



Tuberculosis–Diagnostic Expert System: An architecture for translating patients information from the web for use in tuberculosis diagnosis

Health Informatics Journal
2014, Vol. 20(4) 275–287
© The Author(s) 2013
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1460458213493197
jhi.sagepub.com


**Victor C Osamor, Ambrose A Azeta
and Oluseyi O Ajulo**

Department of Computer and Information Sciences, Covenant University, Nigeria

Abstract

Background: Over 1.5–2 million tuberculosis deaths occur annually. Medical professionals are faced with a lot of challenges in delivering good health-care with unassisted automation in hospitals where there are several patients who need the doctor’s attention.

Objective: To automate the pre-laboratory screening process against tuberculosis infection to aid diagnosis and make it fast and accessible to the public via the Internet. The expert system we have built is designed to also take care of people who do not have access to medical experts, but would want to check their medical status.

Methods: A rule-based approach has been used, and unified modeling language and the client–server architecture technique were applied to model the system and to develop it as a web-based expert system for tuberculosis diagnosis. Algorithmic rules in the Tuberculosis–Diagnostic Expert System necessitate decision coverage where tuberculosis is either suspected or not suspected. The architecture consists of a rule base, knowledge base, and patient database. These units interact with the inference engine, which receives patient’ data through the Internet via a user interface.

Results: We present the architecture of the Tuberculosis–Diagnostic Expert System and its implementation. We evaluated it for usability to determine the level of effectiveness, efficiency and user satisfaction. The result of the usability evaluation reveals that the system has a usability of 4.08 out of a scale of 5. This is an indication of a more-than-average system performance.

Conclusion: Several existing expert systems have been developed for the purpose of supporting different medical diagnoses, but none is designed to translate tuberculosis patients’ symptomatic data for online pre-laboratory screening. Our Tuberculosis–Diagnostic Expert System is an effective solution for the implementation of the needed web-based expert system diagnosis.

Corresponding author:

Victor C Osamor, Bioinformatics Unit, Department of Computer and Information Sciences, College of Science and Technology, Covenant University, P.M.B 1023, Ota, Ogun State, Nigeria.
Email: victor.osamor@covenantuniversity.edu.ng, vcosamor@yahoo.com

Keywords

diagnosis, expert system, pre-laboratory screening, rule based, tuberculosis

Introduction

Mycobacterium is the microbe that is the causative agent of tuberculosis (TB) and infects humans and other animals. This pathogen consists of seven species that cause the disease, and they include the following: *Mycobacterium tuberculosis*, *Mycobacterium bovis*, *Mycobacterium canetti*, *Mycobacterium africanum*, *Mycobacterium pinnipedi*, *Mycobacterium microti*, and *Mycobacterium caprae*.¹ We define TB as an infectious disease of both animals and humans caused by tubercle bacillus bacteria, which is contracted by airborne or by droplet contact with exudates, and symptomatic with characteristic cough and bloody sputum. It is believed that TB affects over 2 billion people or one-third of the world's population with an estimate of 1.5–2 million deaths annually.²

We present an expert system designed to automate the pre-laboratory screening process against tuberculosis infection to aid diagnosis and make it fast and accessible to the public via the Internet.

Trends in automated medical diagnosis

An important breakthrough in the field of medical expert systems was realized in 1976 when Shortliffe published MYCIN used to assist physicians to administer blood infection treatment from diagnosis of infectious blood diseases to subsequent recommendation of antibiotics, with the dosage prescription based on the patient's body weight.³ Other software produced during this early period included Present Illness Program (PIP)—a system for renal disease by Pauker et al.⁴—and in 1978, Close Air Support Network (CASNET), an ophthalmology advisor that was implemented to recommend management for patients with glaucoma⁵ as shown in Table 1. ONCOCIN was developed in the early 1980s with the aim of assisting physicians in the management and treatment of cancer patients.⁶ INTERNIST-1 is a system for diagnosing complex general internal medicine cases and was created in 1982 by Miller and others.⁷ A hybrid medical expert system called XBONE was developed in 1997, and it uses a hybrid representation formalism to integrate rules and an Adaline artificial neural unit for bone disease diagnosis using patients' clinical, demographic, and scintigraphic (nuclear medicine image (NMI)) data.⁸ EpiMAN-TB developed by McKenzie et al.⁹ is a decision-support system using spatial information for the management of TB in cattle and deer, which works by prediction of possum TB hot spots using farm boundary, vegetation, and slope information.

