World Academy of Science, Engineering and Technology
International Journal of Health and Medical Engineering
Vol:12, No:10, 2018

Design of a Pneumonia Ontology for Diagnosis Decision Support System

Sabrina Azzi, Michal Iglewski, Véronique Nabelsi

***Abstract*—**Diagnosis error problem is frequent and one of the most important safety problems today. One of the main objectives of our work is to propose an ontological representation that takes into account the diagnostic criteria in order to improve the diagnostic. We choose pneumonia disease since it is one of the frequent diseases affected by diagnosis errors and have harmful effects on patients. To achieve our aim, we use a semi-automated method to integrate diverse knowledge sources that include publically available pneumonia disease guidelines from international repositories, biomedical ontologies and electronic health records. We follow the principles of the Open Biomedical Ontologies (OBO) Foundry. The resulting ontology covers symptoms and signs, all the types of pneumonia, antecedents, pathogens, and diagnostic testing. The first evaluation results show that most of the terms are covered by the ontology. This work is still in progress and represents a first and major step toward a development of a diagnosis decision support system for pneumonia.

***Keywords*—**Clinical decision support system, diagnostic errors, ontology, pneumonia.

I. Introduction

THE overload of clinical information significantly exceeds the human cognitive capacity and leads often to medical errors including the diagnosis step because physicians can no longer control the whole medical knowledge that makes it possible to recognize, at the proper time, medical problems and determine the right treatment for the patient.

A mistake made in the definition of the patient's diagnosis will have a great impact on all other steps of treatment. Getting the right diagnosis in the right time is the key aspect of health care as it provides an explanation of a patient's health problem, and determines an appropriate and successful therapy [1]. Inaccurate or delayed diagnosis persist throughout all settings of care and continue to harm an unacceptable number of patients for years [2].

In a study conducted in the state of Massachusetts [3], 23% of the people interviewed reported that they had been personally involved in a medical error situation. Almost half of these medical errors are those of diagnosis. In the United Kingdom, the National Health Service reveals that diagnoses - including diagnosis errors - are most often the reason people complain about care, and 35% of these complaints concern misdiagnosis [4].

Pneumonia is among the diseases affected by misdiagnosis [5]. Among 190 cases of discussed diagnostic errors, 7% concerned pneumonia. The authors of [6] assert that several patients received misdiagnosed community acquired pneumonia until the third or fourth clinical consultation. In [7], the authors insist that pneumonia associated with health care represents a major diagnostic challenge due to the low sensitivity and specificity of clinical criteria, radiological findings, and microbiological culture results. They affirm that it is often difficult to distinguish between pneumonia, underlying lung diseases or conditions with pulmonary complications.

According to [7], in most published studies, errors mainly concern missed diagnoses rather than over-diagnosis of pneumonia or diagnosis delays. The authors claim that when clinicians had the correct diagnosis at some point, the case was not categorized as involving an error. Although thoracic radiographs and clinical insight are reasonably sensitive to the detection of underlying abnormal lung pathology, they cannot reliably differentiate pneumonia and other pulmonary pathologies such as insufficiency chronic heart disease, interstitial lung disease or atelectasis. The authors even claim that the clinical diagnosis of pneumonia in inpatients is frequently defective. The paper [8] mentions that there is ambiguity in the diagnostic criteria for heart failure and pneumonia, and there is also an imperfect agreement for the interpretation of the chest x-ray. Radiologists who know the results of a patient interpret radiographs differently than radiologists who do not know the ultimate outcome.

The study conducted in [9] showed that 10 cases of misdiagnosis of pneumonia were found among the 583 cases of ill-diagnosed diseases reported by 310 clinicians from 22 institutions. Diagnostic errors cause 40,000 to 80,000 deaths a year in the United States [10] and a study conducted in [10] states that pneumonia is among the most misdiagnosed diseases.

Efforts to identify and reduce diagnostic errors have not really been of great interest [2]. The result of this lack of

S. A. is with the Department of Computer Science and Engineering, University of Quebec in Outaouais, Gatineau, Canada (phone: 819-595-3900; e-mail: sabrina.azzi@ uqo.ca).

