# Making sense of abbreviations in nursing notes: A case study on mortality prediction

**Jasmine Y. Nakayama, BSN[1], Vicki Hertzberg, PhD[1,2], Joyce C. Ho, PhD[2]**
**[1]Nell Hodgson Woodruff School of Nursing, [2]Department of Computer Science, Emory University, Atlanta, GA**

## Abstract

*Unstructured data from electronic health records hold potential for improving predictive models for health outcomes. Efforts to extract structured information from the unstructured data used text mining methodologies, such as topic modeling and sentiment analysis. However, such methods do not account for abbreviations. Nursing notes have valuable information about nurses' assessments and interventions, and the abbreviation use is common. Thus, abbreviation disambiguation may add more insight when using unstructured text for predictive modeling. We present a new process to extract structured information from nursing notes through abbreviation normalization, lemmatization, and stop word removal. Our study found that abbreviation disambiguation in nursing notes for subsequent topic modeling and sentiment analysis improved prediction of in-hospital and 30-day mortality while controlling for comorbidity.*

## Introduction

Since the Health Information Technology for Economic and Clinical Health Act passed in 2009, health care systems have increasingly implemented electronic health record (EHR) systems to improve communication and coordination among health care teams[1]. Additional insight about providers and recipients of health care can be gained from the large amount of data collected in EHRs[1,2]. Mining such data using machine learning techniques has the potential to provide early notification of adverse patient events[3], and promising results in predicting hospital readmission, personalized disease risk, and mortality have been reported on both publicly available and proprietary clinical datasets[2].

However, such predictive methods primarily rely on structured EHR data, such as demographic information, procedure codes, and administered medications[4]. Unstructured clinical text, a substantial portion of the EHR data, remains relatively untapped, though it often contains important information, such as patients' clinical conditions, plans of care, and social considerations[1]. Some researchers have predicted structured medical codes using different types of unstructured clinical data[5–7]. Other existing works have focused on concept detection and normalization of ontology[8,9]. Yet, these works assume the existence of structured and well-known medical concepts, which is not always true. In particular, nursing progress notes may contain especially meaningful information, as nurses spend significant time with patients and families during health care encounters, perform frequent surveillance, and coordinate care among the interdisciplinary team[10–12]. These nursing notes may offer valuable information about patients beyond what is captured in the structured data and formalized medical concepts[12].

Topic modeling and sentiment analysis are popular text mining methodologies used to extract structured information from clinical notes without necessitating labor-intensive annotations from domain experts[13,14]. In topic modeling, common topics in the corpus are learned, as words that appear together tend to describe similar concepts[13]. This method has been used in predicting health outcomes, such as complications for premature infants[15] and mortality for adults requiring critical care[16–18]. Sentiment analysis is used to determine the emotional expression of words and corpora[19,20]. Studies have found that sentiments measured in clinical notes were associated with mortality[21–23]. However, previous works fail to account for abbreviations during sentiment analysis or topic modeling.

Abbreviations and acronyms are pervasive in clinical text, especially nursing notes[24,25], with the shortened forms often having multiple senses (i.e., meanings) depending on the context and the author[26–30]. As these abbreviations represent some of the most commonly used concepts in health care, word-sense disambiguation adds meaning and accuracy in clinical text analysis[31,32]. In addition, lexicons for sentiment analysis typically do not account for abbreviations, especially those used in health care, as they were developed in other settings (e.g., social media use)[33]. Thus, the true sentiment may not be captured using existing sentiment analyzers. Despite the potential for disambiguation to provide insight, this preprocessing step is rarely done for unstructured notes in risk prediction systems. This may be due to the fact that current state-of-the-art clinical text normalization tools, able to detect and disambiguate shortened

