Journal of Information System Exploration and Research

https://joiser.shmpublisher.com

p-ISSN 2964-1160 | e-ISSN 2963-6361

Smart Expert System for Tuberculosis Diagnosis Using the Naïve Bayes Method

Ulumuddin[1*], Siti Harlina[2], Warjiono[3] , Amin Nur Rais[4] , Arham Arifin[5] , Abdul Ibrahim[6] , James Adam Seo[7]

[1,3,4]*Department of Information Technology, Bina Sarana Informatika University, Indonesia*
[2,5,6] *Department of Information Technology, Dipa Makasar University, Indonesia*
[7]*Department of Information Technology, Citra Bangsa University, Indonesia*

DOI: https://doi.org/10.52465/joiser.v4i3.93
Received 27 June 2026; Accepted 27 July 2026; Available online 29 July 2026

Article Info

Keywords:
Expert system;
Tuberkulosis;
Naive bayes;
Disease diagnosis;
Medical informatics

Abstract

Tuberculosis (TB) remains one of the major infectious diseases worldwide, requiring accurate early diagnosis to reduce transmission and improve treatment outcomes. Although numerous machine learning-based diagnostic systems have been developed, most previous studies have primarily focused on improving classification accuracy without integrating an expert system to support clinical decision-making. This study aims to develop a Smart Expert System for the early diagnosis of tuberculosis using the Naïve Bayes algorithm. The novelty of this research lies in the integration of the Naïve Bayes algorithm with a web-based expert system that not only generates diagnostic predictions but also provides recommendation-based decision support. The proposed system was developed using 200 tuberculosis patient records consisting of demographic data and clinical symptoms. The research process included data preprocessing, model training, system implementation, and performance evaluation using Accuracy, Precision, Recall, and F1-score metrics. The experimental results achieved an Accuracy of 87.00%**,** Precision of 85.00%**,** Recall of 88.00%**,** and an F1-score of 86.00%**,** indicating reliable classification performance. The developed system assists healthcare professionals in conducting preliminary TB screening, accelerating clinical decision-making, and improving public access to early tuberculosis diagnostic information.

1. Introduction

The way that information technology and artificial intelligence have come along has really changed a lot of things in our lives including the way we take care of our health [1], [2]. One of the things that

[*] ***Corresponding Author:***

Ulumuddin,
Department of Information Technology, system information
Bina Sarana Informatika University,dipa makasar university, citra bangsa university,
Tegal central java,south Sulawesi, east nusa tengara, Indonesia
Email:ulumuddin.udn@bsi.ac.id

artificial intelligence has been used for is something called an Expert System. An Expert System is a computer program that is designed to think like an expert and make decisions like they would. It does this by using rules and knowledge that people have already figured out. In the healthcare world Expert Systems are used to help doctors figure out what is wrong with someone suggest what treatment they should get and help them make decisions quickly [3], [4], [5], Using Expert Systems in medicine has really worked out well in places where there are not many doctors or hospitals. Artificial intelligence and Expert Systems are really changing the way that healthcare works. The use of Expert Systems, in healthcare is a deal because artificial intelligence is getting better and better at helping us with our health [6].

Tuberculosis or TB is still a serious disease that affects many people worldwide.TB is caused by a type of bacteria called Mycobacterium tuberculosis.This bacteria mainly affects the lungs.When people, with TB cough or sneeze they can spread it to others through the air.TB is a health concern globally The disease can be spread easily.Mycobacterium tuberculosis is the cause of TB.Lungs are the target of this bacteriaTB continues to be a problem [7]. Tuberculosis has some signs like coughing that will not go away fever, losing weight sweating at night and chest pain. People often think these signs are minor problems so they wait to go to the doctor. This is not good because it means they do not get the help they need away. Tuberculosis can get very bad. People can get other people sick if they do not get medical help. Tuberculosis is an issue and people need to know more, about tuberculosis and its signs [8], [9].

According to the World Health Organization (WHO) Global Tuberculosis Report 2025, tuberculosis (TB) remains one of the world's deadliest infectious diseases. In 2023, an estimated 10.8 million new TB cases were reported globally, resulting in approximately 1.25 million deaths. Indonesia continues to be among the countries with the highest TB burden worldwide, making early detection, accurate diagnosis, and timely treatment essential priorities for improving healthcare services. This situation underscores the urgent need for innovative technological solutions to support TB control efforts. In this context, artificial intelligence (AI)-based technologies, particularly expert systems, offer significant potential to assist healthcare professionals by providing faster, more consistent, and evidence-based preliminary diagnoses, thereby enhancing clinical decision-making and supporting early intervention [10].

