Journal of Biomedical Informatics 53 (2015) 251–260

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Automatic detection of patients with invasive fungal disease from free-text computed tomography (CT) scans

David Martinez [a], Michelle R. Ananda-Rajah [b], Hanna Suominen [c,d,e], Monica A. Slavin [f,g], Karin A. Thursky [f,g], Lawrence Cavedon [h,*]

[a] CIS Department, University of Melbourne, Australia
[b] Infectious Diseases Unit, Alfred Health and The University of Melbourne, Australia
[c] NICTA and The Australian National University, Canberra, Australia
[d] University of Canberra, Canberra, Australia
[e] University of Turku, Finland
[f] Victorian Infectious Diseases Service, Royal Melbourne Hospital, Peter MacCallum Cancer Institute, Australia
[g] Infectious Diseases Department, Peter MacCallum Cancer Institute, Australia
[h] School of Computer Science and IT, RMIT University, Melbourne, Australia

ARTICLE INFO

Article history:
Received 20 June 2014
Accepted 17 November 2014
Available online 24 November 2014

Keywords:
Natural language processing
Data mining
Surveillance
Invasive fungal disease
Aspergillosis

ABSTRACT

Background: Invasive fungal diseases (IFDs) are associated with considerable health and economic costs. Surveillance of the more diagnostically challenging invasive fungal diseases, specifically of the sino-pulmonary system, is not feasible for many hospitals because case finding is a costly and labour intensive exercise. We developed text classifiers for detecting such IFDs from free-text radiology (CT) reports, using machine-learning techniques.

Method: We obtained free-text reports of CT scans performed over a specific hospitalisation period (2003–2011), for 264 IFD and 289 control patients from three tertiary hospitals. We analysed IFD evidence at patient, report, and sentence levels. Three infectious disease experts annotated the reports of 73 IFD-positive patients for language suggestive of IFD at sentence level, and graded the sentences as to whether they suggested or excluded the presence of IFD. Reliable agreement between annotators was obtained and this was used as training data for our classifiers. We tested a variety of Machine Learning (ML), rule based, and hybrid systems, with feature types including *bags of words*, *bags of phrases*, and *bags of concepts*, as well as report-level *structured* features. Evaluation was carried out over a robust framework with separate Development and Held-Out datasets.

Results: The best systems (using Support Vector Machines) achieved very high recall at report- and patient-levels over unseen data: 95% and 100% respectively. Precision at report-level over held-out data was 71%; however, most of the associated false-positive reports (53%) belonged to patients who had a previous positive report appropriately flagged by the classifier, reducing negative impact in practice.

Conclusions: Our machine learning application holds the potential for developing systematic IFD surveillance systems for hospital populations.

1. Introduction

Invasive fungal diseases (*IFDs*) are associated with considerable health [27,38] and economic costs [2,32]. Among the immunocompromised host population, patients with haematological malignancies and haematopoeitic stem cell transplant recipients (HSCTs) carry the greatest burden of IFDs [27,38]. The incidence of IFDs is highly variable, ranging from 8% to 48% in patients with acute myeloid leukaemia and HSCT patients [27,35]. IFDs of the sino-pulmonary system, of which invasive *aspergillosis* (IA) is most common, now comprise the majority of IFDs [38,27]—focus in this paper is exclusively on this class of diagnostically challenging IFDs, including invasive *aspergillosis*. The short-term mortality of invasive pulmonary *aspergillosis* remains unacceptably high at 34–43% in more recent reports among HSCT recipients and patients with a variety of haematological malignancies [37,35,42]. Given the high health and economic burden of IFDs, several authorities and professional

* Corresponding author.
E-mail addresses: davidm@csse.unimelb.edu.au (D. Martinez), m.ananda-rajah@alfred.org.au (M.R. Ananda-Rajah), hanna.suominen@nicta.com.au (H. Suominen), monica.slavin@mh.org.au (M.A. Slavin), karin.thursky@petermac.org (K.A. Thursky), lawrence.cavedon@rmit.edu.au (L. Cavedon).

http://dx.doi.org/10.1016/j.jbi.2014.11.009
1532-0464/

societies advocate that *surveillance* of IFDs should be the standard of care [22,45,53].

Surveillance is a necessary step towards defining the burden of IFDs, informing and evaluating choice of preventative strategies, tracking epidemiological data in response to changing therapeutic advances, host or environmental factors [39], and recognising sporadic but catastrophic healthcare related outbreaks [26]. At present, surveillance of IFDs is not routinely performed in most hospitals for a variety of reasons, including cost and the absence of an easily identifiable laboratory prompt (such as a positive blood culture) for mould diseases. Laboratory-based surveillance for IFDs is not suitable because isolation of infection-causing moulds occurs in less than 50% of cases [15] and/or patients may be too unwell to undergo invasive diagnostic procedures. Further, non-culture based tests are not widely available in hospitals, or may be associated with a delayed turn-around. Clinical review is challenging, requiring multiple data sources (radiology, clinical, laboratory) and multi-disciplinary teams, followed by the application of complicated definitions [13,27,36]. Thus, traditional methods of surveillance using either clinical review, laboratory-based methods, or less commonly administrative data are costly, labour intensive, error prone, and subject to incomplete case findings [9,27].

