Capturing the Patient's Perspective: a Review of Advances in Natural Language Processing of Health-Related Text

G. Gonzalez-Hernandez[1], A. Sarker[1], K. O'Connor[1], G. Savova[2]

[1] Department of Epidemiology, Biostatistics, and Informatics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[2] Boston Children's Hospital and Harvard Medical School, Boston, MA, USA

Summary

Background: Natural Language Processing (NLP) methods are increasingly being utilized to mine knowledge from unstructured health-related texts. Recent advances in noisy text processing techniques are enabling researchers and medical domain experts to go beyond the information encapsulated in published texts (e.g., clinical trials and systematic reviews) and structured questionnaires, and obtain perspectives from other unstructured sources such as Electronic Health Records (EHRs) and social media posts.

Objectives: To review the recently published literature discussing the application of NLP techniques for mining health-related information from EHRs and social media posts.

Methods: Literature review included the research published over the last five years based on searches of PubMed, conference proceedings, and the ACM Digital Library, as well as on relevant publications referenced in papers. We particularly focused on the techniques employed on EHRs and social media data.

Results: A set of 62 studies involving EHRs and 87 studies involving social media matched our criteria and were included in this paper. We present the purposes of these studies, outline the key NLP contributions, and discuss the general trends observed in the field, the current state of research, and important outstanding problems.

Conclusions: Over the recent years, there has been a continuing transition from lexical and rule-based systems to learning-based approaches, because of the growth of annotated data sets and advances in data science. For EHRs, publicly available annotated data is still scarce and this acts as an obstacle to research progress. On the contrary, research on social media mining has seen a rapid growth, particularly because the large amount of unlabeled data available via this resource compensates for the uncertainty inherent to the data. Effective mechanisms to filter out noise and for mapping social media expressions to standard medical concepts are crucial and latent research problems. Shared tasks and other competitive challenges have been driving factors behind the implementation of open systems, and they are likely to play an imperative role in the development of future systems.

Keywords

Natural language processing review; medical terms; social media; mining electronic health records

Yearb Med Inform 2017:214-27
http://dx.doi.org/10.15265/IY-2017-029
Published online August 18, 2017

1 Introduction

The need to embrace the patient's perspective in health-related research and quality of care measures is one point on which all major health organizations around the world agree. Indeed, the World Health Organization in Europe prominently lists "*patient and community participation or direction*" [1] as a practical and proven quality improvement approach in their guidelines to developing quality and safety strategies. Relevant to the approaches we review here, guidelines note that "*work would be needed to find the best ways to introduce* [patient and community participation], *building on local culture and existing structures*." Around the world, numerous initiatives were launched in the last 10 years to incorporate the patient's perspective, including efforts such as patient survey programs by the Care Quality Commission (CQC) in England, the "Better Together" effort of the Scottish Government, the Institute for Quality and Efficiency in Health Care (IQWiG) in Germany, and the Patient-Centered Outcomes Research Institute (PCORI) and the Patient-Focused Drug Development (PFDD) programs in the United States. Patient-reported outcomes (PROs) [2] and measurement instruments, which are extensively validated questionnaires to measure patients' symptoms and quality of life such as the PROMIS [3, 4] set designed by the NIH, have become the standard way to collect the patient's perspective. However, distributing and getting an adequate number of responses to these questionnaires is a constant challenge, and does not necessarily build "*on local culture and existing structures*".

In this paper, we reviewed the recently published literature discussing Natural Language Processing (NLP) methods that could help go beyond structured questionnaires to find, extract, and incorporate information related to the patient's perspective from unstructured fields (text) found in different sources such as social media, patient speech, patient/therapist interactions, and the information recorded in the free text fields of clinical records. This overcomes the problem of the targeted distribution and collection of questionnaires by proactively extracting information reflecting the patient's perspective where it is already present. We also reviewed advances in the closely related field of machine learning, which offered several milestone developments in the recent past, such as cognitively grounded semi-supervised or unsupervised methods. In many cases, the NLP community embraced and adapted these automated learning methods to the specifics of NLP tasks, given that such methods are now capable of utilizing information from labeled and unlabeled data. Relevant to the specific focus of this review, these methods enable the use of

large sets of clinical records and health-related user-generated content from social media microblogs (Twitter, Facebook, Reddit, and Instagram, to name a few) and health forums (such as DailyStrength, WebMD, and others).

The review by Meystre et al. [5] presents an excellent overview of health-related text processing and its applications until 2007. The research presented in the present manuscript includes, for social media and clinical records, advances to fundamental NLP methods such as classification, concept extraction, and normalization published since 2008, mostly omitting what has been included in similar recent reviews [6–8], except when required to complete and highlight recent advances. For each data source, after reviewing advances in fundamental methods, we reviewed specific applications that capture the patient's perspective for specific conditions, treatments, or phenotypes. For the applications, we built on the 2016 review by Demner-Fushman and Elhadad [7] and highlighted major achievements from 2013 to 2016 that are relevant to the patient's perspective focus of this review. The search and selection criteria used were similar to the ones used by Névéol and Zweigenbaum [8], from January 1st, 2013 through December 31st, 2016, resulting in 464 papers. A total of 62 papers focusing on clinical records were selected from this set. For social media, most of the literature has been published within this same time frame; a total of 87 papers were included in this review. Research was included if it was published within the time window of interest and it was, to the best of our judgment, classified as either: (1) an advance to fundamental methods for health-related natural language processing (Health Language Processing, or HLP hereafter) and is applicable to more than a single condition, treatment, or phenotype, or (2) an HLP method applied to specific conditions, treatments, or phenotypes, but focused on capturing the patient's perspective. References dated outside the window of interest were included when required, and their use does not imply they are part of the set of papers defined by this inclusion criterion.

We understand the patient's perspective to include reports of disease progression and manifestation (symptoms) given directly by the patient, which corresponds to the definition of patient-reported outcomes given by the FDA [9], although the level of detail and completeness of information collected using HLP methods will often lag behind what can be collected with patient-reported outcome instruments. However, we also consider of interest for most secondary use of heath data the *collective* patient perspective, "*an understanding of the disease experiences of many patients*" [10] that share specific characteristics. Thus, either written by the patients themselves as in social media, or extracted from clinical notes written by a health provider, unstructured text captures the health narrative from the patient and reflects the characteristics shared by a group of patients (such as suffering from a specific disease or symptom, or smoking status). HLP allows the collection of aspects of the patient's perspective that are more complex than what surveys are designed to collect. Consider for example the nuanced descriptions of quality of life and mental state, such as the ones in Figure 1, which are postings in social media from someone with inflammatory bowel disease/colitis. The postings speak about general routine exercise, well-being, and attitude, in a way that would be difficult to capture in a survey, and they are potentially available in much greater volume, over longer periods of time, and for a large number of patients in social media. Going beyond the structure of surveys greatly increases the number of reports (and derived knowledge) that can be incorporated, aggregating many more patient reports of cohorts with some unifying characteristic, such as a specific condition. Finally, we include time as a unique dimension of the patient's perspective, and review HLP methods tailored to capturing events related in time (for example, the patient reporting a pre-existing condition, or a recent treatment with antibiotics), or specific moments in a "health timeline" (such as when pregnancy was revealed, or when the flu shot was taken).

The paper is organized in two main parts. Part 1 focuses on NLP of electronic health records (EHRs) and Part 2 focuses on NLP of health-related social media text. Within each part, we describe the developments in fundamental tasks and various applications. Our intent is to provide the reader with a scope of the trends and advances in capturing the patient's perspective on health within the last three years, as outlined before.

Jan 12 6:23 am *"This morning got off to a good start with deadlifts. Snatches and rowing will be on the agenda between classes"*

Jan 10 4:16 am *"4:15 AM Let's do this,"*

Jan 4, at 9:08pm *"Resolved: to not let the colitis sidetrack my work and to finally master it."*

Jan 3 6:46 am *"Deadlifts - always humbling."*

Jan 3 4:25 am *"4:25 AM: Good morning everyone. Let's get after it."*

Jan 2 6:25 am *"First day back post flare. After a six week break it's going to be a difficult uphill battle. Swings, presses and squats."*

Fig. 1 Postings in a timeline progressing after a colitis flare-up that lasted 6 weeks.

