International Journal of Medical Informatics 84 (2015) 1039–1047

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.ijmijournal.com

NLP based congestive heart failure case finding: A prospective analysis on statewide electronic medical records

Yue Wang[a,b,1], Jin Luo[b,1], Shiying Hao[b,1], Haihua Xu[d,1], Andrew Young Shin[c,1], Bo Jin[d,1], Rui Liu[b], Xiaohong Deng[f], Lijuan Wang[d], Le Zheng[b], Yifan Zhao[d], Chunqing Zhu[d], Zhongkai Hu[d], Changlin Fu[d], Yanpeng Hao[b], Yingzhen Zhao[b], Yunliang Jiang[b], Dorothy Dai[d], Devore S. Culver[e], Shaun T. Alfreds[e], Rogow Todd[e], Frank Stearns[d], Karl G. Sylvester[b,2], Eric Widen[d,2], Xuefeng B. Ling[b,*,2]

[a] State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou 310027, PR China
[b] Departments of Surgery, Stanford University, Stanford, CA 94305, USA
[c] Departments of Pediatrics, Stanford University, Stanford, CA 94305, USA
[d] HBISolutions Inc., Palo Alto, CA 94301, USA
[e] HealthInfoNet, Portland, ME 04103, USA
[f] Chongqing Key Lab of Catalysis & Functional Organic Molecules, Chongqing Technology and Business University, Chongqing, China

ARTICLE INFO

Article history:
Received 7 February 2015
Received in revised form 23 June 2015
Accepted 25 June 2015

Keywords:
Congestive heart failure
Random forests
Natural language processing
Electronic Medical record
Prospective validation

ABSTRACT

Background: In order to proactively manage congestive heart failure (CHF) patients, an effective CHF case finding algorithm is required to process both structured and unstructured electronic medical records (EMR) to allow complementary and cost-efficient identification of CHF patients.

Methods and results: We set to identify CHF cases from both EMR codified and natural language processing (NLP) found cases. Using narrative clinical notes from all Maine Health Information Exchange (HIE) patients, the NLP case finding algorithm was retrospectively (July 1, 2012–June 30, 2013) developed with a random subset of HIE associated facilities, and blind-tested with the remaining facilities. The NLP based method was integrated into a live HIE population exploration system and validated prospectively (July 1, 2013–June 30, 2014). Total of 18,295 codified CHF patients were included in Maine HIE. Among the 253,803 subjects without CHF codings, our case finding algorithm prospectively identified 2411 uncodified CHF cases. The positive predictive value (PPV) is 0.914, and 70.1% of these 2411 cases were found to be with CHF histories in the clinical notes.

Conclusions: A CHF case finding algorithm was developed, tested and prospectively validated. The successful integration of the CHF case findings algorithm into the Maine HIE live system is expected to improve the Maine CHF care.

1. Introduction

The US Centers for Disease Control and Prevention (CDC) has reported that congestive heart failure (CHF) remains a principle cause of overall hospitalization and its prevalence has not changed significantly between 2000 and 2010 [1]. The estimated heart failure related mortality is approximately 287,000 people per year [2]. In aggregate, heart failure imparts an enormous yearly cost of approximately $31 billion dollars to the US healthcare system [3].

The Centers for Medicare and Medicaid Services (CMS) has proposed CHF readmission rate as a measure of healthcare quality and target for cost control [4]. Many CHF hospitalizations are considered to be preventable if patients were to receive timely and appropriate medical care [5]. Therefore, an effective real-time analytical solution to comprehensively identify CHF cases is needed to help guide targeted interventions and appropriate resource allocation [6].

A traditional method for CHF case finding is based on clinical coding [7] that largely depends on the availability of structured electronic medical record (EMR) datasets. However, this method is flawed resulting in a significant under-reporting of the targeted population [8]. One solution is to find those uncodified CHFs by

* Corresponding author at: Department of Surgery, S370 Grant Bldg., Stanford University School of Medicine, Stanford, CA 94305, USA.
E-mail address: bxling@stanford.edu (X.B. Ling).
[1] Co-first authors.
[2] Co-last authors.

http://dx.doi.org/10.1016/j.ijmedinf.2015.06.007
1386-5056/© 2015 Published by Elsevier Ireland Ltd.

manual review of the narrative EMR clinical notes. However, high labor costs and latency prohibit the practicality of this approach. Therefore, processing both structured and unstructured data for CHF case finding can provide a complementary and cost-efficient way to identify patients and apply targeted care.

Over the last decade, use of Natural Language Processing (NLP) to analyze the EMR narrative texts has been largely confined to clinical research focusing on information extraction for the purpose of EMR enrichment and decision support [10–12]. Alternatively, the applications of NLP to clinical notes, e.g., identification of pneumonia [13], diabetes [14], and CHF [15,16], have shown promise as a case finding method. The reported NLP based case finding studies to date have achieved good *F*-measures [9]. However, these studies utilized relatively small sample size or focused on specific types of clinical notes. The challenge in the current study was to execute CHF case finding utilizing a statewide patient population, where the class distribution is highly imbalanced. Unstructured notes from multiple facilities across the Maine State were found with known narrative expression variability thus impacting information comprehensiveness. To deal with this challenge, the algorithm should have: (1) a comprehensive knowledge base to capture the accumulated domain knowledge from the targeted patient population; (2) a rigorous data model is needed to encompass the unstructured clinical notes of various formats across different facilities; (3) a robust and scalable analytical pipeline is needed to process the vast amount of EMR notes across statewide facilities.

