Artificial/Machine Learning

Automated Extraction of Pain Symptoms: A Natural Language Approach using Electronic Health Records

Amisha D. Dave, BSE[1], Gualberto Ruaño, MD, PhD[1,2], Jonathan Kost, MD[3], and Xiaoyan Wang, PhD[4]

From: [1]University of Connecticut School of Medicine, Farmington, CT; [2]Institute of Living at Hartford Hospital, Hartford, CT; [3]Hartford Hospital Pain Treatment Center, West Hartford, CT; [4]Mount Sinai Genomics Incorporation, Stamford, CT

Address Correspondence: Xiaoyan Wang PhD VP, Healthcare Analytics and Informatics, Mount Sinai Genomics Incorporation 333 Ludlow St, North Tower, 7th fl Stamford, CT 06902 E-mail: xiaoyan.wang@sema4genomics.com

Disclaimer: This study was in part supported by the Partnership in Innovation and Education program.

Conflict of interest: Each author certifies that he or she, or a member of his or her immediate family, has no commercial association (i.e., consultancies, stock ownership, equity interest, patent/licensing arrangements, etc.) that might pose a conflict of interest in connection with the submitted manuscript.

Manuscript received: 06-18-2021 Revised manuscript received: 10-02-2021 Accepted for publication: 11-25-2021

Free full manuscript: www.painphysicianjournal.com

Background: Pain costs more than $600 billion annually and affects more than 100 million Americans, but is still a poorly understood problem and one for which there is very often limited effective treatment. Electronic health records (EHRs) are the only databases with a high volume of granular pain information that allows for documentation of detailed clinical notes on a patient's subjective experience.

Objectives: This study applied natural language processing (NLP) technology to an EHR dataset as part of a pilot study to capture pain information from clinical notes and prove its feasibility as an efficient method.

Study Design: Retrospective study

Setting: All data were from UConn Health John Dempsey Hospital (JDH) in Farmington, CT.

Methods: The JDH EHR dataset contains 611,355 clinical narratives from 359,854 patients from diverse demographic backgrounds from 2010 through 2019. These data were processed through a customized NLP pipeline. A training set of 100 notes was annotated based on focus group-generated ontology and used to generate and evaluate an NLP model that was later tested on the remaining notes. Validation of the model was evaluated externally and performance was analyzed.

Results: The model identified back pain as the most common location of experienced pain with 40,369 term frequencies. Patients most commonly experienced decreased mobility with their pain with 7,375 term frequencies. Pain was most commonly found to be radiating with 26,967 term frequencies and patients most commonly rated their pain as 8/10 with 2,375 term frequencies. All parameters studied had statistical F-scores greater than 0.85.

Limitations: A single-center, pilot study subject to reporting bias, recording bias, and missing patient data.

Conclusions: Our customized NLP model demonstrated good and successful performance in extracting granular pain information from clinical notes in electronic health records.

Key words: Pain ontology, natural language processing, automation, electronic health records, pain location, pain quality, pain quantity, pain symptoms

Pain Physician 2022: 25:E245-E254

Pain is a personal and subjective experience associated with many conditions and presentations, making it difficult to manage (1). Annually, it costs more than $600 billion, affects over 100 million Americans, and is the leading reason for adult outpatient and emergency department medical visits (2). Pain management requires numerous follow-up visits per patient, limiting appointments for others, and acute exacerbations waste countless hours (3). These factors, combined with the push for opioids as the first-line treatment for pain resulted in the opioid crisis, with an average of 100 opioid overdose deaths daily (4). To better understand, quantify, and manage pain symptoms, physicians and researchers need access to granular pain information, which can only be found in electronic health records (EHRs).

EHRs are the only rich database with a high volume of granular pain information, even when considering the comprehensive pain ratings and scales available in pain clinics (5). It allows for documentation of detailed clinical notes where the patient's subjective experience is thoroughly described, often in the patient's own words, along with location, temporality, quality, quantity, past treatments, and other signs and symptoms (6). Therefore, clinical notes embedded within the EHR are crucial to understand and quantify pain in all of its diverse presentations (5). With access to these narratives, we can parse millions of charts in minutes using natural language processing (NLP) (7-10). By going beyond a simple rule-based string search and thoroughly understanding the underlying relationships between words and how they are used, deep learning-based NLP has helped quantify data found in clinical notes. Applications focusing specifically on pain in the free text of clinical records have successfully identified the experience of pain; clinically relevant pain has been detected at higher rates using NLP in EHRs than in patient surveys (8).