2 Mining Electronic Health Records

While one could argue that the EHR does not provide a true "patient perspective" according to the FDA's definition [9], as its contents are invariably captured by a health care provider, we posit that it contains some unique aspects of the patient perspective: the manifestation of symptoms unique to the patient, or a narrative of what the patient expressed (i.e., "patient complains of vertigo…"). In aggregate, EHRs constitute an important source of the collective patient perspective, as defined before. In order to capture this, however, it is necessary to pinpoint the concepts of interest, map them to standardized vocabularies, and put them in temporal context (a timeline) in order to then derive further knowledge. We outline the advances on each of these tasks next.

2.1 Concept Extraction and Normalization

The field has long advanced beyond the simple keyword search functionalities that are widely implemented in commercial EHR systems as well as research platforms such as i2b2 [11]. The limitations of key-

word search strategies are well-known. A keyword-based retrieval will only match documents that contain the keyword exactly as typed (with some limited flexibility allowed – such as suggestions when alternate spellings are found). Advanced NLP enables finding concepts rather than keywords. A concept-based retrieval system can identify a concept expressed in many different ways (including synonyms as well as generalizations or specifications of a concept). For example, if EHRs mentioning adverse reactions to a specific medication are needed, the researcher would have to include keywords to cover all variants and specific types of reactions in order to use a keyword-based search. The situation is even worse if, for instance, a cohort of people involved in a motor vehicle accident is the desired set. In a keyword-based system, the user would need to specify one by one all motor vehicles, while an efficient concept-based system will retrieve them automatically if "motor vehicle" is specified as the query. Only a handful of systems are currently generally available to automatically find concepts in free text from clinical narratives, the most established being MedLEE (Medical Language Extraction and Encoding) [12], and cTAKES [13]. Other systems are Noble Coder [14], CLAMP (http://clamp.uth.edu/), MetaMap Lite (https://metamap.nlm.nih.gov/download/new/), and DNorm adaptation for clinical text [15].

Normalization of concept mentions involves matching the extracted concept to a unique identifier from a medical ontology, usually one from the UMLS Metathesaurus [16]. The UMLS provides a vast coverage of medical terminology in English, and it has been used previously for disease normalization [17–19]. Recently, other lexicons have been developed for normalization in other languages, for example the QUAERO French medical corpus [20], the CLEF eHealth information extraction tasks of 2015 and 2016 (clinical entity and concept recognition from French, ICD-10 coding of causes of death for French), and the NTCIR [21] 2015 and 2016 tasks on ICD-10 coding in Japanese. Regardless of the language, most of the research to date on concept normalization has used some variations of dictionary lookup techniques and string matching algorithms.

With the advances in machine learning techniques and the increased availability of annotated data, recent approaches have used learning-based algorithms to improve basic dictionary matching techniques. For tasks such as gene name normalization, some of these works have involved list-wise learning, which learns the best list of objects associated with a concept and returns the list rather than a single object [22, 23], graph-based normalization [24], conditional random fields [25], regression-based methods [26], and semantic similarity techniques [27]. Leaman et al. [28] applied pairwise learning from a specialized disease corpus for disease name normalization. Mapping free text to concepts in an ontology has been done by Gobbels et al. [29] who used a Naïve Bayes machine learning system to match phrases from clinical records to SNOMED ontology terms, whereas Kate [30] used learned edit distance patterns to normalize clinical text to UMLS IDs. Other approaches have used tools such as MetaMap [31] or cTAKES [13], to extract and map terms to concepts in the UMLS [32, 33]. The tools are generally effective for clinical text, but are not portable: they miss relevant information when applied to colloquial language as the one used in social media [34].

Several shared tasks have taken on the challenge of normalization using clinical texts. Using the ShARe corpus, participants in the 2014 ShARe/CLEF Task 2 were tasked with normalizing semantic modifiers related to disease mentions in clinical texts [35]. The submitted systems obtained accuracies ranging from 0.769 to 0.868 [36]. The SemEval 2014 Task 7 challenged participants to automatically identify and normalize diseases and disorders in clinical texts [37]. The approaches used ranged from rule-based classifiers to hybrid rule-based/machine learning classifiers. The latter approach typically led to a higher performing system. For the top performing system, Zhang et al. [38] developed a machine learning NER module and an ensemble learning module used as a binary classifier to determine if the NER module output was a true positive for disorder entity recognition. For normalization, they developed a Vector Space Model to assign the most appropriate UMLS Concept Unique Identifier (CUI). The 2015 SemEval task related to clinical texts again presented the task of disorder entity recognition and normalization as well as template slot filling [39]. For disorder span recognition, CRF approaches were the most utilized. Pathak et al. [40] developed a CRF system to locate contiguous disorder mentions and for disjointed disorder mentions, a support vector machine (SVM) binary classifier was developed to determine if the two mentions were related or not. This approach allowed them to detect disjointed mentions with about 70% accuracy. For the normalization, they divided the task into three parts: a direct dictionary match, a dictionary match on a modified UMLS that was parsed into phrases, and a string similarity algorithm. Xu et al. [41] extended the work they had done in the prior year [38], including the training of a deep neural network using the unlabeled MIMIC II corpus to obtain word embeddings that were used as a feature in their system. Results are summarized in Table 1.

2.2 Timeline Extraction

The 2015 Clinical TempEval challenge presented the task of extracting temporal information from clinical notes. The participants were tasked with identifying time and event expressions and attributes of those expressions, and the temporal relations between them [42]. Three teams participated, with one participating in all tasks and subtasks. All systems employed supervised classifiers for the challenge and all outperformed the rule-based systems that were used as baselines. The top performing system [43] utilized a combination of CRF, SVM, and rule-based approaches. The 2016 Clinical TempEval presented similar tasks and utilized the same corpus as the prior year. However the number of participants increased [44]. The submitted systems used differing machine learning approaches with the top systems utilizing HMM SVM [45] and CRF [46]. As in the prior year, the top systems achieved a high level of performance for the time and event identification tasks, reaching F-scores close to those of human annotators. The temporal relation tasks proved to be more difficult;

Table 1 Relevant shared tasks results, information extraction, and normalization.

Shared Task Results					
IE and Normalization					
Challenge	Year	Corpus	Task Description	Number of Participants	Results Range / Measurement
ShARe/CLEF	2014	ShARe*	Normalization values of 10 attributes	10	0.769-0.768/Accuracy
Sem-Eval	2014	ShARe	Identification of disorder mentions	21	Strict: 0.787-0.448/F1
					Relaxed: 0.975-0.717/F1
SemEval	2014	ShARe	Normalization of mentions to SNOMED-CT	18	Strict: 0.716-0.299/ accuracy
					Relaxed: 0.923-0.584/ accuracy
SemEval	2015	ShARe	Identification of disorder mentions and normalization to CUI	16	Strict: 0.757-0.093/F1
					Relaxed: 0.787-.0364/F1

* http://share.healthnlp.org/ [accessed: April 1st, 2017]

however, marked progress was made from the prior year's systems. A summary of the results with the ShARe corpus is included in Table 1.

2.3 Knowledge Discovery

The fundamental methods outlined before are generally a basic requirement in any HLP pipeline for secondary use of health data. Recently, interesting directions have been explored where the collective patient perspective data from EHRs has been used to create disease progression and medical decision-making models, to assist in reaching potential diagnoses and assessing risk factors. We review here some of these HLP applications.

Disease Progression Models: To better understand disease evolution patterns, Lui et al. [47] proposed a temporal graph-based representation of patient EHR data. The graph framework's predictive power was tested on the onset risk of heart failure and on the risk of heart-related hospitalizations for COPD patients. On each of these two tasks the system obtained an AUC of 0.72. Using an EHR database of 300,000 patient records, Wang et al. [48] developed an unsupervised disease progression model using a combination of statistical methods. Applying the model to chronic obstructive pulmonary disease (COPD), they identified comorbidities and inferred a progression trajectory model of the disease. Pham et al. [49] created a system, DeepCare, to predict the next stage of disease progression and the risk of unplanned readmissions. The system models illness trajectories and predicts future outcomes. Based on a deep dynamic neural network, it incorporates irregular timing and interventions to infer future prognosis. Tested on a cohort of 12,000 diabetic patients, the system prediction of readmissions achieved an F-score of 0.791, which was an improvement over the baseline method score by 0.077.

Decision Support: For creating a decision-making model, Liang et al. [50] proposed a modified deep belief network to simulate human thinking procedures. Features were extracted through an unsupervised method and a supervised method, SVM, was used for final decision-making. This semi-supervised approach outperformed standard SVM and decision tree methods on EHR data. Wang et al. [51] used clinical notes of the Maine Health Information Exchange (HIE) EMR database to train and test an NLP system to find uncoded cases of congestive heart failure (CHF). The classifier identified 2,411 instances of CHF out of 253,804 cases. The system had and F-score of 0.753, which outperformed prior methods.