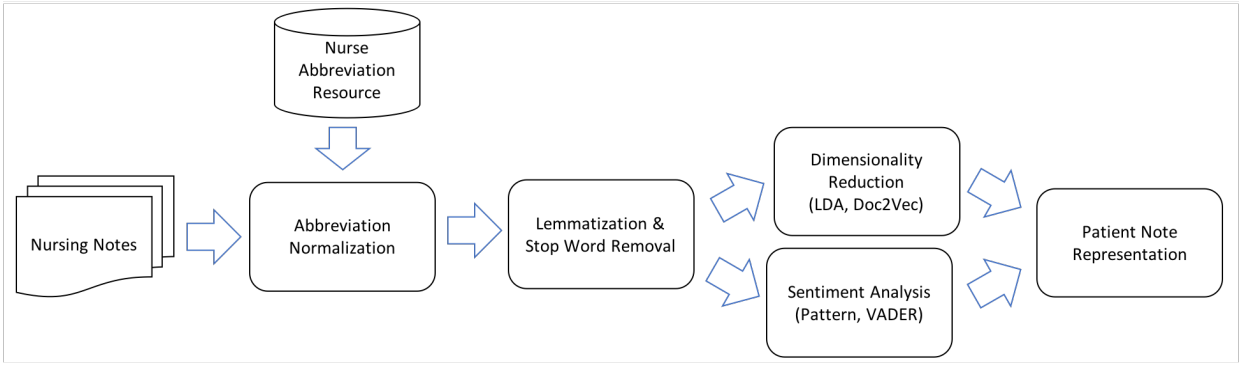


**Figure 1:** An overview of our process to extract structured information from nursing notes.

forms[24,31,32,34], require expert supervision or proprietary software. Utilizing open-source resources for normalizing abbreviations may assist in extracting meaning from clinical text without requiring extensive resources.

We present a new process to extract structured information from nursing notes. Specifically, we propose a simple nursing abbreviation resource that utilizes publicly available resources to disambiguate abbreviations for unstructured notes. We compare our resource to the clinical abbreviation recognition and disambiguation framework, an open-source resource[32]. Our software process includes two additional steps to reduce vocabulary size by removing common words and inflectional forms of words to improve predictive performance. We also introduce the use of an additional sentiment analyzer developed for social media to extract useful patient features. This study uses a novel preprocessing pipeline and shows the value of nursing notes in predicting the outcomes of in-hospital mortality and 30-day mortality after disambiguating common abbreviations used in health care with a simple nursing abbreviation resource in conjunction with topic modeling and sentiment analysis. For reproducibility, our code is published on Github[a].

## Methods

We developed a pipeline that performed simple disambiguation of abbreviations, applied standard preprocessing techniques common in natural language processing, and then utilized dimensionality reduction and sentiment analysis to construct useful features from clinical notes. Figure 1 illustrates the process of extracting structured information from nursing notes through those steps. We briefly describe our data before discussing each step in the pipeline.

*Data Extraction*. This study was a secondary analysis of patient and nursing note data extracted from a database of EHR data for a random sample of 107,433 patients who received care from a health care system in southeastern United States during 2012-2018. Any protected health information was masked prior to data extraction. Patients' International Classification of Diseases-Ninth Revision diagnoses were extracted and used to measure patient comorbidity with the recently enhanced Elixhauser Comorbidity Index[35].

Reflective of nurses' assessments and interventions, free-text nursing progress notes were extracted from the database. Notes were discarded if they did not contain any relevant information. For example, a note was discarded if it only contained "In Error" or "Date Time Correction." Patients without nursing progress notes were excluded from this study, thereby reducing the potential cohort to 4,618 patients. We also required that each patient contained at least one ICD-9 code (to compute the Elixhauser Comorbidity Index), which further reduced our cohort to 3,036 patients.

In-hospital mortality outcomes were defined by discharge dispositions of "expired" for health care encounters (e.g., inpatient admissions and ambulatory surgeries). The 30-day mortality outcomes required additional calculation. While some patients had recorded deaths, patients with unknown deaths were right-censored (i.e., they might be alive or dead). Therefore, we required the presence of a follow-up visit (i.e., an inpatient or outpatient encounter following the index inpatient encounter) within 30 days to determine an alive status for 30-day mortality outcomes. Our sample had 80 deaths among 3,036 patients for predicting in-hospital mortality and 124 deaths among 1,230 patients for predicting 30-day mortality (see Table 1).

[a] https://github.com/joyceho/abbr-norm

**Table 1:** Summary statistics for the two mortality outcomes. For the number of words and unique words, the statistics are the mean and the standard deviation for each patient.

| Outcome | # Deaths | # Patients | # Words | # Unique Words |
|---|---|---|---|---|
| 30-day | 124 | 1230 | $52 \pm 84$ | $35 \pm 44$ |
| In-hospital | 80 | 3036 | $44 \pm 72$ | $31 \pm 38$ |

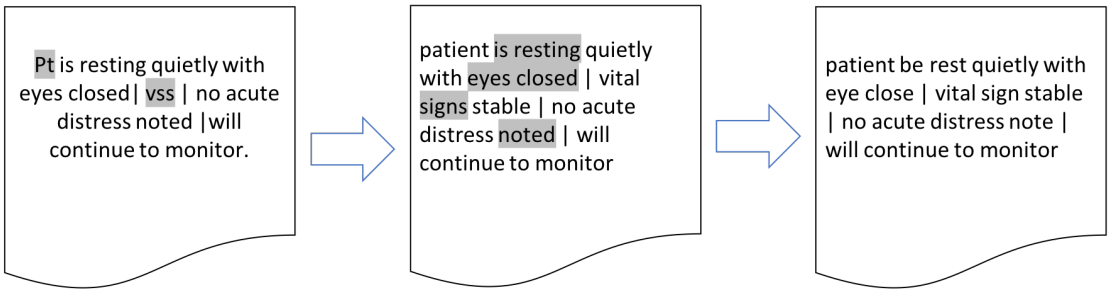


**Figure 2:** An example of the abbreviation normalization and lemmatization process. The left-most note is the original note, the middle note is after the abbreviation normalization process, and the right-most note is after lemmatization. The gray highlighted text are detected abbreviations and identified inflectional forms of base words.

