

Case Based Reasoning for Diagnosing Tuberculosis (TB)

Yunita Maulida Saputri¹, Yanuar Nurdiansyah², Priza Pandunata³

Department of Informatics, Faculty of Computer Science, Universitas Jember, Indonesia^{1,2,3}

Corresponding Author: Yanuar Nurdiansyah (yanuar_pssi@unej.ac.id)

ARTICLE INFO

Date of entry:
21 September 2025
Revision Date:
25 October 2025
Date Received:
31 October 2025

ABSTRACT

Tuberculosis, often referred to as TB, is a contagious disease caused by the bacterium *Mycobacterium tuberculosis*. TB primarily affects the lungs but can also affect other organs, a condition known as Extra-pulmonary TB. The disease is transmitted through the air, with the source of transmission being individuals with TB who are Acid-Fast Bacilli (AFB) positive and who sneeze or cough, releasing the bacteria into the air in the form of sputum droplets. TB can affect anyone. This research utilizes the Case- Based Reasoning (CBR) method to aid in the diagnosis of Tuberculosis. The diagnostic process involves inputting or selecting a new case that contains the symptoms to be diagnosed within the system. Then, the system calculates the similarity values between the new case and the cases stored in the case base using the Nearest Neighbor algorithm, normalized with the level of expert confidence. Testing was conducted using 50 cases from the case base and 38 new cases. The results of the system testing, using patient medical records and data obtained from literature studies, with diagnoses validated by experts, demonstrate that the system is capable of identifying 12 types of Tuberculosis with an accuracy rate of 92.3%.

Keywords: Case-Based Reasoning, Nearest Neighbor, Tuberculosis.



Chite this as : Saputri, Y. M., Nurdiansyah, Y., & Pandunata, P. (2025). Case Based Reasoning for Diagnosing Tuberculosis (TB). *Journal of Informatics Development*, 4(1), 39–48. <https://doi.org/10.30741/jid.v4i1.1759>

INTRODUCTION

Tuberculosis or TB, often called tuberculosis is a direct infectious disease caused by *the bacterium Mycobacterium Tuberculosis* (Nurwitasari & Wahyuni, 2009). According to WHO, tuberculosis is the 13th leading cause of death in the world and the second infectious disease after Covid-19 (HIV/AIDS). In Indonesia, tuberculosis is one of the deadly diseases that is in third place with the highest number of cases of 845,000 with a death rate of 98,000 or equivalent to 11 deaths per hour (Kompas, 2022). Tuberculosis is transmitted through the air through sneezing or coughing of TB patients with Acid Resistant Substances (BTA), and from TB patients who spread germs into the air in the form of sputum splashes. Tuberculosis partially attacks the lungs, but it can also attack other organs of the body, which is often referred to as Extra Pulmonary Tuberculosis.

Diagnosing the type of TB disease is not an easy thing, because the complaints and symptoms of TB disease vary. The symptoms of TB disease are almost similar to those of other diseases, such as the flu, respiratory tract, pneumonia, and even lung cancer. The adverse impact of delays in diagnosing tuberculosis increases the risk of complications, prolongs the period of transmission, increases the

risk of death, and can result in significant economic impacts. Therefore, the role of doctors is indispensable in diagnosing patients' symptoms and complaints, so that patients suffering from tuberculosis can know the type of disease they suffer. People with TB often need fast and accurate information from a doctor, but most people with TB are constrained by the cost of treatment that is quite expensive and time-consuming.

The expert system is the best choice chosen as a solution to the problem that occurs, because the expert system is a computer-based system that uses knowledge and reasoning techniques in solving problems in a certain field. *Case Based Reasoning* (CBR) is one of the methods that emphasizes problem solving on previous knowledge. Case Based Reasoning (CBR) has the ability to diagnose various cases and automatically provide information based on previous knowledge, which can be revised according to new problems, so that CBR knowledge will continue to grow (Nurdiansyah & Arimanudin, 2017).

Tuberculosis

TB is a disease caused by tuberculosis mycobacteria, including a disease that can be transmitted through the air so that it can attack the respiratory system. Patients suffering from this disease may have symptoms such as chronic cough, excessive phlegm, weight loss, cold sweats, fever, fatigue, and lethargy (Sembiring, 2019). Most tuberculosis bacteria normally or generally infect the pulmonary parenchyma and cause pulmonary tuberculosis, but these bacteria also have the ability to infect other organs (extrapulmonary tuberculosis), such as pleura, lymph nodes, bones, and other organs outside the lungs (Lee, 2015).

