

Identifying Heart Failure Symptoms and Poor Self-Management in Home Healthcare: A Natural Language Processing Study

Sena CHAE^{a,1}, Jiyoun SONG^b, Marietta OJO^c and Maxim TOPAZ^{b,c}

^aUniversity of Iowa College of Nursing

^bColumbia University School of Nursing

^cVisiting Nurse Service of New York Center for Home Care Policy and Research

Abstract. The goal of this natural language processing (NLP) study was to identify patients in home healthcare with heart failure symptoms and poor self-management (SM). The preliminary lists of symptoms and poor SM status were identified, NLP algorithms were used to refine the lists, and NLP performance was evaluated using 2.3 million home healthcare clinical notes. The overall precision to identify patients with heart failure symptoms and poor SM status was 0.86. The feasibility of methods was demonstrated to identify patients with heart failure symptoms and poor SM documented in home healthcare notes. This study facilitates utilizing key symptom information and patients' SM status from unstructured data in electronic health records. The results of this study can be applied to better individualize symptom management to support heart failure patients' quality-of-life.

Keywords. Natural language processing, heart Failure, symptoms, self-management, health behavior, self-care

1. Introduction

Approximately 6.2 million adults had heart failure (HF) in the United States (US) in 2020 [14] and around 200,000 patients with HF are discharged from hospitals to home healthcare (HHC) every year [5]. Increasing number of patients with HF require lifelong care with optimal self-management (SM) in home linked with the concept of continuity of care [3]. Under-treatment of symptoms and poor SM is a prevalent and persistent problem among patients with HF [13]. Early recognition of uncontrolled symptoms can lead to better symptom management intervention, hence contributing to improved outcomes for patients with HF [2].

The majority of research studying HF symptoms did not pay attention to patients' symptom experiences and their SM ability documented in HHC agency's electronic health record (EHR). In addition, most existing studies of symptoms using EHRs have mainly used structured data captured in standardized assessments, whereas 80% of EHR data are unstructured, including narrative clinical notes [7]. Clinical notes contain rich

¹ Corresponding Author, Sena CHAE, University of Iowa College of Nursing, Iowa, USA.

information about individual symptom experiences, treatments received, progress of disease, etc. that might be associated with patient outcomes [4].

Although processing unstructured data is time-consuming and labor-intensive, some studies have successfully applied natural language processing (NLP) to identify patients with clinical factors [8; 9; 11]. NLP is a collection of computational algorithms that automatically processes and analyzes narrative resources to extract meaning [1]. Our group has previously developed an NLP algorithm to automatically identify HF patients with ineffective SM in the hospital and primary care setting [12], but this work was not expanded to HHC. To bridge the gap, this study aimed to create and validate an NLP algorithm that identifies HF symptoms and poor SM among HHC patients.

2. Methods

2.1. Data and Study Population

A retrospective, observational study with secondary data analysis was performed using HHC narrative notes from one of the largest HHC agencies in the Northeastern US from the years 2015 to 2017. Our study evaluated two types of HHC clinical notes: 1) visit notes and 2) care coordination notes (total $n=100,000$, average note length = 647 characters, vocabulary size= 267,005 unique words). Visit notes were generated by clinicians to describe the care provided and the patient's status during a HHC visit. Care coordination notes document communication between clinicians and other care-related activities.

2.2. NLP Approach

2.2.1. Step 1: Identifying Lists of Synonyms for HF Symptoms and Poor SM

First, the preliminary lists of symptoms and poor SM status was identified based on literature review. Specifically, 12 common symptom domains and six domains of HF SM were identified based on our previous study [12] (Table 1). This list of synonym terms for symptoms and poor SM behaviors was expanded using an open- source NLP tool called *NimbleMiner* (<https://github.com/mtopaz/NimbleMiner>) [10]. Performance of *NimbleMiner* system was previously validated for identifying clinical terms using narrative notes in EHRs to extract symptoms and important patient information [6; 8-11]. *NimbleMiner's* synonym exploration is based on the word-embedding model (Word2Vec) built from a large body of HHC clinical notes available for the study period (2.3 million clinical notes). Based on the list of symptoms and poor SM categories, *NimbleMiner* suggested synonymous terms for each symptom domain.