In 2005, attention temporarily shifted from diagnosis to public enlightenment when González et al.¹⁰ presented ITS-TB, an e-learning Intelligent Tutoring System for TB by integrating multi-agent systems and the case-based reasoning (CBR) paradigm. Murty et al.¹¹ described the use of the ID3 algorithm in the development of an online rule-based expert system for the identification of disease-causing mosquito species. A Bayesian probabilistic model-based system called BayPAD was developed by Luciani et al.¹² to perform diagnosis of pulmonary embolism. Based on the symptoms and signs that a patient presents, a Bayesian clinical decision support system (CDSS) uses Bayesian inference engine to propose differential diagnoses with their corresponding degree of certainty. Recently, Ocampo et al.¹³ engineered the construction of diagnosis support tools for acute bacterial meningitis (ABM) using a CBR inference engine and also reported a comparative assessment of the quality of a CDSS resulting from the application of CBR to that of system developed using a Bayesian CDSS. TUBERDIAG¹⁴ was developed in 1999 as a fuzzy logic and

Table 1. Automated diagnosis systems. This shows selected and existing automated diagnosis system from 1976 to 2011.

S/N	Automated medical diagnosis	Function	Reference
1	MYCIN	Blood infection diagnosis and treatment	Shortliffe ⁶
2	PIP	Present Illness Program—a system for renal disease	Pauker et al. ⁷
3	CASNET	Ophthalmology advisor—glaucoma management	Weiss et al. ⁸
4	ONCOCIN	Cancer patient management	Shortliffe et al. ⁹
5	INTERNIST-I	Complex general internal medical cases	Miller et al. ¹⁰
6	XBONE	Bone disease diagnosis	Hatzilygeroudis et al. ¹¹
7	EpiMAN-TB	Management of tuberculosis in cattle and deer	McKenzie et al. ¹²
8	TUBERDIAG	Expert system (ES) for tuberculosis (TB) diagnosis	Nguyen et al. ¹⁷
9	ITS-TB	E-learning intelligent tutoring system on TB	González et al. ¹³
10	ES	Identification of disease-causing mosquito species	Murty et al. ¹⁴
11	BayPAD	Diagnosis of pulmonary embolisms	Luciani et al. ¹⁵
12	ABM CBDS	Acute bacterial meningitis (ABM) based diagnosis system	Ocampo et al. ¹⁶
13	FUZZY EXPERT	Decision support system for TB diagnosis	Djam and Kimbi ¹⁸

approximate reasoning system. In 2011, Djam and Kimbi¹⁵ developed a fuzzy expert system for TB diagnosis, which provides decision support platform to TB researchers, physicians and other health-care practitioners in tropical medicine.

Methodology

Basic concepts of rule-based model

The basic model of a rule-based expert system is composed of sets of rules that are regularly applied to the sets of facts in the database. While facts represent circumstances that describe a certain situation in the real world, rules consist of an “if” and a “then” portion or “if-then” statement which represents heuristics that define a set of actions to be carried out in a given situation or circumstance.^{16,17} The “If” portion is on the left-hand side (LHS) and is called premise or predicate, while the “then” portion is on the right-hand side (RHS) and is either a conclusion or action.

- *Action or set of actions* represent the operations of the rule that determine the response of the system.

- *Condition or set of conditions* is expressed as a (simple or complex) Boolean expression which represents the situation under which the actions of the rule are enabled, in case the specified events occur.¹⁹

Algorithmic rules in the TB–Diagnosis Expert System (TB-DES)

Where TB is suspected. The presence of cough with bloody sputum and swollen lymph is probably an indication of *Mycobacterium* infection, which is captured in the following algorithmic rules.

*If <symptom is cough AND
(NOT symptom is headache)
AND (symptom is bloody sputum)
AND symptom is swollen lymph/neck/joint>
then <Notify (Patient), “Tuberculosis is very likely, please go and see your doctor”>.*

Where TB is not suspected. When cough is not observed and sputum occurrence is none or minimal, and where it occurs, it is not bloody, then TB is not likely. This is depicted in the following algorithmic rules.

*If <symptom is cough AND
(NOT Symptom is bloody sputum)
AND (NOT symptom is swollen lymph/neck/joint)>
then <Notify (Patient), “Tuberculosis is not likely but ordinary cough”>.*

Design of TB-DES

To support the design of TB-DES, the unified modeling language (UML) was applied to model the system and facilitate easy implementation.¹⁸ Use case (Figure 1), and Sequence diagrams (Figure 2) were created to describe the system. The architecture of TB-DES shown in Figure 3 consists of a rule base, knowledge base, and patient database. Patient data is acquired over the Internet via a user interface.

Systems implementation

TB-DES was implemented using Java NetBeans Integrated Development Environment (IDE) version 6.0 hosted on the Windows Vista operating system. NetBeans was chosen because it is a robust IDE suitable for rapid and reliable development of enterprise Java web applications. Java was chosen because of its multiplatform and concurrency support.