Risk Assessment: Karmaker et al. [52] used machine learning to analyze EHRs to assess the risk of suicide based on physical illness. After extracting illnesses based on ICD-10 codes, they developed six modules, differentiated by range of time of included illnesses, to predict suicide risk. The maximum AUC, 0.71, was obtained by incorporating illnesses across all time in the model. Overall, improvements in risk assessment were seen over increasing time ranges and all models were an improvement over the AUC of the clinically assessed score. In assessing the risk of coronary artery disease (CAD), Jonnagaddala et al. [53] developed a rule-based text mining system to identify and extract Framingham risk factors from the unstructured text of EHRs in a cohort of diabetic patients. The results were used to calculate the Framingham risk score (FRS). This approach showed the feasibility of extracting such information, however, it did have some limitations including the lack of temporal information and a lack of clinical context for the information extracted. An extension of this study [54] employed machine learning components to assign an indicator and time attributes, where applicable. A Naïve Bayes supervised classifier was developed to assign the time attribute. The system achieved an overall micro-averaged F-score of 0.83. Chen et al. [55] developed a hybrid system based on machine learning and a rule-based approach to identify risk factors for heart disease in the same cohort of patients. A pipeline system that included the use of SSVMs, SVMs, and CRF was used to extract information. Overall, this system achieved an F-score of 0.9268.

Risk Prevention: Due to the readmission penalty program initiated by the Centers for Medicaid and Medicare, providers have attempted to assess patients at high risk for readmissions. These patients are targeted to receive enhanced interventions. In an effort to automate the detection of such patients, Zheng et al. [56] compared neural networks, random forests, and a hybrid model of intelligent swarm heuristics and support vector machines (PSO-SVM) in a

cohort of heart failure patients. The PSO-SVM outperformed the other methods with an accuracy of 78.4%. It improved upon the currently used LACE score which has an accuracy of 43.5%. Futoma et al. [57] tested five models to determine which would perform best at predicting patients at high risk for readmissions. The models were tested on 280 patient cohorts determined by diagnosis-related groups (DRG). They found that deep neural networks outperformed regression models and had consistently better AUCs over the cohorts. They also observed a slightly better AUC for predictions based on individual DRGs rather than on the data as a whole.

Adverse Drug Effects Discovery (ADE): Drug safety studies have also used EHR data. From the free text of psychiatric EHRs, Iqbal et al. [58] used the GATE NLP framework to identify possible extrapyramidal side effects with dictionary matching and a rule-based approach. ADEs were detected with an 85% precision and 86% recall, and the results were used to perform a secondary analysis of the prevalence of ADEs based on subgroups of patients. However, they did not attempt to assess causality to determine the probability of discovering true drug-ADE pairs. To address the problem of confounders, Li et al. [59] used the NLP system MEDLee to structure and encode narrative notes. Temporal information was also extracted. A statistical method used penalized logistic regression to estimate confounder-adjusted ADE associations. This method identified several drug safety signals that warranted further clinical review. Wang et al. [60] developed a discriminative classifier to automatically detect potential drug-ADE pairs. After constructing a set of features, three classifiers were tested and the random fields classifier was determined to be superior. The classifier achieved an AUC of 0.94, which exceeded 0.79, the AUC of the method used by the FDA's Adverse Event Reporting System (FAERS).

Off-label Use Discovery: Clinical notes have also been used to detect off-label drug use by using NLP methods to extract used-to-treat mentions. Jung et al. [61] trained an SVM classifier to detect potential off-label usage from 9.5 million free text clinical notes. After filtering for known drug-indication pairs and drug-adverse event pairs, the system identified 403 novel off-label usages. Drug repurposing signals have also been validated using EHR data. To validate reports that metformin improves cancer survival and reduces cancer risk, Xu et al. [62] automatically extracted information from a cohort of cancer patients, including diabetes status and other covariates including drug exposure. A stratified Cox regression model was used to assess metformin influence on cancer survival probabilities and confirmed reports associating its use with lower cancer mortality.

Cancer-related Information Extraction: A number of studies use various techniques to extract information from cancer-relevant clinical text. A group of studies employ document classification techniques to discover various oncology categories. This approach is taken mainly because there is document-level gold data already available for the purpose of cancer registry abstraction. Yala et al. [63] extracted breast cancer-related information from pathology notes for the following types: (a) Diseases/disorders such as *Ductal Carcinoma In Situ* (*DCIS), Invasive Lobular Carcinoma (ILC), carcinoma, Lobular Carcinoma In Situ (LCIS), Atypical Ductal Hyperplasia (ADH), lobular neoplasia, flat epithelial atypia, blunt adenosis, atypia,* (b) positive and negative lymph nodes, (c) biomarkers/receptors and their values such as *Estrogen Receptors (ER)*, *Progesterone Receptors (PR)*, and *HER2*, and (d) breast side/laterality. They used machine learning where the features were the n-grams from the pathology reports and the classification label was one of (a)-(d). The gold classification labels were created in previous breast cancer studies. Weegar and Dalianis [64] created a pilot rule-based system for information extraction from breast cancer pathology notes in Norwegian – sentinel nodes, axillary nodes, tumor size, histological grade, ER, PR, Ki67, pT. The system was trained and tested on a very small dataset and conceived as a pilot study to generate fodder for a more sophisticated system whose architecture is presented by the authors. Ou and Patrick [65] built a system for information extraction from melanoma pathology reports. The system passed noun phrases produced by the GENIA tagger in a pre-processing step to a conditional random fields classifier that detected named entities. Reported performance was on-par with that of humans. However, it is not clear whether the named entities were further linked by relations, e.g., whether a named entity mention of type *site and laterality* was linked to a specific tumor. The very recent work of the DeepPhe team (cancer.healthnlp.org) moves beyond entity mention-level recognition to episode- and patient- levels over the entire set of patient records (pathology, oncology, clinical, and radiology reports) where multiple tumors associated with types of cancer are described [66]. These tasks require sophisticated extraction techniques such as coreference resolution, relation extraction, and temporal relation extraction to achieve reliable summarization.

Temporal Data (see Table 2): Using a temporal database to discover the possible patterns that might point to cause and effect events, Wang et al. [67] created a system to align and visualize multiple patients' records by event, allowing for a way to more readily find precursor, co-occurring, and after-effect events. To extract relevant temporal and event information from clinical narratives in an EHR, NLP and machine learning techniques have been employed. Nikfarjam et al. [68] proposed using an SVM classifier with a graph-based approach to find temporal relations. Kovecevic et al. [69] proposed to use conditional random fields (CRF) and post-processing rules to identify mentions of clinical events and temporal expressions. Longitudinal data has also been explored as a resource to detect drug safety information. Schumie et al. [70] used Bayesian methods to detect possible drug safety signals in longitudinal observational healthcare records. They reported that the use of temporal data enables their method to distinguish between false drug-event associations, such as a protopathic bias, and genuine adverse effects. To find unreported ADEs in EHRs, Zhao [71] proposed using time-stamped clinical events as a feature in a supervised machine learning algorithm. Assigning weights to temporal relationships between the clinical event and an ADE increased the predictive power of the system over no weighting at all. Chen et al. [72] combined the temporal information

of structured EHRs with that automatically extracted from the clinical text through state-of-the-art NLP techniques [73] to identify rheumatoid arthritis patients with liver toxicity side effects as a result of the administration of methotrexate. This innovation sets their work apart from the computable phenotyping work done within national initiatives such as eMERGE and i2b2 where the phenotypes are temporally non-sensitive.

3 Mining Social Media

Social media has seen a massive growth, perhaps the greatest among all health-related information sources in HLP research in recent years [74], primarily driven by the rapid increase of the number of users. According to the latest Pew report [75], nearly half of the adults worldwide and two-thirds of all American adults (65%) use social media, including over 90% of 18-29-year old persons. The number continues to rise every year as larger numbers of older people commence interacting on social media. Earlier research from the same organization suggests that seven out of ten adult Internet users adults search health-related information on the Internet, one in four read about others' health experiences, and 16% go online to find users with similar health-related experiences [76]. It has been realized that because of the large user-bases that popular social networks have, there is an abundance of health-related knowledge contained within this domain in the form of text. Crucially, social media has opened up unique opportunities in patient-oriented health care—by allowing the access of information directly from users about various health-related topics and empowering the development of data-centric NLP techniques that can take into account patients' perspectives in unprecedented ways [77].

