Heuristic-Based Extraction and Unigram Analysis of Nursing Free Text Data Residing in Large EHR Clinical Notes

Syed Mohtashim Abbas Bokhari[1(✉)], Kriste Krstovski[2,3], Jennifer Withall[1], Rachel Lee[4], Patricia Dykes[5,6], Mai Tran[1], Kenrick Cato[7,8], and Sarah Rossetti[1,4]

[1] Department of Biomedical Informatics, Columbia University, New York, NY, USA
mohtashim_abbas@yahoo.com
[2] Data Science Institute, Columbia University, New York, NY, USA
[3] Columbia Business School, Columbia University, New York, NY, USA
[4] School of Nursing, Columbia University, New York, NY, USA
[5] Harvard Medical School, Brigham and Women's Hospital, Boston, MA, USA
[6] BWH Center for Patient Safety, Research and Practice, Boston, MA, USA
[7] University of Pennsylvania, Philadelphia, PA, USA
[8] Children's Hospital of Philadelphia, Philadelphia, PA, USA

Abstract. Free text in nurses' notes can play an important role in clinical decision-making; however, such information has not been explored to the fullest of its potential as it is hard to extract it from electronic health records (EHRs). Free text is a subset of the information recorded in nursing notes. Automated extraction of free text is challenging due to EHRs' size and structural diversity. Understanding these structural and content-level differences is essential for the extraction. Free text is embedded in other relatively structured texts, which are difficult to detect automatically. Moreover, there is no information indicating whether a note is a free text. As a first step in automating the extraction process, we explore heuristic-based algorithms with the goal of establishing a baseline and developing an annotated dataset, which could then be used for further machine learning-based extraction algorithms for a more scalable solution. In this research, we analyze over 200,000 EHR notes and extract 40,000 free text notes from them. Furthermore, we use the unigram language model to analyze the differences between free and structured texts to better understand the free text content.

Keywords: nursing documentation · health informatics · clinical notes · nursing notes · heuristics · natural language processing · information retrieval · unigram analysis

1 Introduction

Nursing documentation, including the concepts written in nursing notes, can play an integral role in healthcare prediction models to inform effective clinical decision-making [1]. Early warning scores (EWS) are one type of prediction

D. Salvi et al. (Eds.): PH 2023, LNICST 572, pp. 124–136, 2024.
https://doi.org/10.1007/978-3-031-59717-6_9

model implemented as clinical decision support tools in the inpatient setting to identify patients at risk of deterioration, including from events such as cardiac arrest and sepsis which impact approximately 330,000 inpatients per year [2,3]. Early identification of patient deterioration can allow for faster treatment and escalation of care to prevent harmful outcomes, such as inpatient mortality. EWS have had limited impact on clinical outcomes likely due to their primary reliance on vital signs, a late indicator of patient deterioration. When nurses are concerned about the potential for patient deterioration they increase surveillance of the patient and their respective nursing notes documentation in electronic health records (EHRs) [4–7]. Our team has developed an EWS named CONCERN (COmmunicating Narrative Concerns Entered by Registered Nurses) that leverages nursing surveillance and documentation patterns that reflect how nurses observe and monitor subtle changes in patients before deterioration is noted in their physiological conditions' parameters [1]. CONCERN is currently in production at 2 academic medical centers with implementation in progress at 2 more health systems [1].

The data from nursing documentation are large in volume and are structured, semi-structured, and time-varying. The large templated documentation from nursing notes also contains free text data written by nurses. These free text data can be useful as features in EWSs to predict patient health deterioration [1]. These free text data will act as an important feature in our CONCERN EWS [1]. Leveraging free text data can be challenging because of their large volume and clinical diversity [8,24,25]. Nursing EHR data are time-varying, semi-structured, and variable on a content level, which make the identification of the free text portion of notes a cumbersome task.