The various modules as described in Figure 3 are integrated and implemented using Java Server Pages (JSP), Apache Tomcat web server, Apache Derby Version 2.0, and Java NetBeans IDE. JSP technology enabled us to put snippets of Servlet Code directly into a text-based document. A JSP page is a text-based document that contains wireless markup language (WML), extensible markup language (XML) and JSP elements, which determine how the page constructs dynamic content. Apache Tomcat is the Servlet Container used in the official reference implementation for the Java Servlet and JSP technologies. The Java Servlet and JSP specification are developed by Sun under the Java Community Process.

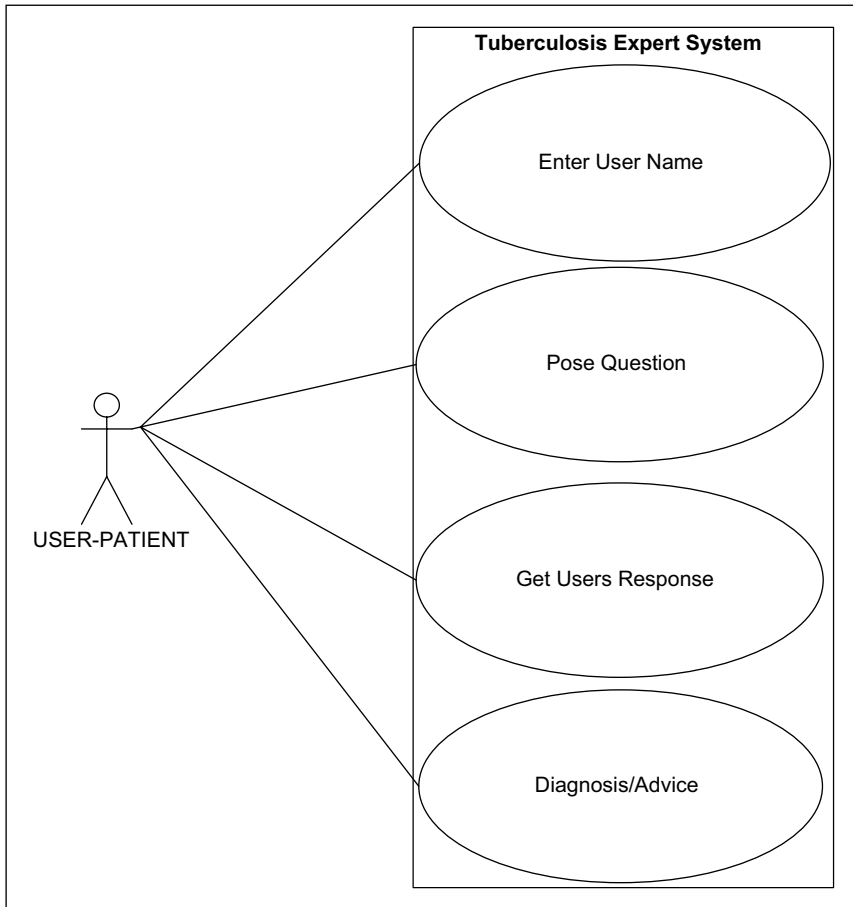


Figure 1. The use case diagrams show actors and their relationships to provide an overview of the functionality. An actor represents a role that a user plays when interacting with TB-DES, as indicated by the patient actor.

Apache Derby is an open source relational database implemented entirely in Java and available under the Apache license version 2.0. Derby provides an embedded JDBC driver that lets one embed Derby in any Java-based solution. Derby also supports the more familiar client/server mode with the Derby Network Client JDBC driver and Derby Network Server. In other words, JDBC is responsible for initiating and maintaining the connection between the interface client and the Derby relational database. Java NetBeans IDE provided the total programming environment where all development took place. The system was tested with varying symptomatic input. The system was able to navigate through various prompted questions relevant to the preclinical screening and finally terminated with a conclusion that defines the infectious or noninfectious status of TB. In some cases, it also includes additional advisory information to help the patient improve his or her health. The sample screen shots of the implemented system are shown in Figures 4 and 5.