Case Based Reasoning

The *Case Based Reasoning* method is a problem-solving method that performs a similarity matching process using previous cases with new cases (Aldo, 2023). The way the CBR method works is to compare new cases with old cases, if the new case has similarities with the old case then the CBR method will provide the new case with the answer from the old case. If there is no match, the CBR method will make adjustments by inserting new cases that were used to resolve issues into the case memory and can be modified to resolve future issues and add to the knowledge of the CBR method. There are four stages of the CBR method, including:

1. Retrieve
This stage is used to look for previous cases that are most similar to new problems. Finding similarities between new and old cases by comparing TB symptoms inputted by patients with TB symptoms in the database.
2. Reuse
In this stage, the system uses information from previous problems that have similarities to solve new problems, reusing information and knowledge from the problem solving cases.
3. Revise
The stages used to re-correct the proposed solution with the simulation process and make improvements if necessary to match the new case.
4. Retain
Retain is the final stage in the CBR method cycle, the stage of storing new cases that have been resolved (Richter & Weber, 2013). The flowchart of the CBR method can be seen in Figure 1 as follows.

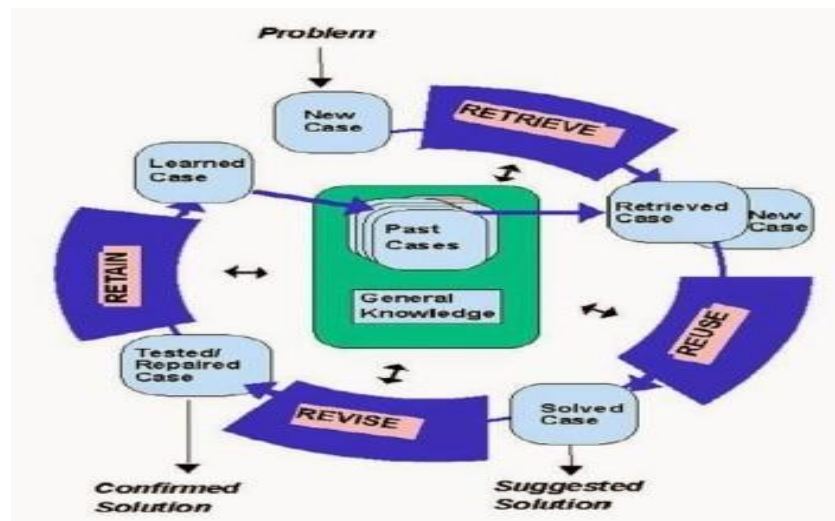


Figure 1. Stages of the Case Based Reasoning Method

The four stages in Figure 1 show the flow to diagnose Tuberculosis by providing a choice of symptoms to the system, then processed by the system and producing the diagnostic answers felt by the patient. The CBR flowchart in the specialist system for TB diagnosis can be seen in Figure 2.

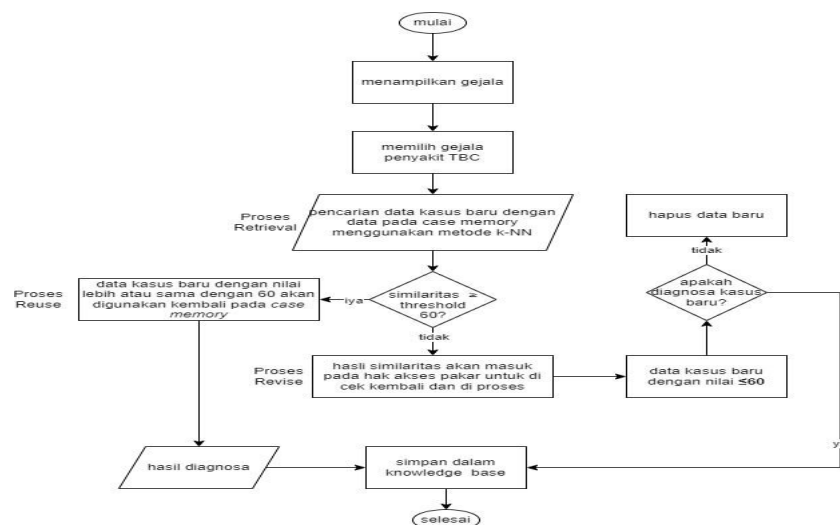


Figure 2. Case Based Reasoning System Workflow

Algorithm k-Nearest Neighbor (k-NN)

