



ONTOLOGY WITH SVM BASED DIAGNOSIS OF TUBERCULOSIS AND STATISTICAL ANALYSIS

Murugavell Pandiyan[†] — Osama El-Hassan² — Amar Hassan Khamis³ — Pallikonda Rajasekaran⁴

¹Research Student, Kalasalingam University, Krishnan koil, India

²Head of eHealth Section, Dubai Health Authority, Dubai, UAE

³Professor, Mohammed Bin Rashid University of Medicine and Health Sciences, UAE

⁴Associate Professor, Kalasalingam University, Krishnan koil, India

ABSTRACT

As per WHO report, Tuberculosis remains one of the world's deadliest communicable diseases. In 2013, an estimated 9.0 million developed TB and 1.5 million died from the disease, 360,000 of which whom were HIV positive. Tuberculosis is still a major problem in advanced countries due to specific socioeconomic factors. From a global perspective, many laboratories use the same methods today that were in use long time ago for the detection of tuberculosis, because most of innovative current technologies for the detection of tuberculosis incurs high cost and cannot be afforded for all the countries. The detection of tuberculosis remains a challenge from the point of diagnosis and confirmation and there is a growing need of accurate diagnosis process. In this research, an ontology based classification of tuberculosis laboratory tests, environmental factors and other vital signs are studied along with support vector machine for the diagnosis of the tuberculosis disease. Through this method, we are able to measure of the weightage of the disease, the future onset of the disease and produce, an alert. Ontology based classification is widely used for knowledge based information grouping and structuring while SVM is used for accurate and fast machine learning algorithm. By combining Ontology and the training data based on various characteristic of the tuberculosis are passed onto linear SVM. The results we are able to achieve with this method are helpful for diagnosis support and future onset.

Keywords: Tuberculosis, Ontology, Machine learning, SVM, Decision support, Symptomatology.

Received: 14 April 2016/ Revised: 14 May 2016/ Accepted: 17 May 2016/ Published: 23 May 2016

1. INTRODUCTION

The field of Ontology is currently used as the universal standard for various classifications of the diseases, treatment procedures and the laboratory codes which are used in data integration, electronic health record and the reporting system. These classifications and the structuring have been extended over to the environmental factors in terms of workplace, epidemic outburst of geographic location and also related information based on the characteristic nature of the disease and the spreading ability. The tuberculosis disease itself is having specific characteristic nature and the vital signs which are classified based on ontology relationships.

The laboratory results, vital signs and, the environmental factors have been plotted as various dimensions on linear SVM and their weightage are being adjusted accordingly. The results which are obtained on the support vector machines based on various training data are used to adjust the bias factor and the weightage which further gives an output a standardized template for decision support system

The statistic study has been performed on individual characteristic of tuberculosis, geographic location and vital signs and their results are used for deriving weightage adjustments and decision alert.

The research work comprises of ontology based classification, statistical analysis and linear SVM for tuberculosis.

A. Challenges

Diagnosing Tuberculosis is very difficult due to its nature

1. Lack of accurate and cost-effective diagnostic method.
2. Although new diagnostic methods are being developed, those tests incur very high cost and are unavailable in certain countries.
3. For the children, even the contemporary diagnosis approaches still fail

To address the aforementioned problems, a mathematical model with the ontology is developed to help in decision support for Tuberculosis diagnosis.

B. Proposed Solution

Within the context of this research work, it offers the design and development of Ontology with SVM [1] based diagnosis using OWL [2]; [3] which as a primary element and backbone for structural representation of Tuberculosis, its nature and diagnosis rules. A Support Vector machine has been used within this context which tends to act as reasoning engine for each rule defined in the ontology based concept.

The research paper is organized in the following manner: within Section 2 it offers the background information and literature survey. This is followed by Section 3 which provides the architecture and implementation details. Finally, Section 4 summarizes the research outcomes and outlines future work.

2. LITERATURE SURVEY

As per El-Sony, et al. [4] it clearly recommends a statistical research based mathematical model for tuberculosis to conduct a detailed research study of the various characteristics of the disease however this mathematical model could be considered far more beneficial and suitable if we build software based reasoning based on the statistics.

This is shared by the views as per Adi, et al. [5] who recommends that a SVM based diagnosis of tuberculosis based on microscope imaging of sputum. However, this research addresses only specific diagnostic tool of tuberculosis based on ziehl-neelsen sputum with imaging analysis.