*Abbreviation Normalization*. To construct a simple abbreviation normalization module that required minimal expert supervision, we leveraged online resources. We scraped nursing abbreviations from Tabers Medical Dictionary[b] and Nurselabs[c] by using Scrapy 1.5, a Python application framework that crawls websites and extracts structured data. To reduce ambiguity, only abbreviations with single senses were collected into our nursing abbreviation resource. Using the compiled resource, our abbreviation normalization module first tokenized the free-text to single words before replacing any occurrences of detected abbreviations with the long-form. Additionally, we compared the abbreviation detection results of our nursing abbreviation resource with those of a readily available framework[d].

*Lemmatization and Stop Word Removal*. As shown in Figure 2, two additional preprocessing steps were performed on the abbreviation normalized text to (1) reduce inflectional forms of the words (e.g., "takes", "took", and "take" all became the base word "take") and (2) remove common words (i.e., stop words). We used WordNet's morphy function[e] (implemented in TextBlob) to obtain the lemma for words tagged as verbs or nouns. This process accounted for plurality and verb tense and reduced the vocabulary size. Common words were also removed using the stop word list in the Natural Language Toolkit (NLTK), a leading Python library for working with text data. Although Onix is the most widely used stop word list, NLTK's stop word list can provide better context[36].

Table 2 summarizes the results of the three preprocessing steps: abbreviation detection and normalization using the scraped nursing abbreviation resource, lemmatization via TextBlob to reduce inflectional forms of the words, and Stop word removal to eliminate common words that will appear in many notes.

**Table 2:** Impact of our preprocessing steps on corpus size (i.e., number of words).

| Outcome | Original | Abbreviation Normalization | Lemmatization | Stop Words |
|---|---|---|---|---|
| 30-day | 4909 | 4976 | 4306 | 4208 |
| In-hospital | 7178 | 7251 | 6333 | 6227 |

*Dimensionality Reduction*. Topic modeling is a popular machine learning technique to structure information from clinical notes[15–18]. Latent Dirichlet Allocation (LDA)[37] is the *de facto* standard for generating latent topic spaces.

[b] https://www.tabers.com/tabersonline/view/Tabers-Dictionary/767492/all/Medical_Abbreviations
[c] https://nurseslabs.com/medical-terminologies-abbreviations-listcheat-sheet/
[d] Only the abbreviation detection module of CARD was able to run on our corpus.
[e] Additional details can be found at https://wordnet.princeton.edu/documentation/morphy7wn.

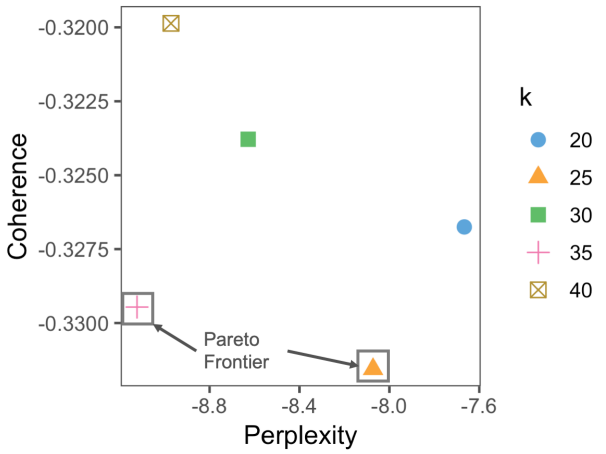


**Figure 3:** The perplexity and coherence on the validation corpus for the 30-day mortality outcome. The two boxed points ($k = 25, 35$) represent the Pareto frontier.

Patients' topic distributions and topic-word distributions were learned on the nursing notes corpus using Gensim[38], a free Python library for extracting semantic topics from documents. For ease of comparison, we used the default setting for the other LDA hyperparameters and only tuned the number of topics ($k$). We created 10 random samples using a 70%-30% train-validation split to assess a range of 20-100 topics. Unlike previous works where $k$ was selected on the predictive performance[17,18], we chose $k$ based on the model's ability to capture the notes and avoid potential overfitting to the validation set. Thus, we used both perplexity and coherence, two common measures of topic models[39].