M. I. is with the Department of Computer Science and Engineering, University of Quebec in Outaouais, Gatineau, Canada (e-mail: michal.iglewski@uqo.ca).

V. N. is with the Department of Administrative Sciences and with the Department of Computer Science and Engineering, University of Quebec in Outaouais, Gatineau (e-mail: veronique.nabelsi@uqo.ca).

attention is that most patients might have at least one misdiagnosis in their lives.

Information gathering is a fundamental step in the diagnosis process [2] as it would not be possible to synthesize clinical information to make a good diagnosis without having a good collection of data including patient history, physical examinations, test results or consultations with other physicians. However, studies carried out in recent decades have asserted that failures in information gathering are often responsible for diagnostic errors [11], [12].

Health professionals or clinicians have recently recognized the benefits of applying information technology (IT) to daily clinical practice [13]. Undeniably, primary objective of medical informatics has always been to assist clinicians in decision-making processes to prevent mistakes that may occur, maximize effectiveness and evidence, and ultimately improve health and care. Tools that support decision-making are referred to as clinical decision support systems [14].

Integration and interaction of the decision support systems (DSS) with the EHR (electronic health record) systems is a main requirement for future DSS. Successful DSS have been developed using their own terminologies [15], [16], which makes it difficult to transfer a successful system to other institutions.

Clinical DSS for pneumonia diagnosis are rare. In [17], a clinical DSS driven by a pneumonia ontology is described. The process of the pneumonia ontology building is not detailed and the ontology is not publically available.

Ontology is an important tool for modeling, sharing and reuse of domain knowledge as it can formally represent sharable domain knowledge for key concepts and their semantic relationships. It enhance clinical DSS by providing a standard vocabulary for biomedical entities, integrating diverse data resources and providing a genuine source of computable domain knowledge that can be exploited for decision support purposes [18], [19].

Our work is part of SEKMED research project (https://formation.sekmed.ca) whose objective is to develop a clinical DSS to help health practitioners in their practice. The focus is to develop an efficient clinical DSS aiming at reduction of pneumonia diagnosis errors.

The objective of the reported work is to build an ontology of pneumonia as a means to reduce diagnosis errors. We analyze the pneumonia diagnosis clinical guidelines with the purpose of extracting all relevant concepts and diagnosis rules. We also use health medical records to extract some other specific concepts used by physicians. We reuse parts of certain biomedical ontologies (see 2.3) to enrich further our ontology. Finally, we apply natural language processing tools to develop the resources needed to build the ontology.

II. Material and Methods

A. Ontology Domain Scope

The ontology domain and scope were defined after meetings with physicians from the Gatineau, Qc.Hospital in order to determine the functional requirements of the ontology. The ontology scope covers all factors relevant to making an appropriate diagnosis decision in the hospital setting including patient symptoms, signs, antecedents and microbiology results, such as gram stain and culture results.

The ontology domain was limited to pneumonia diagnosis. However, it has designed so that it can be easily extended to include treatment and other clinically relevant concepts such as risks factors and anatomical site of infection.

The primary key of our ontology will be the DSS accompanying physicians in their daily practice.

B. Concepts Extraction from Clinical Guidelines

Clinical guidelines contain knowledge frequently consulted by healthcare practitioners to guide diagnostic and management decisions. In the following sections, we describe the steps to get the list of relevant concepts.

1) Search of Relevant Clinical Guidelines

The search for relevant clinical guidelines revealed more than 15 websites related to respiratory diseases clinical guidelines. We identified eight of them, containing 13 pneumonia guidelines that most likely fulfill our ontology requirements. Here is a summary of the key clinical guidelines websites: two major infectious diseases societies that include: - Infectious Disease Society of America(IDSA; http://www.idsociety.org/idsa_practice_guidelines/) and European Society of Clinical Microbiology and Infectious Diseases(ESCMID;https://www.escmid.org/escmid_library/medical_guidelines/escmid_guidelines/), national guidelines repositories from around the world that include:

- Australian Society for Infectious Diseases (ASID; http://www.asid.net.au/resources/clinical-guidelines),
- National Institute for Health and Care Excellence, United Kingdom (NICE; https://www.nice.org.uk/guidance), Canadian Respiratory Guidelines, Canada (CTS; https://cts.lung.ca/guidelines),
- Pulmonary & Critical Care Medicine,USA (PulmCCM; http://pulmccm.org/main/pulmonary-and-critical-care-guidelines-clinical-practice-recommendations-ats-accp-bts-ers/),
- American Thoracic Society, USA (ATS; https://www.thoracic.org/statements/tuberculosis-pneumonia.php),
- British Thoracic Society, United Kingdom (BTS; https://www.brit-thoracic.org.uk/standards-of-care/).

2) Extraction of Concepts

We used Text2Onto [20] for concepts, subclasses, instances and relations extraction. Text2Onto [20] is a framework for ontology learning and data-driven ontology evolution which can be used to automatically generate ontologies from textual resources. Text2Onto is based on Probabilistic Ontology Model, which stores the learned primitives independently of a specific Knowledge Representation language. It features both a graphical user interface and an API that makes it easy to integrate Text2Onto into Java-based applications. It calculates a confidence level for each learned concept. It also updates the learned knowledge each time the corpus is changed.

3) Extraction of Relevant Concepts

Text2Onto provides different algorithms for concept extraction. First, we apply them to all 13 guidelines and compare the results to find the best algorithms for the extraction of different components. The components of interest are concepts, instances, relations and subclasses. The main criterion for the comparison was the relevance value. Next, we eliminate all concepts not related to pneumonia diagnosis such as treatment or prognosis terms. As result, we obtained 750 relevant concepts.

4) Mapping to UMLS

MetaMap [21] identifies biomedical concepts from free-form textual input and maps them into concepts from the Unified Language System (UMLS) Metathesaurus. Applying MetaMap to previously identified relevant concepts will permit to the built ontology to interoperate with existing biomedical ontologies. This allows us to define UMLS identifies (CUI), synonyms, acronyms and abbreviations for 70 % of our list.

C. Other Concepts Sources

In general, the existing guidelines do not deal with some specific terms like different pathogens or different pneumonia subtypes. Those terms can be found in ontology repositories such as NCBO Bioportal, Ontology Lookup Service from the European Bioinformatics Institute, Unified Modeling Language System UMLS or OBO Foundry. We reviewed existing ontologies to identify those that cover pneumonia diseases. We evaluated them for suitability of content coverage and depth of knowledge, as well as potential to support successful inference.

The search for relevant ontologies revealed more than 20 vocabulary resources containing pneumonia disease. There were a number of different types of resources ranging from simple taxonomies, glossaries and thesauri to more formalized clinical vocabularies which are considered to be ontologies. Viral and bacterial terms found in these guidelines, which can be responsible of pneumonia, were imported from the National Center for Biotechnology Information Organismal Classification (NCBITaxon). SNOMED-CT was also used to enrich our ontology with specific concepts related to pneumonia and not mentioned in the clinical guidelines.

D. Ontology Implementation

To maximize the interoperability and reusability of our ontology with existing and future ontologies, the Basic Formal Ontology (BFO) [22] was implemented as the upper ontology. We followed the Foundry’s ontology development guidelines (open, common format, URI/identifier space, versioning, scope, textual definitions, relations, documentation, documented plurality for users, commitment for collaboration, locus of authority, naming conventions, and maintenance) and Cimino’s Desiderata for terminologies [23]. The guidelines of Cimino include: 1) vocabulary content, 2) concept orientation, 3) concept permanence, 4) nonsemantic concept identifier, 5) polyhierarchy, 6) formal definitions, 7) rejection of "not elsewhere classified", 8) multiple granularities, 9) multiple consistent views, 10) context representation, 11) graceful evolution, and 12) recognized redundancy.

We used the Protégé framework to represent the pneumonia ontology.

E. Electronic Health Data for Ontology Improvement and Evolution

The evaluation of ontologies including their evolution and maintenance is a very important step. We link evaluation, evolution and maintenance because we believe that the evaluation is vital for the ontology growth.