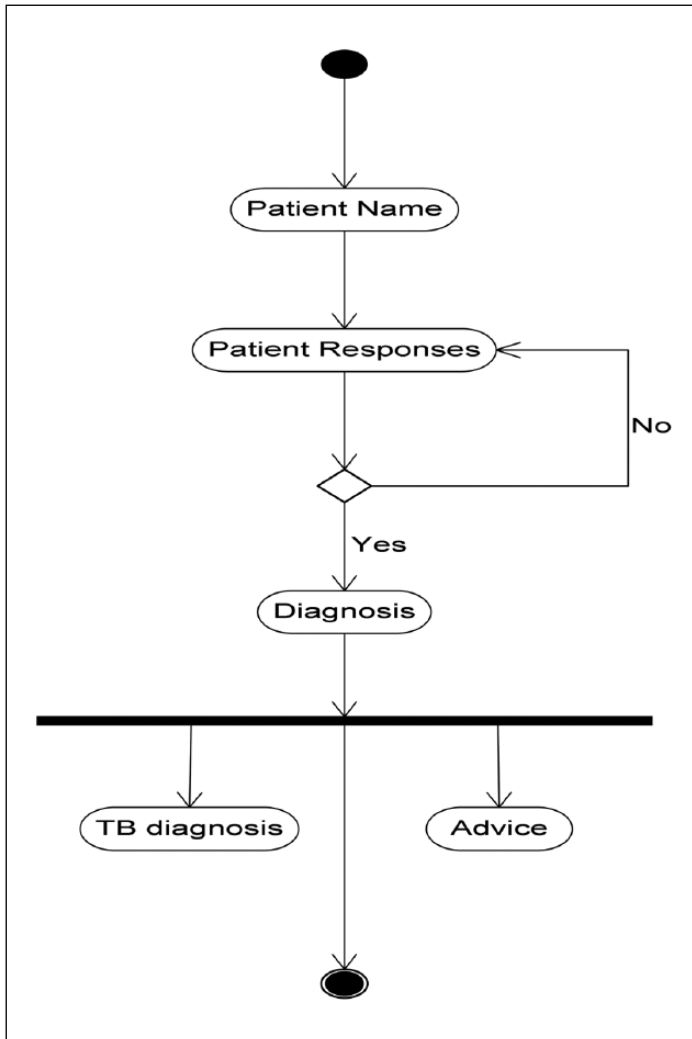


Figure 2. The activity diagram represents the transaction processes of TB-DES.

The minimum hardware requirements for the TB-DES to work at an optimal level are Pentium III (600 MHz) processor or higher, 128 MB of RAM or more, with hard disk size of 2 GB and above, and Operating System Software (Windows XP, 2000 and above).

Results and discussion

The result of the integration of the various parts of the system modeled in Figures 1 to 2 gave rise to the complete architecture as shown in Figure 3. The core of the architecture consists of the patient database, rules base, inference engine and web client–server interface. The patient database contains facts that are the smallest piece of information supported by the inference engine. The rule base contains rules in the form of if-then statements, which represent the knowledge provided by

