Computerized Text Analysis to Enhance Automated Pneumonia Detection

Sylvain DeLisle*[1, 2], Tariq Siddiqui[1, 2], Adi Gundlapalli[3, 4], Matthew Samore[3, 4] and Leonard D'Avolio[5, 6]

[1]VA Maryland Health Care System, Baltimore, MD, USA; [2]Medicine, University of Maryland, Baltimore, MD, USA; [3]VA Salt Lake City Health Care System, Salt Lake City, UT, USA; [4]University of Utah, Salt Lake City, UT, USA; [5]VA Boston Health Care System, Boston, MA, USA; [6]Harvard Medical School, Boston, MA, USA

Objective

To improve the surveillance for pneumonia using the free-text of electronic medical records (EMR).

Introduction

Information about disease severity could help with both detection and situational awareness during outbreaks of acute respiratory infections (ARI). In this work, we use data from the EMR to identify patients with pneumonia, a key landmark of ARI severity. We asked if computerized analysis of the free-text of clinical notes or imaging reports could complement structured EMR data to uncover pneumonia cases.

Methods

A previously validated ARI case-detection algorithm (CDA) (sensitivity, 99%; PPV, 14%) [1] flagged VAMHCS outpatient visits with associated chest imaging (n = 2737). Manually categorized imaging reports (Non-Negative if they could support the diagnosis of pneumonia, Negative otherwise; kappa = 0.88), served as a reference for the development of an automated report classifier through machine-learning [2]. EMR entries related to visits with Non-Negative chest imaging were manually reviewed to identify cases with Possible Pneumonia (new symptom(s) of cough, sputum, fever/chills/night sweats, dyspnea, pleuritic chest pain) or with Pneumonia-in-Plan (pneumonia listed as one of two most likely diagnoses in a physician's note). These cases were used as reference for the development of the EMR-based CDAs. CDA components included ICD-9 codes for the full spectrum of ARI [1] or for the pneumonia subset, text analysis aimed at non-negated ARI symptoms in the clinical note [1] and the above-mentioned imaging report text classifier.

Results

The manual review identified 370 reference cases with Possible Pneumonia and 250 with Pneumonia-in-Plan. Statistical performance for illustrative CDAs that combined structured EMR parameters with or without text analyses are shown in the Table. Addition of the "Text of Imaging Report" analyses increased PPV by 38-70% in absolute terms. Despite attendant losses in sensitivity, this classifier increased the F-Measure of all CDAs based on a broad ARI ICD-9 codeset. With the possible exception is CDA 6, whose F-measure was the highest achieved in this study, the text analysis seeking ARI symptoms in the clinical note did not add further value to those CDAs that also included analyses of the chest imaging reports.

Conclusions

Automated text analysis of chest imaging reports can improve our ability to separate outpatients with pneumonia from those with a milder form of ARI.

CDA Number	1	2	3	4	5	6	7	8	9	10	11	12
	Possible Pneumonia						Pneumonia-in-Plan					
CDA Components												
(Pneumonia ICD-9 Codes)	●	●					●	●				
(ARI ICD-9 Codes)			●	●	●	●			●	●	●	●
OR Text of Clinical Notes				●	●					●	●	
AND Chest Imaging Obtained	●	●	●	●	●	●	●	●	●	●	●	●
AND Text of Imaging Reports		●		●		●		●		●		●
Sensitivity (%)	36.8	28.4	85.9	58.4	99.7	66.2	52	40.8	93.6	68.8	100	74.8
Specificity (%)	95.4	99.7	29.8	98.5	2.2	98	95.4	99.6	29.8	96.8	2.3	95.7
PPV (%)	55.3	93.8	16	86.1	13.7	83.3	52.8	91.1	12	68.5	9.3	63.6
NPV (%)	91	90	93.2	93.8	98.1	95	95.2	94.4	98	97	100	97.4
F-Measure	44.2	43.6	27	69.6	24.1	73.8	52.4	56.4	21	68.6	17	68.7

Keywords

situational awareness; influenza; surveillance; electronic medical record; pneumonia

References

[1] DeLisle S, South B, Anthony JA, Kalp E, Gundlapalli A, et al. Combining Free Text and Structured Electronic Medical Record Entries to Detect Acute Respiratory Infections. PLoS ONE (2010) 5(10): e13377.

[2] D'Avolio L, Nguyen T, Goryachev S, Fiore L. Automated Concept-Level Information Extraction to Reduce the Need for Custom Software and Rules Development. Journal of the American Medical Informatics Association 2011 18(5): 607.

*Sylvain DeLisle
E-mail: sdelisle@umaryland.edu