From the perspective of health, there are two broad categories of social media sources—generic social networks such as Facebook, Twitter, and Instagram, and domain-specific social networks such as PatientsLikeMe[1] and DailyStrength[2]. While generic social networks (GSNs) contain information about a range of topics, domain-specific networks, often referred to as online health communities (OHCs), are dedicated exclusively for discussions associated with health. Thus, GSNs typically provide access to data from large groups of users and may provide access to patients' perspectives that are not available from any other sources, on a wide range of topics [78], and in distinct languages [79]. For example, GSNs such as Twitter provide unique windows for researchers to study population-level attitudes and behaviors regarding prescription and illicit drug abuse and misuse [80], data that may not be available from traditional sources such as published literature, surveys, or EHRs. OHCs, in contrast, have much smaller user bases, but they include users with common interests/problems/objectives and usually provide cleaner and more targeted data (e.g., breast cancer forums).

Despite the profusion of health-related information available from social media, automatic processing of this data has made relatively slow progress. Social media text is generally noisy and unwieldy [81], given the presence of domain-specific terminologies, semantic information and complex language usage [82–84], the challenges are exacerbated when extracting health information from social media [85, 86]. Recent NLP research suggests that when used to solve specific clinical or public health associated tasks, only small proportions of social media data selected for the given task are useful. A significant proportion of the data is noise or unreliable, even when collected using carefully designed queries [87]. Health chatter on social media is also jam-packed with colloquialisms and misspellings, making detection and machine-level understanding of important concepts difficult. Furthermore, particularly for GSNs, user posts often lack contextual information, aggravating the difficulty of machine-level semantic understanding of the texts. NLP research in this domain has thus primarily been targeted

Table 2 Results from shared tasks related to temporal relation extraction from clinical records.

Temporal Extraction					
Challenge	**Year**	**Corpus**	**Task Description**	**Number of Participants**	**Results Range / Measurement**
i2b2	2012	i2b2[1]	Identification of time expressions	10	0.92-0.83/F1[α]
			Identification of event expressions	10	0.66-0.45/F1[α]
			Identification of temporal relations	10	0.69-0.43/F1
SemEval	2015	THYME[2]	Identification of time expressions	3	0.725-0.404/F1[α]
			Identification of event expressions	1	0.875/F1[α]
			Identification of temporal relations	1	Document Time: 0.702/F1
					Narrative Containers: 0.102
SemEval	2016	THYME[3]	Identification of time expressions	10	0.795-0.118/F1[α]
			Identification of event expressions	10	0.903-.0755/F1[α]
			Identification of temporal relations	10	Document Time: 0.756-0326/F1
					Narrative Containers: 0.479-0.017/F1

[1] https://ww-healthnlp.org/ [accessed: April 1st, 2017]
[2] https://github.com/stylerw/thymedata/ [accessed: April 1st, 2017]
[3] http://thyme.healthnlp.org/ [accessed: April 1st, 2017]
[α] Results reported are for span extraction only.

[1] https://www.patientslikeme.com/ [accessed: April 1st, 2017]
[2] https://www.dailystrength.org/ [accessed: April 1st, 2017]

towards discovering knowledge from the abundance of noisy, imbalanced data. Social media NLP research has focused on a variety of techniques such as query formulation and keyword selection for the detection of targeted content, generative techniques such as Latent Dirichlet Allocation (LDA) and variants of topic modeling techniques to identify common themes across large data sets, and supervised text classification techniques for filtering out noise and irrelevant data. Innovative entity recognition and extraction techniques have been proposed for health concept detection/identification tasks. More recent research contributions have made progress in concept normalization, signal generation, and predictive analytics via the application of sophisticated text mining pipelines. We review the most recent of these advances in this section and discuss some of the data, resources, and tools that are currently available for NLP of social media health data. We begin by discussing some of the most commonly addressed topics that social-media-based health text mining has addressed.

3.1 Social Media Sources, Topics, and Data Acquisition

Within the two broad categories of social networks (i.e., GSNs and OHCs), there are notable differences in many key attributes of the individual sources. Within GSNs, for example, Facebook is a broad coverage social networking site, Twitter is for microblogging, and Instagram is for photo sharing [88]. The data sharing and privacy policies of these media influence how widely they are used for research, with Twitter being the most popular because of its public streaming API. A recent systematic review by Sinnenberg et al. [74] identified 137 peer-reviewed research articles, which utilized Twitter for health-related research, with 57% of them focusing on contents while the rest focused on recruitment or interventions. Over 80% of the identified publications were either focused on lexical content analysis or surveillance, and 108 of the articles that focused on the contents, represented 5.1 billion tweets—a monumental amount of information. Population/public health topics are most commonly addressed within the Twitter articles identified and the same trend can be seen for Facebook and Instagram, although one key difference we observed is that a significant portion of public health research using Facebook focused on communication rather than on lexical content processing [89–91]. For monitoring and surveillance research from social media, the most common topic has been influenza surveillance [92, 93]. Data collection strategies for this task and other similar tasks (e.g., identifying health threats [94], drug interactions [95], smoking patterns [96, 97], pharmacovigilance [98, 99]) have been typically employing simplistic NLP techniques, relying on keywords and/or hashtags as queries or on the direct selection of users via network links (e.g., followers of a brand of e-cigarettes). While for some research tasks such querying techniques suffice, other recent studies, particularly in the field of medication-effect analysis where the recall is generally low, have devised techniques to address the social media specific challenge of misspellings [100]. Pimpalkhute et al. [101] proposed a phonetic spelling variant generator that automatically generates common misspellings given a term. While the system has been used for collecting medication-related chatter from Twitter [102] and personal health messages [103], it may be applied to a variety of other data collection tasks. Data collection from OHCs has not faced similar challenges, as posts are usually categorized/structured, and the strategies employed have been simpler. However, because of the privacy policies of such sources, publicly available NLP data sets are scarce.

3.2 Content Analysis and Text Classification

Over half of the studies involving social media data that are cited in this paper employ lexical content analysis. A large subset of these studies has merely relied on collecting data in bulk using appropriate queries, and then deriving conclusions using simple statistical models directly based on the volume of data [104]. Sometimes, the information is coupled with available metadata (e.g., geolocation for flu surveillance) [105,106]. In some cases, studies involving big data from social media have overcome the challenges faced by studies using big data from other sources (e.g., Google Flu Trends, its shortcomings, and the use of Twitter for flu surveillance [107]).

Early health-related NLP research utilizing social media data suggests that rule-based approaches perform particularly poorly in this domain because of the nature of the data, resulting in a shift towards machine-learning-based approaches in recent years [100]. Text classification techniques have been applied to extract high quality information from noisy, health-related social media data [108] or for other downstream tasks in both GSNs (e.g., [109]) and OHCs (e.g., [110]). These downstream tasks include, but are by no means limited to, sentiment analysis [111–113], adverse drug reaction detection [114, 115], antibiotic discussion characterization [116], prescription medication abuse detection [117], substance use (e.g., tobacco products) classification [96, 118], personal event detection [119–121], and user behavior analysis [122]. Sentiment analysis/classification is a well explored NLP task, and, because of its particular suitability to large social media data and the availability of annotated resources (e.g., data from the SemEval task [123]), it has found applications in an assortment of inter-domain tasks. For example, recently, sentiment analysis/classification techniques have been applied to understanding user sentiments associated with drugs [124, 125], including non-medical use of drugs [126] and vaccinations [127], treatments [111, 128], and diseases [129].

In terms of methods for content analysis and filtering, supervised learning techniques that incorporate informative features have been the most popular, although many studies still rely on a manual analysis for deriving final conclusions [130]. Many early studies discovered that the high amount of noise present in social media data posed an important problem for mining knowledge, and they attempted to address the issue by designing supervised classification techniques [131]. Traditional text classification features such as bag-of-words and n-grams are most commonly used and social media specific features such as emoticons have also

been found to be informative for particular tasks [132]. Sarker and Gonzalez showed that social media text classification, particularly for short text nuggets such as Twitter posts, benefits from the generation of large numbers of semantic features representing information about the texts from multiple aspects such as topics, polarities, sentiments, and others. The short nature of Twitter posts along with the presence of frequent misspellings commonly result in extremely sparse vectors, and thus generating many features to represent diverse properties of the text enables classifiers to learn better fitting models. Another interesting aspect of social-media-based health text classification is that medical domain-specific tools such as MetaMap [133], which have been used in the past for generating features for text classification, are not very useful when it comes to social media text. This is primarily because such tools are designed for formal medical text only, and are incapable of understanding social media expressions. Unsurprisingly, SVMs have obtained the best performance for a long time and are still very difficult to beat in terms of performance [108, 114, 117, 134–136]. Very recently, deep neural-network-based models have been designed and employed for social media health text classification, with very promising results [137].