Nursing notes include: 1) templated documentation, which are structured data entered by nurses elsewhere in the chart, and 2) narrative (free text) information written by nurses in their own words. The free text may represent nurses' concerns about patients and can be useful in predictive modeling [1]. However, to leverage information from the free text documentation by nurses we first need to be able to identify where this free text resides within semi-structured nurses' notes and how to retrieve it. Often the narrative free text can be found embedded in other relatively structured texts, which is difficult to detect. Such data are not explicitly labeled as free text and can often be found intertwined within relatively structured texts, thereby making the detection difficult. The absence of clear distinctions between documents' information such as document headers and metadata, further adds up to the problem. This ultimately poses challenges in the automatic extraction of the free text data, which may contain important signals for improved clinical decision-making [1,9].

This research study is focused on HTML-based nursing notes from an academic medical center in the Northeastern United States. The dataset contains more than 200K notes with all free text (with no structured data in it), structured data with no free text, and free text embedded in structured data. Our study aimed to 1) identify and retrieve all narrative (free text) notes data, and 2) distinguish and retrieve the free text embedded in the nursing notes. In this

regard, this research uses a heuristic-based approach to extract free text data and utilizes unigram analysis to gain deeper insights into the nursing free text. Unigrams are the elementary subset of the n-gram language models, which is a subfield in natural language processing [10,11]. Based on our prior work we know that nursing free text has signals of nurses 'concerns about a patient. Such concerns can help detect patient health deterioration early even before the vital signs start to appear [1].

While existing research [12–16] has applied machine learning and NLP algorithms directly to free text datasets, and there has been an attempt [17] to recognize tables within free text data, our research uniquely focuses on first establishing the ground truth regarding the location and nature of free text to build a training dataset. This training dataset can be used in the future by machine learning algorithms to identify the free texts dynamically independent of the heuristic-based approach, which is extremely important for the scalability of the system given the potential variation in syntax across different sites. Since we see the problem as a classification task; a training set is required, which aims to establish the foundation to use machine learning approaches for predictive modeling in the future. Moreover, the fraction of the free text in the structured portions can be less than 1%, in addition, there is no metadata to differentiate. Without an annotated dataset, machines may struggle to distinguish the relevant portions. Since the relevant free text portions are so small and as an embedded part (free text) in the structured text, all look the same. Therefore, our heuristic approach is useful for creating a training set for the supervised learning approaches to make the solution heuristic-independent in the future for scalability purposes. Moreover, this HTML format is coming from the Epic EHR which is a widely used system in many hospitals within the US which also makes our heuristic-based approach potentially generalizable across hospitals that use the Epic EHR. Furthermore, unigram analysis of both structured and free text data in this research gives us more insights into the difference in the nursing free text compared to the structured data.

This research is foundational to developing an automated framework for identifying free text containing nurses' concerns from nursing notes through machine learning approaches in the future. In addition, our framework also needs to be a scalable component of the CONCERN EWS that is already being spread to multiple sites.

2 Methods

2.1 Description of Data

In our study, we used more than 200K nursing clinical notes that originated from the Epic©EHR system. Epic is one of the most widely used EHRs in the United States. The notes were retrieved using the Fast Healthcare Interoperability Resources' standard (FHIR) document service. FHIR is a set of rules and specifications for exchanging electronic healthcare data. The notes data was stored in SQL in base64 [18] encoded format, which was decoded into HTML

notes files. HTML notes files contain both free texts, as well as structured data. Figure 1 shows different stages of our dataset: notes SQL data in base64 [18] encoded format, decoded HTML notes files, retrieved text from HTML notes, and HTML text transformed into JSON documents.

