Published in final edited form as:

Proceedings (IEEE Int Conf Bioinformatics Biomed). 2017 November ; 2017: 1296–1299. doi:10.1109/BIBM.2017.8217848.

Automatic Methods to Extract New York Heart Association Classification from Clinical Notes

Rui Zhang,
Institute for Health Informatics, and College of Pharmacy, University of Minnesota, Minneapolis, MN, USA

Sisi Ma,
Institute for Health Informatics, and Department of Medicine, University of Minnesota, Minneapolis, MN, USA

Liesa Shanahan,
Medtronic, Inc., Minneapolis, MN, USA

Jessica Munroe,
Medtronic, Inc., Minneapolis, MN, USA

Sarah Horn, and
Medtronic, Inc., Minneapolis MN, USA

Stuart Speedie
Institute for Health Informatics, University of Minnesota, Minneapolis, MN, USA

Abstract

Cardiac Resynchronization Therapy (CRT) is an established pacing therapy for heart failure patients. The New York Heart Association (NYHA) classification is often used as a measure of a patient's response to CRT. Identifying NYHA class for heart failure patients in an electronic health record (EHR) consistently, over time, can provide better understanding of the progression of heart failure and assessment of CRT response and effectiveness. However, NYHA is rarely stored in EHR structured data such information is often documented in unstructured clinical notes. In this study, we thus investigated the use of natural language processing (NLP) methods to identify NYHA classification from clinical notes. We collected 6,174 clinical notes that were matched with hospital-specific custom NYHA class diagnosis codes. Machine-learning based methods performed similar with a rule-based method. The best machine-learning method, support vector machine with n-gram features, performed the best (93% F-measure). Further validation of the findings is required.

Disclosures

J. Munroe, L. Shanahan and S. Horn are employees of Medtronic, Inc. that manufactures and markets Cardiac Resynchronization Therapy devices. The research reported in this article was performed in collaboration with the named Medtronic, Inc. employees by R. Zhang, S. Ma and S. Speedie, faculty members at the University of Minnesota under a contract with the University of Minnesota funded by Medtronic, Inc. The University of Minnesota researchers did not receive any additional compensation for their work.

Keywords

Clinical Notes; Electronic Health Records; New York Heart Association (NYHA); Natural Language Processing

I. Introduction

Approximately 8% of the population over the age of 75 [1] is affected by Heart Failure (HF) in developed countries. An estimated 5.7 million Americans have heart failure and account for more than 1 million hospital admissions annually [2]. HF is a progressive condition associated with high morbidity and mortality rates. Cardiac Resynchronization Therapy (CRT) is an established pacing therapy for patients with heart failure and a proven adjunct therapy for systolic, symptomatic patients with intraventricular conduction abnormalities [3–6]. Despite multiple, randomized, double-blind trials confirming the safety and efficacy in treating HF patients with CRT, an estimated 25–30% of patients do not derive clinical benefits [7–9].

Effectiveness of CRT is often measured using a composite score of several clinical outcomes, including mortality, heart failure hospitalization, ejection fraction (EF) measurements, and NYHA functional classification [8–11]. As a CRT response measure, NYHA is a classification system for evaluating the severity of functional limitations resulting from a patient's heart failure condition [9].

EHR data are usually stored in both structured formats (such as diagnosis, medication) and unstructured formats (such as notes, imaging). Identifying NYHA status in an EHR system could change clinical practice in several ways. First, capturing NYHA class consistently over time may inform a better understanding of the progression of HF to assess CRT effectiveness and its benefit-risk profile versus other therapies. Second, examining the change of a patient's NYHA class from HF diagnosis over their cardiac care continuum in real-world data sources could also provide greater insights on how to improve the response rate and effectiveness of CRT. A patient's NYHA class assessment over time (from diagnosis over their treatment pathway) provides the potential to better understand why some patients benefit from CRT while others do not [7–9]. Third, if NYHA class can be reliably determined using the derived values, it may mitigate intra- and inter-physician assessment variability by providing more consistent and objective assessment of an HF patient's functional status.