Without NLP, it would be impossible to process millions of clinical narratives to reach a meaningful understanding of pain (10). Therefore, EHRs and NLP provide a unique opportunity to assess and manage pain symptoms, ultimately helping clinicians better understand pain. This will reduce the burden on patients, solve an expensive health care issue, and alleviate factors contributing to the opioid crisis. Previous studies using NLP and EHRs have focused on better understanding pain in specific disease processes, such as breast or prostate cancer, however this study is the first to use these tools to understand pain as a whole (8,11). Additionally, very few NLP studies have used customized NLP pipelines and no studies to our knowledge have extracted comprehensive characteristics of pain from clinical notes. Customization is an essential component for NLP-driven studies because it allows for a tailored understanding of the ontology of the subject and the development of relationships between concepts. This allows for greater granularity and dimensionality of the topic at hand, enhancing overall understanding of pain through an automated process. It requires strong development of ontology of pain to obtain granular and accurate pain information.

In this study, we applied NLP technology to an EHR dataset from the UConn Health John Dempsey Hospital to better understand the relationship between pain symptoms and their various identifying parameters and treatment.

Methods

Step 1: Retrieve Data

The dataset is a collection of the EHRs from about 600,000 deidentified patients at UConn Health John Dempsey Hospital. The records accessed were documented from 2010 through 2019. A word2vec query expansion algorithm was implemented with the seed term 'pain' and 'ache' to identify the pain cohort. Among the total 359,854 patients, 611,355 notes were identified with pain symptoms.

Step 2: Implement the NLP Pipeline to Extract Pain Symptoms (Fig. 1)

Step 2a: Generate Pain Ontology: Entities and Semantics

A focus group of physicians was used to generate pain ontology to identify which parameters of pain were important to capture in the clinical notes to best characterize and understand pain (Fig. 2). The focus group annotated sample notes individually and these annotations were discussed as a group until consensus was achieved. Ultimately the focus group established the following entities and attributes of pain necessary to capture in the notes: location, onset, quality, quantity, severity, radiation, alleviating and aggravating factors, frequency, and prior treatment received.

Location was defined as the body location of the pain. Onset was defined as when the pain began. Quality was defined as descriptive words associated with the pain. Quantity was defined as a number on a scale from 0 to 10 where 0 is no pain at all and 10 is the worst pain the patient has ever experienced. Severity was defined as either mild, moderate, or severe levels of pain. Radiation was defined as movement of the pain in a way that causes the patient to experience it in places other the original location. Alleviating and aggravating factors were defined as medication or other forms of attempted treatment that make the pain better or worse respectively. Frequency is defined as how often the pain is present. Prior treatment received was defined as

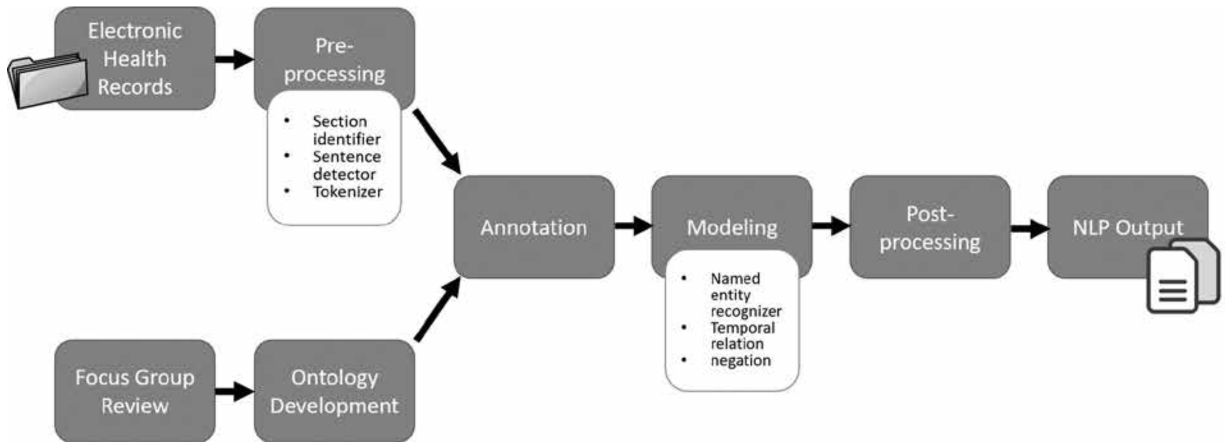

Fig. 1. *Methods Overview. As records were being extracted from the EHR and notes were being preprocessed, a focus group review was conducted to develop pain ontology. Using the ontology, the preprocessed notes were then annotated. This allowed for generation of the model, subsequent postprocessing, and the final output.*