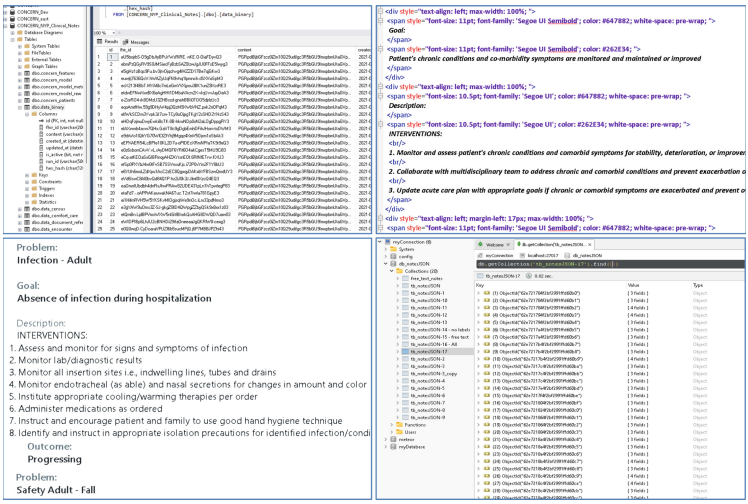

Fig. 1. Example notes originated from the Epic© EHR system: Encoded SQL notes data (top-left), HTML notes file (top-right), notes text (bottom-left), and notes JSON documents (bottom-right).

We distribute our dataset into two parts: 1) sections with structured text and 2) free text. The task is to differentiate and extract the free text. And it is important to be aware of the structural and content-level differences in the data to identify and detach free text from the structured portion. Structural differences refer to the way specific data are stored, which is important for retrieval of relevant data while the content level differences are important to uniquely identify relevant texts, which is crucial for the dynamicity of the solution.

2.2 Segregate and Retrieve the Narrative Embedded in the Nursing Notes

We started by traversing through the HTML nursing notes. We found significant variability in the formats, some of the examples are shown in Fig. 2. The figure shows different layouts of the nursing clinical notes, including plain text, different tabular and other formats, demonstrating a high level of variability. However, we observed that the independent divs in HTML notes files were primarily the place where the narrative data were stored. The 'div' tag defines a division or a section in an HTML document. In automating the extraction process, it is important to determine: 1) which div contained the relevant information (free texts), 2) how

to locate the relevant divs, 3) how to differentiate the relevant divs from the other div information, and 4) where to cut and extract the information given that no nested divs exist to indicate the ideal spot to cut. Again, there were no explicit labels to differentiate parts of the notes such as document header, div name, or any other metadata, as well as to differentiate between different divs and analyze div text accordingly. Importantly, there can be hundreds of lines in a note while there may be only a few free text words present in the note.

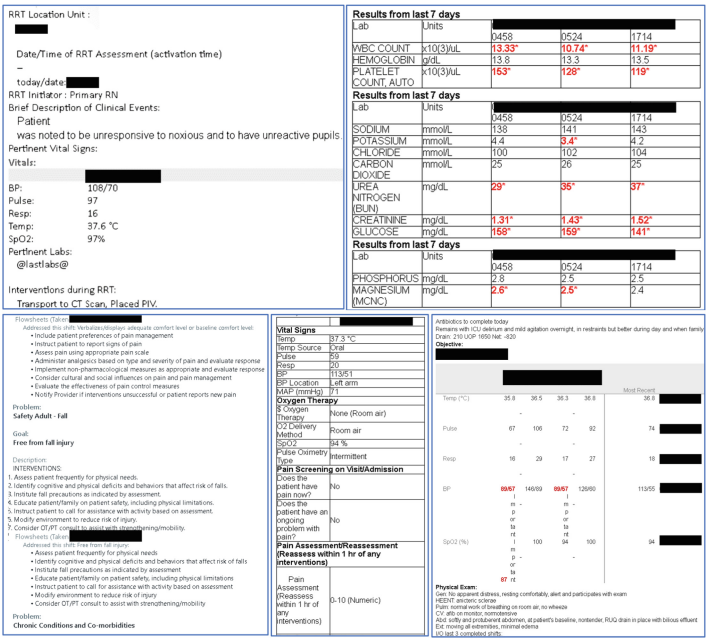

Fig. 2. Example formats of nursing notes originated from the Epic© EHR system

The approach we used was to manually review thousands of files to develop a heuristic-based algorithm, based on the identified static (prespecified) rules to retrieve the relevant portions of the data from HTML tags. For instance, we observed that if certain indicators such as tokens, formats, and headings, exist in certain locations, the data are likely to be free text. Also, it is important to determine which are the relevant and also the irrelevant tokens since token names may overlap between free text and structured portions of a note (see Fig. 3). In this case, we take into consideration other indicators to determine the relevant information, such as the location of the token in the document. The ultimate goal of this approach is to help build a free text dataset that can be used to identify such narrative texts automatically independent of the syntactical differences, which, as aforementioned, is important for the scalability of the system.