In this study, we set to develop and integrate a real-time NLP-based CHF case finding algorithm into the Maine HIE care flow (Fig. 1).

2. Methods

2.1. Ethics statements

No PHI was released for the purpose of this clinical research. Because this study analyzed de-identified data, the Stanford University Institutional Review Board considered it exempt (October 16, 2014).

2.2. Data source

The health information exchange in Maine (HealthInfoNet, HIN) is an independent nonprofit organization initiated in 2009 that contains records for nearly the entire population of the state of Maine residents and is connected to the majority of health care facilities in the state. There are currently 35 hospitals, 384 federally qualified health centers, and over 400 ambulatory practices connected to the Maine HIE. HIN maintains an opt-out consent process with a patient opt-out rate of slightly over 1%; and certain behavioral health and HIV related information is excluded from the database as required by Maine law. To identify the CHF cohort from a statewide population, all categories of clinical notes from the connected facilities were included. There were 2,139,299 notes in the Maine HIE EMR database covering a period from July 1, 2012 to June 30, 2014, with more than 100 different types of clinical reports including history/physical reports, discharge summaries and emergency reports.

2.3. Experimental design (Fig. 2)

CHF cases were identified utilizing the Clinical Classification Software (CCS) single-level diagnosing group (#108 Congestive heart failure; nonhypertensive) [17], and the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes including 398.91, 428.0, 428.1, 428.20, 428.21, 428.22, 428.23, 428.30, 428.31, 428.32, 428.33, 428.40, 428.41, 428.42, 428.43, and 428.9. The CHF case finding algorithm consisted of two phases: (1) the EMR database analyses of patient encounters with the CHF ICD9 codes; (2) NLP-based case finding analyses based on our knowledge base and de-noised dictionary (see below). Total of 8 datasets as outlined in Fig. 2 were utilized throughout this study: retrospective A and A1 to A4, prospective B and B1 to B2 datasets. The NLP engine was trained with data set A1, analyzed with dataset A2, finalized with manually chart-reviewed gold standard dataset A3, and evaluated with manually chart-reviewed gold standard dataset A4 in a retrospective timeframe from July 1, 2012 to June 30, 2013. Our case finding algorithm was prospectively deployed to the HIE live system (Supplementary Table 1). The algorithm's prospective performance was gauged using another chart-reviewed gold standard dataset B1 of uncoded encounters within the prospective testing period from July 1, 2013 to June 30, 2014. The clinical notes of the NLP identified cases were further profiled to explore unique clinical patterns associated with these previously uncoded but genuine CHF cases.

2.4. Workflow to construct the gold standard dataset A3, A4 and B1

Clinical notes of samples were randomly selected and manually reviewed by two physician curators. When there was a disagreement on diagnosis that could not be resolved by the curators, the sample was excluded. The resultant datasets was used as the gold standard to validate our NLP case finding method.

2.5. NLP knowledge base

The developed knowledge base has two modules: (1) a controlled vocabulary consisting of CHF related clinical terms; and (2) the extracted rules combining vital signs and comorbidities in the clinical notes (Supplementary Fig. 1).

The clinical terms in our NLP knowledge base were derived from the following sources: (1) ICD-9-CM code string descriptions and corresponding synonyms; (2) the comprehensive clinical terminology within the Systematized Nomenclature of Medicine—Clinical Terms (SNOMED CT) [18]; (3) a mapping between ICD-9-CM and SNOMED CT proposed by the US National Library of Medicine (NLM) [19]; and (4) a controlled vocabulary thesaurus named Medical Subject Headings (MeSH) used by NLM for article indexing [20]. These clinical terms in the knowledge base were further tokenized, combined and filtered to derive our controlled vocabulary of single and dual tokens. If those controlled vocabularies contain stop words, e.g. "the", "a", "of", provided by the text mining (tm) package [21], they were removed. A total of 148 final NLP terms were compiled, and 52/148 were found to be significantly (Mann–Whitney test P value <0.05) associated with CHF.

The vital sign/comorbidity, including BMI, CHF standard markers, obesity, fasting blood glucose level, smoking history, and alcohol use status, can provide important cues of being with CHF. To compile the knowledge base that enabled structured information to be derived, a series of regular expressions representing the rules to enable information unification and design of different feature categories were compiled. As an example, BMI was presented directly in some notes, but could also be calculated based on height and weight. Therefore, BMI information was unified from two sources and normalized into four categories: underweight, normal, overweight and obesity according to the BMI classification of the World Health Organization (WHO) [22]. The blood pressure and fasting blood glucose levels were classified according to related standards from American Heart Association and American Diabetes Association, respectively [23,24].