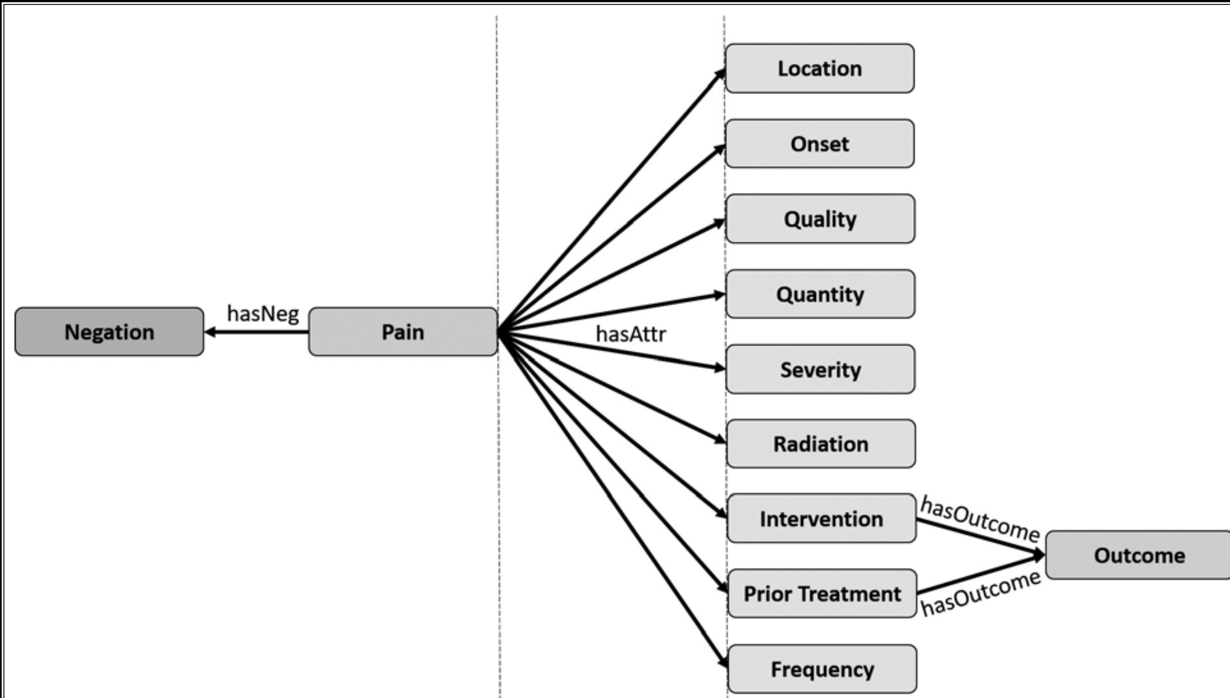

Fig. 2. *Pain ontology generated through physician focus group. The physician focus group identified the following 9 parameters of pain: location, onset, quality, quantity, severity, radiation, intervention, prior treatment, and frequency. The parameters "intervention" and "prior treatment" had a potential outcome that could be identified. They also identified the lack of pain, represented by negating words.*

interventions the patient has received in the past from a physician not necessarily related to the pain.

Each note contained information from a patient's history of present illness as documented by the clinician. It included some combination of the above parameters.

Step 2b: Preprocessing and Parsing

Each individual clinical note was tokenized with identification of sections and individual sentences. This allowed Clinical Language Annotation, Modeling, and Processing (CLAMP), a user-friendly NLP toolkit that allows researchers to build customizable NLP pipelines, to process the notes more efficiently, while also providing individual output files for each note (12). Preprocessing included normalization of the entities. For example, quantity of pain was most commonly represented as "x/10," "x out of 10," "x of 10," and "x over 10" in the clinical notes and were standardized to "x/10."

At the suggestion of members of the clinician focus group, pain-specific quality words were provided and used to capture the quality of pain. These were entered into the NLP dictionary and tagged as quality entities to assist in the development of the algorithm. The discrete recognition of these words helps to train the algorithm to recognize these words as well as other words that may be similar when parsing through the clinical notes. This makes it easier for the algorithm to put these words in context with the clinical notes and better understand the relationship between these words and other entities within the note.

Step 2c: Annotation and Model Training

After generating the ontology of pain, 250 notes were annotated and processed to create a model to capture pain in clinical notes. These notes served as the training set for the model and were processed. A pain-customized named entity recognition pipeline of dictionary lookup and condition random field model was built using the CLAMP toolkit. This pipeline tags entities, or keywords, in the notes, that the model was trained to recognize, and categorizes it into a semantic (Fig. 3). The semantic categories were the parameters of pain identified during the clinician focus group. The NLP model additionally recognized relationships between entities. For example, it can understand if an entity is negated or not depending on the context of the words in the note. It can also understand that a particular name brand medication is considered a certain class of medication for a particular disease (Fig. 3). To appropriately train the model, there were a few iterations of testing that allowed for all the necessary parameters to be identified. The notes were then run using the trained model on 3,500 randomly selected notes.