According to [24], there are six levels of ontology evaluation: 1) lexical, vocabulary, concept, and data, 2) hierarchy taxonomy, 3) other semantic relations, 4) context, application, 5) syntactic, and 6) structure, architecture, design. There are also four evaluation approaches that can be or not applied to each level: golden standard, application-based, data-driven, and assessment by human. Currently, we focus on lexical, vocabulary, concept and data level using the data-driven evaluation approach. To achieve this, we used EHR used by physicians during their practice in a real-life setting. We use the publically available Multi-parameter Intelligent Monitoring in Intensive Care II (MIMIC II) database [25] (http://physionet.org/mimic2) to improve our ontology. It contains clinical data collected between 2001 and 2008 from a variety of Intensive Care Units (ICUs) patients. MIMIC II database contains patients admitted to medical units who typically have a high incidence of infections, as well as patients admitted to neonatal, post-surgical and cardiothoracic intensive care units who have a lower incidence and a more limited range of infections. MIMIC II is a diverse EHR database of 32,535 critically ill patients. We divided all patients with infective pneumonia into two random groups (90%, 10%). We use the first one for the ontology improvement and the other one for the evaluation purposes. The first results show that most of the terms are currently covered by the ontology. However, since the validation process is done manually, it is very time consuming.

III. Results

Our pneumonia ontology [26] mainly focuses on the diagnostic criteria. It covers symptoms and signs, all the types of pneumonia (Community-Acquired Pneumonia, Health-Care Associated Pneumonia, and Hospital-Acquired Pneumonia), antecedents, pathogens, and diagnostic testing. It also includes other diseases that nearly may have the same clinical signs and similar diagnostic test results. Our complete ontology contains 550 classes and 160 object properties. "Organism" hierarchy includes all the pneumonia pathogens including viruses, bacteria, fungi, and parasites. "Diagnostic testing" hierarchy covers all the diagnostic laboratory and imaging tests. "Antecedents" class contains all the diseases that a patient may have at the moment of consultation and are related to pneumonia causes or has had in his past. We created "Potential diseases" class to include diseases that have similar clinical signs and diagnostic test results to get the physician's attention in order to avoid errors.

IV. CONCLUSION

Improving the diagnosis represents a moral, professional, and public health imperative since diagnosis errors are frequent and their consequences may be unfortunate and irreparable. Pneumonia is among the diseases concerned by this serious problem.

The use of ontologies in the medical domain is increasing because it offers many advantages such as sharing a consistent understanding of what information means and separating domain knowledge from operational knowledge. To our best knowledge, there is no publically available pneumonia ontology. The main goal of our paper is to present the process of semi-automatically building of pneumonia ontology representing the medical diagnostic knowledge. This ontology will be the core of a decision support system for pneumonia diagnosis. We mainly focus on pneumonia clinical guidelines to extract relevant concepts.

In future work, we will extract diagnosis rules from the clinical guidelines in order to integrate them in the pneumonia ontology. Then, we plan to assess the system both on the usability dimension and on the quality of the recommendations provided in a hospital setting.

ACKNOWLEDGMENT

We are indebted to thank the CISSSO (Centre Intégré de la Santé et des Services Sociaux de l'Outaouais) and specially Dr Sylvain Croteau, physician at Gatineau, Qc. Hospital.