Therefore, we traverse through all the files and their respective divs and select only those divs which are relevant, i.e., div text contains specific free text

tokens, e.g., 'Assessments/Comments', 'Additional Comments', 'Comments', 'Other Comments', 'Comment', 'Nursing Note', 'Progress Note', 'Treatment Note', and 'Note'. Algorithm 1 retrieves the text from the selected divs of the HTML and removes leading and trailing spaces to check if the first few words contain a heading, i.e., a title containing a colon. Having traversed through all the divs in a document, only relevant divs are selected based on the aforementioned free text tokens.

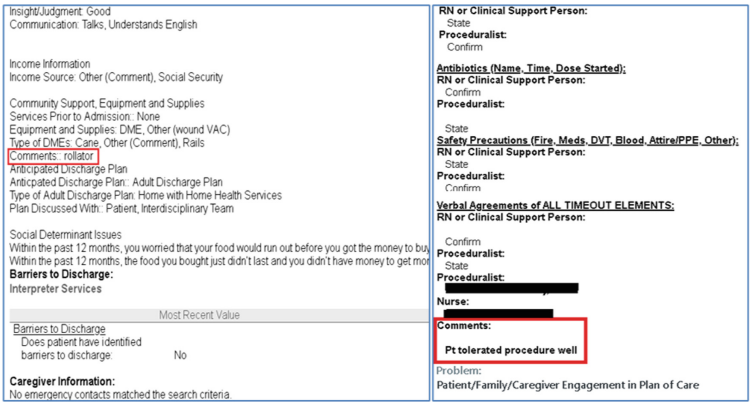

Fig. 3. Example of overlapping token: structured text (left) and free text (right)

If a relevant (containing tokens) heading is found in a div (in certain locations), the algorithm extracts the surrounding text of other divs as free text, otherwise ignores it. If the heading contains specific tokens such as 'comment' and 'comments' then it checks the location of the div within the HTML since not all comments are free text, but the comments in the later part of the documents are likely to be free text. We identified the free text nursing notes from the respective divs based on such identified static rules to build our approximate free text dataset. In the process, we ignore the divs containing the plan of care or discharge notes. The plan of care notes primarily contain structured text reflecting future plans, as opposed to immediate concerns about a patient's state. Discharge notes are documented at the end of patients' hospital stay and therefore would not be available to our algorithm because we are interested in predicting deterioration during a patient's hospital stay in real-time.

2.3 Identify and Retrieve All Narrative Notes Data

Upon examination, we observed that if no heading exists in a document, the structured text is unlikely to be present, rather, the document content is likely

Algorithm 1:

```
check_label (selected_div, div_count)
    boolean = False
    div_txt = selected_div.text.strip() # Remove leading and trailing empty spaces from div text

    # No 'plan of care' or discharge note'
    if (!(div_txt.lower().contains("plan of care" or "discharge note"))):
        words = div_txt.split()
        for w in words[:5] # Check the first five words to check if they are heading/title
            if ((w.istitle() or w.isuper()) and (w.endswith(":"))):
                if (!w.equals("Comment" or "Comments")):
                    boolean = True
                    break
                else:
                    # Only consider comments from the later divs of the document
                    if ((selected_div.index()/ div_count)>0.95):
                        boolean = True
                        break
    return boolean
```

to be all free text. We depict this in Fig. 4. Algorithm 2 detects all free text by checking if there are no headings in the HTML file. Overall, the algorithm works in this way if there exist relevant tokens, headings, and other indicators, then the document is a mix of structured and free text. If no tokens are present, then the document is likely all free text with no structured portions in it and we annotate it as all free text notes accordingly. If the relevant indicators (tokens) exist in a document and the pre-specified (static) rules are met, the algorithm identifies and extracts free text from relevant locations and what is left behind is merely the structured portion.