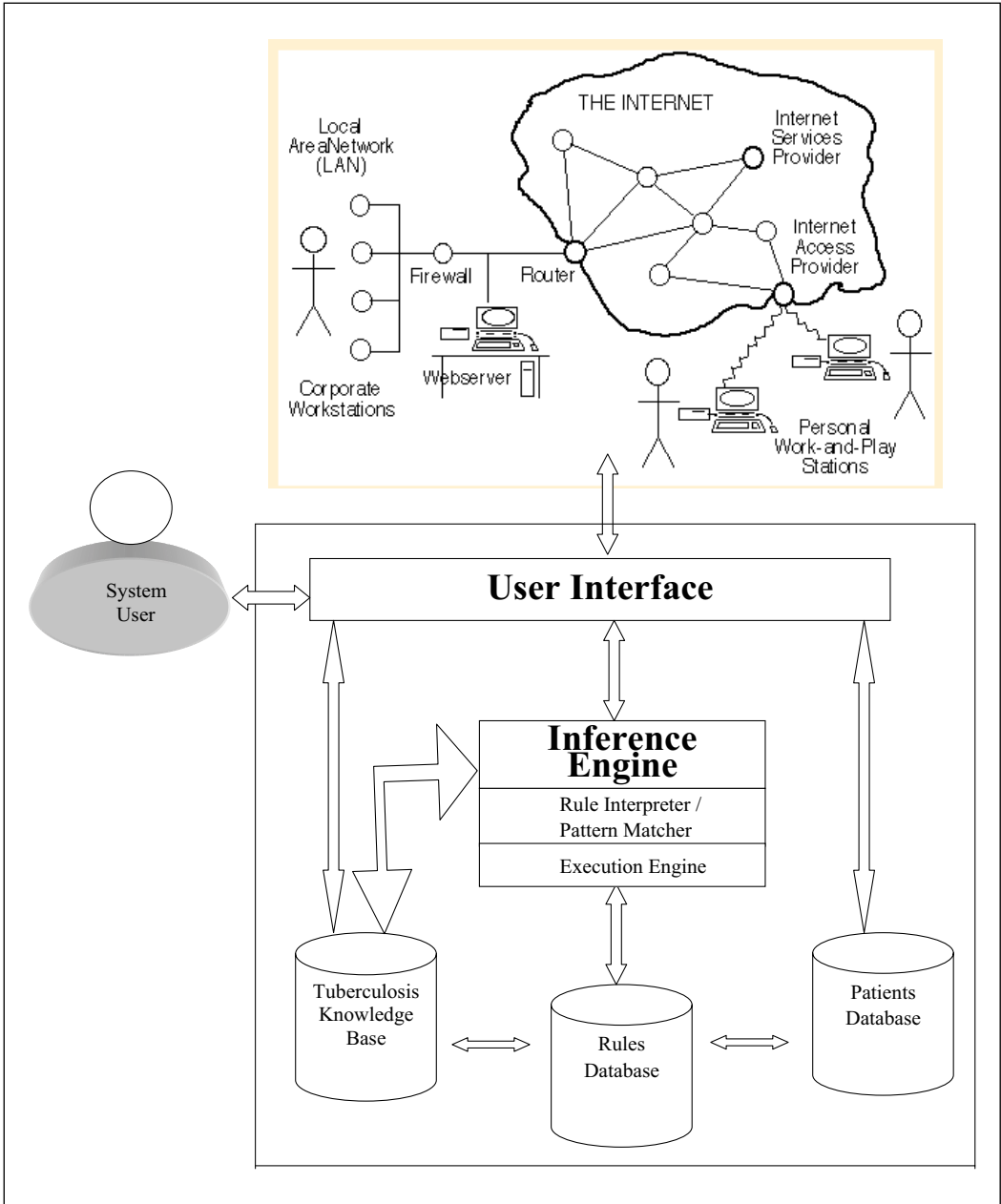


Figure 3. The TB-DES architecture consists of a rule base, knowledge base, and patient database. These units interact with the inference engine that receives patient data through the Internet via a user interface.

the user and/or an expert in the problem domain. The inference engine matches facts in the patient database against rules in the rule base, and it determines which rules are applicable according to the reasoning method adopted by the engine.

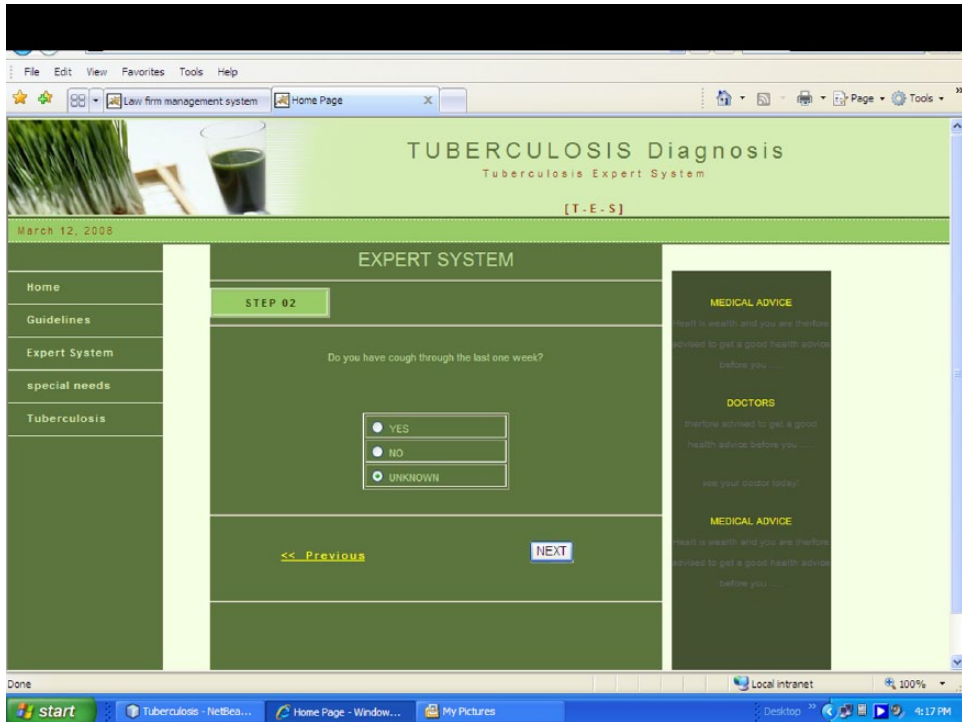


Figure 4. Question about any recent cough symptoms. The user is asked whether he or she had cough during the past week; the user selects an option and then moves to the next page.