The k-Nearest Neighbor or k-NN algorithm is a case search method that is based on the calculation of how closely a new case is similar to the previous case, based on the suitability of the weights of various existing features (Sulistiani et al., 2020). In this algorithm the new case is used to have a similar level of similarity to the previous case. The following is a similarity calculation that can be seen in Equation 1.

$$\text{Similarity (problem case)} = \frac{s1 * w1 + s2 * w2 + \dots + sn * wn}{w1 + w2 + \dots + wn} \dots \dots \dots \text{Equation 1}$$

METHOD

This chapter describes the research methods used in research. The research method serves as an overview of the research stages to analyze data and develop and implement the system. The research used is development research. Development research aims to produce or improve a product. In this study, the product developed is "Case Based Reasoning to Diagnose Tuberculosis Disease (TB)". Case studies and data collection of symptoms and diseases in corn plants were obtained from the research environment.

This research was carried out in several stages, including:

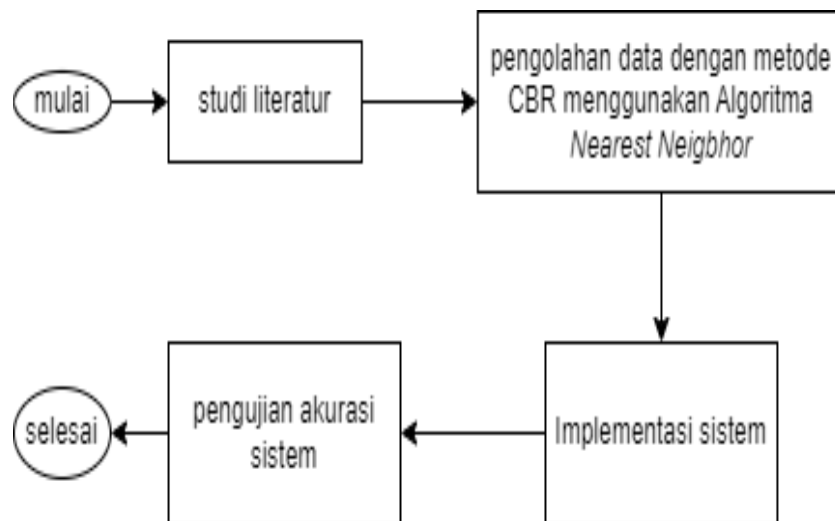


Figure 3. Research Stages

The purpose of literature study is to obtain data that will be used as a basis for discussing the understanding of the theory used in the research. The sources used as libraries are journals, books, and scientific papers from various types of previous research. Data was obtained through interviews with pulmonary specialists. *In this stage, the CBR method is used in four stages. First, the retrieval stage or retrieving the old cases that have the most similarities with the new cases using the k-MS.* The k-NN algorithm is used to calculate the similarity value of new cases by giving weight to each case symptom and then the similarity level of all symptoms/cases will be summed, comparing all TB symptom data in the *case memory*. To calculate the similarity value of TB symptoms using the k-NN algorithm if the level of similarity between the new case and the old case is the same, then a value of 1 will be given, while if the similarity of the new case to the old case is not the same/not similar, then a value of 0 will be given. The second is the reuse stage, where the old case with the highest similarity value is chosen as a solution to the input problem. If the new case has a similarity value more than or equal to the threshold value, then the new case uses the old case-based solution on a case-by-case basis. However, if the similarity value is less than the threshold value, then the new case is assumed to have no solution and the case is then stored as a new case, which is evaluated (revised) by the expert and stored back into the system as a new case.