Another effective and popular social media lexical content analysis approach has been topic modeling. Unlike text classification algorithms, which require supervision/annotated data, topic models utilize large volumes of unlabeled data to identify topics that represent that data. Therefore, topic modeling techniques can be readily employed to identify and analyze the contents of targeted social media health chatter. Paul and Dredze [138], for example, proposed the Ailment Topic Aspect Model (ATAM) to identify health-related topics from Twitter. Using this technique, the authors discovered some of the most popular health-related topics of Twitter conversations, such as influenza, allergies, and obesity. The approach has been replicated, modified, and other topic modeling techniques have been employed to find health-related topics in languages other than English (e.g., Chinese [139]), and for focused tasks such as analyzing suicide-related content on social media [140] and characterizing discussions about vaccines [141]. Li and colleagues [142, 143] proposed topic models for identifying clusters of 'life events', and using them for building user timelines and other downstream applications. Other variants of topic models have also been applied for deriving task-specific knowledge from large, unlabeled social media data sets—such as the use of keyword-biased topic models for predicting the safety of dietary supplements [144] and time-sensitive probabilistic topic models for capturing changing user interests and topics over time from health-related chat logs [145]. Recent advances in deep neural-network-based models have seen the application of such models for the generation of topical content, which may then be applied to downstream tasks [146].

3.3 Information Extraction and Normalization

It was obvious early on in information extraction research from social media health text that traditional approaches performed poorly in this domain. Social media and health are both complex lexical domains and the intersection of the two particularly aggravates challenges associated with text mining. Rare and non-standard contents are specifically difficult for systems to extract and aggregate [78,79] particularly for pharmacovigilance, via the use of NLP. Because of the relatively new research focus on social media mining for health-related tasks, challenges associated with the extraction of pertinent, task-specific content have only been discovered/realized in the recent past.

In line with early biomedical NLP approaches, information extraction approaches specific to the social media domain mostly employed lexicon-based techniques to solve many problems, such as detecting adverse drug reactions [147], identifying users making pregnancy announcements [148], and mining opinions [149]. However, with the evolving nature of language usage on social media and the unconstrained number of ways in which the same information can be expressed in this domain, developing thorough lexicons is a difficult task. Additionally, even the most thorough lexicons may get outdated within a relatively short period of time. Recent efforts have attempted to combine multiple task-specific lexicons for improving information detection/extraction performance on target datasets, but these lexicon-based approaches are easily outperformed by learning-based approaches in the presence of expert annotated data [99]. Denecke's qualitative analysis [150] on concept extraction from social media details the specific problems associated with the use of standardized lexicons in this domain. The author explains that knowledge-bases/lexicons such as MetaMap and cTakes fail to identify common language or consumer health vocabulary. In particular, verbs, personal pronouns, adjectives, and connecting words present problematic contents.

With emerging efforts in the preparation of annotated social media based datasets for a variety of health-related and public health monitoring tasks, recent benchmark systems have employed supervised learning for concept extraction/sequence labeling. Conditional random fields (CRFs) currently produce the best performance on annotated data for the task in English and other languages [99, 112, 151, 152]. The primary reason behind the success is the ability of CRFs to incorporate contextual information when determining whether a given token should be classified as relevant or not. For example, concerning disease, disorders, or adverse drug reaction mentions in social media, while users may use a variety of creative terms to express their minds, similar concepts are likely to occur in similar contexts. In addition to context terms, such context-incorporating learning-based algorithms have shown improvements when generalized representations of the context tokens are provided. Recent advances in learning semantic representations from large unlabeled datasets [153] have aided social media mining research by allowing systems to identify semantically similar terms. Nikfarjam et al. [99], for example, learnt distributed representations of social media terms, as used within a domain, clustered the vectors, and used cluster numbers as features. The clusters contain the terms that are found to be close to each other in the

semantic space, and so, terms that are used in the same contexts are clustered together. Such features are particularly attractive because they don't require human annotations or the manual creation of resources, but can be learned automatically in the presence of large amounts of unlabeled data—of which there is abundance in social media. Therefore, such approaches are easily portable to other tasks or even to other languages. Morlane-Hondère et al. [152], for example, used the same features for French social media data.

One of the least explored topics in social media health language processing is perhaps normalization. As discussed in the previous sections, normalization is the task of grouping together lexical items that essentially represent the same concepts. Within the domain of medical NLP, tools based on lexicons and knowledge bases such as MetaMap [133] have been used for identifying and grouping distinct lexical representations of the same concepts. Learning-based approaches for normalization have been proposed and employed for tasks within the medical domain [21, 22, 23, 25, 27, 147, 148, 149], but research on this topic within the domain of social media is still in its infancy. While medical text is itself complex, social media text presents additional challenges due to non-standard expressions and spellings. Typos, ad hoc abbreviations, phonetic substitutions, use of colloquial language, ungrammatical structures, and even the use of emoticons make social media text significantly different from texts from other sources [154]. While these properties present NLP challenges, these also constitute the primary motivation for building normalization systems.

Some research on normalization of social media text focused at the lexical level, and has similarities to spell checking techniques with the primary difference that *out-of-vocabulary* terms in social media text are often intentionally generated. Text messages have been used as input data for normalization models, and various error models have been proposed, such as Hidden Markov Models [155] and noisy channel models [156]. Similar lexical normalization techniques have been evaluated on social media texts as well [157, 158]. For concept normalization of social media texts, approaches still mostly rely on custom-built lexicons [159, 160] and lexical matching. Metke-Jimenez and Karimi [159] presented a large, compiled lexicon of adverse drug reaction terms and employed term weighting techniques for retrieving relevant concepts. Very recently, with the development of new, efficient techniques for learning vector representations of terms/phrases, new mechanisms for performing normalization using vector representations of words have been proposed. In a typical normalization pipeline, word vector representations are generated from large unlabeled data sets and then similarity measurements, such as cosine similarity, are utilized to identify semantically similar concepts [103, 161]. In terms of methods, techniques adapted from the field of machine translation [161] and neural-network-based approaches [103, 162] are currently being explored. Although progress in this field is very limited, the importance of such techniques has been realized and there is currently a drive from the research community to release more data sets for evaluating social-media-based health text normalization systems [87, 159]. Considering the necessity of high-performance normalization systems for health text mining tasks from social media, the severe lack of research on this topic, and the recent advances in semantic representations of text nuggets, NLP research in the near future should, and inevitably will, focus on the development of more advanced concept normalization techniques.

3.4 Individual and Temporal Data in Social Media

Very recently, the detection of personal health-related life events and their extraction have received some attention as researchers are starting to focus on using social media data for precision medicine. Simple bag-of-words and n-gram models have been combined with temporal information from individuals' social media timelines to perform personal life event detection [119]. Unsupervised LDA models and human annotations have also been used to detect personal events [142]. Li and Cardie [143] proposed using an unsupervised Dirichlet Process Mixture Model to capture temporal and personal event mentions to create a chronological timeline from a user's Twitter feed. Kiciman and Richardson [120] combined several supervised learning methods to classify and extract personal event information mentions in Twitter. Combining this information with temporal information, they propose to create a system for creating action—outcome relationships from a user's timeline. Wen et al. [121] designed a temporal tagger to extract temporal expressions from sentences containing predefined event words from a patient's profile on an online cancer support community. A trained classifier was then used to normalize dates and sort mentions of the event in a chronological timeline. Chandrashekar et al. [163] attempted to combine query formulation, supervised classification, and rule-based information extraction techniques to identify and study pregnancy timelines. These studies have suggested that individuals belonging to specific cohorts can be detected from social media, and specific health-related events over time can be mined from the information posted by individual users. Future research will inevitably build on these studies and utilize social media data for personalized tasks.

