Decision Support Systems 84 (2016) 104–116

Contents lists available at ScienceDirect

Decision Support Systems

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

A case-based reasoning system for aiding detection and classification of nosocomial infections

H.J. Gómez-Vallejo [a], B. Uriel-Latorre [b], M. Sande-Meijide [b], B. Villamarín-Bello [b], R. Pavón [a], F. Fdez-Riverola [a], D. Glez-Peña [a,*]

[a] Informatics Department, Universidad de Vigo, Campus Universitario As Lagoas s/n, 32004 Ourense, Spain
[b] Preventive Medice Service, Complexo Hospitalario Universitario de Ourense, Rúa Ramón Puga 52-56, 32004 Ourense, Spain

ARTICLE INFO

Article history:
Received 18 June 2015
Received in revised form 29 January 2016
Accepted 13 February 2016
Available online 20 February 2016

Keywords:
Healthcare-associated infections
Automatic surveillance
Infection detection and classification
Case-based reasoning
Clinical decision support system

ABSTRACT

Nowadays, it is recognized worldwide that healthcare-associated infections are responsible for an increase in patient morbidity, mortality, and higher costs related to prolonged hospital stays. As electronic health data are increasingly available today, there is a unique opportunity to implement real-time decision support systems for automating the surveillance of healthcare-associated infections. As a consequence, different electronic surveillance systems have been implemented to date with varying degrees of success. However, there have been few instances in which clinical data and physician narratives with the potential to significantly improve electronic surveillance alternatives have been adopted. In this context, the present work introduces a case-based reasoning system for the automatic surveillance and diagnosis of healthcare-associated infections. The developed system makes use of different machine learning techniques in order to (*i*) automatically extract evidence from different types of data including clinical unstructured documents, (*ii*) incorporate static *a priori* knowledge handled by infection preventionists, and (*iii*) dynamically generate new knowledge as well as understandable explanations about the system's decisions. Results obtained from a real deployment in a public hospital belonging to the Spanish National Health System trained with 2569 samples belonging to 1800 patients during more than 10 consecutive months recognize the usefulness of the system.

1. Introduction

Nosocomial infections (NIs), more commonly known as healthcare-associated infections (HAIs), refer to the infections acquired by a patient when treated in a hospital environment. HAIs or infections acquired in health-care settings are the most frequent adverse effect of health-care delivery worldwide. As is the case for many other patient safety issues, HAIs create additional suffering and come at a high cost for patients and their families. Infections prolong hospital stays, create long-term disability, increase resistance to antimicrobials, represent a massive additional financial burden for health systems, generate high costs for patients and their family, and cause unnecessary deaths.

Presently, HAIs are accepted as a major problem for both patients (causing an increment in morbidity and mortality) [1] and hospitals (with economic and reputational implications) [2] in almost all countries. Generally, at any given time, the prevalence of healthcare-associated infections in developed countries varies approximately between 3.5% and 12% when it is considered the entire hospital (i.e., all hospital units) and all types of infection. However, the percentage may vary significantly depending on both hospital unit and type of infection. In the USA alone, it was estimated that during 2011, there were 648,000 patients with 721,800 HAIs in acute care hospitals [3] (with one in twenty-five patients suffering at least one HAI). The number of deaths occurring in the USA as a direct consequence of these infections is estimated to be at least 99,000 per year. Similarly, the European Centre for Disease Prevention and Control (ECDC) reports an average prevalence of 7.1% in European countries, where approximately 4,100,000 patients are estimated to acquire an HAI every year, accounting for 37,000 deaths. However, approximately 20–30% of HAIs are considered to be preventable by intensive hygiene and control programs [4]. Taking into consideration the increasing emphasis placed by national administrations on benchmarking and public reporting of surveillance data, applicable procedures and methods need to be efficient, reliable, and transferable.

In this context, world-wide health services have established active prospective surveillance programs which should be strictly followed by infection preventionists and hospital epidemiologists with the goal of ultimately decreasing rates of infection. Although these programs have proven to be adequate, different factors actually hinder the use of traditional time-consuming surveillance methods applied by infection control specialists [5]: (*i*) inability to find cases via manual surveillance, (*ii*) use of partially subjective clinical criteria to identification of

* Corresponding author. Tel./fax: +34 988 387001.
E-mail address: dgpena@uvigo.es (D. Glez-Peña).
URL: http://sing.ei.uvigo.es (D. Glez-Peña).

http://dx.doi.org/10.1016/j.dss.2016.02.005
0167-9236/

HAIs, (*iii*) widely variable sensitivity depending on the site of infection, intensity of surveillance and/or application of site definitions, or (*iv*) the lack of valid and reliable metrics for inter-hospital comparisons.

In such a situation, the development of electronic surveillance tools based on information technologies, which are able to provide a more rapid and accurate detection of HAIs at a lower cost, remains a significant challenge [6]. Existing approaches range from systems focused on infection control to antimicrobial management and mathematical modeling of the underlying dynamics of pathogen transmission [7–8]. However, publicly available surveillance systems are scarce, not exhaustive, and their drawbacks and architecture have not been adequately defined and/or documented [5], which makes their adoption very difficult from a technical and economical point of view. Additionally, there are novel sources of clinical information available in the form of electronically unstructured documents that can be definitively taken into consideration to boost the effectiveness and comprehensiveness of the automatic surveillance of HAIs.