Table 1. State of the art

Study	Metode	Research result	weakness
Yanti & Firdaus, in 2022 [11]	Naïve bayes	To develop an expert system for tuberculosis (TB) diagnosis using the Naïve Bayes algorithm and evaluate its diagnostic performance based on the testing dataset	The dataset used in this study consisted of only 22 patient records, which limits the model's ability to generalize to a broader populatio
Orjuela-Cañón et al. in 2022 [12]	Machine learning	Demonstrates that machine learning can enhance tuberculosis (TB) diagnostic support through the analysis of clinical data	The study focused on developing a machine learning model and did not extend to the development of an expert system that could be directly utilized by healthcare professionals
Mauliana et al. 2025 [13]	Naïve bayes	To develop an Android-based tuberculosis (TB) diagnosis application using the Naïve Bayes algorithm to support early detection among the community.	The implementation focused on a mobile application and did not comprehensively evaluate the system using performance metrics such as

Study	Metode	Research result	weakness
			accuracy, precision, recall, and F1-score.
Vats et al. in 2024 [14]	Deep Learning (Incremental Learning)	The proposed model achieved an accuracy of approximately 93.2% in detecting tuberculosis (TB) from chest X-ray images	The study focused on the analysis of medical images, requiring radiological examinations, and did not support symptom-based diagnosis through an expert system
Vats et al. in 2023 [15]	Systematic Review Machine Learning	The study demonstrated that Support Vector Machine (SVM) and Random Forest achieved high performance in biomarker-based tuberculosis (TB) diagnosis	The study focused on algorithmic analysis and did not develop a web-based expert system to support clinical decision-making.

Based on the review of previous studies, numerous researchers have applied various machine learning algorithms and expert systems for tuberculosis (TB) diagnosis. However, these studies generally exhibit several limitations, including the use of relatively small datasets, a primary focus on improving classification accuracy without integrating expert systems as decision support tools, and the lack of diagnostic recommendations that can be directly utilized by healthcare professionals. Furthermore, most studies have evaluated model performance solely in terms of accuracy, without conducting a more comprehensive assessment using standard performance metrics such as accuracy, precision, recall, and F1-score. These research gaps highlight the need to develop an early TB diagnosis system that not only achieves high classification performance but also provides practical decision support to assist healthcare professionals in clinical practice.

Based on the identified research gaps, this study is guided by the following research questions:RQ1. How can a web-based Smart Expert System using the Naïve Bayes algorithm be developed to support the early diagnosis of tuberculosis (TB)?RQ2. How does the proposed system perform in diagnosing tuberculosis (TB) when evaluated using the standard performance metrics of Accuracy, Precision, Recall, and F1-score?

This study makes three primary contributions. First, it develops a web-based Smart Expert System that integrates the Naïve Bayes algorithm as the classification method for the early diagnosis of tuberculosis (TB). Second, the proposed system not only generates diagnostic predictions but also provides recommendation-based decision support to assist healthcare professionals in the diagnostic process. Third, the study utilizes a dataset comprising 200 tuberculosis patient records and performs a comprehensive evaluation of the proposed system using standard performance metrics, including Accuracy, Precision, Recall, and F1-score. These contributions result in a more practical and clinically applicable diagnostic system compared with those reported in previous studies.

This study is, about creating a web-based system that helps doctors diagnose tuberculosis early.The system uses a method called Bayes.It looks at the symptoms a patient has to figure out if they might have tuberculosis.Users go to the system select their symptoms and the system calculates the chances of them having tuberculosis [16]. The system then gives a diagnosis.It also provides information on tuberculosis. Suggests what to do next.The goal is to make people more aware of how important it's to get medical help early. The system aims to help increase awareness of tuberculosis.It does this by giving health information and recommendations for treatment.The system uses patient symptom data to determine the likelihood of tuberculosis infection.The Naive Bayes method is used to perform calculations.The system generates a diagnostic result based on these calculations.

The Smart Expert System we are working on is supposed to help people and doctors find tuberculosis early. This system will make it easier for them to do their job [17]. The Smart Expert System will also make healthcare better by using intelligence to diagnose diseases. This study can help people who want to work on similar systems that use machine learning to help doctors. The Smart Expert System can get even better if we use complex medical information and find ways to make the system work more accurately. The Smart Expert System is what we are focusing on to improve healthcare services [18].

2. Literature Review

Different expert system techniquies have been created to help diagnose tuberculosis. Theses methodes include. Certainty factor. Dempster- shafer, fuzzy logic, decision tree, random forest, logistic regression and naïve bayes. Each method has its own features, advantages, and drawbacks, which depend on how complicated the medical data is, how much computing power is needed, and how well it can handle uncertainty in making a diagnosis.

Each method offers distinct advantages and limitations depending on the characteristics of the data and the complexity of the problem. Certainty Factor is simple to implement and effectively incorporates expert knowledge; however, it relies on subjective confidence values assigned by experts. Dempster–Shafer theory is capable of handling uncertainty by combining multiple pieces of evidence, although its evidence combination process becomes increasingly complex as the number of hypotheses grows. Fuzzy Logic is effective in representing vague and imprecise symptoms, but it requires the development of numerous fuzzy rules, making the system more difficult to design and maintain.

Decision Tree provides an interpretable classification model but is susceptible to overfitting, particularly when trained on limited or noisy datasets. Random Forest generally achieves high classification accuracy by combining multiple decision trees; however, it demands greater computational resources. Logistic Regression offers a probabilistic approach for binary classification but assumes a linear relationship between predictor variables and the outcome. In contrast, the Naïve Bayes classifier is computationally efficient, simple to implement, and well suited for medical diagnosis because it estimates class probabilities effectively even with relatively small datasets. Although it assumes conditional independence among features, previous studies have demonstrated that Naïve Bayes can still achieve competitive performance in many medical classification applications, making it an appropriate method for the proposed tuberculosis expert system.

The study conducted by [19] developed an expert system for diagnosing pulmonary tuberculosis using a combination of the Certainty Factor (CF) and Dempster–Shafer (DS) methods. The system utilized tuberculosis patient data as the basis for the diagnostic process and achieved an accuracy of 88.2%. The findings demonstrated that the integration of these two methods effectively supported disease identification. However, the proposed system did not provide a web-based decision support mechanism, limiting its accessibility and ease of use.

