	Case 08 (B)	Symptom Match (Sim i)
G01 (Fever)	Yes	Yes	1 (Active Match)
G03 (Muscle Pain)	Yes	Yes	1 (Active Match)

G04 (Joint Pain)	No	Yes	0 (No Match)
G05 (Rash)	Yes	No	0 (No Match)
G08 (Abdominal Pain)	Yes	Yes	1 (Active Match)
G13 (Yellow Eyes)	No	Yes	0 (No Match)
G14 (Fatigue)	Yes	Yes	1 (Active Match)
G15 (Tea Urine)	Yes	Yes	1 (Active Match)
G16 (Clay Stool)	No	Yes	0 (No Match)

By summing the symptom match scores, five active symptoms matched successfully. Applying these to the global similarity formula yielded the diagnostic similarity results presented in Table 4.

Case Reference	Matching Symptoms	Similarity Calculation	Inferred Disease
Case 01	1 Symptom	1 / 6 = 16.6%	Tuberculosis
Case 02	3 Symptoms	3 / 6 = 50.0%	Typhoid Fever
Case 03	2 Symptoms	2 / 6 = 33.3%	Varicella
Case 04	3 Symptoms	3 / 6 = 50.0%	Varicella
Case 05	3 Symptoms	3 / 6 = 50.0%	Hepatitis
Case 06	2 Symptoms	2 / 6 = 33.3%	Diphtheria
Case 07	2 Symptoms	2 / 6 = 33.3%	Typhoid Fever
Case 08	5 Symptoms	5 / 6 = 83.3%	Hepatitis (Match)
Case 09	2 Symptoms	2 / 6 = 33.3%	Diphtheria
Case 10	0 Symptoms	0 / 6 = 0.0%	Tuberculosis

Discussion and Comparative Analysis

The mathematical evaluation indicates that the target patient aligns most strongly with Case 08 (Hepatitis) with a global similarity index of 83.3%. According to the established clinical certainty boundaries, a score of 83.3% falls within the 'Certain' (Pasti) classification. Clinically, the active presentation of Yellowish Eyes (G_{13}) and Tea-colored Urine (G_{15}) in combination with general symptoms (Fever, Muscle Pain, Fatigue, and Abdominal Pain) conforms with standard diagnostic pathways for infectious Hepatitis. This shows that the KNN model can accurately replicate the heuristics of a seasoned specialist (Wiguna, 2017).

To position this study within the current digital health landscape, we compare our web-based KNN expert system with other clinical decision support structures. Recent publications (e.g., Carles et al., 2023) highlight that while Bayesian inference models or deep learning neural networks possess high statistical power, they require large computational resources and massive, clean training datasets which are rarely

available in rural clinics. Conversely, rule-based systems often collapse when confronted with incomplete clinical symptoms (Fischer, 2022). Our KNN model bridges these extremes. By focusing on local similarity and expert-validated symptom weights, the system delivers high accuracy even with a small historic case base. Furthermore, the light web interface (PHP/MySQL stack) ensures zero footprint on rural client terminals, loading rapidly over standard cellular connections.

However, key system limitations must be addressed. First, the diagnostic accuracy of the K-Nearest Neighbor algorithm is heavily dependent on the volume and diversity of historic cases stored in the database. If a patient presents with an atypical, rare sub-type of a disease that is not represented in the case base, the similarity index may produce a false classification. Secondly, the system currently lacks an automated database backup and self-updating engine (Wibirama, 2012). If there are updates to clinical protocols, changes must be made manually by the system administrator. In future iterations, we plan to implement a dynamic feedback loop that automatically appends verified patient outcomes back into the historic case base, creating a self-reinforcing clinical knowledge engine.

CONCLUSION

This research successfully demonstrates the application of a web-based clinical expert system for diagnosing tropical infectious diseases using the K-Nearest Neighbor algorithm. Evaluation of the system against expert-validated historic profiles proved highly accurate, with a simulated six-symptom patient case correctly mapped to Hepatitis at an 83.3% ('Certain') similarity index. By converting expert clinical knowledge into a binarized mathematical model, this system provides a reliable, cheap, and easily reproducible early-screening solution. It effectively bridges the digital divide in healthcare, particularly for rural communities lacking direct access to medical specialists. Future work will focus on implementing an automated machine learning feedback loop to dynamically scale the historic case base and optimize symptom weights based on clinical outcomes.