3. ONTOLOGY WITH SVM

A. Design

In our approach we have considered the entire factor that affects the tuberculosis cause.

We have divided our research area into the following modules: Building ontology model on tuberculosis vital signs its characteristics, patient geographic location, statistical analysis of the various components of tuberculosis, SVM engine to notify the inference result based on various inputs of vital signs/lab results. A reasoning engine is designed to combine the ontology rule and the SVM.

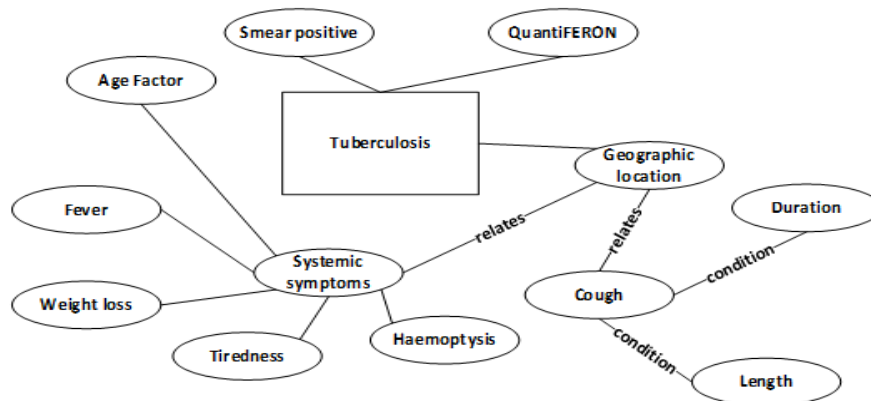


Fig-1. Ontology model relates various vital signs and lab tests

- *Ontology Model*

```

<vitalsigns:tb>
<symptom:cough>
<icd10:code >
<rdf:Bag>
<rdf:li>R05</rdf:li>
</rdf:Bag>
</icd10>
<rule:condition maxDays="10">
</rule:condition>
<rule:condition relatedTo="#geo">
</rule:condition>
<rule:condition intersectWith="#fever">
</rule:condition>
</symptom:cough>
<symptom:fever>
<icd10:code >
R50.9
</icd10>
<rule:condition relatedTo="#geo">
</rule:condition>
<rule:condition interSectWith="#cough">
</rule:condition>
</symptom:hemoptysis>
</vitalsigns:tb>
<lab:tests>
<lab:test>
<lab:loinc>
45323-3
</lab:loinc>
<lab:result>
<rule:condition interSectWith="#symptom">
</rule:condition>
<result:code>
    
```

```
40023700
</result:code>
<result:status>
<result:apply weightage="100">
</result:apply>
</lab:result>
</lab:test>
</lab:tests>
<geo:location>
<geo:population refersTo="geoservice" regioncode="@code" >
</geo:population>
</geo:location>
</rdf:RDF>
```

- *SVM Engine*

Each vital sign and the laboratory code act as a plane, the results are plotted and linearly separated. The weightage of each component is adjusted based on the statistical analysis of the trained data with IBM.SPSS.Statistics [6].