Unfortunately, the multi-criteria measures did not yield a single optimal value of $k$. Therefore, we employed the notion of Pareto optimality, used in engineering and economics, to find the best trade-offs between the different criterion. We found the Pareto frontier (or set) by identifying values of $k$ that were not dominated in both perplexity and coherence by other values of $k$. Thus, each value in the Pareto frontier represented a trade-off in perplexity or coherence. Figure 3 illustrates the Pareto frontier selection process for the 30-day mortality outcome.

Another option for topic modeling is document-level embeddings, where each document is represented using a unique vector. Unlike LDA, where the model is learned on an unordered collection of words, doc2vec (also known as paragraph2vec) preserves the semantics of the words and remembers the current context[40]. Doc2vec builds on word2vec, which uses neural networks to learn word vectors that represent the sense of the word. Similarly, doc2vec uses the same concept at the document level to capture the topic of the paragraph. We use the Gensim implementation of doc2vec and only tuned the dimensional representation of the documents (also denoted as $k$). The model is evaluated on the self-similarity for all the training notes[f]. Self-similarity is evaluated based on the number of documents that were self-ranked in the top 10, 25, 50, and 100. Based on these four criteria, the Pareto frontier was selected as the optimal dimensional representation.

*Sentiment Analysis*. Given the descriptive nature of the nursing notes, we employed two different sentiment analyzers to extract sentiment-related features: Pattern for Python[41] and Valence Aware Dictionary and sEntiment Reasoner (VADER)[42]. An algorithm implemented in TextBlob, Pattern for Python tokenized the text, tagged the part-of-speech, and used the SentiWordNet lexicon[43] to classify sentiment polarity and subjectivity. This has been used in previous works for mortality prediction[21–23]. Designed for social media text, the VADER algorithm was implemented in NLTK and produced four sentiment metrics when given a list of words[42]. The first three represented the portions of the text that were positive, neutral, and negative. The last metric, a compound score, summed the lexicon ratings.

*Experimental Setup*. Variables were concatenated so that each patient had three sets of structured clinical features: Elixhauser score, topics of the nursing notes ($k$), and two sets of sentiment-related features of the nursing notes (i.e.,

[f]Introduced in the doc2vec tutorial on Gensim `https://github.com/RaRe-Technologies/gensim/blob/develop/docs/notebooks/doc2vec-lee.ipynb`.

**Table 3:** Counts of the top 11 abbreviations and their long-forms in nursing notes for 30-day and in-hospital mortality.

| Abbreviations | Long-Form | 30-Day Count | In-hospital Count |
|---|---|---|---|
| pt,Pt,pt.,Pt.,pt' | patient | 2704 | 5593 |
| Dr.,MD,Dr,md,dr,MD.,md.,m.d,dr. | physician | 813 | 1660 |
| RN,RN.,R.N. | registered nurse | 261 | 503 |
| AM,am.,a.m.,AM.,A.M. | morning | 188 | 414 |
| c/o,C/O,C/o,c/o. | complaint of | 183 | 434 |
| hr,hr. | hour | 182 | 521 |
| VSS.,VSS,vss,vss. | vital signs stable | 176 | 370 |
| BP,bp,BP.,b/p,bp.,b/p.,B.P. | blood pressure | 150 | 344 |
| CT,CT. | computed/computerized tomography | 141 | 259 |
| IV,IV. | intravenous | 140 | 324 |
| lab,lab.,Lab. | laboratory | 139 | 284 |

polarity and subjectivity from Pattern and positive, neutral, negative, and compound metrics from VADER). Thus, the patient feature vector $v$ was constructed from these three sets, where each element represented the value of the feature $f$ of the patient. No additional normalization or standardization was performed on the features. A nested cross-validation framework was used to learn a regularized logistic regression model for each of the mortality outcomes. A separate $\ell_2$ regularized logistic regression model was individually built for each outcome, and each set of model features was evaluated. Five-fold cross-validation was used to split the data into five groups of train-test splits. For each fold, the $\ell_2$ regularization parameter was selected using five-fold cross-validation of the training data to determine the optimal values with the area under the Receiver Operating Characteristic Curve (AUC) as the objective. The parameter was then used to learn a model for the entire training set, and the predictions were evaluated on the test data.

**Results**

*Effect of Abbreviation Normalization*. There were 526 abbreviations detected in the nursing notes. Table 3 summarizes the top 11 abbreviations and normalized long-forms in nursing notes for 30-day and in-hospital mortality cohorts. The most common abbreviations describe individuals (e.g., "patient," "physician," "registered nurse') and patients' care (e.g., "blood pressure," "vital signs," and "complaint of").