The study by [20] developed an expert system application for tuberculosis diagnosis using the Certainty Factor (CF) method based on patient medical records collected from Palmerah Community Health Center. The proposed system achieved an accuracy of 85%, demonstrating its effectiveness in supporting the diagnostic process. However, the Certainty Factor method relies heavily on expert-assigned confidence values, making the diagnostic results susceptible to subjective judgment.

The study by [21]proposed an expert system for the early diagnosis of pulmonary tuberculosis by integrating the Forward Chaining and Certainty Factor (CF) methods. The proposed approach achieved a diagnostic confidence level of up to 98%, demonstrating its effectiveness in identifying the disease based on predefined rules. However, the system relied on a rule-based approach, limiting its flexibility to learn new patterns from data and adapt to more complex diagnostic scenarios.

The study by [22] implemented the Naïve Bayes algorithm with Laplace Correction in an expert system for the diagnosis of pulmonary tuberculosis using tuberculosis patient data. The application of Laplace Correction was intended to address the zero-probability problem during the classification process, thereby improving the stability of the diagnostic results. However, the study did not provide a quantitative evaluation of the classification performance, making it difficult to objectively compare the proposed model with other existing methods.

The study by [23] analyzed the application of the Naïve Bayes method for diagnosing tuberculosis based on clinical symptom data. The method was selected due to its computational simplicity and its ability to generate diagnostic results using a probabilistic approach. However, the system was implemented as a desktop-based application, which limited user mobility and reduced the accessibility of the system.

The study by [24] developed an expert system for diagnosing tuberculosis in adolescents using the Dempster–Shafer method. This approach effectively combined multiple pieces of symptom evidence to address uncertainty during the diagnostic process. However, the proposed method did not incorporate a probabilistic learning mechanism, limiting the model's ability to learn from historical data and improve its diagnostic performance over time.

The study by [25] implemented an expert system for the early diagnosis of pulmonary tuberculosis using the Fuzzy Tsukamoto method based on patient symptom data. The fuzzy-based approach was able to accommodate the inherent uncertainty and ambiguity of clinical symptoms, resulting in a more flexible inference process. However, the method required the development of a large number of fuzzy rules, making the system more complex to develop and maintain.

The study by [26] developed SitusParu, an expert system for the early detection of pulmonary tuberculosis using the Certainty Factor (CF) method. The proposed system achieved an accuracy of 81.25%, indicating that the method is effective in supporting the early diagnosis of tuberculosis. However, the system evaluation was limited to accuracy alone and did not include other important performance metrics, such as precision, recall, and F1-score. Consequently, the overall performance of the proposed model was not comprehensively assessed.

Based on the findings of previous studies, this research develops a web-based expert system using the Naïve Bayes algorithm with a dataset of 200 tuberculosis patient records. Unlike earlier studies, this research provides a more comprehensive evaluation of the proposed model by employing accuracy, precision, recall, and F1-score as performance metrics. This comprehensive evaluation enables a more objective assessment of the system's effectiveness in supporting the early diagnosis of tuberculosis.

3. Method

This chapter describes the methodology employed to develop and evaluate a web-based Smart Expert System for the early diagnosis of pulmonary tuberculosis using the Bernoulli Naïve Bayes algorithm. The study adopted a quantitative experimental approach, in which the proposed model was designed, implemented, and evaluated using real clinical data. The methodology was organized into several systematic stages to ensure the reliability and validity of the developed system.

The research began with the design of the overall research workflow, followed by the definition of the research design and the collection of pulmonary tuberculosis patient data. The collected dataset then underwent a series of preprocessing procedures, including data cleaning, transformation, feature encoding, and feature selection, to improve data quality and prepare the data for the classification process. Since the Bernoulli Naïve Bayes algorithm requires binary input features, all symptom variables were transformed into binary values before model training.

To assess the robustness and generalization capability of the proposed classification model, K-Fold Cross-Validation was employed during the evaluation stage. The Bernoulli Naïve Bayes algorithm was then trained using the processed dataset by calculating prior and conditional probabilities for each disease class, followed by posterior probability estimation during the prediction process. Laplace Smoothing was applied to overcome zero-probability problems that may occur during model training. The classification performance was evaluated using a confusion matrix and standard performance metrics, including Accuracy, Precision, Recall, and F1-score. Finally, the trained model was integrated into a web-based Smart Expert System, enabling users to input symptoms and receive diagnostic predictions along with initial recommendations for pulmonary tuberculosis.