3.5 Shared Tasks

There has been a move in recent times to design shared tasks and release data for social media health NLP tasks. Although, compared to the general health NLP domain, such efforts in the social media domain are very limited and relatively new. They present some of the first opportunities to compare distinct approaches and systems on social media tasks. Table 3 presents four such tasks that we decided to include in this review. A fifth task, on normalization of social media concepts, was proposed in 2016, but did not receive any participating systems [87]. Three of the tasks were on text classification and one on information extraction. Class-specific F-scores were typically used as the evaluation metrics, except in one task (Depression and Post-traumatic Stress Disorder (PTSD) detection). Unsurprisingly, the performances of the best performing systems on social media texts

Table 3 Performance of NLP systems on common data sets: best performances and types of tasks are shown.

Task	Task type	Score type	Performance upper limit/ range
Triaging mental health posts [164]	Classification	F-score (macro-averaged)	Shared task best: 0.42 [165]
Depression and PTSD detection [166]	Classification	AUC	Shared task best: 0.84 [167]
Adverse drug reaction mention detection [87]	Classification	F-score	Shared task best: 0.42 [168] Upper limits: 0.54 (Twitter) 0.68 (DailyStrength)
Adverse drug reaction extraction [87]	Information Extraction	F-Score	Shared task best: 0.61 [169] Upper limits: 0.72 (Twitter) 0.82 (DailyStrength)

are generally lower than the performances of similar systems in generic health-related texts. Despite this, with the growth of public release of annotated and targeted, unlabeled data sets, the development of better performing and more applicable systems is inevitable.

4 Conclusion

In this survey, we reviewed recent main developments in the field of health text processing. The survey focused on fundamental NLP tasks as well as higher-level applications related to extracting the patient's perspective from clinical records and social media.

One of the main challenges in the field is the availability of data that can be shared and which can be used by the community to push the development of methods based on comparable and reproducible studies. Shared data and open-source state-of-the-art methods go hand in hand in the quest of reproducible scientific discoveries.

From the applications point of view, NLP approaches that use EMR data would benefit from the direct use by clinical investigators for biomedical discoveries such as very large scale PheWAS/GWAS studies based on patient cohorts collected through automatic computable phenotypes. From a methods point of view, the field is now well aligned with the developments in the general NLP area (which was not the case a decade ago). State-of-the-art technologies from machine learning have infiltrated the domain. The next frontier tasks are the ones that require higher-level discourse processing such as co-reference and temporal relation extraction, as well as normalization.

With respect to the use of social media, despite the great interest it has elicited in the research community, when used to solve specific clinical or public health associated tasks, only small proportions of the data selected for the given task are actually useful. A significant proportion of the data is noise or unreliable, even when collected using carefully designed queries [87]. In order to obtain enough data, information retrieval methods with adaptable queries are needed, along with greater flexibility in concept extraction, and reliable normalization. While other state-of-the-art information extraction/identification approaches have been proposed in the machine learning and NLP literature, the scarcity of annotated social media data is still a major obstacle in assessing the true value of social media for the various health-related tasks. Novel, generic extraction algorithms for various health-related tasks have been proposed in the recent past [170], but their performance in real-life social media data have not been evaluated. This obstacle has been discussed in recent reviews [100] and there has been a greater urgency for creating and releasing annotated datasets and targeted unlabeled data sets in distinct languages [102, 159, 171].

Not until all of the fundamental tasks are readily available will knowledge discovery applications that use social media as a source flourish and be reliable enough to be used by regulatory agencies.

Acknowledgments

The authors are partially supported by funding from the US National Institutes of Health (1U24CA184407-01, R01LM10090, R01GM114355). The content is solely the responsibility of the authors and does not necessarily represent the official views of the US National Institutes of Health.

References

1. The World Health Organization Regional Office for Europe. Guidance on Developing Quality and Safety Strategies with a Health System Approach; 2008. http://www.euro.who.int/__data/assets/pdf_file/0011/96473/E91317.pdf. Accessed January 15, 2017.
2. Snyder CF, Jensen RE, Segal JB, Wu AW. Patient-Reported Outcomes (PROs): Putting the Patient Perspective in Patient-Centered Outcomes research. Med Care 2013;51(803):S73-S79.
3. Witter JP. The Promise of Patient-Reported Outcomes Measurement Information System-Turning Theory into Reality. A Uniform Approach to Patient-Reported Outcomes Across Rheumatic Diseases. Rheum Dis Clin North Am 2016;42(2):377-94.
4. Broderick JE, DeWitt EM, Rothrock N, Crane PK, Forrest CB. Advances in Patient-Reported Outcomes: The NIH PROMIS(®) Measures. EGEMS (Washington, DC) 2013;1(1):1015.
5. Meystre SM, Savova GK, Kipper-Schuler KC, Hurdle JF. Extracting Information from Textual Documents in the Electronic Health Record: a Review of Recent Research. Yearb Med Inform 2008:128-44.
6. Névéol A, Zweigenbaum P. Clinical Natural Language Processing in 2015: Leveraging the Variety of Texts of Clinical Interest. Yearb Med Inform 2016;(1):234-9.
7. Demner-Fushman D, Elhadad N. Aspiring to Unintended Consequences of Natural Language Processing: A Review of Recent Developments in Clinical and Consumer-Generated Text Processing. Yearb Med Inform 2016;(1):224-33.
8. Névéol A, Zweigenbaum P. Clinical Natural Language Processing in 2014: Foundational Methods Supporting Efficient Healthcare. Yearb Med Inform 2015;10(1):194-8.
9. Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. 2006:301-827. http://www.fda.gov/cder/guidance/index.htm. Accessed January 16, 2017.
10. NCI Dictionary of Cancer Terms - National Cancer