The determination function for the various linear component is given as

$$f(x) = w \cdot x + b$$

$$x_i \in R_n \quad Y_i \in \{0,1\}$$

$$w^T \cdot x_i + b \leq v \quad Y_i = 0$$

$$w^T \cdot x_i + b \geq v \quad Y_i = 1$$

- *Result Analysis*

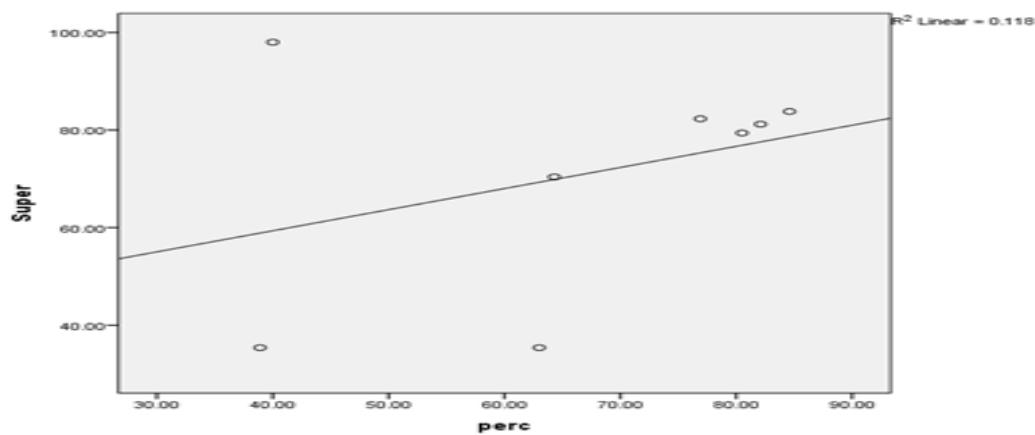
When analyzed 2007 Sudan clinical data of Tuberculosis for all entire regions with the confirmed smear positive with the following symptoms Cough, Lose of Weight etc. as tabulated below along with the Southern Region of patients' data, we derived the results using an IBM SPSS tool and the results were helpful in calculating the weighted average for each symptom and their intersection relationship to model the ontology rules.

Table-1. Clinical Data of Tuberculosis

Symptoms	Super Set TB%	Southern Region %
Cough	82.1	81.2
Lose of Weight	84.6	83.8
fever	76.9	82.3
Tiredness	80.5	79.4
Nigh sweating	64.3	70.4
Chest pain	63	35.4
Shortness breath	38.9	35.4
Loss appetite	40	98
Hemoptysis	16.5	13.4
Diff. Swallowing	12.8	9

Source: National Tuberculosis Program, Khartoum, Sudan (July 2002) [4]

Table 1 Tabular column of patient’s data having TB and the patients from southern region



Graph-1. shows the linear separation of the patients from southern region having the fore-mentioned symptoms

Table 2 shows the technologies used for the implementation of the solution. Fig.2 shows the “Architecture of the solution”.

Table-2. Technologies Used

Technologies	Brand
Database & Language	MySQL [7]
Statistical Analysis	IBM.SPSS.Statistics.
OWL	Jena [8]
SVM	LIBSVM [9]

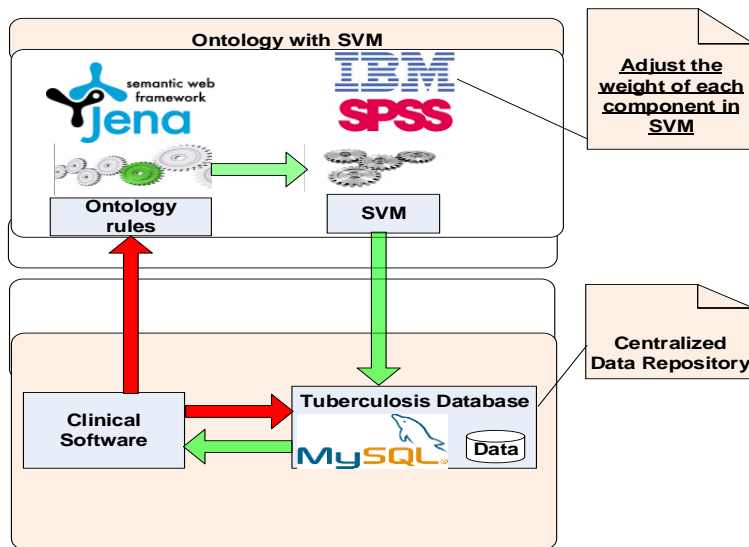


Fig-2. Architecture of Ontology and SVM engine

B. Dataflow

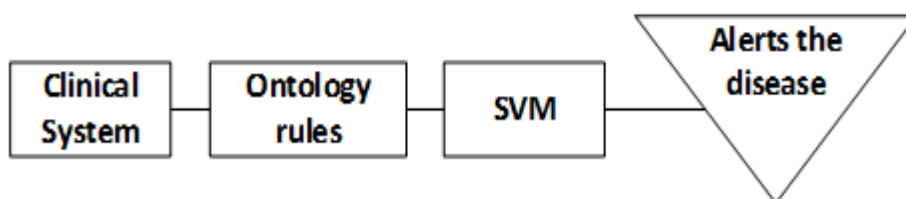


Fig-3. Data Flow

In a situation when a patient is encountered, the clinical system would capture the essential vital signs such as cough, fever, duration, length of the cough, tiredness, weight loss and geographic location to fulfill the initial requirements of collating basic information of the patient. Depending on the data captured the system queries the ontology model for the rule and the data is fed into the SVM engine to infer the result either 0 or 1 based on the trained data already plotted with the weighting factor.