3.1. Research workflow

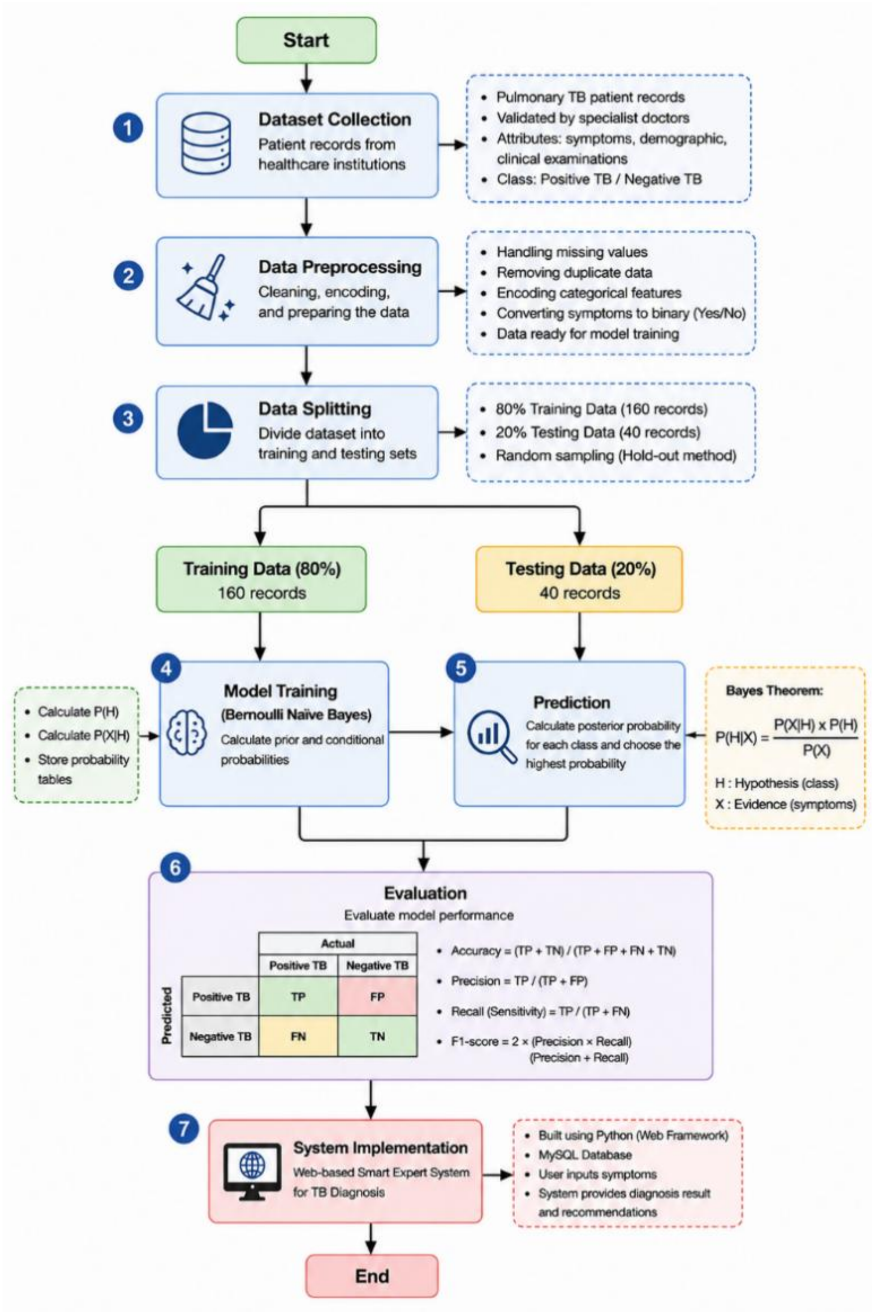

Figure 1. reseacrh workflow

3.2. Research design

This study used aquantitative experimental methode to create a web – based smart expert system that can help detect pulmonary tuberculosis early. Using the Bernoulli naïve bayes algorithm. The research method involved different steps like gathering data, preparing the data, building the model, checking the model's accuracy with K-Fold Cross-Validation, assessing how well the model performs, creating a web-based system, and thoroughly testing the system.

3.3. Dataset collection

This section provides a detailed description of the data source used in this study.The dataset consisted of 200 pulmonary tuberculosis patient records obtained from the medical records of a healthcare facility (adjust according to the actual data source). Each record was validated by a pulmonologist to ensure that the diagnostic labels used in this study were clinically reliable and valid.

The diagnosis for each patient was established based on the clinical assessment of a pulmonologist and was used as the ground truth. Therefore, all data included in this study were assigned medically validated diagnostic labels.

3.4. Data preprocessing

Before the model training process, the dataset underwent several preprocessing steps to improve data quality and ensure the reliability of the classification model. The preprocessing stage included the following procedures: Data Cleaning, Removing duplicate records, Handling missing values, Ensuring the consistency of data attributes.

3.4.1. Data transformation

Since the Bernoulli Naïve Bayes algorithm requires binary input data, all symptom attributes were converted into binary values as follows:Yes → 1 No → 0.

3.4.2. Feature encoding

All categorical attributes were transformed into numerical representations using Binary Encoding.

3.4.3. Feature selection

All symptom attributes were used as predictor variables because each of them has a clinically relevant association with the diagnosis of pulmonary tuberculosis.

3.5. K-fold cross validation

To obtain a more robust and reliable evaluation of the model, this study employed K-Fold Cross-Validation. The dataset was divided into k subsets (e.g.**,** k = 10). During each iteration: One fold was used as the testing set, The remaining nine folds were used as the training set.This process was repeated ten times, ensuring that each data instance served as the testing set exactly once.The model performance was evaluated by calculating the average values of Accuracy**,** Precision, Recall, and F1-score across all folds.