NYHA classification (I–IV) is based on patient symptoms relative to activity level and as a result is most often documented in dictation/notes thereby limiting computational access. Fortunately, informatics methods, specifically natural language processing (NLP) techniques, have been developed to extract information from unstructured clinical narratives. Studies have demonstrated that unstructured notes contain more detailed information than structured data [12, 13].

Despite its relevance, to the best of our knowledge, there has been no study investigating the identification of NYHA class from clinical notes. The objective of this study is to test the feasibility of leveraging NLP techniques to extract specific outcome data (NYHA class) for patients with a CRT device from electronic health records. If feasible, deriving NYHA classification using NLP or machine learning may offer an opportunity to gain insights into predictors of CRT effectiveness more efficiently and ultimately provide better patient care.

II. Methods

A. Retrive EHR for heart failure patients

The clinical notes used in this study were retrieved from the University of Minnesota's Clinical Data Repository (CDR). IRB approval was obtained for accessing the clinical notes. In this study, we focused on a heart failure patient cohort identified using diagnosis codes associated with a heart failure. All patients in the CDR were retrieved if they had at least one diagnosis of heart failure according to relevant International Classification of Diseases, Ninth and Tenth Revisions (ICD9 and ICD10 codes). We then collected all encounters with their associated diagnoses, and clinical notes for those patients.

Select diagnosis codes with NYHA classification—We further narrowed down our patient cohort to those having encounters with NYHA information embedded in local custom diagnosis codes. We retrieved these custom code names containing NYHA information, and then developed rules to group the Fairview diagnosis codes based on the NYHA classification (I – IV) stated in the code name. For example, a Fairview custom code named "Diastolic congestive heart failure with preserved left ventricular function, NYHA class 2 (H)" indicated the NYHA class II.

Retrieve clinical notes with NYHA mentions—In addition, we used keyword and concept searching through the NLP web-based search engine, patient information extraction from researchers (PIER) for clinical notes [14]. To retrieve clinical notes with NYHA mentions, we searched for the CUIs that were obtained using the UMLS Metathesarus browser. We also searched note texts using case insensitive keywords including "New York heart association", "NYHA", "New York heart classification", "NY classification", "NY class" to avoid NYHA mentions in notes without any of the NYHA CUIs.

B. Collect reference standard

To train and evaluate the NLP algorithms for extracting NYHA class from notes, we used the date on which the notes were created as well as patient identifiers to match clinical notes to encounters with at least one custom NYHA Class diagnosis code. We assumed that within a single patient encounter the notes should mention the same NYHA classification as associated NYHA custom diagnosis code. We thus assigned the NYHA class associated with Fairview custom codes (rule developed in section B) to the clinical notes in the same encounter. The notes with the assigned NYHA were collected as a corpus for development and evaluation of our NLP methods. We collected a total of 6,174 notes with an associated NYHA classification.

C. Develop and evaluate NLP methods

We developed and compared a rule-based method with applied machine learning methods to identify NYHA classification from clinical notes. To train and evaluate the machine learning methods, we split the corpus *at a ratio of approximately 3:2.* into a training data set of 3,700 notes and a testing set of 2,474 notes.

Rule-based method—We created rules by reviewing the notes through an iterative process to cover frequently occurring patterns that appeared in the training set. We used the training set to refine and generalize the patterns to capture most information related to NYHA class. There were also co-occurring NYHA classes such as “NYHA II/III”. We identified classifications that appeared within 40 characters after an NYHA mention (such as “NYHA”, “New York heart association”, “NY classification”, *etc.*) with indicator words including single digit (1, 2, 3, and 4), and single Roman numeral (I, II, III, and IV). We also considered two digits or two Roman numeral (between range 1 to 4, inclusive), separated by single character ‘−‘ or ‘/‘. Only patterns related to NYHA 1–2, NYHA 2–3, NYHA 3–4 are allowed.

