MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI231327

Symbolic Artificial Intelligence to Diagnose Tuberculosis Using Ontology

Napthaline GERARD [a,1], Sarah BEN OTHMAN [a], Pajanivel RANGANDIN[b], Marc BROUCQSAULT [c], Slim HAMMADI [a]
[a]*Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France*
[b]*Mahatma Gandhi Mediacl COllege & Research Institute, Puducherry, India*
[c]*ALTAO Compagny, Lille, France,*
ORCiD ID: Napthaline Gerard https://orcid.org/0009-0008-5050-9063, Sarah Ben Othman https://orcid.org/0000-0002-2648-2031, Pajanivel Rangandin https://orcid.org/0000-0002-2608-1633, Slim Hammadi https://orcid.org/0000-0002-5187-4870

Abstract Pulmonary Tuberculosis (PTB) is an infectious disease caused by a bacterium called Mycobacterium tuberculosis. This paper aims to create Symbolic Artificial Intelligence (SAI) system to diagnose PTB using clinical and paraclinical data. Usually, the automatic PTB diagnosis is based on either microbiological tests or lung X-rays. It is challenging to identify PTB accurately due to similarities with other diseases in the lungs. X-ray alone is not sufficient to diagnose PTB. Therefore, it is crucial to implement a system that can diagnose based on all paraclinical data. Thus, we propose in this paper a new PTB ontology that stores all paraclinical tests and clinical symptoms. Our SAI system includes domain ontology and a knowledge base with performance indicators and proposes a solution to diagnose current and future PTB also abnormal patients. Our approach is based on a real database of more than four years from our collaborators at Pondicherry hospital in India.

Keywords: PTB diagnosis, ontology, domain ontology, knowledge base

1. Introduction

Pulmonary Tuberculosis (PTB) is a disease caused by bacteria called Mycobacterium Tuberculosis [1]. It is estimated that over one-quarter of the world's population has been infected with Mycobacterium tuberculosis rivaling the impact of HIV/AIDS [2]. TB disease in the lungs or throat is infectious and can be contagious. TB in other parts of the body, such as the kidney or spine, is usually not contagious. There are many types of testing to diagnose TB namely Symptom screening, Chest X-ray, etc. [3]. It is challenging to diagnose TB to date in a hospital because most lung diseases have similar clinical symptoms, and the early stage of TB is not specific. It is very difficult for doctors to identify the issues due to various similar lungs disease, initially diagnosis takes a long time for each patient.The automatic TB diagnosis system is developed to reduce the burden, reduce time and increase accuracy. Still, the automatic diagnosis of TB is one of the most essential and difficult health problems in the real world. Therefore, early diagnoses of TB are very important to effectively treat patients before getting high levels

[1] Corresponding Author: Napthaline Gerard, napthalineg@gmail.com.

of risk. Timely diagnosis and initiation of therapy are also affected by patient delay and health system delays [4]. Data insufficiency may affect the accuracy of diagnosis on patients. Thus, how to train a robust TB diagnosis model using insufficient data is a more challenging task. Fortunately, clinical knowledge is composed of relationships between different medical events (i.e., symptoms or diseases). The introduction of clinical knowledge into diagnosis provides additional information. So, taking these relationships into account improves the reliability and accuracy of the diagnosis results [5].

To address these issues, we present an SAI system that contains a domain ontology with a diagnostic module and a knowledge base. In this context, we propose domain ontology to establish a biological diagnosis system for PTB to avoid data misclassification. In the literature, there are some existing ontologies for domain knowledge, knowledge base, and diagnosis such as Diabetes ontology (DO) for diabetes disease diagnosis for case base knowledge [6]. The existing TB ontology is a draft ontology for TB clinical data that had already been developed by the clinical data interchange standards consortium. They generated a set of terms and their definitions related to TB diagnosis and treatment based on usage in various TB databases and from several TB studies. This work was later completed under the critical path to TB drug regimens initiative [7]. To the best of our knowledge, there is no complete TB diagnosis ontology already published. Thus, the proposed ontology is a novel approach to diagnosing PTB using clinical and paraclinical data for accurate and original results. In addition, our ontology acts as an intelligent system using logical expression and reasoning as an SAI system. We conduct this work as an international project among Mahatma Gandhi Medical College & Research Institute, Pondicherry-India, and Lille university - France.