3.6. Bernoulli naïve bayes classification

This study employed the Bernoulli Naïve Bayes algorithm because all input attributes were represented as binary variables (Yes/No).Bernoulli Naïve Bayes estimates the probability of a class based on the Bernoulli distribution, assuming that each attribute is conditionally independent of the others.The Bayes theorem used in this study is as follows:

$$P(HIX) = \frac{P(HIX).P(H)}{P(X)} \tag{1}$$

Description:

1. P(H|X) = the posterior probability of the hypothesis given the observed evidence.
2. P(X|H) = the likelihood, representing the probability of observing the symptoms given the disease.
3. P(H) = the prior probability of the disease.
4. P(X) = the probability of the observed evidence

3.6.1. Independence assumption

Bernoulli Naïve Bayes assumes that each symptom is conditionally independent of the other symptoms given the disease class. Although this assumption is relatively simple, numerous studies have demonstrated that Naïve Bayes can still achieve strong classification performance in disease diagnosis due to its computational efficiency and its ability to effectively handle high-dimensional medical datasets.

3.6.2. Laplace smoothing

To prevent zero probabilities when a symptom does not appear in a particular class during the training process, this study applied Laplace Smoothing. The Laplace Smoothing formula is expressed as follows:

$$P(XiIC) = \frac{Nic+1}{Nc+2} \quad (2)$$

3.7. Training process

During the training phase, the following steps were performed:

1. The prior probability of each class was calculated, for example: P(TB)
2. The conditional probability of each symptom given each class was computed, for example: P(Cough/TB), P(fever/TB) and so forth for all symptom attributes.
3. All calculated probabilities were stored in the Bernoulli Naïve Bayes model for use during the classification process.

3.8. Prediction process

During the testing phase, users selected the symptoms they were experiencing.The model then computed the posterior probability for each class using all probabilities obtained during the training phase.The class with the highest posterior probability was selected as the final diagnostic result. Example P (TB) = 0.52, P (Cought / TB)=0.84, P (Fever /TB) = 0.73 , Eat All. The posterior probabilities were then calculated until the final probability for each class was obtained.

3.9. Perormance evaluation the performance of the model was evaluated using a confusion matrix.

Table 2. performance evaluation

	Positif	Negative
Positive	TP	FP
Negative	FN	TN

The calculate Accuracy, Precision, Recall, F1-score using the standard equation.

3.10. System implementation

The Smart Expert System was developed as a web-based application using Python as the programming language and MySQL as the database management system. Users can input the symptoms they are experiencing, after which the system calculates the probability of pulmonary tuberculosis using the Bernoulli Naïve Bayes algorithm. The system then provides the diagnostic result along with initial recommendations to support early clinical decision-making.

4. Results and Discussion

This section presents the experimental results of the proposed web-based Smart Expert System for pulmonary tuberculosis diagnosis. It describes the dataset, preprocessing, experimental setup, model performance, and discussion of the findings.

Table 3. Description of dataset used in the experiment

No	Feature / Variable	Description	Data Type	Value
1	Prolonged Cough	Cough lasting more than two weeks	Binary	0 = No, 1 = Yes
2	Fever	Presence of persistent fever	Binary	0 = No, 1 = Yes
3	Night Sweats	Excessive sweating during sleep	Binary	0 = No, 1 = Yes
4	Weight Loss	Significant unexplained weight loss	Binary	0 = No, 1 = Yes
5	Chest Pain	Pain in the chest area	Binary	0 = No, 1 = Yes
6	Shortness of Breath	Difficulty breathing	Binary	0 = No, 1 = Yes
7	Loss of Appetite	Decreased appetite	Binary	0 = No, 1 = Yes
8	Hemoptysis	Coughing up blood	Binary	0 = No, 1 = Yes
9	Target Class	Tuberculosis diagnosis result	Binary	0 = Negative, 1 = Positive

Before the classification process using the Bernoulli Naïve Bayes algorithm was performed, the dataset underwent several preprocessing stages to ensure data quality and improve the classification model's performance. The dataset consisted of 200 patient records, containing clinical symptom information and tuberculosis diagnosis outcomes as the target class

The first stage was data cleaning, which involved checking data consistency, removing duplicate records, and ensuring that no missing values were present in any attribute. In addition, the dataset was examined for inconsistencies in data formatting to ensure that all records followed a uniform structure and were suitable for the classification process.

The next stage was data encoding. All symptom attributes, originally represented as categorical values (*Yes* and *No*), were converted into binary numerical values, where 1 indicated the presence of a symptom (*Yes*) and 0 indicated its absence (*No*). Similarly, the target variable was encoded as 1 for patients diagnosed with tuberculosis and 0 for patients without a tuberculosis diagnosis. This encoding process was necessary because the Bernoulli Naïve Bayes algorithm is specifically designed to operate on binary-valued features.

After the encoding process, additional preprocessing was performed by validating the data types to ensure that all attributes were stored as binary numerical values. The dataset was then divided into 160 records (80%) for training and 40 records (20%) for testing. The training set was used to build the classification model, while the testing set was used to evaluate the model's ability to classify previously unseen data. These preprocessing steps produced a clean, consistent, and well-structured dataset that met the requirements of the Bernoulli Naïve Bayes algorithm, thereby enabling the classification process to be carried out effectively and reliably.