Machine-learning based methods—We pre-processed the text and generated features, which were used in the machine learning methods. The classic stopwords [15] (such as “of”, “as”) were removed from the texts. lexical variant generation (LVG) was used to normalize lexically different forms of the same term as equivalent. We only kept the terms with their frequencies within top 1%. We selected two feature sets: (1) “bag-of-words”, and (2) *n*-gram features, where *n* was from 2 to 5 words. Support Vector Machine (SVM) was applied to both feature sets extracted from the training dataset to determine NYHA classification for each encounter.

Evaluation—We applied both rule-based and machine-learning based methods to the testing dataset and generated the most likely NYHA classification for each encounter based on the features in the associated set of clinical notes. We compared the resulting NYHA classes with the NYHA Classes from the reference standard encounters. Precision, recall, and F-measure were reported for each NYHA class for each of the two features sets and three machine learning methods.

III. Results

A. NYHA subgroup

The matching process used to create the Reference Standard resulted in a collection of 6,174 clinical notes from 6,039 encounters for 1,370 patients. Table I lists the number of clinical notes for each NYHA classification as well as the numbers assigned to the training and testing sets. NYHA class II had the largest and NYHA class IV the smallest number of notes.

B. Model performance

The rule based method performed well overall (F-measure: 93.37%). The best machine learning method with *n*-gram features (overall F-measure: 93.52%) slightly outperformed the rule-based method. Machine learning algorithm with *n*-gram features significantly outperformed that with bag-of-word features. The best method was SVM with the *n*-gram feature set, whose overall F-measure reaches 93.52% (Table II).

IV. Discussion

Text analysis using NLP pattern matching and machine learning algorithms applied to those associated notes indicated that in over 95% of those encounters, there was evidence found in the clinical notes that matched the custom code NYHA Class. This is not surprising given that these custom codes are assigned by expert medical records personnel who review clinical notes to assign those codes.

The rule-based method performed well on overall NYHA classes, similar to machine-learning based methods. Performance of machine-learning algorithms with bag-of-words did not perform well due to the fact that bag-of-words does not consider the order of words like *n*-grams. With the ordering information of words, the performance generally improved about 8%.

While the machine learning models with *n*-gram features collectively perform exceptionally well within the dataset used, there are some limitations to the generalizability of this result across the larger body of clinical notes available in a typical EHR. The model will need further testing with a larger and more varied collection of clinical notes associated with a clinical encounter where those notes may not contain any specific mention of an NYHA class. Further work with these types of notes will require expert annotation to determine a gold standard of classification as opposed to the “silver” reference standard used in this work.

As was discovered in exploring this EHR data, information relating to NYHA class exists not only in clinical notes, but also in various text-based comment fields throughout the record. An extension of this work would include those fields in the analysis to explore their value in further improving the strength of this model. In addition, many notes with NYHA mentions were not associated with encounters that had NYHA custom diagnosis codes. Notes without direct NYHA mentions may also contain additional information such as symptoms (e.g., fatigue, shortness of breath) which can be used to infer NYHA class.

V. Conclusions

NYHA is an important indicator of CRT response in heart failure patients. NYHA classification is not specifically documented in structured data; much information is stored in unstructured clinical notes. We developed a rule-based method and compared it to machine learning methods to identify NYHA class from those notes. The support vector machine method with n-gram features performed the best in identifying NYHA Class, even it is comparable with rule-based method. Further validation of these results and methods is required.

Acknowledgments

The clinical data was provided by the University of Minnesota's Clinical Translational Science Institute (CTSI) Informatics Consulting Service. The authors thank John Herre, MD of Cardiology and Electrophysiology, Sentara Norfolk General Hospital, Norfolk Virginia for providing clinical expertise, Gretchen Sieger and CTSI Best Practice Informatics Consulting (BPIC) team for their data support for this study.

Research supported by Medtronic through a contract with the University of Minnesota and by the Clinical and Translational Science Institute grant from National Center for Advancing Translational Sciences of the National Institutes of Health (UL1TR000114, Blazar).