2. Methods

Symbolic AI solves the problem by using logical thinking rules, and reasoning. The reasoning and the rules created through human intervention also refer to human-readable and explainable processes. Symbolic AI contains domain ontology, knowledge base, and diagnostic modules are shown below (Figure 1).

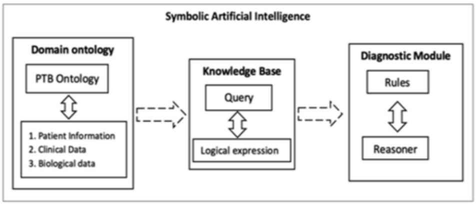

Figure 1. SAI system and the process of the system

2.1. Domain Ontology

Ontologies mean storing and exploiting the knowledge of different domains, and research in several aspects of ontologies is extensive. The field of medicine, describes the concepts of medical terminologies and the relation between them, the sharing of medical knowledge. Researchers in AI first developed ontologies to facilitate knowledge sharing and reuse. It is used in different forms: Database Management Systems, Knowledge Management, Domain Knowledge, Knowledge Representation, Semantic Web, etc. Domain ontology is more specific all the data and information stored are related to the concerned subject.

It is very important to choose the right methodology for the new ontology. To define the domain concepts, we use the standard medical manual of PTB from India to build a healthcare domain ontology. The domain ontology in PTB also converts our raw data into structured data. PTB specialists have validated the PTB ontology and each element defined and the relations between classes, then we describe all its elements with the OWL 2 language in ontology form. Table 1 presents an overview of the properties defined in PTB ontology.

Table 1. The table classifies the order and information contained in our ontology

Demographic information	Clinical information	Preliminary Diagnose	Types of Diagnosis
Patient Id	Fever	Sputum NAAT	Arterial Blood Gas
Patient Age	Cold	CB NAAT	Bio-Chemistry
Patient gender	Bleeding Cough	Sputum AFB	Electrolytes
		Sputum line Probe	HIV
		Sputum Liquid Culture & Sensitivity	Lipid Profile
		Sputum Solid Culture & sensitivity	Pathology
			Liver Function Test
			Microbiological investigation

2.2. Knowledge base

A knowledge-based system (KBS) is a form of artificial intelligence (AI) that aims to capture the knowledge of human experts to support decision-making. The techniques are defined by explicit symbolic methods, such as formal methods and programming languages, and are usually used for deductive knowledge [8]. It is usually associated with knowledge bases and expert systems and is a continuation of the von Neumann and Turing machines [9]. In particular, rule-based systems have the advantage of rule modularity, as the rules are discrete and autonomous knowledge units that can easily be inserted or removed from a knowledge base [10]. Finally, the maintenance of rule bases is difficult as it requires complex verification and validation.

After creating the domain ontology, we define the reasoning model to realize a functional ontology and integrate it into a decision-support system. We use the SWRL (Semantic Web Rule Language), a rule language for the semantic web, combining the OWL-DL language and RuleML (Rule Markup Language) to create a range for each test for all possible scenarios of the PTB. These are integrated into the resulting ontology to reason semantically based on rules, for example:

PTBontology:Patient_ID(?X)^has_Gender(?Z,Female)^has_CPK_MB_creatine_kinase_myocardial_band(?X,?Y)^swrlb:greaterThan(?Y,"26"^^xsd:integer)^swrlb:greaterThan(?Y,"192"^^xsd:integer) -> Normal(?X)

In addition, this model provides rules that assess abnormal cases and TB cases by proposing solutions to help healthcare workers. Thus, each rule is described with the properties and instances defined in the ontology. It is based on performance indicators to give an accurate assessment. We also categorized each paraclinical test into three categories: average range, above average, and below average. The defined rule base has been validated by the medical PTB specialist. The exploitation of semantic rules can also be used to manage missing data.

3. Results

To implement the PTB domain ontology, we used Protégé 5.5 (Figure 2). It is an open-source software [11] that provides a plug-and-play environment for rapid application development. Protégé is the most familiar tool for developing ontology. Developers can create the ontology easily through its graphical interface without having to worry about the syntax language. Thus, we integrate the SWRLTAB tool into protégé for creating and exploiting SWRL rules. The resulting ontology contains 250 classes, 38 object properties, 167 data properties, and 678 patients. Each paraclinical test is stored in the form of object property and data property. Our PTB ontology system is shown in figure 2. This is a visual representation of ontology in protégé.