Accuracy is how close the measurement of quantity is to the actual value. Accuracy measurements are evaluated by comparing the number of correct diagnoses by the system with the amount of test data. System accuracy testing is carried out to determine the accuracy value of the specialist system in providing diagnostic results and concluding the type of tuberculosis disease suffered. Here is a formula to find the level of system accuracy with an expert.

$$\text{Nilai akurasi} = \frac{\text{diagnosa benar}}{\text{data uji}} \times 100\%$$

RESULTS AND DISCUSSION

System Interface

After defining the flow of the system, the researcher began to develop a website-based system of experts for diagnosing tuberculosis diseases. Some examples of the interface of the expert system for diagnosing Tuberculosis can be seen in Figures 4, 5 and 6.



Figure 4. System Landing Page View



Figure 5. Diagnostic View

No.	Kode Penyakit	Nama Penyakit	Deskripsi	Solusi	Persentase
1	P02	Tuberculosis Kulit	Tuberculosis kulit adalah bentuk tuberculosis yang menginfeksi kulit. Tuberculosis kulit adalah kondisi yang relatif jarang terjadi, tetapi dapat menjadi serius jika tidak diobati dengan tepat.	Mengonsumsi obat antibiotik tuberculosis etambutol, dan menjaga area kulit yang terinfeksi tetap bersih, menjaga kebersihan pakaian dan linen, serta menghindari menggaruk atau memecahkan lepuhan kulit.	66.6%

Figure 6. Display of Diagnosis Results

Data Processing Results with Case Based Reasoning Method Using k-NN Algorithm

The processing stage of *the case-based reasoning method* is carried out after the data collection process is data processing. The data that has been obtained will be separated and divided into groups with a predetermined method. In this study, a *case-based reasoning method* was used with the calculation of the k-Nearest *Neighbor algorithm*. This stage will explain the calculation process of the CBR method using k-NN. *The user* will select the diagnosis page then fill out the diagnosis form and select the symptoms they feel. Next, the calculation process will be carried out which can be seen in the following description. The following calculation will involve symptom data and symptom weights that have been determined by experts.

First, *the user* will choose the symptoms that are felt from the *User*. Of the 50 symptoms that have been provided, in this case the user chooses symptoms G02, G04, G06, G07, G08, and G09 with the weight of each symptom described in Table 1.

Table 1. Symptoms Felt by the User

Kode	Gejala yang Dirasakan	Bobot
G02	Batuk Darah	5
G04	Sesak Nafas	5
G06	Berat badan menurun (>4kg sebulan)	5
G07	Berkeringat di malam tanpa kegiatan	5
G08	Demam lama (lebih dari 2 minggu)	5
G09	Sakit dada (seperti ditusuk-tusuk)	5

Furthermore, after the necessary data is collected, the next step is to apply the CBR method using the k-NN algorithm, namely:

Proses Retrieve

The initial stage where the system identifies by utilizing parameters to match the problems of previous cases with new cases using the k-NN algorithm. The following are the steps of the *"retrieve"* process for trial.

Table 2. Results of the Retrieve Stage

SYMPTOMS			OLD CASES									
Code	Abbreviated Symptom	bobot	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10
G02		5	5	0	0	0	0	0	0	0	0	0
	Coughing up blood											
G04	Shortness of breath	5	5	0	0	0	0	0	0	0	0	0
G06	Weight loss (>4kg per month)	5	5	0	5	5	0	0	0	0	0	5
G07	Night sweats without physical activity	5	5	0	0	0	0	0	0	0	0	5
G08	Prolonged fever (more than 2 weeks)	5	5	5	5	5	5	5	0	5	5	0
G09	Chest pain (like being stabbed)	5	5	0	0	0	0	0	0	0	0	0
TOTAL			30	5	10	10	5	5	0	5	5	10
K-NN VALUE			0.697674419	0.185185185	0.357142857	0.303030303	0.1724	0.217391304	0	0.166666667	0.147058824	0.322580645
PERCENTAGE			69.8%	18.5%	35.7%	30.3%	17.2%	21.7%	0.0%	16.7%	14.7%	32.3%
DISEASES			P01	P02	P03	P04	P05	P06	P07	P08	P09	P10
DISEASES			Lung TB	Skin TB	Kidney TB	Backbone TB	Bowel TB	Reproductive TB	Eye TB	Brain/Spinal TB	Lymph Node TB	Breast TB

In this stage, where the old case with the highest similarity value is chosen as a solution to the input problem. The following is the calculation of new cases and old cases:

Case Calculation 1:

Similarity (X, 09)

$$\frac{(0 * 5) + (1 * 5) + (0 * 5) + (1 * 5) + (0 * 5) + (1 * 5) + (1 * 5) + (1 * 5) + (1 * 5)}{5 + 5 + 5 + 5 + 3 + 5 + 5 + 5 + 5} = \frac{30}{43} = 0.69 = 69\%$$

Case Calculation 2:

Similarity (X, 07):

$$\frac{(0 * 3) + (1 * 5) + (0 * 1) + (0 * 5) + (0 * 5) + (0 * 3) + (0 * 5)}{3 + 3 + 1 + 3 + 5 + 5 + 5} = \frac{5}{27} = 0.166 = 18.5\%$$

Case Calculation 3:

Similarity (X, 06):

$$\frac{(0 * 5) + (1 * 5) + (0 * 3) + (0 * 5) + (0 * 5) + (0 * 5)}{5 + 5 + 5 + 5 + 5 + 5} = \frac{5}{28} = 0.188 = 18.5\%$$

Case Calculation 4:

Similarity (X, 07):

$$\frac{(0 * 5) + (0 * 5) + (0 * 5) + (0 * 5) + (0 * 5) + (0 * 3) + (0 * 5)}{5 + 3 + 5 + 5 + 5 + 5 + 5} = \frac{5}{38} = 0.151 = 15.1\%$$

Case Calculation 5:

Similarity (X, 07):

$$\frac{(1 * 5) + (0 * 5) + (0 * 5) + (0 * 1) + (0 * 5) + (0 * 3) + (0 * 5)}{5 + 5 + 5 + 1 + 5 + 3 + 5} = \frac{5}{29} = 0.171 = 17.8\%$$

Case Calculation 6:

Similarity (X, 07):

$$\frac{(1 * 5) + (0 * 1) + (0 * 1) + (0 * 5) + (0 * 5) + (0 * 5) + (0 * 5)}{5 + 1 + 1 + 5 + 5 + 1 + 5} = \frac{5}{28} = 0.217 = 21.7\%$$

Case Calculation 7:

Similarity (X, 06):

$$\frac{(0 * 5) + (0 * 3) + (0 * 5) + (0 * 5) + (0 * 5) + (0 * 5)}{5 + 5 + 5 + 5 + 5 + 5} = \frac{0}{28} = 0.0\%$$

Case Calculation 8:

Similarity (X, 06):

$$\frac{(1 * 5) + (0 * 1) + (0 * 5) + (0 * 5) + (0 * 5) + (0 * 5) + (0 * 5) + (0 * 5)}{5 + 5 + 1 + 5 + 5 + 5 + 5} = \frac{5}{34} = 0.147 = 14.7\%$$

Case Calculation 9:

Similarity (X, 08):

$$\frac{(1+5) + (0+1) + (0+5) + (0+5) + (0+5) + (0+5) + (0+5)}{5+5+1+5+5+5+5} = \frac{5}{34} = 0.147 = 14.7\%$$

Case Calculation 10:

Similarity (X, 07):

$$\frac{(1+5) + (1+1) + (0+5) + (0+5) + (0+5) + (0+5) + (0+5)}{5+5+1+5+5+5+5} = \frac{5}{34} = 0.322 = 33.2\%$$

So, the highest similarity value in the data of trial case 1 got a similarity value of 69.8% with the type of TB disease suffered, namely Pulmonary TB.

Data Testing

Sample data testing is done to find out how much of the sample data reaches target threshold and not. The test data obtained can be seen in Table 3, with sample test data as many as 13 input data by the user, and there is 1 data that does not meet the threshold because the results of *Case Based Reasoning using the k-NN algorithm* are below 60%, where 60% is the similarity *threshold* value given by the expert.