The user who is ready to undertake the TB diagnosis pre-laboratory prediction connects through the Interface layer via a browser client, sending his or her personal (patient) data and associated symptoms as prompted by the system. The web server handles the requests by sending it to the expert system through same Interface layer. The required diagnostics prediction is processed by the expert system and delivered to the client. This is made possible by the internal working of the inference engine, which matches the facts (the symptoms supplied by user) in the patient database against the rules defined and sourced by the TB expert and placed in the knowledge base.

The TB-DES starts by asking series of questions from patients and data submitted to the system are taken as symptoms, for example, cough, loss of appetite, or any other symptoms. These symptoms may be associated with several illnesses like common cold, asthma, and so on, but if the patient is predicted to have TB, then lab confirmation is required. Possible symptoms of TB recognised by TB-ES are shown in Table 2.

The symbolic representation of expert knowledge in a computer can be represented using a number of schemes, but the rule-based approach is perhaps the most straightforward. Rules are conditional “if-then” statements that indicate circumstances under which conclusions can be drawn or actions taken, while categorical rule may be cause-to-effect, effect-to-cause, associational, or definitional rule.³ In some cases, knowledge represented in expert systems may include indications of the degree of certainty associated with the rule.

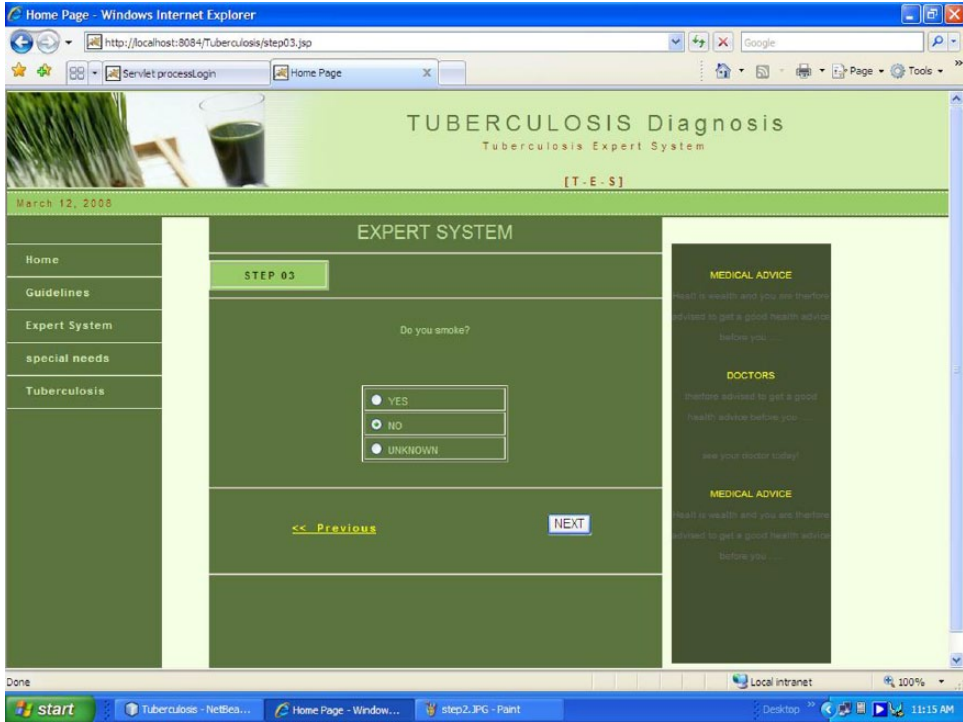


Figure 5. Smoking habits. The user is asked whether he or she smokes, because such behavior may aggravate TB; the user selects an option and then clicks on the next button to move to the next page..

Table 2. Symptoms of tuberculosis.

Symptoms	Description of manifestations
Cough	Common cough with a progressive increase in production of mucus Coughing up blood or bloody sputum Coughing that lasts longer than 2 weeks with green, yellow, or bloody sputum
Swollenness	Swollen lymph glands Swollen joints Swollen neck
Miscellaneous	Difficulty breathing or shortness of breath Fever or chills Loss of appetite Weight loss Night sweats Chest pain Weakness

This shows the possible symptoms of tuberculosis ranging from cough, swollenness, and other miscellaneous symptoms.

Table 3. Descriptive statistical analysis of questionnaire data.