4.1. Experimental setup

The proposed Smart Expert System was evaluated using the Bernoulli Naïve Bayes classification algorithm implemented in Python. The experimental evaluation employed 10-Fold Cross Validation to obtain a robust estimation of the model's performance. The dataset consisted of 200 pulmonary tuberculosis patient records that had undergone preprocessing, including data cleaning, binary encoding, and feature transformation. Each fold used 90% of the data for training and the remaining 10% for testing, ensuring that every record was used once as testing data and nine times as training data. The final performance values were obtained by averaging the evaluation results across all folds.The system performance was evaluated using Accuracy, Precision, Recall, F1-score, and Confusion Matrix, which are commonly used metrics for binary classification in medical diagnosis.

4.2. Model performance evaluation

The classification performance of the proposed Bernoulli Naïve Bayes model is presented using the Confusion Matrix shown in table

4.2.1. Accuracy

Accuracy measures the proportion of correctly classified instances among the total number of testing samples.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

The proposed Bernoulli Naïve Bayes model achieved an Accuracy of 87.00%, indicating that the classifier was able to correctly identify tuberculosis and non-tuberculosis cases with a high level of reliability.

4.2.2. Precision

Precision measures the proportion of correctly predicted tuberculosis cases among all predicted positive cases.

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

The obtained Precision value of 85.00% indicates that most patients predicted as tuberculosis cases were correctly classified by the proposed model.

4.2.3. Recall

Recall evaluates the ability of the classifier to identify actual tuberculosis patients

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

The Recall value reached 88.00%, demonstrating that the proposed model successfully detected the majority of tuberculosis patients while minimizing false-negative diagnoses.

4.2.4. F1- Score

The F1-score represents the harmonic mean of Precision and Recall

$$F1 = \frac{2\ X\ Precision\ x\ Recall}{Precision + Recall} \tag{6}$$

The proposed model achieved an F1-score of 86.00%, indicating a balanced classification performance between Precision and Recall.

4.2.5. Summary of model performance

Table 4. performance evaluation bernoulli naïve bayes

Evaluation matrix	Result
Accuracy	87.00%
Precision	85.00%
Recall	88.00%
F1- Score	86.00%

4.3. Discussion

This section discusses the experimental findings obtained from the proposed web-based smart expert system for pulmonary tuberculosis diagnosis. The discussion focuses on the classification performance of the Bernoulli Naïve Bayes algorithm, compares the obtained results with previous studies, identifies the limitations of the proposed approach, and highlights its practical implications for primary healthcare services. These discussions provide a comprehensive interpretation of the results and demonstrate the potential contribution of the proposed system to early tuberculosis screening.

4.3.1. Performance analysis of bernoulli naïve bayes

The experimental results demonstrate that the Bernoulli Naïve Bayes classifier provides satisfactory performance for pulmonary tuberculosis diagnosis. The model achieved consistent classification performance across all validation folds, indicating good generalization capability.

The high classification performance can be attributed to the nature of the clinical dataset, in which most predictor variables represent binary symptoms (e.g., cough, fever, weight loss, night sweats, chest pain, and dyspnea). Since Bernoulli Naïve Bayes is specifically designed for binary feature representation, it effectively models the probability distribution of tuberculosis symptoms.

Furthermore, the probabilistic framework enables the classifier to estimate the likelihood of tuberculosis even when only partial symptom information is available, making it particularly suitable for early disease screening.

4.3.2. Comparison with previous studies

Table 5. Comparison with previous studies

Study	Method	Dataset	Accuracy	Remarks
[20]	Certainty Factor	Medical record (TB patients)	85.00%	Depends on expert confidence values
[19]	Dempster–Shafer	TB patient data	88.20%	Good uncertainty handling but computationally complex
[25]	Fuzzy Tsukamoto	Symptom dataset	89.10%	Requires many fuzzy rules
This Study	Bernoulli Naïve Bayes	200 TB patient records	87.00%	Web-based system with Accuracy, Precision, Recall, and F1-score evaluation

Unlike rule-based methods such as Certainty Factor and Dempster-Shafer, the proposed Bernoulli Naïve Bayes classifier learns probability distributions directly from patient data. Consequently, the model is capable of adapting to different data distributions while maintaining computational efficiency.

4.3.3. Limitations

Although the proposed model demonstrates promising performance, several limitations should be acknowledged.First, the dataset consists of only 200 patient records collected from a limited healthcare setting, which may affect the generalizability of the model. Second, the diagnosis relies solely on clinical symptoms without incorporating radiographic images or laboratory biomarkers. Third, the Naïve Bayes algorithm assumes conditional independence among symptoms, an assumption that may not fully reflect complex clinical relationships.Future studies should involve larger multicenter datasets and investigate more advanced machine learning techniques to further improve diagnostic performance.

4.3.4. Implications for primary healthcare

The proposed Smart Expert System has practical implications for primary healthcare services, particularly in community health centers with limited access to pulmonologists. By providing rapid preliminary diagnosis based on patient symptoms, the system can assist healthcare workers in identifying suspected tuberculosis cases at an earlier stage. Early screening may contribute to reducing diagnostic delays, improving treatment initiation, and supporting national tuberculosis control programs.

5. Conclusion

This research effectively created a web-based Smart Expert System for the early detection of pulmonary tuberculosis utilizing the Bernoulli Naïve Bayes classifier. The suggested system can aid healthcare providers in making initial tuberculosis diagnoses based on the clinical symptoms exhibited by patients. The use of the Bernoulli Naïve Bayes algorithm allows the system to efficiently determine the probability of tuberculosis, while employing 10-Fold Cross Validation offers a dependable assessment of the model's classification effectiveness. Moreover, the system has the capability to enhance public knowledge about tuberculosis by delivering early diagnostic details and suggestions for subsequent medical consultation.