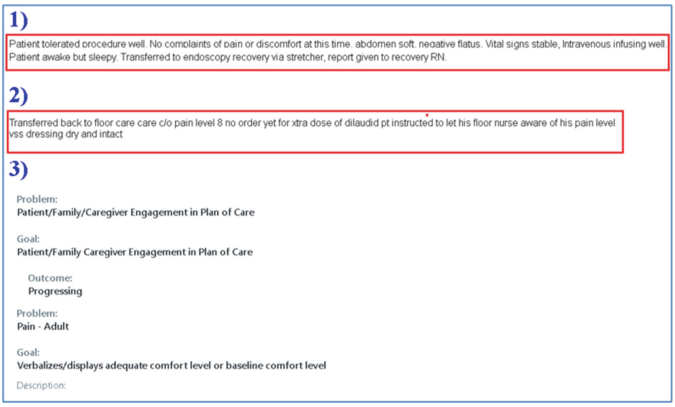

Fig. 4. Free-text examples

In this way, we extracted the relevant information and then categorized the identified free texts into different categories based on the identified tokens. For efficient analysis, the extracted information was stored in a JSON. In addition, we aimed to understand the unique characteristics of free text compared to structured text in order to inform the creation of an automated dynamic system

Algorithm 2:

```
1 func_check_no_label (all_divs):

2       check_no_label = True
3       for div in all_divs:
4               if (div not empty):
5                       if (check_label(div) = True):
6                               check_no_label = False
7                               break

8   return check_no_label
```

that identifies free text. To do so, we conducted a thematic analysis using the unigram language model [10] to identify and compare the recurring domains with the aim to understand the clinical context in which the notes were likely written. Two registered nurses (RL, JW) who have training in informatics research and clinical experience, served as the subject matter experts and individually interpreted the unigram results in Table 1 and 2 to gain more insight about the difference between the contents of free text and structured data. They then met with the primary author (SMAB) to iteratively discuss and reach a consensus on the interpretation of the results related to clinical context and nurse documentation workflow.

3 Results

In our analysis of over 200K documents, we retrieved (based on the pre-specified rules in the algorithm) 40K free text notes in total, out of which 33K were identified as all free text records and 7K free text records found embedded in structured data. We detached free text from the structured portion through our aforementioned heuristics-based approach. A large portion (160K) of the notes consists of only the structured data while the percentage of narrative free text in a note was found to be 1–3%. We found high levels of redundancy in the structured portion of the note as compared to the narrative portion. The same words/blocks of the structured portion of the note are repeated several times. The contents of the structured and free text differed sufficiently; we detected 15K unique words in the narrative text that are not present in the structured portion of the text and 7K unique words in the structured text that are not present in the narrative portion of the text. There were 14.5K overlapping words found.

Figure 5 shows the word clouds for the free text and structured portion indicating the difference between the two where the size of each word indicates its frequency. Table 1 shows the top 20 most frequently occurring free-text terms exclusive to narrative free text while Table 2 shows the top 20 most frequently found terms unique to structured data. Tables show the frequency each of these words appears across the entire dataset (word frequency) and the number of documents in which each of these terms appears (document frequency). Again, the free text is written narratively by registered nurses while the structured

portions are generated from structured (templated) data entered by nurses elsewhere in the EHR. Therefore, our approach does not convert these words into their respective root words in order to preserve any grammatical structures that exist and may be helpful in the future to detect such text automatically using machine learning approaches.

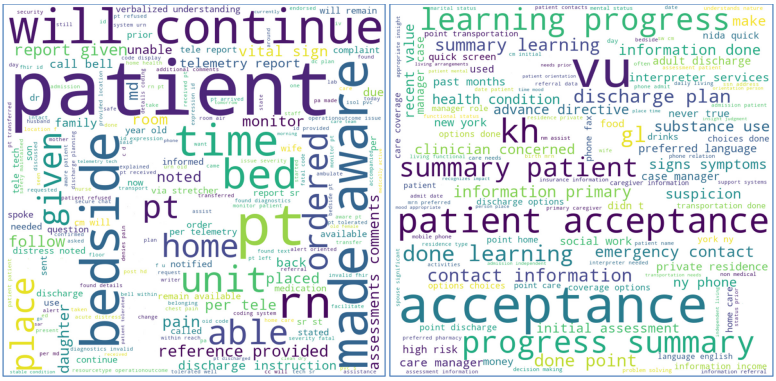

Fig. 5. Free text (left) and structured data (right) word clouds.