Step 2d: Postprocessing

Each key entity was tagged in its corresponding semantic category after modeling. The relationships between different semantic groups were identified via arrows and words defining the relationship. The key entities were highlighted in respective colors to code each semantic.

For each entity, postprocessing included encoding and postcoordination. The output data identified the start index, end index, semantic category, concept unique identifier (CUI) code, assertion value, and the entity itself. The CUI code corresponds to a specific Unified Medical Language System category, allowing for grouping of different diagnoses. The assertion category consists of either present, absent, or null, clarifying whether the entity was being referred to as being present, absent, or not applicable.

Step 2e: Evaluation of NLP Performance

Two domain experts externally validated the NLP model output. Fifty randomized notes were randomly selected from the testing set and given to the experts to analyze independently. These notes were simultaneously run using the trained CLAMP model. The model output was then compared to that of the 2 experts and validity was calculated. Precision, recall, and F1-score were calculated to evaluate the performance of NLP pipeline.

Step 3: Statistical Analysis

Statistical analysis was conduct in R (R Core Team) (13) to calculate summary statistics and evaluate model performance. Recall, precision, and F1 measure were used to assess the performance of the NLP pipeline. Recall was calculated as the ratio of the number of entities that were identified by the pipeline over the total number of the corresponding entities in the gold standard (i.e., TP/(TP + FN)). Precision was measured as the ratio of the number of distinct entities returned by the pipeline that were correct according to the gold standard divided by the total number of entities found by our pipeline (i.e., TP/(TP + FP)). The F1 score was calculated as the harmonic mean of precision and recall (i.e., 2 x Precision x Recall / (Precision + Recall)

Results

This study included 611,355 clinical notes of patient encounters written by UConn Health John Dempsey Hospital clinicians.

Body location, quality of pain, quantity of pain,

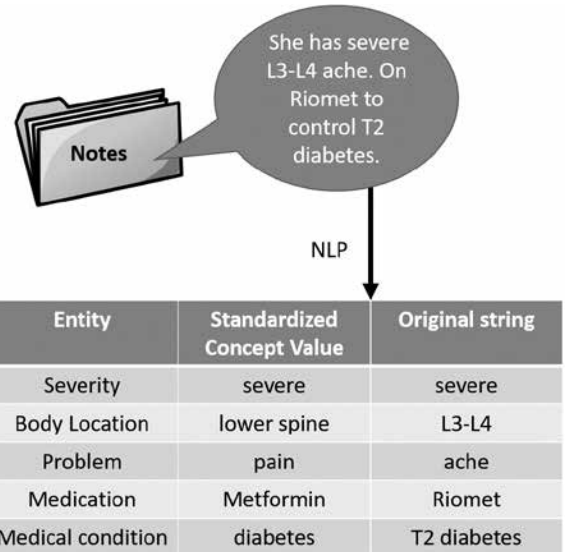

Fig. 3. *NLP symptom and medication extraction in clinical notes. In the example given, the NLP model is able to detect specific words such as "frontal," "headache," severe," and "Oxaydo" as part of standardized concepts that it has been trained to identify. These standardized concepts are encoded as broader categories, assisting the model in understanding relationships between the words.*

and co-occurring symptoms were the 4 parameters of pain found to be the most important in characterizing the pain in the clinical notes from this study. As a result, these were the 4 semantics categories that were analyzed for frequency counts. Body location frequency data is displayed in Table 1. Mention of back pain as a general category occurred 40,369 times within the 611,355 notes, of which 18,660 occurrences are attributed specifically to lower back pain. Following the back, the knee was the next most common location where patients experienced pain with 30,913 mentions. Of these, the right knee accounted for 10,908 tags and left knee accounting for 9,946 tags. The shoulder was the third most common body location for pain, with shoulder pain being mentioned 28,365 times, with the right shoulder accounting for 9,620 tags and the left shoulder accounting for 8,136 tags.

Co-occurring symptoms of pain are displayed in Table 2. These are symptoms that the patient was experiencing along with the pain that assisted the physician in determining a differential diagnosis for the pain. Decreased mobility occurred as the most frequent co-occurring symptom, appearing 7,375 times. Following that, back pain and nocturnal pain were the second and third most common co-occurring symptoms, appearing 6,940 times and 5,557 times respectively.