This research's scientific contribution is the incorporation of the Bernoulli Naïve Bayes algorithm into an online expert system for diagnosing pulmonary tuberculosis. In contrast to traditional rule-based expert systems, the suggested method utilizes a probabilistic machine learning model that can learn from patient information and facilitate objective clinical decision-making. Bernoulli Naïve Bayes is especially appropriate since the clinical symptoms are depicted as binary variables, enabling effective and precise probability assessment for early tuberculosis detection. Even with these encouraging outcomes, numerous limitations need to be recognized. Initially,

The suggested model was created utilizing a dataset containing 200 patient records, potentially restricting the generalizability of the classification model. Secondly, the diagnosis is based solely on clinical symptom information, excluding radiological images, laboratory test results, or additional supporting clinical data. Additionally, while the dataset was created for model development, wider clinical validation with various healthcare facilities and expert physicians is still necessary to assess the practical applicability of the suggested system in actual healthcare settings.

Future studies should aim to enhance the dataset by incorporating multicenter clinical data to better model generalization and robustness. Extra clinical characteristics, including chest X-ray images, lab test results, and electronic health records, could.

Acknowledgements

The authors want to thank everyone who helped with this research.They are grateful to all the people and organizations that gave them advice and support.The institution where the study was done provided a place to work and the tools needed and the authors appreciate that.The authors also want to thank the validators and respondents who helped collect data.They are thankful, to all the parties who helped develop and improve the system.The authors hope that their research will help add to what we know about expert systems and artificial intelligence in healthcare.

Credit Authorship Contribution Statement

Ulumuddin: Conceptualization, Methodology, Software, Data curation, Formal analysis, Investigation, Validation, Visualization, Writing – original draft, Project administration. **Siti Harlina:** Methodology, Investigation, Data curation, Validation, Writing – review & editing. **Warjiono:** Validation, Formal analysis, Resources, Writing – review & editing. **Amin Nur Rais:** Investigation, Validation, Supervision, Writing – review & editing. **Arham Arifin:** Software, Validation, Visualization, Writing – review & editing. **Abdul Ibrahim:** Supervision, Validation, Project administration, Writing – review & editing. **James Adam Seo:** Supervision, Validation, Writing – review & editing.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data will be made available on request.

Use of Artificial Intelligence

The authors used artificial intelligence (AI)-assisted tools solely to improve the clarity, grammar, and readability of the manuscript. All research design, data collection, data analysis, interpretation of results, and scientific conclusions were performed entirely by the authors. The authors reviewed and approved all AI-assisted content and take full responsibility for the accuracy and integrity of the final manuscript.

References

[1] P. Teknologi, “Perkembangan Teknologi Informasi di Indonesia ∗),” 2002.
[2] T. M. Mitchell, “Really Work ?,” vol. 18, no. 3, pp. 11–20, 1997.
[3] D. Maulina, A. M. Wulanningsih, F. I. Komputer, and U. A. Yogyakarta, “METODE CERTAINTY FACTOR DALAM PENERAPAN SISTEM PAKAR DIAGNOSA PENYAKIT ANAK Abstraksi Pendahuluan Tinjauan Pustaka Metode Penelitian,” vol. 1, no. 2, pp. 23–32, 2020.
[4] P. R. A. S. Kucing, I. P. Sari, L. Trisnawati, and N. Silviana, “Penerapan sistem pakar berbasis,” vol. 8, no. 1, pp. 59–68, 2023.