The impact of our abbreviation normalization process was assessed using the predictive power of the learned topics. For comparative purposes, we used the same $k$ for both original notes and the abbreviation normalized notes over a range of 10-50 topics. Table 4 summarizes the difference in predictive power between the topics learned from the abbreviation normalized nursing notes versus the non-abbreviation normalized nursing notes. We observed that abbreviation normalization improves the AUC and is statistically significant. The table also suggests that the normalization process is more beneficial for predicting in-hospital mortality and when used with the LDA model. This may be due to the significant reduction in corpus size (i.e., number of words) from our preprocessing steps (see Table 2).

**Table 4:** The AUC difference from abbreviation normalized nursing notes to the original nursing notes and their associated statistical significance from a paired t-test under the hypothesis that the difference is greater than 0.

| Outcome | Model | AUC Difference | p-value |
|---|---|---|---|
| 30-day | LDA | 0.103 | 0.017 |
| 30-day | Doc2Vec | 0.027 | 0.001 |
| In-hospital | LDA | 0.118 | 0.006 |
| In-hospital | Doc2Vec | 0.054 | 0.007 |

We also analyzed the difference in abbreviations between our nursing abbreviation resource and the clinical abbreviation recognition and disambiguation (CARD) framework[32]. CARD detected 1,042 abbreviation candidates from the nursing notes, of which 848 candidates were not detected using our normalization process. Similarly, our abbreviation

process identified 332 (out of 526) abbreviations that were not flagged as candidates from CARD. Table 5 shows the top 10 abbreviations based on occurrences that were unique to CARD and our process. Although CARD detected almost twice as many abbreviations, many of them were noisy and difficult to disambiguate. For instance, the words "TO," "NO," "a," and "AT" can each have multiple senses or may not be abbreviations at all.

**Table 5:** Top 10 abbreviations that were uniquely detected by the CARD framework and our framework.

| CARD Abbrs. | Count | Our Abbrs. | Count |
|---|---|---|---|
| PAS | 751 | Dr. | 515 |
| TO | 495 | pt. | 264 |
| NP | 444 | Pt. | 159 |
| PAS=0 | 379 | lab | 136 |
| a | 368 | post | 134 |
| NO | 347 | RN. | 107 |
| hrs | 294 | VSS. | 100 |
| PA | 276 | am | 84 |
| meds | 230 | pre | 71 |
| AT | 230 | med | 57 |

*Effect of Sentiment and Topic Modeling*. Next, we evaluated the predictive power of the topic-based features and the sentiment-based features for predicting mortality. Based on the general performance improvements from abbreviation normalization, only the preprocessed notes were used for this purpose. For this study, the number of topics ($k$) was identified using the Pareto frontier method. We then selected the best performing curve on the test set (i.e., based on highest AUC across all features) for each outcome and topic model algorithm.

Table 6 summarizes the AUC for 5 different feature sets: (1) only the Elixhauser score; (2) Elixhauser score and topic-based features; (3) Elixhauser, topic-based features, and sentiment analysis with Pattern; (4) Elixhauser, topic-based features, and sentiment analysis with Vader; and (5) Elixhauser, topic-based features, and both sentiment analysis with Pattern and Vader (or all). There is generally a noticeable performance improvement from the topic-based features and sentiment-based features, with the lone degradation observed for in-hospital mortality trained on doc2vec topics. Additionally, LDA performed better than doc2vec on the 30-day task, as there may not be sufficient examples to properly learn the document-embedding representation. The results suggest that the VADER sentiment was particularly useful for predicting 30-day mortality but not necessarily for in-hospital mortality. Similar trends were observed for the other values of $k$ in the Pareto set.

*Effect of Learning Topics from Larger Note Corpus*. Given the relatively small progress note corpus associated with each task, we explored the effect of using all the progress notes from our database (without requiring an inpatient encounter or a follow-up). This resulted in 5,129 patients with a total corpus size of 9,145 words, which were then preprocessed to 8,015 words. Topic models were learned using the entire patient corpus with the test patients removed. The topics were then inferred for the test patients. Figure 4 shows the boxplot of the AUC difference between using all the progress nursing notes and just the progress notes associated with each mortality outcome (i.e., 30-day and in-hospital) with only the Elixhauser and topic-based features. While the larger corpus marginally helped LDA learn better topics, it had negligible effect for the doc2vec model.