References

1. Redfield MM, Jacobsen SJ, Burnett JC, Mahoney DW, Bailey KR, Rodeheffer RJ. Burden of systolic and diastolic ventricular dysfunction in the community: appreciating the scope of the heart failure epidemic. JAMA. Jan; 2003 289(2):194–202. [PubMed: 12517230]
2. Mozaffarian D, et al. Heart Disease and Stroke Statistics-2016 Update: A Report From the American Heart Association. Circulation. Jan; 2016 133(4):e38–360. [PubMed: 26673558]
3. Cleland JG, et al. The effect of cardiac resynchronization on morbidity and mortality in heart failure. N Engl J Med. Apr; 2005 352(15):1539–49. [PubMed: 15753115]
4. Moss AJ, et al. Cardiac-resynchronization therapy for the prevention of heart-failure events. N Engl J Med. Oct; 2009 361(14):1329–38. [PubMed: 19723701]
5. Daubert JC, et al. 2012 EHRA/HRS expert consensus statement on cardiac resynchronization therapy in heart failure: implant and follow-up recommendations and management. Heart Rhythm. Sep; 2012 9(9):1524–76. [PubMed: 22939223]
6. Auricchio A, Prinzen FW. Non-responders to cardiac resynchronization therapy: the magnitude of the problem and the issues. Circ J. 2011; 75(3):521–7. [PubMed: 21325727]
7. Ghio S, et al. Long-term left ventricular reverse remodelling with cardiac resynchronization therapy: results from the CARE-HF trial. Eur J Heart Fail. May; 2009 11(5):480–8. [PubMed: 19287017]
8. Linde C, et al. Long-term impact of cardiac resynchronization therapy in mild heart failure: 5-year results from the REsynchronization reVErses Remodeling in Systolic left vEntricular dysfunction (REVERSE) study. Eur Heart J. Sep; 2013 34(33):2592–9. [PubMed: 23641006]
9. Packer M. Proposal for a new clinical end point to evaluate the efficacy of drugs and devices in the treatment of chronic heart failure. J Card Fail. Jun; 2001 7(2):176–82. [PubMed: 11420770]
10. Zannad F, et al. Clinical outcome endpoints in heart failure trials: a European Society of Cardiology Heart Failure Association consensus document. Eur J Heart Fail. Oct; 2013 15(10): 1082–94. [PubMed: 23787718]
11. Neaton JD, Gray G, Zuckerman BD, Konstam MA. Key issues in end point selection for heart failure trials: composite end points. J Card Fail. Oct; 2005 11(8):567–75. [PubMed: 16230258]
12. Ford E, Carroll JA, Smith HE, Scott D, Cassell JA. Extracting information from the text of electronic medical records to improve case detection: a systematic review. J Am Med Inform Assoc. Sep; 2016 23(5):1007–15. [PubMed: 26911811]
13. Wei WQ, Teixeira PL, Mo H, Cronin RM, Warner JL, Denny JC. Combining billing codes, clinical notes, and medications from electronic health records provides superior phenotyping performance. J Am Med Inform Assoc. Apr; 2016 23(e1):e20–7. [PubMed: 26338219]
14. McEwan R, et al. NLP-PIER: A Scalable Natural Language Processing, Indexing, and Searching Architecture for Clinical Notes. AMIA 2016 Joint Summits on Translational Informatics. 2016:150–159.
15. Stopword List Available: http://www.textfixer.com/resources/common-english-words.txt

Table I

Number of clinical notes in Training and Testing set.

NYHA Classification	Training Set	Testing Set	Total
I	843	524	1367
II	1506	996	2512
III	1045	745	1780
IV	306	209	515
Total	3700	2474	6174

Table II

Performance of Methods.

Methods	**Precision**	**Recall**	**F-Measure**
Rule-based method	94.99%	92.13%	**93.37%**
SVM, bag of words	85.34%	83.28%	84.23%
SVM, *n*-gram	93.84%	93.21%	**93.52%**