Summary of study variable				
Usability Measures	Responses from doctors	Responses from patient	Total number of respondents	Total mean rating
Effectiveness	4	6	10	4.00
Efficiency	4	6	10	4.10
Satisfaction	4	6	10	4.15
Average mean rating = 4.08				

A usability study was carried out using the specification in ISO 9241–11¹⁹

To conduct the study, a questionnaire was designed and administered to doctors and patients. The survey questionnaire was divided into five sections, namely, background information, effectiveness of the system, efficiency of the system, and user satisfaction. Each section on effectiveness, efficiency, and user satisfaction contains four questions represented by Q1, Q2, Q3, and Q4. A total of 12 questionnaires were administered, but only 10 responses were received, analyzed and reported (see Table 3). The questions were designed using a 5-point Likert scale, where 1 = strongly disagree, 2 = disagree, 3 = undecided, 4 = agree, and 5 = strongly agree.

The average ratings, standard deviation (SD) and variance (VAR) are presented in Table 4. In terms of questions such as “Would you support the use of computer diagnosis for TB in your hospital?,” 6 users said “yes” and 4 said “no.” Responses to Q4 on effectiveness – “I feel comfortable using the application” – were lower compared to other questions. Notwithstanding, the user survey results were encouraging, particularly, the responses to questions on user satisfaction. It was proposed by Sauro and Kindlund²⁰ that “Good Usability” should have a mean rating of 4 on a 1–5 scale and 5.6 on a 1–7 scale. Therefore, it can be concluded that the application has “Good Usability” based on the average total rating of 4.08. The usability analysis graph for each survey variable tested is shown in Figure 6 and Kindlund³² that “Good Usability” should have a mean rating of 4 on a 1–5 scale and 5.6 on a 1–7 scale. Therefore, it can be concluded that the web-based application developed has “Good Usability” based on the average total rating of 4.08. The usability analysis graph for each survey variable tested is shown in Figure 7.

Conclusion

Following the vast information harvested from the clinical biology of *Mycobacterium sp.*, we applied a rule-based algorithm for the development of the TB-DES TB prediction system. Our knowledge base is inspired largely by the depth of information gathered on the clinical biology of *Mycobacterium sp.* A rule-based approach has been applied and knowledge stored as rules and facts to help medical practitioners reduce patient–doctor ratio in hospitals and ultimately consultation time in the long run. This system further aids in acquiring patients’ medical information via web for use in TB diagnosis, thereby ensuring medical consultation without walls.

The result of the usability evaluation reveals that the prototype application developed has “Good Usability” rating of 4.08 out of 5. This shows that the application will not only complement existing manual TB diagnosis systems but will be of benefit to doctors and patients.

Table 4. User survey results N = 10.

Survey questions	Average	SD	VAR
<i>Background information</i>			
Would you support the use of computer diagnosis for TB in your hospital? (yes/no)	6/4		
Do you need more computing skills/training/time to be able to use the system? (yes/no)	5/5		
<i>Effectiveness</i>			
Q1 I was able to complete my task successfully and correctly using the application	4.10	0.57	0.32
Q2 The system did not show error message(s) while using it	4.00	0.82	0.67
Q3 I was able to recover from my mistakes easily	4.00	0.47	0.22
Q4 I feel comfortable using the application	3.90	0.32	0.10
Average	4.00	0.55	0.33
<i>Efficiency</i>			
Q1 Using the system saves me time	4.10	0.57	0.32
Q2 I was able to complete my task on time	4.00	0.47	0.22
Q3 I was well able to navigate the user interface on time when using the system	4.20	0.42	0.18
Q4 I didn't have to carry out too many/difficult steps before completing my task	4.10	0.57	0.32
Average	4.10	0.51	0.26
<i>User satisfaction</i>			
Q1 The system was easy to learn	4.00	0.67	0.44
Q2 The system was easy to use and user-friendly	4.40	0.52	0.27
Q3 I am satisfied using the system	3.90	0.57	0.32
Q4 I feel the system met my need	4.30	0.48	0.23
Average	4.15	0.56	0.32
Resultant average rating	4.08	0.54	0.30

TB: tuberculosis; SD: standard deviation; VAR: variance.

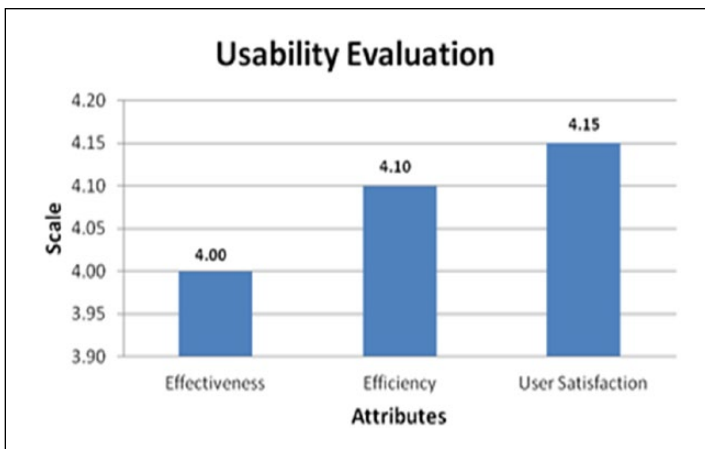


Figure 6. Usability analysis. The effectiveness, efficiency and user satisfaction with the system

Declaration of conflicting interests

There is no conflict of interest for this work.