**Table 6:** Predictive performance on the five different feature sets using the combination of the Elixhauser scores (Elix), topic features (Topic), sentiment features with Pattern (Pattern), and sentiment features from VADER (Vader).

| Outcome | Topic Model | Elix | Elix+Topic | Elix+Topic+Pattern | Elix+Topic+Vader | All |
|---|---|---|---|---|---|---|
| 30-day | LDA | 0.7760 | 0.7961 | 0.7955 | 0.7961 | **0.7963** |
| 30-day | Doc2Vec | 0.7760 | 0.7719 | 0.7748 | 0.7807 | **0.7809** |
| In-hospital | LDA | 0.7600 | 0.7945 | 0.7889 | 0.8014 | **0.8129** |
| In-hospital | Doc2Vec | **0.7600** | 0.7574 | 0.7561 | 0.7593 | 0.7574 |

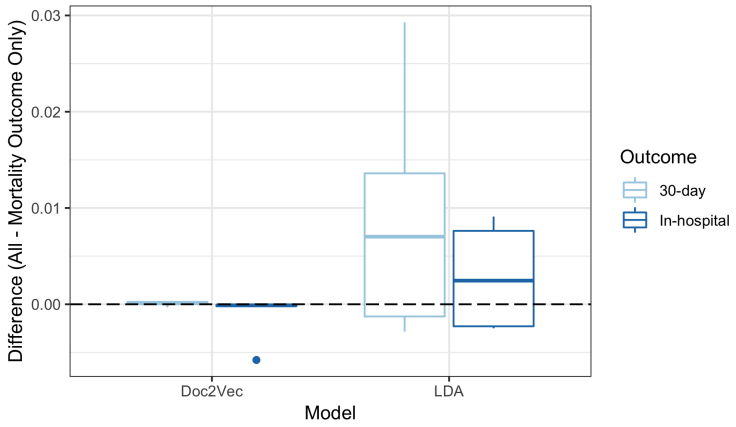


**Figure 4:** The difference in predictive performance of learning topics using all the progress notes available in the database compared to just the progress notes from those with mortality outcomes.

## Discussion

Overall, disambiguating abbreviations in nursing notes used for topic modeling and sentiment analysis improves prediction of in-hospital and 30-day mortality. In this secondary analysis of EHR data, we extracted International Classification of Diseases-Ninth Revision diagnoses and nursing notes to predict outcomes of in-hospital and 30-day mortality. The Elixhauser Comorbidity Index was calculated from the diagnoses. We preprocessed nursing notes through abbreviation detection and normalization, lemmatization, and stop word removal. Topic modeling with LDA and doc2vec and sentiment analysis with Pattern for Python and VADER were then applied to the nursing notes. Predictive power was assessed for models with and without abbreviation normalization while controlling for comorbidity. Compared to non-abbreviation normalized nursing notes, abbreviation normalized nursing notes resulted in an improvement in AUC. This may be due to the high prevalence of abbreviations in nursing note and the potential to provide better context (i.e., more meaningful words) in short nursing notes.

Topic modeling improved prediction for both outcomes, with LDA performing better than doc2vec. Sentiment analysis only improved 30-day mortality prediction, with VADER sentiment performing better than Pattern. Interestingly, VADER was specifically designed for social media text, and these results indicate that this lexicon can be adapted for another context. A possible explanation for VADER's success is that a nursing note conveys a story in a few words, similar to social media texts. Increasing corpus size resulted in marginal improvement for LDA but poorer performance for doc2vec. We suspect the overall lack of improvement is due to the increased corpus size without a significant jump in patient observations. Unlike LDA, which uses a bag-of-word representation, the document-embedding representation is more complex and thus requires more training samples. However, the performance improvement is negligible for LDA and may not be worth the additional computational complexity.

Limitations of our study include the short length and restricted context of notes in our sample (i.e., we had an average of $\sim 50$ words per note). The short lengths potentially impacted the ability of the topic model algorithms (i.e., LDA and doc2Vec) to learn more differentiating topics, and they made it difficult to use existing word sense disambiguation software without additional expert annotation. Our simple nursing abbreviation resource was better suited to identify common abbreviations in our nursing notes compared to other resources that identified ambiguous abbreviations with multiple senses. While our nursing abbreviation resource could be expanded to incorporate other abbreviation resources, we found that scraping the two online resources was sufficient after inspecting a small random sample of notes. We experimented with a more comprehensive medical abbreviation dictionary, but common words (e.g., "AND") were replaced with inappropriate senses. We also explored adding other categories of free-text nursing notes (e.g., care plans and observations about vital signs) but the predictive performance impact was negligible.

We also note the limitations from our selection of topic model algorithms and testing of our model. Since LDA uses a bag-of-words model (i.e., unordered collection of words for patient representation)[37], it can lose the inherent context in

the nursing notes. While doc2vec preserves the semantics of the words using a deep-learning model, it has been well-documented that the results can be difficult to reproduce[44] and may not perform as well in specialized deep learning models[45]. Furthermore, doc2vec does not capture negations or long-range dependencies across notes. However, our limited corpus was insufficient to adequately learn the doc2vec parameters and thus did not outperform LDA. Therefore, we do not expect specialized deep learning models to perform significantly better without more patients and more extensive notes. Additionally, these clinical notes were written over multiple years at various facilities, and it is unclear how institutional procedures and policies affected documentation practices. Finally, we were unable to run multiple prediction tasks to verify our findings with other health outcomes.