REFERENCES

[1] S. Holmboe Eric and J. Durning Steven, "Assessing clinical reasoning: moving from in vitro to in vivo," in *Diagnosis* vol. 1, ed, 2014, p. 111.
[2] E. National Academies of Sciences, and Medicine, N. A. Press, Ed. *Improving Diagnosis in Health Care*. 2016.
[3] H. S. o. P. Health, "The Public's Views On Medical Error In Massachusetts," Betsy Lehman Center for Patient Safety and Medical Error Reduction Health Policy Commission2014.
[4] "Complaints about Acute Trusts 2013-14 and Q1, Q2 2014-15," Parliamentary and Health Service Ombudsman.
[5] "Diagnostic Errors: Technical Series on Safer Primary Care," Geneva: World Health Organization 2016.
[6] R. M. Centor and R. Kumar, "Diagnostic error in community-acquired pneumonia," *Diagnosis,* vol. 1, no. 2, pp. 145-145, 2014.
[7] Z. Kubilay, A. J. Layon, H. Baer, and L. K. Archibald, "When is pneumonia not pneumonia: a clinicopathologic study of the utility of lung tissue biopsies in determining the suitability of cadaveric tissue for donation," *Cell and tissue banking,* vol. 17, no. 2, pp. 205-210, 2016.
[8] M. Robert McNutt, M. Richard Abrams, and M. Scott Hasler. (2005). *Diagnosing Diagnostic Mistakes*. Available: https://psnet.ahrq.gov/webmm/case/95.
[9] G. D. Schiff *et al.*, "Diagnostic error in medicine: analysis of 583 physician-reported errors," *Archives of internal medicine,* vol. 169, no. 20, pp. 1881-1887, 2009.
[10] D. E. Newman-Toker and P. J. Pronovost, "Diagnostic errors—the next frontier for patient safety," *Jama,* vol. 301, no. 10, pp. 1060-1062, 2009.
[11] H. Singh *et al.*, "Exploring situational awareness in diagnostic errors in primary care," *BMJ Qual Saf,* vol. 21, no. 1, pp. 30-38, 2012.
[12] L. Medford-Davis, E. Park, G. Shlamovitz, J. Suliburk, A. N. Meyer, and H. Singh, "Diagnostic errors related to acute abdominal pain in the emergency department," *Emerg Med J,* vol. 33, no. 4, pp. 253-259, 2016.
[13] H. Huang, *PACS and imaging informatics: basic principles and applications*. John Wiley & Sons, 2011.
[14] R. Nelson and N. Staggers, *Health Informatics-E-Book: An Interprofessional Approach*. Elsevier Health Sciences, 2016.p.170.
[15] R. Gonzales *et al.*, "A cluster randomized trial of decision support strategies for reducing antibiotic use in acute bronchitis," *JAMA internal medicine,* vol. 173, no. 4, pp. 267-273, 2013.
[16] R. S. Evans, S. Pestotnik, D. Classen, and J. Burke, "Development of an automated antibiotic consultant," *MD computing: computers in medical practice,* vol. 10, no. 1, pp. 17-22, 1993.
[17] P. J. Haug *et al.*, "An ontology-driven, diagnostic modeling system," *Journal of the American Medical Informatics Association,* vol. 20, no. e1, pp. e102-e110, 2013.
[18] O. Bodenreider, "Biomedical ontologies in action: role in knowledge management, data integration and decision support," *Yearbook of medical informatics,* p. 67, 2008.
[19] V. Kashyap, A. Morales, and T. Hongsermeier, "On implementing clinical decision support: achieving scalability and maintainability by combining business rules and ontologies," in *AMIA Annual Symposium Proceedings*, 2006, vol. 2006, p. 414: American Medical Informatics Association.
[20] P. Cimiano and J. Völker, "A framework for ontology learning and data-driven change discovery," in *Proceedings of the 10th International Conference on Applications of Natural Language to Information Systems (NLDB), Lecture Notes in Computer Science, Springer*, 2005, vol. 3513, pp. 227-238: Springer.
[21] A. R. Aronson, "Metamap: Mapping text to the umls metathesaurus," *Bethesda, MD: NLM, NIH, DHHS,* pp. 1-26, 2006.
[22] P. Grenon, B. Smith, and L. Goldberg, "Biodynamic ontology: applying BFO in the biomedical domain," *Studies in health technology and informatics,* pp. 20-38, 2004.
[23] J. J. Cimino, "Desiderata for controlled medical vocabularies in the twenty-first century," *Methods of information in medicine,* vol. 37, no. 4-5, p. 394, 1998.
[24] J. Brank, M. Grobelnik, and D. Mladenić, "A survey of ontology evaluation techniques," 2005.
[25] M. Saeed *et al.*, "Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II): a public-access intensive care unit database," *Critical care medicine,* vol. 39, no. 5, p. 952, 2011.
[26] (2018). *Pneumonia Ontology*. Available: http://larip.uqo.ca/Pneumonia.owl.