The quality of pain data is displayed in Table 3. These were the most common pain descriptors used by patients when describing their pain to their clinician. The most common descriptor was radiation, which

Table 1. *Most common body location of experienced pain.*

Body Location	Frequency
Back (Low Back)	40,369 (18,660)
Knee (Right Knee, Left Knee)	30,913 (10,908, 9,946)
Shoulder (Right Shoulder, Left Shoulder)	28,365 (9,620, 8,136)
Leg	15,381
Joint	14,017
Bladder	11,957
Bowel	8,275
Neck	6,920
Abdominal	5,474
Arms	5,470

Table 2. *Most common symptoms of experienced pain.*

Symptoms	Frequency
Decreased mobility	7,375
Back pain	6,940
Nocturnal pain	5,557
Bruising	5,553
Locking	5,147
Ache	5,069
Popping	5,009
Crepitus	4,915
Rash	4,907
Tingling in legs	4,688

appeared 26,967 times. Following that, numbness appeared 25,827 times and tingling appeared 21,020 times.

The distribution of pain quantity in the clinical note is shown in Fig. 4. The most frequent pain scale rating in the 611,355 notes was 8/10, which appeared 2,375 times. Following that, 5/10 appeared 1,588 times and 6/10 appeared 1,581 times.

Internal validation data for the model can be found in Table 4. In identifying body location, the model has a precision value of 0.78 and a recall value of 1, resulting in an F-score of 0.87. In identifying co-occurring symptoms within the clinical note, the model has a precision value of 1 and a recall value of 1, resulting in an F-score of 1. In identifying the quality of a patient's pain, the model has a precision value of 1 and a recall value of 0.99, resulting in an F-score of 0.99. Quantity of pain was not assessed through internal validation because these values were identified by the model via a direct dictionary look-up.

Table 3. *Most common qualities of experienced pain.*

Quality	Frequency
Radiation	26,967
Numbness	25,827
Tingling	21,020
Ache	15,428
Sharp	9,205
Tender	6,407
Burning	4,259
Constant	3,925
Dull	3,382
Throb	3,330

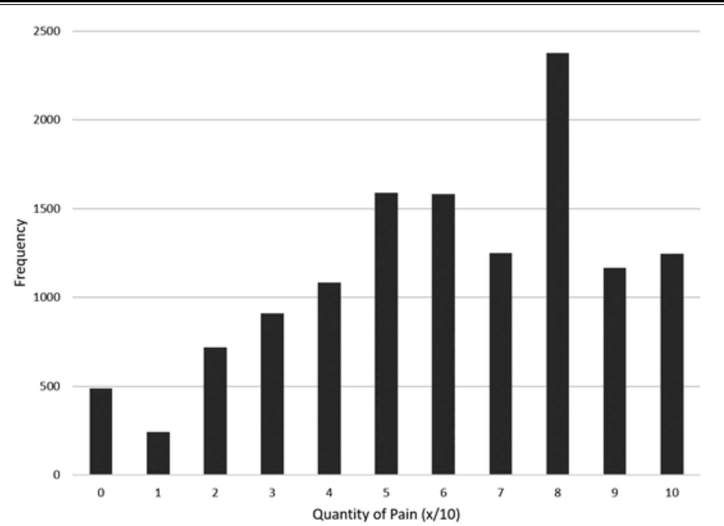

Fig. 4. *Frequency of pain quantity values. This histogram represents how often patients rate their pain on a scale of 0-10 out of 10.*

Discussion

The pain extraction model created using the aforementioned NLP tool was successful in extracting granular pain parameters from the clinical notes of EHRs of patients at UConn Health. This is an advanced step toward better understanding and quantifying pain in terms of location, quality, quantity, and co-occurring symptoms by helping clinicians, researchers, and other interested parties access otherwise difficult to analyze pain data. With this information, clinicians will be able to better manage their patients with pain, learning from the vast amount of data to be analyzed within clinical notes. Currently, pain management is a personalized approach that involves some degree of trial-and-error to find options that work well for the patient (14). By utilizing NLP to analyze the various etiologies of pain described by patients in EHRs and understand which interventions were successful and which were not in unique scenarios, researchers and clinicians can provide an even more tailored approach to the management of pain for patients (14). This will result in faster control of the pain, fewer appointments for the patient, and a significant improvement in the patient's quality of life, reducing the burden of their condition (3).

Chronic pain within the United States costs up to $635 billion annually in both loss of productivity and medical costs (14). By arming clinicians with research from the thorough data that can be found within EHRs, we can bring these costs down by decreasing the number of follow-up appointments to adjust treatment regimens. Treatment can be personalized beyond etiology and can take factors such as gender, race, pain location, and severity at presentation, to ensure that management is being catered to the patient. Patients can get back to work and live more active lives, decreasing their risk of other comorbidities that drive up the cost of health care in the United States. A greater wealth of information about pain management with respect to opioids will help clinicians better understand patient and pain-specific indications for opioids, reducing the number of unnecessary prescriptions. This type of predictive modelling will keep these medications out of vulnerable patients' hands and reduce the number of prescribed tablets overall, decreasing the risk of overdose (14).