Institute. https://www.cancer.gov/publications/dictionaries/cancer-terms. Accessed April 26, 2017.
11. i2b2: Informatics for Integrating Biology and the Bedside. https://www.i2b2.org/. Accessed January 15, 2017.
12. Friedman C. A Broad-coverage Natural Language Processing System. Proc AMIA Symp 2000:270-4.
13. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): Architecture, Component Evaluation and Applications. J Am Med Informatics Assoc 2010;17(5):507-13.
14. Tseytlin E, Mitchell K, Legowski E, Corrigan J, Chavan G, Jacobson RS. NOBLE - Flexible Concept Recognition for Large-scale Biomedical Natural Language Processing. BMC Bioinformatics. 2016;17:32.
15. Leaman R, Khare R, Lu Z. Challenges in Clinical Natural Language Processing for Automated Disorder Normalization. J Biomed Inform 2015;57:28-37.
16. Bodenreider O. The Unified Medical Language System (UMLS): Integrating Biomedical Terminology. Nucleic Acids Res 2004;32(90001):267D-270.
17. Kang N, Singh B, Afzal Z, van Mulligen EM, Kors JA. Using Rule-based Natural Language Processing to Improve Disease Normalization in Biomedical Text. J Am Med Inform Assoc 2013;20(5):876-881.
18. Jimeno A, Jimenez-Ruiz E, Lee V, Gaudan S, Berlanga R, Rebholz-Schuhmann D. Assessment of Disease Named Entity Recognition on a Corpus of Annotated Sentences. BMC Bioinformatics 2008;9 Suppl 3:S3.
19. R. Leaman, C. Miller GG. Enabling Recognition of Diseases in Biomedical Text with Machine Learning: Corpus and Benchmark. In: 3rd International Symposium on Languages in Biology and Medicine Jeju Island, South Korea. 2009:82 89.
20. Névéol A, Grouin C, Leixa J, Rosset S, Zweigenbaum P. The Quaero French medical corpus: A ressource for medical entity recognition and normalization. PROC BIOTEXTM, REYKJAVIK. 2014. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.672.946. Accessed January 16, 2017.
21. Data/Tools | NTCIR. http://research.nii.ac.jp/ntcir/data/data-en.html. Accessed January 16, 2017.
22. Huang M, Liu J, Zhu X. GeneTUKit: a Software for Document-level Gene Normalization. Bioinformatics 2011;27(7):1032-3.
23. Huang M, Névéol A, Lu Z. Recommending MeSH Terms for Annotating Biomedical Articles. J Am Med Inform Assoc 18(5):660-7.
24. Sullivan R, Leaman R, Gonzalez G. The DIEGO Lab Graph based gene Normalization System. In: Proceedings - 10th International Conference on Machine Learning and Applications, ICMLA 2011; 2011.
25. Buyko E, Tomanek K, Hahn U. Resolution of Coordination Ellipses in Biological Named Entities Using Conditional Random Fields. smtp.bootstrep.org.
26. Tsuruoka Y, McNaught J, Tsujii J, Ananiadou S. Learning String Similarity Measures for gene/protein Name Dictionary Look-up using Logistic Regression. Bioinformatics 2007;23(20):2768-774.
27. Wermter J, Tomanek K, Hahn U. High-performance Gene Name Normalization with GeNo. Bioinformatics 2009;25(6):815-821.
28. Leaman R, Islamaj Dogan R, Lu Z. DNorm: Disease Name Normalization with Pairwise Learning to Rank. Bioinformatics 2013;29(22):2909-2917.
29. Gobbel GT, Reeves R, Jayaramaraja S, et al. Development and Evaluation of RapTAT: a Machine Learning System for Concept Mapping of Phrases from Medical Narratives. J Biomed Inform 2014;48:54-65.
30. Kate RJ. Normalizing Clinical Terms using Learned Edit Distance Patterns. J Am Med Inform Assoc July 2015:ocv108.
31. Aronson AR. Effective Mapping of Biomedical Text to the UMLS Metathesaurus: the MetaMap Program. Proc AMIA Symp January 2001:17-21.
32. Sohn S, Kocher JPA, Chute CG, Savova GK. Drug Side Effect Extraction from Clinical Narratives of Psychiatry and Psychology Patients. J Am Med Inform Assoc 2011;(18):144-9.
33. Jonnagaddalaa J, Liaw ST, Rayb P, Kumarc M, Dai HJ. TMUNSW: Identification of Disorders and Normalization to SNOMED-CT Terminology in Unstructured Clinical Notes. In: SemEval-2015: 394.
34. Denecke K. Extracting Medical Concepts from Medical Social Media with Clinical NLP Tools: a Qualitative Study. In: Proceedings of the Fourth Workshop on Building and Evaluation Resources for Health and Biomedical Text Processing; 2014.
35. Kelly L, Goeuriot L, Suominen H, Schreck T, Leroy G, Mowery D, et al. Overview of the ShARe/CLEF eHealth Evaluation Lab 2014. Available at: http://doras.dcu.ie/20109/
36. Mowery DL, Velupillai S, South BR, Christensen L, Martinez D, Kelly L, et al. Task 2: ShARe/CLEF eHealth Evaluation Lab 2014. Available at: http://doras.dcu.ie/20112/1/invited_paper_10.pdf
37. Pradhan S, Elhadad N, Chapman W, Manandhar S, Savova G. SemEval-2014 Task 7: Analysis of Clinical Text. In: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Dublin, Ireland, August 23-24, 2014. p. 54-62.
38. Zhang Y, Wang J, Tang B, Wu Y, Jian M, Chen Y, et al. UTH_CCB: A Report for SemEval 2014 – Task 7 Analysis of Clinical Text. . In: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Dublin, Ireland, August 23-24, 2014. p. 802-6.
39. Elhadad N, Pradhan S, Gorman SL, Manandhar S, Chapman W, Savova G. SemEval-2015 Task 14: Analysis of Clinical Text. In: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Denver, Colorado, June 4-5, 2015. p. 303-10.
40. Pathak P, Patel P, Panchal V, Soni S, Dani K, Choudhary N, et al. ezDI: A Supervised NLP System for Clinical Narrative Analysis. In: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Denver, Colorado, June 4-5, 2015. p. 412-6.
41. Xu J, Zhang Y, Wang J, Wu Y, Jiang M, Soysal E, et al. UTH-CCB: The Participation of the SemEval 2015 Challenge–Task 14. In: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Denver, Colorado, June 4-5, 2015. p. 311-4.
42. Bethard S, Derczynski L, Savova G, Pustejovsky J, Verhagen M. SemEval-2015 Task 6: Clinical TempEval. . In: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Denver, Colorado, June 4-5, 2015. p. 806-14.
43. Velupillai S, Mowery DL, Abdelrahman S, Christensen L, Chapman WW. BluLab: Temporal Information Extraction for the 2015 Clinical TempEval Challenge. In: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Denver, Colorado, June 4-5, 2015. p. 815-9.
44. Bethard S, Savova G, Chen W-T, Derczynski L, Pustejovsky J, Verhagen M. SemEval-2016 Task 12: Clinical TempEval. In: Proceedings of SemEval-2016, San Diego, California, June 16-17, 2016. p. 1052-62.
45. Lee H-J, Zhang Y, Xu J, Moon S, Wand J, Wu Y, et al. UTHealth at SemEval-2016 Task 12: an End-to-End System for Temporal Information Extraction from Clinical Notes. In: Proceedings of SemEval-2016, San Diego, California, June 16-17, 2016. p.1292-7.
46. Khalifa A, Velupillai S, Meystre S. UtahBMI at SemEval-2016 Task 12: Extracting Temporal Information from Clinical Text. . In: Proceedings of SemEval-2016, San Diego, California, June 16-17, 2016. p. 1256-62.
47. Liu C, Wang F, Hu J, Xiong H. Temporal Phenotyping from Longitudinal Electronic Health Records: a Graph Based Framework. doi:10.1145/2783258.2783352.
48. Wang X, Sontag D, Wang F. Unsupervised Learning of Disease Progression Models. doi:10.1145/2623330.2623754.
49. Pham T, Tran T, Phung D, Venkatesh S. DeepCare: a Deep Dynamic Memory Model for Predictive Medicine. In: PAKDD 2016: Advances in Knowledge Discovery and Data Mining; 2016. p. 30-41.
50. Liang Z, Zhang G, Huang JX, Hu QV. Deep Learning for Healthcare Decision Making with EMRs. In: 2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE; 2014. p. 556-9.
51. Wang Y, Luo J, Hao S, Xu H, Shin AY, Jin B, et al. NLP based Congestive Heart Failure Case Finding: a Prospective Analysis on Statewide Electronic Medical Records. Int J Med Inform 2015;84(12):1039-47.
52. Karmakar C, Luo W, Tran T, Berk M, Venkatesh S. Predicting Risk of Suicide Attempt Using History of Physical Illnesses From Electronic Medical Records. JMIR Ment Health 2016;3(3):e19.
53. Jonnagaddala J, Liaw S-T, Ray P, Kumar M, Chang N-W, Dai H-J. Coronary Artery Disease Risk Assessment from Unstructured Electronic Health Records using Text Mining. J Biomed Inform 2015;58:S203-S210.
54. Jonnagaddala J, Liaw S-T, Ray P, Kumar M, Dai H-J, Hsu C-Y. Identification and Progression of Heart Disease Risk Factors in Diabetic Patients from Longitudinal Electronic Health Records. Biomed Res Int 2015; 2015:636371.
55. Chen Q, Li H, Tang B, Wang X, Liu X, Liu S, et al. An Automatic System to Identify Heart Disease