The optimal screening method for IFD surveillance is undefined but the choice of screening method is critically important to minimise effort while maximising case capture. Computed tomography (CT) is appealing as a screening method for IFD surveillance. CT is a key diagnostic test for IFDs stipulated in internationally recognised guidelines [13], and lung involvement is present in the overwhelming majority (90–100%) of patients with IFDs [27,38,36]. CT is a non-invasive test uniformly performed when IFD is suspected and it is widely available in hospitals with results reported within hours rather than days. Although the radiologic features of IFDs are not specific for IFD [29], CT remains a valuable diagnostic adjunct [13].

In this paper, we address the challenge of surveillance of sino-pulmonary IFDs using the technique of directly processing free-text radiology reports, specifically CT scans, as a means of screening patients for features supportive of IFD. Text Mining techniques over scans have been previously proposed to support detection and surveillance, including for infectious diseases and their symptoms (e.g., fever) [11,46,17,41,34,25,30], but not previously to the challenge of detecting IFDs. The approach has the potential to identify patients with suspected IFDs in real-time, delivering to hospitals a feasible, sustainable and cost-effective solution to the task of IFD surveillance with minimal interruption to routine clinical workflow.

The described approach is a novel methodological contribution to the task of identifying cases of a specific condition from CT scan reports. Text classifiers were developed for CT scan reports at the sentence, individual report, and patient level. Classification at the report level enables the potential for real-time detection and monitoring of incidence of IFDs, while patient-level classification enables a surveillance and reporting mechanism for IFD. Sentence-level classification is shown to facilitate improved classification performance at the report- and patient-levels, and has the added benefit of indicating "supporting evidence" for a positively-classified report. The collection of scan reports used was drawn from three major Melbourne hospitals. A subset of these reports was manually annotated by infectious disease physicians with domain expertise in the area of IFDs. Manual annotation was performed *at the sentence level*, where the physicians arbitrated over several categories according to pre-specified annotation guidelines, of which presence or absence of language supportive of the presence of an IFD was most discriminatory. Note that we use a light annotation scheme, requiring much less annotation effort than is generally required for machine-learning approaches to biomedical text mining tasks (which may require annotation of phrases and relationships between entities). Moreover, we ensured that the machine-learning classifiers have no hand-coded intervention: i.e., they were constructed in a fully automated fashion. This makes the techniques more transferable to other problem tasks, and means even better performance could potentially be achieved on deployment in specific contexts, via hand-crafted tuning.

Our ultimate aim is to develop an automated system that can be trained to detect a new condition by having an expert in that condition analyse and annotate data directly. In particular, other than data-annotation, the models require no human input during training.

2. Related work

As noted by Demner-Fushman et al. [14], radiology reports are a rich source of knowledge and were used in early applications of still-influential clinical NLP systems (e.g., [24,20]). The types of reports examined have varied (e.g., X-ray, CT), and tasks have ranged from specific classification tasks to more general named entity recognition (e.g., [52]), coding, and information extraction (e.g., [18,19]), across a broad range of tasks. Applications have included disease/infection detection and surveillance from radiology reports, with pneumonia being the disease of greatest focus. A number of authors have demonstrated favourable performance of more sophisticated NLP techniques for identifying/classifying radiology reports for specific purposes, in comparison to simpler techniques. For example, Solti et al. [40] demonstrated the efficacy of machine-learning (ML) NLP over keyword search for identifying X-ray reports of cases of acute lung injury; Womack et al. [51] showed that the NegEx system [10] compares favourably to keyword search for identifying cases of acute fracture. Table 1 summarises the main works discussed in this section.

Work associated with LDS Hospital from Salt Lake City (Utah, US) has resulted in a series of systems for coding concepts and performing disease and infection surveillance, particularly the identification of pneumonia cases from X-ray reports. Haug et al. [24] used Augmented Transition Networks (ATNs) for syntactic-level processing of X-ray reports and a Bayesian Network model for semantic-level representation and constraining the interpretation of concept terms. The system was evaluated on the task of identifying (three) pneumonia-related concepts and inferring the presence (or absence) of acute bacterial pneumonia from X-ray reports, and was found to perform to a level comparable to physicians: 0.95/0.78/0.85 recall/precision/specificity respectively for the system vs. 0.94/0.87/0.91 for the team of three physicians employing majority vote.

More recently, Tinoco et al. [43] compared the effectiveness of a more general version of LDS Hospital's computerised surveillance system (CSS) against manual methods (i.e., chart review) on the tasks of detecting reportable adverse drug effects (ADE) and hospital-acquired infections (HAIs). They found that the automated system detected substantially more HAIs than did manual review (92% vs. 34%) while the two methods did not differ substantially on detecting ADEs, demonstrating the value of automated surveillance techniques for HAIs in general. One significant source of HAIs *missed* by the CSS was information in physician's narratives—58% of the HAIs missed by the CSS were explicitly described in the narratives: i.e., adding text mining capabilities would have improved performance of HAI-detection up to approximately 97%.

The general clinical NLP system MEDLEE has been used for a number of tasks involving processing radiology reports, including in early work by its authors [20]. Of specific relevance to our context, Mendonça et al. [31] described the application of MEDLEE to