## Conclusion

Analyzing unstructured clinical data from EHRs provides opportunity for deeper insight into patient health outcomes. In particular, nursing notes contain information about patients, such as clinical conditions and factors impacting health, that can be used for prediction. We demonstrated that predictive models can be improved by extracting structured information from nursing notes using abbreviation normalization, topic modeling, and sentiment analysis. This process of incorporating unstructured text may assist earlier identification of at-risk patients who may, in turn, benefit from early intervention. Future opportunities for research include assessing predictive value for other health outcomes, utilizing clinical texts with increased variety and volume, and exploring other avenues of abbreviation disambiguation in the presence of multiple senses.

## Acknowledgements

This work was supported by the National Institute of Health award 1K01LM012924-01 and the Robert Wood Johnson Foundation's Future of Nursing Scholars program.

## References

1. Maddox TM, Matheny MA. Natural language processing and the promise of big data: small step forward, but many miles to go. Circ Cardiovasc Qual Outcomes. 2015;8(5):463–465.

2. Westra BL, Sylvia M, Weinfurter EF, Pruinelli L, Park JI, Dodd D, et al. Big data science: a literature review of nursing research exemplars. Nursing Outlook. 2017;65(5):549–561.

3. Rochefort CM, Buckeridge DL, Forster AJ. Accuracy of using automated methods for detecting adverse events from electronic health record data: a research protocol. Implementation Science. 2015;10(1):5.

4. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet. 2012;13(6):395.

5. Mullenbach J, Wiegreffe S, Duke J, Sun J, Eisenstein J. Explainable prediction of medical codes from clinical text. In: Proc of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics; 2018. p. 1101–1111.

6. Baumel T, Nassour-Kassis J, Cohen R, Elhadad M, Elhadad N. Multi-label classification of patient notes a case study on ICD code assignment. In: Proc of the Workshops at the 32nd AAAI Conference on Artificial Intelligence; 2017. p. 409–416.

7. Rios A, Kavuluru R. EMR coding with semi-parametric multihead matching networks. In: Proc of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics; 2018. p. 2081–2091.

8. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, et al. Clinical information extraction applications: A literature review. J Biomed Inform. 2018;77:34–49.

9. Sohn S, Wang Y, Wi CI, Krusemark EA, Ryu E, Ali MH, et al. Clinical documentation variations and NLP system portability: a case study in asthma birth cohorts across institutions. J Am Med Inform Assoc. 2018;25(3):353–359.

10. Tower M, Chaboyer W, Green Q, Dyer K, Wallis M. Registered nurses decision-making regarding documentation in patients progress notes. Journal of Clinical Nursing. 2012;21(19pt20):2917–2929.

11. Hyun S, Johnson SB, Bakken S. Exploring the ability of natural language processing to extract data from nursing narratives. Computers, Informatics, Nursing: CIN. 2009;27(4):215–223.

12. Collins SA, Vawdrey DK. Reading between the lines of flowsheet data: nurses' optional documentation associated with cardiac arrest outcomes. Applied Nursing Research: ANR. 2012;25(4):251.

13. Blei DM. Probabilistic topic models. Communications of the ACM. 2012;55(4):77–84.

14. McCoy TH, Castro VM, Roberson AM, Snapper LA, Perlis RH. Improving prediction of suicide and accidental death after discharge from general hospitals with natural language processing. JAMA Psychiatry. 2016;73(10):1064–1071.

15. Saria S, McElvain G, Rajani AK, Penn AA. Combining structured and free-text data for automatic coding of patient outcomes. AMIA Ann Symp Proc. 2010;p. 881–892.

16. Saeed M, Long W, Lee J. Risk stratification of ICU patients using topic models inferred from unstructured progress notes. AMIA Annu Symp Proc. 2012;2012:505–511.

17. Ghassemi M, Naumann T, Doshi-Velez F, Brimmer N, Joshi R, Rumshisky A, et al. Unfolding physiological state: mortality modelling in intensive care units. In: Proc. of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2014. p. 75–84.

18. Lehman Lw, Long W, Saeed M, Mark R. Latent topic discovery of clinical concepts from hospital discharge summaries of a heterogeneous patient cohort. In: Proc of the 36th International Conference of the IEEE Engineering in Medicine and Biology Society; 2014. p. 1773–1776.

19. Hu M, Liu B. Mining and summarizing customer reviews. In: Proc of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2004. p. 168–177.

20. Denecke K, Deng Y. Sentiment analysis in medical settings: new opportunities and challenges. Artif Intell Med. 2015;64(1):17–27.