Vocabulary expansion was implemented by three researchers with PhDs in nursing who have expertise in symptoms and health informatics. The researchers independently reviewed, added, or deleted synonym terms. Then the lists of synonyms from three researchers was combined to generate a final synonym list. *NimbleMiner* negation module was refined for the project: negations are words indicating negated synonyms (e.g., “denies”, “not”, “ruled out”, etc.).

2.2.2. Step 2: NLP Algorithm Refinement

In this step, the NLP algorithm was refined to accurately identify symptoms and poor SM. To accomplish that, the NLP algorithm was applied on a random sample of HHC notes ($n = 5,000$) and reviewed whether symptoms and poor SM were identified correctly.

2.2.3. Step 3: NLP Algorithm's Precision Evaluation

The NLP software was applied on 100,000 additional random clinical notes. For each category of HF symptoms and poor SM, 20 clinical notes were randomly extracted where symptoms or poor SM were identified by the NLP software. Each note was reviewed by a member of the study team to evaluate whether the NLP software was correct or if there was an error in the identified instance. Precision (also known as positive predictive value) was calculated for each category and calculated overall mean precision. Then, the NLP algorithm was applied on a large set of clinical notes (2.3 million) and found that 6.87 % ($n = 67,683$) of clinical notes in HHC had documentation of HF symptoms or poor SM.

3. Results

Table 1. NLP performance for identifying patients with symptoms and poor SM.

Symptoms and poor SM category	Examples of synonym terms identified using NLP	Precision
Anorexia/decreased appetite	Eating very little, not eaten	0.913
Chest pain	Cp status, chest tightness	0.739
Confusion	Forgetfulness, disoriented	1
Cough	Still coughing, sob cough	0.928
Dizziness	Lightheadedness, vertigo	0.781
Weight loss	Losing weight, lb wt loss	0.64
Dyspnea	Fluid overload, crackles	0.804
Fatigue	Generalized weakness, poor endurance	0.964
Nausea	Feeling nausea, nauseated	0.631
Palpitation	Tachycardic, feeling palpitations	0.761
Peripheral edema	Pitting edema noted, lle edema	0.851
Weight gain	Wt gained, pound weight gain	0.863
Mean Precision for Symptoms		0.817
Unspecified non-adherence	Noncompliance, uncontrolled	1
Poor diet adherence	Not drinking enough, ate alot	0.9
Poor medication adherence	Medication nonadherence, ran out of lasix	0.952
Poor exercise physical activity adherence	Not exercise, limited exercise	0.95
Issues with other self-care activities	Does not check bs	1
Missed medical encounters	Missed appointment, not follow up	0.85
Mean Precision for Poor SM		0.942
Overall Mean Precision		0.859

The additional synonym terms for each symptom and poor SM domain beyond the preexisting vocabulary list was identified using the NLP system we developed (Table 1). Table 1 summarizes results for identifying HF symptoms and poor SM. The average precision for all symptoms and poor SM categories was 0.86 (range 0.64 - 1). Highest

precision for symptoms was for the symptom of “confusion” (1) and the lowest precision for symptom was for “nausea” (0.63) and “weight loss (0.64).” Highest precision for poor SM category was “unspecified nonadherence” (e.g., “remains noncompliant”) and issues with other self-care activities (e.g., “does not check blood pressure”) (1) and the lowest precision for poor SM was “Missed medical encounters” (e.g., “missed follow up appointment”) (0.85). Overall, the number of positively labeled notes for poor SM domain was less than the number of for symptom identification, but the precision for poor SM (0.94) was higher than precision for symptom identification (0.81).

4. Discussion

The novel data science approaches using NLP were applied to discover previously hidden symptoms and poor SM quickly and accurately within previously underexplored HHC clinical notes. Importantly, our study is the first to identify clinical notes for HF symptoms and poor SM documented in routinely collected clinical notes using NLP in HHC.

To improve the performance of NLP algorithm to accurately identify symptoms and poor SM, we expanded vocabularies of terms, added irrelevant terms and additional negation terms, and increased the negation distance. When applied on the full sample of clinical notes, only a relatively small fraction of clinical notes had vocabulary related to poor SM. These results were consistent with the previous literature in the hospital setting [12].