The free text unigrams in Table 1 revealed two primary clinical areas: 1) wound ostomy continence (WOC) care [associated terms in yellow] and 2) cardiac patient care [associated terms in cyan]. Interestingly, despite the existence of specific structured fields in the EHRs to capture concepts aligned with wound ostomy continence and cardiac care, we observed unique terms associated with these domains captured only in the free text. There may be several reasons that nurses choose to document this information in free text, ranging from the usability or data granularity issues with the structured forms to a preference for narrative forms of documentation to describe particular clinical phenomena. Future research should seek to explain the characteristics of terms found only in narrative text versus those found only in structured data.

As for the terms exclusive to structured text in Table 2, the most frequently appearing word was 'VU' which is short for "verbalized understanding". The documentation of patient verbalized understanding in structured fields indicates the nurses' evaluation of the level of comprehension regarding the patient's education. Besides, these unigram results for the structured portion comprise the terms primarily found in two nursing documentation templates, 1) care management assessment [associated terms in green], 2) substance use history screening [associated terms in pink] while to a lesser extent, the others belong to templates for nursing care plans and the associated action(s) to achieve the stated care planning goals [associated term in grey] and time-out documentation for the operating room [associated term in peach]. This may infer that these specific templates are more frequently used than others; further research should explore the portion of structured documentation that are focused on these aspects of patient care.

Table 1. Unigram analysis of the terms exclusive to narrative-free text data

Rank	Free Text Word	Word Freq.	Document Freq.
1	isol - [isolation]	599	598
2	flange	510	171
3	hollister	315	219
4	couplets	293	284
5	kpouch	279	114
6	peristomal	269	228
7	midabdominal	255	187
8	pacs [premature atrial contractions]	241	231
9	drainable	202	167
10	endo	197	177
11	budded	195	190
12	ceraring	180	159
13	urinal	171	152
14	padded	157	153
15	mf [multiform]	154	144
16	convexity	151	137
17	incont [incontinence]	149	124
18	sterility	138	138
19	phenylephrine	138	108
20	apcs [atrial premature complexes]	138	127

Also, we noticed that some aspects of the templates are not always relevant or useful in all patient cases. For instance, a prevalent term found was the word "element", which often appeared as part of a templated structured field as N/A, thereby indicating that the specific data element was not applicable. The frequent occurrences of the terms such as "element" being "N/A" in the documentation suggest that such information is continuously being recorded, even when a specific data element does not apply to the patient's situation. The need to document each aspect of patient care, even when certain data elements are not applicable, may contribute to the documentation burden specifically related to reviewing and synthesizing data, as well as "note bloat" [19,20]. This may impact the workload of clinicians, thereby affecting the time spent on direct patient care [21–23]. Furthermore, understanding the rationale of nurses regarding their decision to document certain aspects of clinical care in narrative free-text notes rather than structured flowsheet fields, could also be an area of future research. Use of our heuristic approach to detect and leverage concerning clinical concepts documented in narrative nursing notes, and subsequently incorporating this as a feature into the predictive model can help improve clinical deterioration prediction.