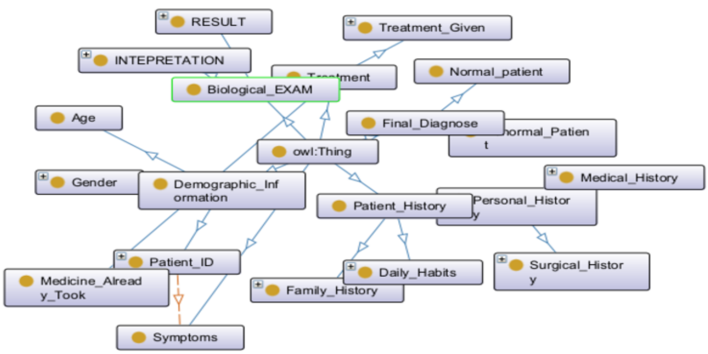

Figure 2. PTB Ontology – Graphical representative of our system

The SWRL rules that are created by the SWRLTAB and Protégé tools are generated with the inference engine. The rule base contains 519 semantic rules (Figure 3).

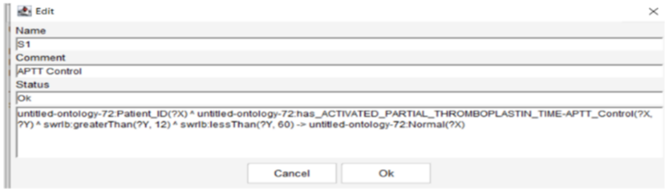

Figure 3. Implementation of SWRL rule

4. Discussion

Ontology evaluation requires formal and appropriate evaluation criteria. Therefore, there is no standard TB diagnosis ontology to measure its similarity to our ontology. Ontologies in the same domain are necessary, but in the literature, there is no complete ontology to diagnose Pulmonary Tuberculosis using all the paraclinical tests already published and which we could use. Our system contains eight categories of paraclinical tests, containing each different test type. In total there are 173 tests. By evaluating our PTB ontology, most patients who did not get TB positive by the microbiological test, have negative results for other types of tests. This ontology was validated by a general physician and a PTB specialist. The proposed method is also very useful in accurately diagnosing PTB.

5. Conclusions

This paper, we proposed an SAI system to diagnose TB in ontology with biological data. In collaboration with 2 doctors and to know deep knowledge about biological data and their impact, we designed PTB ontology by gathering all paraclinical tests without losing any information about the patients. Our ontology system is the best solution to reduce the issue with laboratory TB test complexity also our physicians are satisfied with our ontology. The performance is good also, we are processing to train in the hospital for a better result. In future work, combining clinical reports and X-rays of the same patient list with help of time sequencing such as the ODE – RNN (Ordinary Differential Equation Recurrent Neural Network) method which predicts due to the complexity of data and loss of accuracy.

References

[1] Tuberculosis fact sheet. World Health Organization; 2014.
[2] World Health Organization. Global tuberculosis report 2015. Geneva: WHO, 2015.
[3] Diagnostics for pulmonary tuberculosis. USA, Yale University School of Medicine; 2016.
[4] Sreeramareddy CT, Panduru KV, Menten J, et al. Time delays in diagnosis of pulmonary tuberculosis: a systematic review of the literature. BMC Infect Dis 2009; 9:91.
[5] Edward Choi, Mohammad Taha Bahadori, Le Song, Walter F. Stewart, Jimeng Sun, and Gram: Graph-based attention model for healthcare representation learning in Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017.
[6] H S. El-Sappagh, M. Elmogy. A fuzzy ontology modeling for case base knowledge in diabetes mellitus domain, in ES and Technology an International Journal (2017)
[7] The critical path to TB drug regimens. 2014.
[18] A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. de Melo, C. Gutierrez, J. E. L. Gayo, S. Kirrane, S. Neumaier, A. Polleres, et al., Knowledge graphs, arXiv preprintarXiv:2003.02320 (2020)
[9]E.N.Benderskaya, S.V.Zhukova, Multidisciplinary trends in modern artificial intelligence: Turing's way, in Artificial Intelligence, Evolutionary Computing, and Metaheuristics, Springer, 2013, pp. 319– 343.
[10] I.Hatzilygeroudis, J.Prentzas, Neuro symbolical approaches for knowledge representation in expert systems, International Journal of Hybrid Intelligent Systems 1 (2004) 111–126.
[11] Liu Y, Chen S, Wang Y. SOFERS: scenario ontology for emergency response system. JNW. 2014;9(9):2529–2