Prior to this study, NLP has been used to identify symptom information from EHRs to classify diseases in various fields of medicine (15). Within the field of neurology and psychology, NLP excelled in efficiently identifying patients with delirium based on clinical notes (16). This type of data can be used to identify patients at greater risk of developing delirium, allowing for prevention and early treatment (16). In rheumatology, integration of NLP with narrative data from EHRs allowed for improved identification of axial spondyloarthritis, a painful condition with low prevalence (17). Identification of these patients early in their disease course has potential to decrease their disease burden and improve their quality of life as they live with a chronic inflammatory disease. Within the realm of pain, NLP has been used with EHRs of patients with breast cancer undergoing treatment and metastatic prostate cancer to longitudinally track their pain and associated symptoms (8,11). These studies can help track associated pain with various chemotherapies, helping clinicians make personalized decisions about treatment while taking pain management into consideration. It can also help identify various symptoms of pain in those with metastatic cancer, some that may present in unique ways. This information can help guide pain and symptom management in patients with cancer and other terminal diseases.

Our study investigates the problem of pain in a broader setting by analyzing pain of many etiologies and how it is being documented in clinical notes to expand accessibility to this swath of data for future research. This study offers a more in-depth extraction of pain symptoms and allows for specific classification of the pain presented. With the success of this study, there is opportunity for identification of unsuccessful treatment with the ability to personalize pain management for patients.

Body location frequency, seen in Table 1, shows the breakdown of the most common locations of pain to be present in patients within our dataset. It also shows the power of customizing the NLP pipeline to pick up granular pain information because it is able to isolate the most common body locations to their respective anatomical locations. This granularity allows for even deeper understanding of both the cause and treatment of pain by clinicians because many conditions localize to one side over another, helping to identify one disease process over the other.

Table 4. *Internal validation for parameters of pain.*

Parameter	Precision	Recall	F-Score
Body location	0.78	1	0.87
Duration	0.93	1	0.97
Frequency	0.88	0.97	0.92
Prior treatment	1	1	1
Onset	1	0.97	0.98
Co-occurring symptoms	1	1	1
Outcome of interventions	1	0.99	0.99
Quality	1	0.99	0.99
Severity	0.82	1	0.9

Looking at the frequency of co-occurring symptoms of pain, many of the most frequent co-occurring symptoms are those that are typically found with musculoskeletal pain (Table 2). For instance, "back pain," "locking," "popping," and "crepitus" are all symptoms that are typically seen in joint pain. Joint pain, as seen in Table 1, is the fifth most common type of pain in the dataset, showing that many of the parameters explored in this study are dependent on one another. These types of relationships are what strengthen the power of NLP and allow for deeper understanding of the pain being experienced by the patient.

With quality data, we can see that the top 3 most common qualities associated with patients' pain is "radiation," "numbness," and "tingling," all mentioned more than 20,000 times (Table 3). This finding is likely due to the fact that these terms are not mutually exclusive with other quality terms, allowing them to be present in many different presentations of pain. For example, pain can be "sharp" and radiate at the same time.

In particular, the extraction of quantity data gives insight into the psychology of rating pain that patients experience when asked to rate their pain on a scale of 0 to 10. Of note, neither of the extremes are the most common pain rating. Instead, 8/10 is present most commonly at 2,375 times. While it is possible that most patients aren't experiencing the worst pain of their lives, there is also an interesting patient-clinician dynamic in which patients are less likely to report severe pain as the worst pain they have ever experienced. These patients defer to 8/10, which signifies severe pain without a sense of pain over exaggeration. This could explain why the first 4 most common pain ratings are all found in the middle of the pain scale, as opposed to the extremes. Additionally, we see that the 4 lowest pain scale ratings, 0-3, are all present in the bottom 4 ratings in Fig. 4. This is likely due to the fact that patients often do not seek medical care when they have having minimal to no pain, resulting in decreased frequencies for these ratings.

Internal validation of this model supports its ability to successfully identify important parameters of pain from clinical notes with all parameters having an F-score of 0.85 and greater (Table 4). Of note, the model performed best in identifying prior treatment and other conditions the patient may have had with both parameters having an F-score of 1. This is likely due to the wide variety of prior treatments and other conditions present in the training notes, giving the model breadth in identifying variants of these parameters. These parameters were also more distinctly found in the notes with less overlap with other parameters and were less likely to be subjective when they were present. The model's ability to identify body location resulted in the lowest F-score of 0.87. While this is still good performance, there is room for improvement in proper identification. Body location is likely to be more difficult to identify and discern from other distractors because it is often mentioned more than once per note, with not all mentions referring to the location of pain. Often clinicians included a review of systems within the note that also mentioned various parts of the body. This made it difficult to discern at which location the patient was actually experiencing pain, driving down the precision value. Note that the identification of body location has a recall value of 1, which means that the model has no false negatives identified. Instead, the precision value of 0.87 shows that the model needs further training in differentiating between true and false positives.