Table 2. Unigram analysis of the terms exclusive to structured data

Rank	Structured Text Word	Word Freq.	Document Freq.
1	vu [verbalized understanding]	192822	758
2	discipline	4632	772
3	latino	3972	1324
4	solving	3747	1100
5	element	3474	526
6	grass	3357	767
7	implement	2946	245
8	opium	2457	763
9	hydrocodone	2451	753
10	mushrooms	2394	758
11	hallucinogens	2394	758
12	ecstasy	2394	758
13	stimulants	2373	759
14	sedatives	2370	758
15	dexedrine	2361	759
16	concerta	2361	759
17	ritalin	2361	759
18	ghb	2358	758
19	serepax	2358	758
20	introductions	2352	778

4 Discussion

Literature suggests that when nurses optionally decide to write free text the contents may be a strong signal for information that the nurse wants to communicate to the rest of the healthcare team [1,9]. In this regard, this research analyzed over 200K EHR notes and extracted 40,000 free text notes from them. The problem is that such free text is often found embedded in large datasets, which are hard to retrieve given a lack of clear distinctions between the data. Furthermore, it was challenging to extract such data because of their structural diversities.

This paper describes a heuristic-based extraction and unigram analysis approach to identify as well as understand free text residing in larger EHR nursing notes. We analyzed the data by identifying the unigrams unique to free text data to determine the difference between the two datasets (structured and free text documentations). Because if there were no major differences between the two texts then it would be harder to detect such texts dynamically as both could be labeled essentially the same. Our research found the difference between free text and structured data is statistically significant; there are many clinical terms

that were only recorded in the free text by nurses. The choice of nurses to exclusively document this information in the free text could be attributed to several hypotheses. It could potentially result from usability concerns or limitations with the granularity of data accommodated by structured forms. Alternatively, it may reflect a preference for narrative documentation when conveying specific clinical phenomena. Further research is needed to understand the characteristics and implications of terms present in either free text or structured data from nurses' notes. Typically, free text notes give a summarized and up-to-date picture of a patient's current state. Such free text data may be used in EWS to predict health deterioration early before changes in vital signs appear [1].

To the best of our knowledge, this is a unique contribution to the NLP literature, namely, to extract free text from the primary formats of nursing documentation (structure, semi-structured, free text) and subsequently use unigram analyses to attain deeper insights into the free text. The HTML format notes used for this research are coming from Epic which itself is a widely used system in US hospitals, implying a common HTML format. We understand the limitations of the heuristic-based approaches though; however, we see the problem as a text classification problem, which relies on annotated datasets for training purposes. Our heuristic-based approach helped annotate the data to train ML algorithms in the future for a more scalable solution CONCERN EWS system at other hospitals.

Acknowledgments. This work is supported by the National Institute for Nursing Research (NINR) CONCERN Study #1R01NR016941 and the American Nurses Foundation Reimagining Nursing Initiative (RN Initiative). JW is a postdoctoral research fellow supported by the Reducing Health Disparities through Informatics training grant (T32NR007969).

References

1. Rossetti, S.C., Knaplund, C., Albers, D., et al.: Healthcare process modeling to phenotype clinician behaviors for exploiting the signal gain of clinical expertise (HPM-ExpertSignals): development and evaluation of a conceptual framework. J. Am. Med. Inform. Assoc. **28**, 1242–51 (2021)
2. Merchant, R.M., Yang, L., Becker, L.B., et al.: Incidence of treated cardiac arrest in hospitalized patients in the United States. Crit. Care Med. **39**(11), 2401–2406 (2011)
3. Liu, V., Escobar, G.J., Greene, J.D., et al.: Hospital deaths in patients with sepsis from 2 independent cohorts. JAMA **312**, 90–2 (2014)
4. Collins, S.A., Vawdrey, D.K.: Reading between the lines of flowsheet data: nurses optional documentation associated with cardiac arrest outcomes. Appl. Nurs. Res.: ANR **25**(4), 251 (2012)
5. Collins, S.A., Fred, M., Wilcox, L., et al.: Workarounds used by nurses to overcome design constraints of electronic health records. In: NI 2012: 11th International Congress on Nursing Informatics, June 23-27, 2012, Montreal, Canada, vol. 2012. American Medical Informatics Association (2012)
6. Collins, S.A., Bakken, S., Vawdrey, D.K., et al.: Agreement between common goals discussed and documented in the ICU. J. Am. Med. Inform. Assoc. **18**, 45–50 (2011)
7. Collins, S., Hurley, A.C., Chang, F.Y., et al.: Content and functional specifications for a standards-based multidisciplinary rounding tool to maintain continuity across acute and critical care. J. Am. Med. Inform. Assoc. **21**, 438–47 (2014)