21. McCoy TH, Castro VM, Cagan A, Roberson AM, Kohane IS, Perlis RH. Sentiment measured in hospital discharge notes is associated with readmission and mortality risk: an electronic health record study. PLoS ONE. 2015;10(8):e0136341.

22. Tran N, Lee J. Using multiple sentiment dimensions of nursing notes to predict mortality in the intensive care unit. In: 2018 IEEE EMBS International Conference on Biomedical & Health Informatics; 2018. p. 283–286.

23. Waudby-Smith IER, Tran N, Dubin JA, Lee J. Sentiment in nursing notes as an indicator of out-of-hospital mortality in intensive care patients. PLoS ONE. 2018;13(6):e0198687.

24. Moon S, Pakhomov S, Liu N, Ryan JO, Melton GB. A sense inventory for clinical abbreviations and acronyms created using clinical notes and medical dictionary resources. J Am Med Inform Assoc. 2014;21(2):299–307.

25. Shilo L, Shilo G. Analysis of abbreviations used by residents in admission notes and discharge summaries. QJM: An International Journal of Medicine. 2017;111(3):179–183.

26. Long WJ. Parsing free text nursing notes. AMIA Annu Symp Proc. 2003;2003:917.

27. Kuhn IF. Abbreviations and acronyms in healthcare: when shorter isn't sweeter. Pediatric nursing. 2007;33(5):392–398.

28. Xu H, Stetson PD, Friedman C. A study of abbreviations in clinical notes. AMIA Annu Symp Proc. 2007;2007:821–825.

29. Walsh KE, Gurwitz JH. Medical abbreviations: writing little and communicating less. Archives of Disease in Childhood. 2008;93(10):816–817.

30. Shilo L, Shilo G. Analysis of abbreviations used by residents in admission notes and discharge summaries. QJM: An International Journal of Medicine. 2018;111(3):179–183.

31. Finley GP, Pakhomov SVS, McEwan R, Melton GB. Towards comprehensive clinical abbreviation disambiguation using machine-labeled training data. AMIA Annu Symp Proc. 2016;2016:560–569.

32. Wu Y, Denny JC, Trent Rosenbloom S, Miller RA, Giuse DA, Wang L, et al. A long journey to short abbreviations: developing an open-source framework for clinical abbreviation recognition and disambiguation (CARD). J Am Med Inform Assoc. 2017;24(e1):e79–e86.

33. Kolchyna O, Souza TT, Treleaven P, Aste T. Twitter sentiment analysis: lexicon method, machine learning method and their combination; 2015.

34. Wu Y, Denny JC, Rosenbloom ST, Miller RA, Giuse DA, Xu H. A comparative study of current clinical natural language processing systems on handling abbreviations in discharge summaries. AMIA Annu Symp Proc. 2012;2012:997–1003.

35. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Medical Care. 2005;43(11):1130–1139.

36. Bird S, Loper E. NLTK: the natural language toolkit. In: Proceedings of the ACL 2004 on interactive poster and demonstration sessions. Association for Computational Linguistics; 2004. p. 31.

37. Blei, David M, Ng, Andrew Y, Jordan, Michael I. Latent Dirichlet allocation. Journal of Machine Learning Research. 2003;3:993–1022.

38. Řehůřek R, Sojka P. Software framework for topic modelling with large corpora. In: Proc. of the LREC 2010 Workshop on New Challenges for NLP Frameworks. Valletta, Malta; 2010. p. 45–50.

39. Newman D, Lau JH, Grieser K, Baldwin T. Automatic evaluation of topic coherence. In: Proc of the 2010 Annual Conference of the North American Chapter of the Association for Computational Lingustics; 2010. p. 100–108.

40. Le Q, Mikolov T. Distributed representations of sentences and documents. In: Proc of the 31st International Conference on Machine Learning; 2014. p. 1188–1196.

41. De Smedt T, Daelemans W. Pattern for Python. Journal of Machine Learning Research. 2012;13:2063–2067.

42. Hutto CJ, Gilbert E. VADER: A parsimonious rule-based model for sentiment analysis of social media text. In: Proc of the 8th International Conference on Weblogs and Social Media; 2014. p. 216–225.

43. Baccianella S, Esuli A, Lrec FS, 2010. Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. Lrec. 2010;10:2200–2204.

44. Lau JH, Baldwin T. An empirical evaluation of doc2vec with practical insights into document embedding generation. Proc of the 1st Workshop on Representation Learning for NLP. 2016;p. 78–86.

45. Grnarova P, Schmidt F, Hyland SL, Eickhoff C. Neural document embeddings for intensive care patient mortality prediction. In: NIPS Workshop on Machine Learning for Health; 2016. p. 1–5.