Funding

This work was carried out using the authors' own funds

References

1. Thoen CO and Bloom BR. Pathogenesis of mycobacterium bovis. In: Thoen CO and Steele JH (eds) *Mycobacterium bovis infection in animals and humans*. 1st ed. Ames, IA: Iowa State University Press, 1995, pp. 3–14.
2. Tiruvilumala P and Reichman LB. Tuberculosis. *Annu Rev Public Health* 2002; 23: 403–426.
3. Shortliffe EH. *Computer based medical consultations: MYCIN*, 1976. New York: American Elsevier, 1976.
4. Pauker SG, Gorry GA, Kassirer JP, et al. Towards the simulation of clinical cognition. Taking a present illness by computer. *Am J Med* 1976; 60: 981–995.
5. Weiss SM, Kulikowski CA, Amarel S, et al. A model-based method for computer-aided medical decision making. *Artif Intell* 1978; 11: 145–172.
6. Shortliffe EH, Scott AC, Bischoff MB, et al. ONCOCIN: an expert system for oncology protocol management. In: *Proceedings of 7th international joint conference on artificial intelligence*, Aug. 24–28, 1981. Los Altos, CA, 1981, pp. 876–881. Morgan Kaufmann, pp. 876–881. Publisher: The Conference.
7. Miller MA, Pople HE and Myers JD. INTERNIST-1: an experimental computer-based diagnostic consultant for general internal medicine. *N Engl J Med* 1982; 307: 468–476.
8. Hatzilygeroudis I, Vassilakos PJ and Tsakalidis A. XBONE: a hybrid expert system supporting diagnosis of bone diseases. In: *Proceedings of the Medical Informatics Europe* (eds C Pappas, N Maglaveras and J-R Scherrer), Thessaloniki, Greece, May 1997, pp. 259–299. IOS Press.
9. McKenzie JS, Morris RS, Tutty CJ, et al. EpiMAN-TB, a decision support system using spatial information for the management of tuberculosis in cattle and deer in New Zealand. In: *Proceedings of the second annual conference of GeoComputation & SIRC*, University of Otago, New Zealand, 26–29 August 1997.
10. González C, Burguillo JC, Vidal JC, et al. An intelligent tutoring system to provide e-learning in public health. In: *Proceedings of the 16th EAEEIE conference*, Lappeenranta, Finland, 6th–8th June 2005.
11. Murty US, Kumar DVRS, Rao MS, et al. Rapid identification of female Culex mosquito species using expert system in the South East Asian region. *Bioinformatics* 2005; 1: 40–41.
12. Luciani D, Cavuto S, Antiga L, et al. Bayes pulmonary embolism assisted diagnosis: a new expert system for clinical use. *Emerg Med J* 2007; 24: 157–164.
13. Ocampo E, et al. Comparing Bayesian inference and case-based reasoning as support techniques in the diagnosis of Acute Bacterial Meningitis. *Expert Syst Appl* 2011; 38: 10343–10354.
14. Nguyen PH, Dang HH, Nguyen CV, et al. TUBERDIAG: an expert system for pulmonary tuberculosis diagnosis. *Int J Uncertain Fuzz* 1999; 7(4): 371–381.
15. Djam XY and Kimbi YH. A decision support system for tuberculosis diagnosis. *Pac J Sci Technol* 2011; 12(2): 410–425.
16. Daniele LM, Dockhorn CP and Ferreira PL. Towards a rule-based approach for context-aware applications. In: *Proceedings of the 13th Open European Summer School and IFIP TC6.6 workshop, EUNICE*, July 18–20, 2007, Enschede, The Netherlands, 2007.
17. Soe SM and Zaw MP. Design and implementation of rule-based expert system for fault management. *Proc World Acad Sci Eng Tech* 2008; 48: 779.
18. Whitten JL, Bentley LD and Dittman KD. *System analysis and design methods*. 5th ed. New York: Irwin/McGraw-Hill, 2001.

19. International Organization for Standardization (ISO) 9241-11: 1998. Ergonomic requirements for office work with visual display terminals (VDTs)—part 11: guidance on usability.
20. Sauro J and Kindlund E. A method to standardize usability metrics into a single score. In: *CHI*, Portland, OR, 2-7 April 2005. pp. 401-409. ACM press.