8. Kang, M.J., Rossetti, S.C., Knaplund, C., et al.: Nursing documentation variation across different medical facilities within an integrated health care system. Comput. Inf. Nurs. **39**, 845 (2021)
9. Rossetti, S.C., Dykes, P.C., Knaplund, C., et al.: The communicating narrative concerns entered by registered nurses (CONCERN) clinical decision support early warning system: protocol for a cluster randomized pragmatic clinical trial. JMIR Res. Protoc. **10**, e30238 (2021)
10. Xu, W., Xu, D., Alatawi, A., et al.: Statistical unigram analysis for source code repository. Int. J. Semant. Comput. **12**, 237–60 (2018)
11. Martin, J.H.: Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Pearson/ Prentice Hall, New Jersey (2009)
12. Afzal, Z., Schuemie, M.J., van Blijderveen, J.C., et al.: Improving sensitivity of machine learning methods for automated case identification from free-text electronic medical records. BMC Med. Inform. Decis. Mak. **13**, 1–11 (2013)
13. Zuccon, G., Wagholikar, A.S., Nguyen, A.N., et al.: Automatic classification of free-text radiology reports to identify limb fractures using machine learning and the SNOMED CT ontology. AMIA Summits Transl. Sci. Proc. **2013**, 300 (2013)
14. Wrenn, J.O., Stetson, P.D., Johnson, S.B.: An unsupervised machine learning approach to segmentation of clinician-entered free text. In: AMIA Annual Symposium Proceedings. vol. 2007, p. 811. American Medical Informatics Association (2007)
15. Koleck, T.A., Dreisbach, C., Bourne, P.E., et al.: Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. J. Am. Med. Inform. Assoc. **26**, 364–79 (2019)
16. Jensen, K., Soguero-Ruiz, C., Oyvind Mikalsen, K., et al.: Analysis of free text in electronic health records for identification of cancer patient trajectories. Sci. Rep. **7**, 46226 (2017)
17. Ng, H.T., Lim, C.Y., Koo, J.L.T.: Learning to recognize tables in free text. In: Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics, 443–450 (1999)
18. Josefsson, S.: The base16, base32, and base64 data encodings. Tech. rep. (2006)
19. Moy, A.J., Schwartz, J.M., Chen, R., et al.: Measurement of clinical documentation burden among physicians and nurses using electronic health records: a scoping review. J. Am. Med. Inform. Assoc. **28**, 998–1008 (2021)
20. Bakken, S., Dykes, P.C., Rossetti, S.C., et al.: Patient-Centered Care Systems, pp. 575–612. Computer Applications in Health Care and Biomedicine. Springer, Biomedical Informatics (2021)
21. Tran, B., Lenhart, A., Ross, R., et al.: Burnout and EHR use among academic primary care physicians with varied clinical workloads. AMIA Summits Transl. Sci. Proc. **2019**, 136 (2019)
22. Gregório, J., Cavaco, A.M., Lapao, L.V.: How to best manage time interaction with patients? Community pharmacist workload and service provision analysis. Res. Soc. Adm. Pharm. **13**(1), 133-47 (2017)
23. Morris, R., MacNeela, P., Scott, A., et al.: Reconsidering the conceptualization of nursing workload: literature review. J. Adv. Nurs. **57**, 463–71 (2007)
24. Bokhari, S.M.A., Basharat, I., Khan, S.A., Qureshi, A.W., Ahmed, B.: A framework for clustering dental patients' records using unsupervised learning techniques. In: 2015 Science and Information Conference (SAI), pp. 386–394. IEEE (2015)
25. Bokhari, S.M.A., Khan, S.A.: Applying supervised and unsupervised learning techniques on dental patients' records. In: Emerging Trends and Advanced Technologies for Computational Intelligence: Extended and Selected Results from the Science and Information Conference 2015, pp. 83–102. Springer (2016)