In this context, our InNoCBR system is a completely automated surveillance system which effectively integrates data from multiple sources to assist infection control specialists in their daily activity (i.e. from monitoring to diagnosing and controlling HAIs), following the ECDC surveillance definitions and criteria for all specific types of infections. The main component of InNoCBR is its case-based reasoning system (CBR), which is able to construct and manage its own independent knowledge base that makes use of clinical-, laboratory-, and medico-administrative-based data. These raw data are enhanced with prior expert knowledge represented as executable rules that drive the autonomous operation of the system. Moreover, a natural language processing (NLP) module is able to handle specific electronic physician narratives and daily comments from nursing staff in order to provide additional clues about the occurrence of underlying HAIs.

The remainder of the paper is structured as follows. Section 2 presents an overview of previous related systems devoted to the surveillance and control of HAIs. Section 3 introduces the general architecture of our CBR system, documenting its life cycle and providing details about the internal knowledge base and the different modules that make up the system. Section 4 demonstrates the suitability of the developed approach though a set of different scenarios and cases working with real data from a public hospital belonging to the Spanish National Health System (NHS). Finally, Section 5 briefly outlines the main conclusions and identifies future research work.

2. Previous work: Electronic surveillance/control of HAIs and use of CBR in the health sciences

Given the global importance of HAIs to both public and private health systems and their negative effects on patients, a fair number of electronic surveillance initiatives have been published in the scientific literature over the last decade. In order to acknowledge all these works, we outline below four systematic reviews that conveniently summarize the current state of the art, together with a list of specific commercial software packages currently available in the market.

The work of [9] reported an up-to-date compilation of 26 publications comprising discussions on 27 different electronic HAI surveillance systems from January 2001 to December 2011. The study was focused on data use and effectiveness, and its main conclusions showed (*i*) the great importance of microbiology data in HAI detection, (*ii*) the limited adoption of clinical data and physician narratives to date, and (*iii*) the fact that the use of heterogeneous data sources may result in higher system sensitivity at the expense of specificity. From another complementary perspective, the work of [10] discussed 24 studies ranging from 1980 to 30 September 2007 with the goal of comparing the utility of electronic and conventional surveillance methods. The main conclusion of the study was that, despite the explosion in database systems and availability of electronic data, electronic surveillance is not in widespread usage. However, reviewed systems demonstrated that electronic surveillance has moderate to excellent utility when compared to conventional alternatives.

In addition to what precedes, the work of [11] analyzed 44 studies published between January 2000 and December 2011 with the goal of evaluating system effectiveness in the detection of different types of infection. Although the findings of this systematic review endorsed previous recommendations about the convenience of implementing electronic surveillance systems, it also recognized that the architecture of existing systems greatly varies and that individual hospitals need to identify and assess available information sources before making the decision to develop a new, or adopt an existing, electronic surveillance system. Finally, the work of [12] recently updated and summarized existing electronic surveillance systems currently in use for the identification and surveillance of HAIs up to August 21, 2014. The focus of this systematic review was to describe primary data sources, data elements and validation methods, comparing their usage with recommended standards. The primary conclusion of this study was that the vast majority of the analyzed systems use standard definitions, but the lack of widespread internal data, denominator, and external validation compromise the reliability of their findings.

Although the majority of the previously published studies demonstrated the utility of electronic surveillance systems developed within public institutions rather than privative available software [13], it is a fact that there are also several commercial applications that give complementary support to different aspects related to infection tracking. From this perspective, but taking into consideration that these alternatives are often more expensive to purchase and maintain and less validated, some examples of privative software packages and applications are the following: MedMined Surveillance Advisor[1] from CareFusion, TheraDoc[2] from Hospira, Sentri7[3] from Pharmacy OneSource, 3M Clintrac Infection Control Manager[4] from 3M, Infection MonitorPro Software[5] from rL Solutions and SafetyAdvisor[6] from Premier Inc.

However, despite the existence of many implanted surveillance systems, there is not yet a definitive consensus on relevant questions concerning fundamental HAI issues like (*i*) the fact that clinical details associated with the definition of HAIs are not available in hospital information systems, (*ii*) the subjective importance of multiple data sources, (*iii*) the accuracy and quality of available clinical data and data linkage procedures, and (*iv*) the use of different computer algorithms and queries for specific electronic surveillance tasks.

For its part, and in parallel with the previously commented advances made and challenges existing in e-surveillance and control of HAIs, since 1970 electronic Decision Support Systems (DSS) in healthcare became more common, feasible, and necessary [14]. Although there are multiple examples of how DSS can improve and reduce costs in healthcare delivery, such as predicting the length of stay [15], evidence reporting widespread use of CDSS (Clinical Decision Support Systems) is, in general, limited [16]. However, and more concretely, multipurpose medical CBR systems have also experienced a significant growth in healthcare management over the past decade [17]. In fact, health sciences and CBR create an ideal partnership, the former offering a wide variety of complex tasks difficult to solve using traditional approaches, while the latter providing key benefits regarding the true integration of different types of knowledge [18] and its medically accepted reasoning process [19]. In this line, the work of Begum and colleagues [20] recently compiled up to 34 successfully developed CBR applications

[1] http://www.carefusion.com/medical-products/carefusion-brands/medmined/ (last accessed: Aug 27, 2015).
[2] http://www.theradoc.com/ (last accessed: Aug 27, 2015).
[3] http://www.pharmacyonesource.com/Sentri7 (last accessed: Aug 27, 2015).
[4] http://solutions.3m.com/wps/portal/3M/en_US/3M_Health_Information_Systems/HIS/Products/ClinTrac_Infection_Control/ (last accessed: Aug 27, 2015).
[5] http://www.rlsolutions.com/RLSolutions/media/pdf/whitepapers/TheCaseForInfectionPreventionAndControlSoftware.pdf (last accessed: Aug 27, 2015).
[6] https://www.premierinc.com/transforming-healthcare/healthcare-performance-improvement/patient-safety/safetyadvisor/ (last accessed: Aug 27, 2015).