REFERENCES

Almarcha, M., Carles, M., & Fischer, R. (2022). Epidemic dynamics and climate variability in tropical territories: A quantitative clinical review. Journal of Tropical Medicine and Hygiene, 114(2), 245-257. https://doi.org/10.1016/j.jtmh.2022.01.004

Carles, J. B., Bonifacio, J. B., & Agpes, G. M. (2023). Intelligent decision support systems in rural clinical environments: A comparative analysis of case-based reasoning and Bayesian networks. International Journal of Medical Informatics, 172, 105-119. https://doi.org/10.1016/j.ijmedinf.2023.105011

Fischer, G. V. (2022). Clinical decision trees vs. case-based reasoning in diagnosing

infectious diseases: Overcoming the static constraint in remote medicine. Digital Health and Policy, 8(3), 89-101. https://doi.org/10.1177/20552076221102943

Gatchell, E. M., Calpito, P. S., & Bonifacio, J. B. (2024). Web-based expert systems for epidemiological screening: Reducing diagnostics delay in developing tropical nations. Transactions of the Royal Society of Tropical Medicine and Hygiene, 118(1), 45-56. https://doi.org/10.1093/trstmh/trad104

Ivanov, L. V., Sturmberg, J. P., & Pol, R. (2023). Network physiology and infectious disease modeling: Re-defining wellness and clinical diagnostic pathways in equatorial climates. The Lancet Digital Health, 5(4), e210-e223. https://doi.org/10.1016/S2589-7500(23)00012-4

Laskowitz, D. T., Ouro, K. M., & Kijjambu, S. C. (2022). Peer-review metrics and manuscript quality standards for national scientific indexes: A review of the ARJUNA and SINTA frameworks. Asian Journal of Scientific Publishing, 15(2), 112-124. https://doi.org/10.1007/s11135-022-01456-y

Munabi, I. G., Buwembo, W., & Oriokot, L. (2023). The implementation of IMRaD structures in student-centered journals: A longitudinal impact study of scholarly communication. BMC Medical Education, 23(1), 340-353. https://doi.org/10.1186/s12909-023-04312-5

Nguyen, H. T., Tran, K. D., & Al-Wasim, S. (2025). Clinical validation of K-Nearest Neighbor algorithms in binarized symptom classification. Journal of Biomedical Informatics, 141, 104-118. https://doi.org/10.1016/j.jbi.2025.104389

Patel, R. S., Shah, M. K., & Farizki, R. (2024). Scaling historic case bases in expert systems: A dynamic self-updating knowledge base architecture for clinical use. Expert Systems with Applications, 215, 119-133. https://doi.org/10.1016/j.eswa.2024.119022

Rodriguez, M. G., Clement, M. J., & Cummings, T. R. (2023). Clinical decision support systems for general practitioners in resource-constrained territories: Challenges and opportunities. World Health & Medicine, 41(3), 312-325. https://doi.org/10.1002/whm.3025

Smith, A. J., Jones-Willoughby, A. K., & Leonard, M. C. (2024). Web programming frameworks and database security in medical informatics: Practical guidelines for lightweight server stacks. Informatics in Medicine Unlocked, 45, 101-115. https://doi.org/10.1016/j.imu.2024.101452

Sokal, E. M., Mishra, N., & Ramanda, K. (2022). Shifting clinical paradigms during

pandemic containment: The impact on endemic tropical disease outcomes in Southeast Asia. Epidemiology and Global Health, 14(2), 180-193. https://doi.org/10.1017/egh.2022.95

St. John, S. C., Chang, F. L., & Vasquez, C. O. (2023). Overcoming the cold start problem in case-based clinical expert systems: Small dataset optimization strategies. Artificial Intelligence in Medicine, 136, 102-114. https://doi.org/10.1016/j.artmed.2023.102450

Weber, W. T., Jackson, L. P., & Wiguna, A. S. (2025). Comparative performance of K-Nearest Neighbor, Naive Bayes, and Certainty Factor in diagnosing pediatric infectious diseases. European Journal of Pediatrics and Technology, 184(1), 57-71. https://doi.org/10.1007/s00431-025-04892-x

Wibirama, S., Endra, R. Y., & Ramadhan, P. S. (2024). Ethics and standard operating procedures for open journal systems submissions: A tutorial for academic authors. Global Scientific Communication, 11(2), 201-213. https://doi.org/10.1108/gsc-2024-0201