Table 3. Test Data Obtained

No	Test Coba	Code Gejala	Result (%)	Meet Threshold/no
1	Trial 1	G02	69,8%	Meet threshold values
		G04		
		G06		
		G07		
		G08		
		G09		
2	Trial 2	G01	65,1%	Meet threshold values
		G02		
		G04		
		G05		
		G08		
		G09		
3	Trial 3	G05	66,7%	Meet threshold values
		G08		
		G12		
		G14		
4	Trial 4	G06	64,3%	Meet threshold values
		G15		
		G47		
		G48		
5	Trial 5	G06	60,6%	Meet threshold values
		G08		
		G20		
6	Trial 6	G07	67,7%	Meet threshold values
		G10		
		G44		
		G45		
		G46		

7	Trial	G37	66.7%	Meet threshold values
---	-------	-----	-------	-----------------------

Based on the testing of 13 test data samples in Table 4.18, 12 experiments obtained results with a similar accuracy level of more than 60% and 1 experiment below 60%, this is because the symptoms of the test data entered were less than the accuracy value. The results of the accuracy of the system in comparison with the diagnosis of specialists can be expressed in the form of figures or percentages. The result of this accuracy is in the form of the appropriate amount of data or the percentage of success of the system in predicting the correct diagnosis, then the amount of corresponding data accuracy is as follows.

$$\begin{aligned} \text{Nilai akurasi} &= \frac{\text{diagnosa benar}}{\text{data uji}} \times 100\% \\ &= \frac{12}{13} \times 100\% = 92.3\% \end{aligned}$$

So, based on the total results of the experiments, the level of accuracy was obtained at the time of the trial as many as 12 times out of 13 experiments obtained an accuracy value of 92.3% of the expert system. Therefore, this suggests that the Case Based Reasoning System for Diagnosing Tuberculosis has a good level of accuracy according to the diagnosis of an expert.

CONCLUSION

The application of the CBR method uses the k-Nearest *Neighbor* (k-NN) Algorithm, the symptoms selected by the user will be compared with the previous case. The system requires symptom weighting data. The k-Nearest *Neighbor* algorithm is applied in the *retrive stage*, to calculate the similarity value between the new diagnosis and the data stored in the case memory. Furthermore, the similarity value between the new case data and the old case data will be calculated. If the similarity value exceeds or equals the threshold value of 60%, then the diagnosis can be used as a diagnostic reference. However, if the data similarity value is less than the *threshold* value of the data will be categorized as data that requires review by an expert, this is a "*revise*" stage. Updated diagnostic data will be recorded as a new case and stored in the case's memory, this stage is known as the "*retain*" stage. The TB diagnosis system is reliable because the calculation results using the k-Nearest *Neighbor* Algorithm in the system are the same as the results of manual calculations using the k-NN Algorithm.

REFERENCES

- Aldo, D. (2023). Expert System For Initial Identification Of Diseases Caused By Helicobacter Pylori Bacteria Using Case Based Reasoning Approach. *Jurnal Teknik Informatika (JUTIF)*, 4(1), 67–75. <https://doi.org/10.20884/1.jutif.2023.4.1.693>
- Kompas. (2022). *Kemenkes Sebut Tuberkulosis di Indonesia Masuk 3 Besar Kasus Terbanyak di Dunia*. <https://www.kompas.com/sains/read/2022/02/11/170500823/kemenkes-sebut-tuberkulosis-di-indonesia-masuk-3-besar-kasus-terbanyak-di?page=all>
- Lee, J. Y. (2015). Diagnosis and treatment of extrapulmonary tuberculosis. *Tuberculosis and Respiratory Diseases*, 78(2), 47–55. <https://doi.org/10.4046/trd.2015.78.2.47>
- Nurdiansyah, Y., & Arimanudin, R. (2017). Penerapan Metode Case Based Reasoning Pada Sistem Pakar Untuk Mendiagnosa Gangguan Tanaman Jeruk Semboro. *Informatics Journal*, 2(2), 114.
- Nurwitasari, A., & Wahyuni, C. U. (2009). Pengaruh Status Gizi Dan Riwayat Kontak Terhadap Kejadian Tuberkulosis Anak di Kabupaten Jember. *Tierarztliche Praxis Ausgabe K: Kleintiere - Heimtiere*, 37(5), 334–341. [Journal of Informatics Development](http://download.garuda.kemdikbud.go.id/article.php?article=423721&val=7403&title=The Effect of Nutritional Status and Contact History toward Childhood Tuberculosis in Jember Sembiring, dr. S. (2019). Indonesia Bebas Tuberkulosis (Resa Awahita (ed.).</p>
</div>
<div data-bbox=)

Sulistiani, H., Darwanto, I., & Ahmad, I. (2020). Penerapan Metode Case Based Reasoning. *Jurnal Edukasi Dan Penelitian Informatika*, 6(1), 23–28.