With NLP, enhancements to the EHR could serve as a platform for clinical data integration to allow personalized health care delivery and outcomes research. Pain management requires a comprehensive medical informatics application to integrate toxicology, pharmacogenetics, and drug dependence risk, and correlate these data to outcomes and prescriptions. A patient's functionality, mental status, side effects or complications to treatments, well-being and any other pertinent data are collected on any interval basis. Comprehensive screening and risk assessment of patients is time-consuming but vital for proper evaluation of their chronic pain conditions. NLP can improve the quality of patient care by empowering patients to share and obtain health care information, participate in decision-making, and communicate effectively with their physicians regarding treatment outcomes. Also, NLP applied to EHRs should enable physicians proactively to access and optimize patient treatments (procedures, drugs).

In the future, NLP should enable personalization of pain management in screening, treatment, and monitoring. Personalizing pain management requires screening for drug usage and substance abuse. Personalized medicine will also help in avoiding drug interactions and factor in genetic variability for opioid selection and dosing. NLP could pinpoint patients who will benefit the most from cytochrome p450 genotyping to improve drug safety (18). Necessary for this personalized pain medicine is the data storage and health informatics enabled by NLP. This resource is particularly necessary in the case of multimodal treatments in pain management with multiple drugs and procedures. Such NLP-enabled presentation of outcomes and test results will unambiguously facilitate and make clinically actionable the translation of complex and disaggregated EHR information to personalized pain management.

The use of NLP in extracting the vast amount of data and entities in this study lends itself easily to the implementation of artificial intelligence in the field of pain. This will allow for the creation of robust artificial intelligence networks that can look at the characteristics of diagnoses and correlate them with pain parameters such as duration, severity, and type of pain. This can help facilitate the development of disease markers and treatment for otherwise difficult to diagnose conditions. For instance, hypermobile Ehlers-Danlos syndrome is a condition that is severely underdiagnosed despite having a prevalence of 1:5,000 patients (19). It commonly is underdiagnosed by clinicians because many of the presenting symptoms, including pain, are nonspecific (20). With extraction of symptoms stored in the EHR of patients who have the diagnosis, NLP can be used to train a model to identify hypermobile Ehlers-Danlos syndrome using clinical markers and phenotypes. This improvement in diagnostics is just one potential way that the use of NLP on clinical notes can benefit patients directly.

This study is limited by limitations typically present when working with EHRs. This includes reporting bias in which clinicians are trusted to report their conversation with the patient in its entirety, recording bias in which clinicians are trusted to record the patient's clinical narrative as accurately as possible, and missing data, which is hard to track down after the fact (21). Comparing the results of this model to that of the validity-confirming physicians, this dataset size is was confirmed to be sufficient in order to determine the feasibility of this study. This methodology did require multiple iterations of annotations and dictionary additions, allowing the end model to efficiently identify the majority of pain parameters of interest. The model generated is consistent in identifying quality and quantity markers as well as body location and co-occurring symptoms, which is difficult to achieve exclusively with NLP, requiring preprocessing and postprocessing. This study looked at 611,355 notes to determine the feasibility of this methodology in accurately extracting pain symptoms. In general, this approach to extracting granular pain information from clinical notes does show evidence of limitations that are inherently found when generalizing and scaling up the protocol (21).

Additional work in this study will focus on improving the NLP model to accurately extract more parameters of pain and develop better understanding of the relationships between the parameters and the remainder of the clinical notes. This can be done by building upon the ontology developed in this study to capture deeper relationships between entities. This will involve decreasing dependence on the dictionary feature to generate an updated model for extracting pain parameters. We also hope to integrate the use of clinical notes from other institutions to make the model more generalizable as well as use more notes for training purposes to increase the capture of note variability in the model. We would like to make use of the CUI code feature present within the NLP software to allow for better categorization of notes based on improved understanding of the clinical note as a whole. An additional step to move from the NLP pipeline software to within the EHR software will help make research in this field more accessible to researchers. Further research could also be done to link pain symptoms to treatments and outcomes that have been tried by patients to analyze treatment effectiveness and understand underlying prescribing patterns among clinicians. In particular, building a link between the pain parameters described in this study and the prescription of particular medications can create a safer pain management protocol and identify situations in which an inappropriate opioid prescription is taking place.