[5] D. Menggunakan, M. Forward, and C. Berbasis, “Forward chaining,” vol. 3, no. 1, pp. 35–46, 2018.
[6] A. Pura and D. Alamsyah, “Sistem Pakar Diagnosa Penyakit Ginjal,” vol. 6, pp. 53–74, 2019, doi: 10.36079/lamintang.ijai-0601.32.
[7] R. A. Dachi, L. Hakim, and T. Wandra, “Sosialisasi Tentang Penyakit Tuberkulosis Paru Di Rumah Sakit Putri Hijau Medan,” *J. Abdimas Mutiara*, vol. 3, pp. 367–374, 2022.
[8] J. S. Sipayung, W. Hidayat, and E. M. Silitonga, “Faktor Risiko yang Memengaruhi Kejadian Tuberkulosis (TB) Paru di Wilayah Kerja Puskesmas Perbaungan Risk Faktors Affecting the Incident of Pulmonary Tuberculosis (TB) in the Working Area of Perbaungan Public Health Center,” vol. 15, no. 2, 2023.
[9] “The Indonesian Journal of Health Promotion MPPKI Media Publikasi Promosi Kesehatan Indonesia Analisis Implementasi Strategi Promosi Kesehatan dalam Pencegahan Penyakit Tuberkulosis (TB) (Studi Kasus di Wilayah Kerja Puskesmas Kalumata Kota Ternate),” 2022, doi: 10.31934/mppki.v2i3.
[10] who, “Tuberculosis resurges as top infectious disease killer,” web who. Accessed: Jul. 16, 2026. [Online]. Available: https://www.who.int/news/item/29-10-2024-tuberculosis-resurges-as-top-infectious-disease-killer?.
[11] T. Firdaus and F. Yanti, “Implementasi Metode Naïve Bayes Pada Sistem Pakar Diagnosa Penyakit Gigi Berbasis Web,” *J. Din. Inform.*, vol. 14, no. 2, pp. 87–98, 2022, doi: 10.35315/informatika.v14i2.9201.
[12] A. D. Orjuela-Cañón, A. L. Jutinico, C. Awad, E. Vergara, and A. Palencia, “Machine learning in the loop for tuberculosis diagnosis support,” *Front. Public Heal.*, vol. 10, no. 2, 2022, doi: 10.3389/fpubh.2022.876949.
[13] P. Mauliana, W. Wiguna, and D. Anisa, “Aplikasi Diagnosis Penyakit Tuberkulosis Menggunakan Naïve Bayes di Dinas Kesehatan Kota Bandung,” *J. Infortech*, vol. 7, no. 2, pp. 161–170, 2025, doi: 10.31294/infortech.v7i2.11507.
[14] S. Vats *et al.*, “Incremental learning-based cascaded model for detection and localization of tuberculosis from chest x-ray images,” *Expert Syst. Appl.*, vol. 238, no. PD, p. 122129, 2024, doi: 10.1016/j.eswa.2023.122129.
[15] S. A. P. Vimala Balakrishnan, Yousra Kherabi, Ghayathri Ramanathan and C. K. Tiong, “Machine learning approaches in diagnosing tuberculosis through biomarkers - A systematic review,” vol. 179, no. tuberculosis, pp. 16–25, 2023, doi: https://doi.org/10.1016/j.pbiomolbio.2023.03.001.
[16] C. Lange *et al.*, “STATE OF THE ART Tuberculosis in migrants in low-incidence countries : epidemiology and intervention entry points,” vol. 21, no. November 2016, pp. 624–636, 2017.
[17] D. Fitriati and I. Gibran, “Sistem Pakar Diagnosis Penyakit Meningitis MFitriati, D., & Gibran, I. (2021). Sistem Pakar Diagnosis Penyakit Meningitis Menggunakan Metode Forward Chaining. Jurnal UMJ, 12(1), 46–50. https://jurnal.umj.ac.id/index.php/just-it/index menggunakan Metode Forw,” *J. UMJ*, vol. 12, no. 1, pp. 46–50, 2021.
[18] A. Maroosi and B. Amiri, “A new clustering algorithm based on hybrid global optimizationbased on a dynamical systems approach algorithm,” *Expert Syst. Appl.*, vol. 37, no. 8, pp. 5645–5652, 2010, doi: 10.1016/j.eswa.2010.02.047.
[19] A. A. Prameswaty, M. H. P. Swari, and W. S. J. Saputra, “Perancangan Sistem Pakar Diagnosis Penyakit Tbc Paru Dengan Metode Certainty Factor Dan Dempster Shafer,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 5, pp. 8658–8663, 2024.
[20] N. Siti Fatonah, “Membangun Aplikasi Sistem Pakar Menggunakan Metode Certainty Factory Dalam Mendiagnosa Tuberculosis Studi Kasus Di Puskesmas Palmerah,” *Integr. Perspect. Soc. Sci. J.*, vol. 2, no. 2, p. 3030, 2025.
[21] I. G. Ngurah, A. Praja, A. Wahyudi, and O. Gama, “PERANCANGAN SISTEM PAKAR DIAGNOSIS DINI TUBERKULOSIS PARU MENGGUNAKAN METODE FORWARD CHAINING DAN CERTAINTY FACTOR,” vol. 10, no. 3, pp. 4552–4559, 2026.
[22] Yanti Apriyani, I. D. I. Iskandar, Mira Kusmira, Melisa Winda Pertiwi, Imam Amirulloh, and Taufik Wibisono, “Implementasi Sistem Pakar dengan Algortima Naïve Bayes dengan Laplace Correction untuk Diagnosis Tuberkulosis Paru,” *Inf. (Jurnal Inform. dan Sist. Informasi)*, vol. 13, no. 1, pp. 24–46, 2021, doi: 10.37424/informasi.v13i1.72.
[23] M. R. Syahwana and R. M. Simanjorang, “Analisa Sistem Pakar Metode Bayes Dalam Mendiagnosa Penyakit Tubercolosis,” *J. Sist. Informasi, Tek. Inform. dan Teknol. Pendidik.*, vol. 1, no. 2, pp. 57–66, 2022, doi: 10.55338/justikpen.v1i2.7.

[24] R. Simamora, A. Alhafiz, and S. Julianita, "Sistem Pakar Mendiagnosis Tuberkulosis Pada Remaja Menggunakan Metode Dempster Shafer," *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 3, no. 5, pp. 713–723, 2024, doi: 10.53513/jursi.v3i5.8276.
[25] D. Abdullah, M. Rahmi, and Z. Yunizar, "Implementasi Sistem Pakar Diagnosa Awal Penyakit Tuberculosis Paru Menggunakan Fuzzy Tsukamoto," *VARIASI Maj. Ilm. Univ. Almuslim*, vol. 13, no. 3, pp. 153–157, 2022, doi: 10.51179/vrs.v13i3.860.
[26] R. Surya and D. Gunawan, "Situsparu: Sistem Pakar Untuk Deteksi Penyakit Tuberkulosis Paru," *Ultim. J. Tek. Inform.*, vol. 10, no. 1, pp. 41–47, 2018, doi: 10.31937/ti.v10i1.781.