Available online at www.sciencedirect.com

Public Health

journal homepage: www.elsevier.com/puhe

Original Research

ID-Viewer: a visual analytics architecture for infectious diseases surveillance and response management in Pakistan

M.A. Ali [a,*], Z. Ahsan [b], M. Amin [a], S. Latif [a], A. Ayyaz [a], M.N. Ayyaz [a]

[a] Centre of System Simulation and Visual Analytics Research, University of Engineering and Technology, GT Road, Lahore, Pakistan

[b] University Medical and Dental College, Sargodha Road, Faisalabad, Pakistan

ARTICLE INFO

Article history:
Received 27 October 2014
Received in revised form
31 December 2015
Accepted 11 January 2016
Available online 13 February 2016

Keywords:
Scan statistics
Disease forecast
Infectious diseases
Outbreak detection
Dynamic syndromic classification

ABSTRACT

Objectives: Globally, disease surveillance systems are playing a significant role in outbreak detection and response management of Infectious Diseases (IDs). However, in developing countries like Pakistan, epidemic outbreaks are difficult to detect due to scarcity of public health data and absence of automated surveillance systems. Our research is intended to formulate an integrated service-oriented visual analytics architecture for ID surveillance, identify key constituents and set up a baseline for easy reproducibility of such systems in the future.

Study design: This research focuses on development of ID-Viewer, which is a visual analytics decision support system for ID surveillance. It is a blend of intelligent approaches to make use of real-time streaming data from Emergency Departments (EDs) for early outbreak detection, health care resource allocation and epidemic response management.

Methods: We have developed a robust service-oriented visual analytics architecture for ID surveillance, which provides automated mechanisms for ID data acquisition, outbreak detection and epidemic response management. Classification of chief-complaints is accomplished using dynamic classification module, which employs neural networks and fuzzy-logic to categorize syndromes. Standard routines by Center for Disease Control (CDC), i.e. c1-c3 (*c1-mild, c2-medium and c3-ultra), and* spatial scan statistics are employed for detection of temporal and spatio-temporal disease outbreaks respectively. Prediction of imminent disease threats is accomplished using support vector regression for early warnings and response planning. Geographical visual analytics displays are developed that allow interactive visualization of syndromic clusters, monitoring disease spread patterns, and identification of spatio-temporal risk zones.

Results: We analysed performance of surveillance framework using ID data for year 2011–2015. Dynamic syndromic classifier is able to classify chief-complaints to appropriate syndromes with high classification accuracy. Outbreak detection methods are able to detect the ID outbreaks in start of epidemic time zones. Prediction model is able to forecast dengue trend for 20 weeks ahead with nominal normalized root mean square error of 0.29. Interactive geo-spatiotemporal displays, i.e. heat-maps, and choropleth are shown in respective sections.

* *Corresponding author*. Tel.: +92 3244067030.
E-mail addresses: asif.ali@kics.edu.pk, asif6827@gmail.com (M.A. Ali).

http://dx.doi.org/10.1016/j.puhe.2016.01.006

Conclusion: The proposed framework will set a standard and provide necessary details for future implementation of such a system for resource-constrained regions. It will improve early outbreak detection attributable to natural and man-made biological threats, monitor spatio-temporal epidemic trends and provide assurance that an outbreak has, or has not occurred. Advanced analytics features will be beneficial in timely organization/formulation of health management policies, disease control activities and efficient health care resource allocation.

Introduction

In Pakistan, epidemic outbreaks are difficult to detect due to unavailability of historical ID data, as currently there is no automated detection and surveillance system in place. However, in the research community, the detection of adverse health events has recently focused on prediagnostic information in a bid to improve the outbreak detection and response time. This type of detection, also called '*Syndromic Surveillance*' involves collection and analysis of statistical health data. Earliest research on syndromic surveillance and control of disease outbreaks dates back to 1998 when CDC initiated plans to develop programs for early detection and investigation of ID outbreaks. Later, in the 21st century after the anthrax outbreak, the US accelerated the nationwide implementation and development of syndromic surveillance systems.[1]

National Electronic Disease Surveillance System (NEDSS)[2] was developed to facilitate the transfer of health data, laboratory test information, and clinical data to state and local health departments efficiently and securely using data standards (such as HL-7[c]), which is currently serving as mainstay of prevailing surveillance systems. In 2010, the director general of the World Health Organization (WHO) called a meeting to address upcoming global health issues; the key focus of the meeting was to improve global health governance, and development of state-of-the-art health surveillance architectures in order to cope with imminent epidemic threats.[3] Sylvie Briand et al.[4] alongside many others[5–12] addressed key challenges faced in global health surveillance, lack of standards defined for disease reporting, importance and need for automated mechanisms of ID threats management. They emphasized the need for model-based surveillance frameworks for public health policy making and in turn judge their effectiveness and preparedness for prevention of IDs and efficacy of control activities.[13–15] These systems should support well-timed assessment of ID data, permit access to health surveillance services,[16–18] and help in efficient planning/management of disease control activities.

Globally, many syndromic surveillance systems have been designed and developed for ED data collection, integration, analysis and early warning of ID outbreaks. Most renowned systems in this domain include BioSense,[19,20] CDC's Early Aberration Reporting System (EARS),[21] Real-time Outbreak and Disease Surveillance (RODS),[22,23] BioPortal, and the Electronic Surveillance System for Early Notification of Community Based Epidemics (ESSENCE).[24] Other noteworthy systems include: New York City Emergency Department (NYC-ED), the Bio-event Advanced Leading Indicator Recognition Technology (BioALIRT),[25] Biological Spatio-temporal outbreak reasoning module (BioStorm),[26] National Bioterrorism Syndromic Surveillance Demonstration Program (NBSSD),[27] Indiana Public Health Emergency Surveillance System,[28] Oak Ridge Bio-surveillance Toolkit (ORBiT)[29] and Integrated Forecast and Early Enteric Outbreak (INFERNO).[30] Above-mentioned systems use over-the-counter (OTC) drug sales, ambulatory records, 911 calls, physician's data, chief complaints data, school absentees and laboratory records for early outbreak detection. They have suitable syndromic classification modules, to timely process chief complaints, based on text processing and machine-learning techniques for classification, e.g. BioSense provides monitoring support for eleven syndromic, and seventy-eight sub-syndromic categories, RODS uses Naïve–Bayes algorithm to classify chief complaints data into seven syndromic categories, and EARS provides surveillance for forty-two syndromes. In addition to syndromic classification modules, these systems have customizable data visualization and analysis tools to help public health officers monitor, analyse, and control unusual trends in health care data.

Pakistan is located in a sub-tropical zone experiencing diverse weather conditions; it undergoes four seasons in a year starting with winter (December–March), early-summer (April–June), late-summer (July–September) and post-monsoon (October–November). Changes in yearly weather conditions have deep impacts on public health and ID spread patterns, causing the region to undergo heavy burden of IDs throughout the year. Late summer provides optimum environmental conditions for vector-borne diseases, winter season is most suited for spread of air-borne infections, whereas post-monsoon is favourable for food-borne, vector-borne and water-borne IDs. Despite the catastrophic situations caused by ID epidemics, floods and other biological threats, until now, there is no nationwide health surveillance system in place for Pakistan, which puts it among the list of heavily ID burdened regions declared by WHO.[31,32] Fluctuating weather conditions alongside other natural and/or man-made catastrophic events pose the need for public health surveillance system keeping

[c] Health Level Seven (HL7), is an organization involved in the development of international health care informatics interoperability standards.