The weighting factor is determined with the help of IBM. SPSS. Statistics and LIBSVM to produce the linearly separated component.

4. DISCUSSION

Within the context of this research paper, we can observe that the architecture presented illustrates the foundation for the development of a decision support for Tuberculosis diagnosis taking into account all the factors with the lab results, we have to expand the design and test the system in the infrastructure of Infectious disease reporting system [10] and extend the statistical analysis in real-time to adjust the weightage factor.

5. CONCLUSION AND FUTURE WORK

There was a dire need to work on such a research paper which has been driven by a great deal of hard work and it even poses a lot of challenges which support the need for diagnosing tuberculosis for the underprivileged or developing countries where they are unable to simply afford the contemporary technically advanced laboratory tests. Furthermore, the research also serves as an exhaustive approach for the diagnosis of Tuberculosis. This work has to be extended in different arenas, such as integrating the Infectious disease reporting system with the policy-based framework in order to be compliant to healthcare regulators' rules and regulations. Such an approach specified by El-Hassan, et al. [11] allows specifying rules for accessing resources (e.g. patients data) in both normal and emergency situations.

Funding: This study received no specific financial support.

Competing Interests: The authors declare that they have no competing interests.

Contributors/Acknowledgement: All authors contributed equally to the conception and design of the study.

REFERENCES

- [1] C. Hsu, C. Chang, and C. Lin, "A practical guide to support vector classification, 2003." Available: <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>. [Accessed: 22- May- 2016], 2015.
- [2] W3C, "OWL web ontology language reference, W3.org, 2004." Available: <http://www.w3.org/TR/owl-ref/>. [Accessed 22- May- 2016], 2004.
- [3] M. K. Smith, C. Welty, and D. McGuinness, "'OWL web ontology language guide, W3C, 2009." Available: <http://www.w3.org/TR/owl-guide>. [Accessed 22- May- 2016], 2015.
- [4] A. I. El-Sony, S. A. Mustafa, A. H. Khamis, S. Sobhi, D. A. Enarson, O. Z. Baraka, and G. Bjune, "Symptoms in patients attending services for diagnosis of pulmonary tuberculosis in Sudan," *International Journal of Tuberculosis and Lung Disease*, vol. 7, pp. 550-555, 2009.
- [5] K. Adi, R. Gernowo, A. Sugiharto, K. Sofjan, P. Adi, and B. Ari, "Tuberculosis (TB) identification in the Ziehl-Neelsen sputum sample in Ntsc channel and support vector machine (SVM) classification," *International Journal of Innovative Research in Science, Engineering and Technology*, vol. 2, pp. 5030-5035, 2013.
- [6] Analytics, "IBM SPSS - IBM analytics, IBM." Available <http://www-01.ibm.com/software/analytics/spss>. [Accessed 22- May- 2016], 2016.
- [7] MYSQL, "Database,2015: MYSQL. "MySQL", Mysql.com." Available <https://www.mysql.com>. [Accessed 22- May- 2016], 2016.

- [8] Jena, "Jena a semantic web framework for Java. Available <https://jena.apache.org>," n.d.
- [9] C. Chang and C. Jin, "LIBSVM – A library for support vector machines, Csie.ntu.edu.tw." Available <http://www.csie.ntu.edu.tw/~cjlin/libsvm/#download>. [Accessed 22- May- 2016]. 2016.
- [10] M. Pandiyan, O. El Hassan, Z. Maamar, and P. Rajasekaran, "Semantic interoperability for infectious diseases reporting system," in *FEDCSIS*, 2011, pp. 363–367.
- [11] O. El-Hassan, J. Fiadeiro, and R. Heckel, "Managing socio-technical interactions in healthcare systems," in *Business Process Management Workshops*, 2008, pp. 347-358.

Views and opinions expressed in this article are the views and opinions of the authors, International Journal of Medical and Health Sciences Research shall not be responsible or answerable for any loss, damage or liability etc. caused in relation to/arising out of the use of the content.