Possible explanation of the relatively low precision observed for some domains (e.g., nausea or weight loss), were that it might be due to 1) inaccurate text-level terms (e.g., term “lower extremities” was included in the domain of edema, but patients might have other symptom such as tremor or weakness in their “lower extremities”), 2) small size of selected sample notes for symptom domain, 3) rare records in notes in case of SM domain, 4) synonym term being relatively far from a negation term (e.g., denies headache dizziness chest pain sob “dyspnea”), or 5) negation term being not detected (e.g., tomorrowdenies “sob”). For the next step of this study, the built NLP algorithm will be applied to the clinical notes for patients with HF extracted from EHR and investigate the association between HF symptoms and poor SM and poor outcomes (e.g., ER visits and unplanned readmission).

5. Limitations

Reliable NLP symptom identification performance using HHC clinical notes was observed. However, only partial NLP performance was evaluated using precision and we plan to conduct a more complete performance evaluation using additional metrics, including recall and F-score. In addition, clinical notes extracted from one HHC agency was used, and we need to test the list of symptom vocabulary with multiple HHC institutions and different patient population.

6. Conclusion

This study showed that it is feasible to develop NLP system to find HF symptoms and poor SM documented in HHC notes. HHC providers' early assessments for HF symptoms and their SM status have the potential to improve patients' quality of life and decrease health care utilization in future study.

Acknowledgements

This study was funded by Agency for Healthcare Research and Quality (AHRQ) #R01HS027742: "Building risk models for preventable hospitalizations and emergency department visits in HHC (Homecare- CONCERN)." The content is solely the responsibility of the authors and does not necessarily represent the official views of the AHRQ.

References

- [1] Fleuren WW, Alkema W. Application of text mining in the biomedical domain. *Methods* 74 (2015), 97-106.
- [2] Grady KL, Dracup K, Kennedy G et al. Team management of patients with heart failure. *Circulation* 102 (2000), 2443-2456.
- [3] Jaarsma T, Brons M, Kraai I, Luttik ML, Stromberg A. Components of heart failure management in home care: A literature review. *European Journal of Cardiovascular Nursing* 12 (2013), 230-241.
- [4] Jensen K., Soguero-Ruiz C, Oyvind Mikalsen K et al. Analysis of free text in electronic health records for identification of cancer patient trajectories. *Scientific Reports* 7 (2017), 46226-46226.
- [5] Jones CD, Ginde AA, Burke RE, Wald HL, Masoudi FA, Boxer RS. Increasing home healthcare referrals upon discharge from US hospitals: 2001–2012. *Journal of the American Geriatrics Society* 63 (2015), 1265-1266.
- [6] Koleck TA, Tatonetti NP, Bakken S et al. Identifying symptom information in clinical notes using natural language processing. *Nurs res* (2020).
- [7] Murdoch TB, Detsky AS. The inevitable application of big data to health care. *Jama* 309 (2013), 1351-1352.
- [8] Topaz M, Lai K, Dowding D et al. Automated identification of wound information in clinical notes of patients with heart diseases: Developing and validating a natural language processing application. *International Journal of Nursing Studies* 64 (2016), 25-31.
- [9] Topaz M, Murga L, Bar-Bachar O, Cato K, Collins S. Extracting alcohol and substance abuse status from clinical notes: The added value of nursing data. *Studies in Health Technology and Informatics* 264 (2019), 1056-1060.
- [10] Topaz M, Murga L, Bar-Bachar O, McDonald M, Bowles K. NimbleMiner: An open-source nursing-sensitive natural language processing system based on word embedding. *Computers, informatics, nursing: CIN* 37 (2019), 583-590.
- [11] Topaz M, Murga L, Gaddis KM et al. Mining fall-related information in clinical notes: Comparison of rule-based and novel word embedding-based machine learning approaches. *Journal of Biomedical Informatics* 90 (2019), 103103.
- [12] Topaz M, Radhakrishnan K, Blackley S, Lei V, Lai K, Zhou L. Studying associations between heart failure self-management and rehospitalizations using natural language processing. *West J Nurs Res* 39 (2017), 147-165.
- [13] Toukhsati SR, Driscoll A, Hare DL. Patient self-management in chronic heart failure - establishing concordance between guidelines and practice. *Cardiac failure review* 1 (2015), 128-131.
- [14] Virani SS, Alonso A, Benjamin EJ et al. Heart disease and stroke statistics-2020 update: A report from the American Heart Association. *Circulation* 141 (2020), e139-e596.