This study successfully and efficiently extracted granular pain information from clinical notes in EHRs. By generating the ontology of pain and using that to customize the NLP pipeline, this study makes a major contribution to the NLP and pain community at large. This will help to increase the accessibility and the feasibility of large scale pain research to help researches better understand pain and its presentation in various disease states. This information can be used to help patients achieve better pain management and help providers prescribe safer and more effective analgesics. The ability to generate an individualized NLP model to extract pain symptoms from clinical notes from clinician-supported ontology is a major innovation that can be used to study other conditions with primary presentations found in clinical notes.

We hope that this study will provide the pain community at large with potential strategies and tools to better understand and manage pain in its various forms.

Acknowledgments

The authors would like to thank Dr. Reinhard Laubenbacher (University of Florida, formerly at the University of Connecticut) for the insight discussion, and Drs. Rebecca Andrews and Marilyn A. Katz (University of Connecticut) for clinical guidance.

References

1. Fillingim RB. Individual Differences in pain: Understanding the mosaic that makes pain personal. *Pain* 2017; 158:S11-S118.
2. Volkow ND. Characteristics of opioid prescriptions in 2009. *JAMA* 2011; 305:1299-1301.
3. Phillips CJ. The cost and burden of chronic pain. *Rev Pain* 2009; 3:2-5.
4. Ellis RJ, Wang Z, Genes N, Ma'ayan A. Predicting opioid dependence from electronic health records with machine learning. *BioData Min* 2019; 12:3.
5. Fodeh SJ, Finch D, Bouayad L, et al. Classifying clinical notes with pain assessment using machine learning. *Med Biol Eng Comput* 2018; 56:1285-1292.
6. Carrell DS, Cronkite D, Palmer RE, et al. Using natural language processing to identify problem usage of prescription opioids. *Int J Med Inform* 2015; 84:1057-1064.
7. Cook MJ, Yao L, Wang X. Facilitating accurate health provider directories using natural language processing. *BMC Med Inform Decis Mak* 2019; 19:80.
8. Heintzelman NH, Taylor RJ, Simonsen L, et al. Longitudinal analysis of pain in patients with metastatic prostate cancer using natural language processing of

medical record text. *J Am Med Inform Assoc* 2013; 20:898-905.

9. Wang X, Chused A, Elhadad N, Friedman C, Markatou M. Automated knowledge acquisition from clinical narrative reports. *AMIA Annu Symp Proc* 2008; 2008:783-787.
10. Wang X, Hripcsak G, Markatou M, Friedman C. Active computerized pharmacovigilance using natural language processing, statistics, and electronic health records: A feasibility study. *J Am Med Inform Assoc* 2009; 16:328-337.
11. Forsyth AW, Barzilay R, Hughes KS, et al. Machine learning methods to extract documentation of breast cancer symptoms from electronic health records. *J Pain Symptom Manage* 2018; 55:1492-1499.
12. Soysal E, Wang J, Jiang M, et al. CLAMP - a toolkit for efficiently building customized clinical natural language processing pipelines. *J Am Med Inform Assoc* 2018; 25:331-336.
13. R: The R Project for Statistical Computing. Available from: www.r-project.org/
14. Cohen SP, Vase L, Hooten WM. Chronic pain: An update on burden, best practices, and new advances. *Lancet* 2021; 397:2082-2097.
15. Koleck TA, Dreisbach C, Bourne PE, Bakken S. Natural language processing of symptoms documented in free-text narratives of electronic health records: A systematic review. *J Am Med Inform Assoc* 2019; 26:364-379.
16. Fu S, Lopes GS, Pagali SR, et al. Ascertainment of delirium status using natural language processing from electronic health records. *J Gerontol A Biol Sci Med* 2020 Available from: https://academic.oup.com/biomedgerontology/advance-article/doi/10.1093/gerona/glaa275/5943765?login=false
17. Zhao SS, Hong C, Cai T, et al. Incorporating natural language processing to improve classification of axial spondyloarthritis using electronic health records. *Rheumatology (Oxford)* 2020; 59:1059-1065.
18. Kost JA, Ruaño G. Pharmacogenetics and the personalization of pain management: A potential role in precision opioid treatment. *CT Medicine* 2020; 84:13-18.
19. Malfait F, Castori M, Francomano CA, Giunta C, Kosho T, Byers PH. The Ehlers–Danlos syndromes. *Nat Rev Dis Primers* 2020; 6:1-25.
20. Chopra P, Tinkle B, Hamonet C, et al. Pain management in the Ehlers–Danlos syndromes. *Am J Med Genet C Semin Med Genet* 2017; 175:212-219.
21. Shoenbill K, Song Y, Gress L, Johnson H, Smith M, Mendonca EA. Natural language processing of lifestyle modification documentation. *Health Informatics J* 2020; 26:388-405.