Risk Factors in Clinical Texts over Time. J Biomed Inform 2015;58:S158-63.
56. Zheng B, Zhang J, Yoon SW, Lam SS, Khasawneh M, Poranki S. Predictive Modeling of Hospital Readmissions using Metaheuristics and Data Mining. Expert Syst Appl 2015;42(20):7110-20.
57. Futoma J, Morris J, Lucas J. A Comparison of Models for Predicting early Hospital Readmissions. J Biomed Inform 2015;56:229-38..
58. Iqbal E, Mallah R, Jackson RG, Ball M, Ibrahim ZM, Broadbent M, et al. Identification of Adverse Drug Events from Free Text Electronic Patient Records and Information in a Large Mental Health Case Register. PLoS One 2015;10(8):e0134208.
59. Li Y, Salmasian H, Vilar S, Chase H, Friedman C, Wei Y. A Method for Controlling Complex Confounding Effects in the Detection of Adverse Drug Reactions using Electronic Health Records. J Am Med Inform Assoc 2014;21(2):308-14.
60. Wang G, Jung K, Winnenburg R, Shah NH. A Method for Systematic Discovery of Adverse Drug Events from Clinical Notes. J Am Med Informatics Assoc 2015;22(6):1196-204.
61. Jung K, LePendu P, Chen WS, et al. Automated Detection of off-label Drug Use. PLoS One. 2014;9(2):e89324. doi:10.1371/journal.pone.0089324.
62. Xu H, Aldrich MC, Chen Q, Iyer SV, Readhead B, Dudley JT, et al. Validating Drug Repurposing Signals using Electronic Health Records: a Case Study of Metformin Associated with Reduced Cancer Mortality. J Am Med Inform Assoc 2014;9(2):e89324.
63. Yala A, Barzilay R, Salama L, Griffin M, Sollender G, Bardia A et al. Using Machine Learning to Parse Breast Pathology Reports. Breast Cancer Res Treat 2017 Jan;161(2):203-11.
64. Weegar R, Dalianis H. Creating a Rule-based System for Text Mining of Norwegian Breast Cancer Pathology Reports. In: Sixth International Workshop on Health Text Mining and Information Analysis; 2015.
65. Ou Y, Patrick J. Automatic Population of Structured Reports from Narrative Pathology Reports. In: Proceedings of the Seventh Australasian Workshop on Health Informatics and Knowledge Management - Volume 153. HIKM '14. Darlinghurst, Australia, Australia: Australian Computer Society, Inc.; 2014:41–50.
66. Hochheiser H, Castine M, Harris D, Savova G, Jacobson RS. An Information Model for Computable Cancer Phenotypes. BMC Med Inform Decis Mak 2016;16(1):121.
67. Wang TD, Plaisant C, Quinn AJ, Stanchak R, Murphy S, Shneiderman B. Aligning Temporal Data by Sentinel Events. In: Proceeding of the Twenty-Sixth Annual CHI Conference on Human Factors in Computing Systems - CHI '08. New York, New York, USA: ACM Press; 2008:457.
68. Nikfarjam A, Emadzadeh E, Gonzalez G. Towards Generating a Patient's Timeline: Extracting Temporal Relationships from Clinical Notes. J Biomed Inform 2013;46 Suppl:S40-7.
69. Kovacevic A, Dehghan A, Filannino M, Keane JA, Nenadic G. Combining Rules and Machine Learning for Extraction of Temporal Expressions and Events from Clinical Narratives. J Am Med Inform Assoc 2013;20(5):859-66.
70. Schuemie M. Methods for Drug Safety Signal Detection in Longitudinal Observational Databases: LGPS and LEOPARD. Pharmacoepidemiol Drug Saf 2011 Mar;20(3):292-9.
71. Zhao J. Temporal Weighting of Clinical Events in Electronic Health Records for Pharmacovigilance. In: 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE; 2015:375-81.
72. Lin C, Karlson EW, Dligach D, Ramirez MP, Miller TA, Mo H, et al. Automatic Identification of Methotrexate-induced Liver Toxicity in Patients with Rheumatoid Arthritis from the Electronic Medical Record. J Am Med Inform Assoc 2015;22(e1):e151-61
73. Chen L, Dligach D, Miller T, Bethard S, Savova G. Layered Temporal Modeling for the Clinical Domain. J Am Med Inf Assoc 2015.
74. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a Tool for Health Research: a Systematic Review. Am J Public Health 2017;107(1):143.
75. Shannon Greenwood, Andrew Perrin, Maeve Duggan. PEW Research Center Social Media Update 2016.
76. Fox S. The social life of health information; Pew Research Center.
77. Househ M. The use of Social Media in Healthcare: Organizational, Clinical, and Patient Perspectives. Stud Health Technol Inform 2013;183:244-8.
78. Elhadad N, Gravano L, Hsu D, Balter S, Reddy V, Waechter H. Information Extraction from Social Media for Public Health. In: KDD at Bloomberg: The Data Frameworks Track; 2014.
79. Prieto VM, Matos S, Álvarez M, Cacheda F, Oliveira JL. Twitter: a Good Place to Detect Health Conditions. PLoS One 2014;9(1):e86191.
80. Seaman I, Giraud-Carrier C. Prevalence and Attitudes about Illicit and Prescription Drugs on Twitter. In: 2016 IEEE International Conference on Healthcare Informatics (ICHI). IEEE; 2016:14-7.
81. Baldwin T, Cook P, Lui M, MacKinlay A, Wang L. How Noisy Social Media Text, How Different Social Media Sources? 2013. p. 356-64.
82. Sarker A, Mollá D, Paris C. An Approach for Query-focused Text Summarisation for Evidence Based Medicine. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Vol 7885 LNAI. Springer Verlag; 2013. p. 295-304.
83. Zillner S, Neururer S, Zillner S, Neururer S. Big Data in the Health Sector 10.2 Analysis of Industrial Needs in the Health Sector. doi:10.1007/978-3-319-21569-3_10.
84. Cohen K, Demner-Fushman D. Biomedical Natural Language Processing. 1st ed. (Cohen K, Demner-Fushman D, eds.). Amsterdam/Philadelphia: John Benjamins Publishing Company; 2014.
85. Sarker A, Gonzalez G. Data, Tools and Resources for Mining Social Media Drug Chatter. In: Fifth Workshop on Building and Evaluating Resources for Biomedical Text Mining (BIOTXTM). Osaka; 2016:99-107.
86. Torii M, Tilak SS, Doan S, Zisook DS, Fan J-W. Mining Health-Related Issues in Consumer Product Reviews by Using Scalable Text Analytics. Biomed Inform Insights 2016;8(Suppl 1):1-11.
87. Sarker A, Nikfarjam A, Gonzalez G. Social Media Mining Shared Task Workshop. Pac Symp Biocomput 2016;21:581-92.
88. Wong CA, Merchant RM, Moreno MA. Using Social Media to Engage Adolescents and Young Adults with their Health. Healthc (Amst) 2014;2(4):220-4.
89. Gittelman S, Lange V, Gotway Crawford CA, et al. A New Source of Data for Public Health Surveillance: Facebook Likes. J Med Internet Res 2015;17(4):e98.
90. Kite J, Foley BC, Grunseit AC, Freeman B. Please Like Me: Facebook and Public Health Communication. PLoS One 2016;11(9):e0162765.
91. Platt T, Platt J, Thiel DB, Kardia SLR. Facebook Advertising Across an Engagement Spectrum: a Case Example for Public Health Communication. JMIR Public Health Surveill 2016;2(1):e27.
92. Broniatowski DA, Dredze M, Paul MJ, Dugas A. Using Social Media to Perform Local Influenza Surveillance in an Inner-City Hospital: A Retrospective Observational Study. JMIR Public Health Surveill 2015;1(1)e5.
93. Sharpe JD, Hopkins RS, Cook RL, Striley CW. Evaluating Google, Twitter, and Wikipedia as Tools for Influenza Surveillance Using Bayesian Change Point Analysis: a Comparative Analysis. JMIR Public Health Surveill 2016;2(2):e161.
94. Ofoghi B, Mann M, Verspoor K. Towards Early Discovery of Salient Health Treats : a Social Media Emotion Classification Technique. Pac Symp Biocomput 2016;21:504-15.
95. Correia RB, Li L, Rocha LM. Monitoring Potential Drug Interactions and Reactions via Network Analysis of Instagram User Timelines. Pac Symp Biocomput 2016;21:492-503.
96. Aphinyanaphongs Y, Lulejian A, Brown DP, Bonneau R, Krebs P. Text Classification for Automatic Detection of e-Cigarette use and use for Smoking Cessation from Twitter: a Feasability Pilot. Pac Symp Biocomput 2016;21:480-91.
97. Guillory J, Kim A, Murphy J, Bradfield B, Nonnemaker J, Hsieh Y. Comparing Twitter and Online Panels for Survey Recruitment of E-Cigarette Users and Smokers. J Med Internet Res 2016;18(11):e288.
98. Coloma PM, Becker B, Sturkenboom MCJM, van Mulligen EM, Kors JA. Evaluating Social Media Networks in Medicines Safety Surveillance: Two Case Studies. Drug Saf 2015;38(10):921-30.
99. Nikfarjam A, Sarker A, O'Connor K, Ginn R, Gonzalez G. Pharmacovigilance from Social Media: Mining Adverse Drug Reaction Mentions using Sequence Labeling with Word Embedding Cluster Features. J Am Med Inform Assoc 2015;22(3):671-81.
100. Sarker A, Ginn R, Nikfarjam A, et al. Utilizing Social Media Data for Pharmacovigilance: a Review. J Biomed Inform 2015;54:202-12.
101. Pimpalkhute P, Patki A, Nikfarjam A, Gonzalez G. Phonetic Spelling Filter for Keyword Selection in Drug Mention Mining from Social Media. AMIA Jt Summits Transl Sci Proc 2014 Apr 7;2014:90-5.
102. Sarker A, Gonzalez G. A Corpus for Mining Drug-related Knowledge from Twitter Chatter: Language Models and their Utilities. Data Brief 2016;Nov 23;10:122-131.

103. Limsopatham N, Collier N. Towards the Semantic Interpretation of Personal Health Messages from Social Media. In: Proceedings of the ACM First International Workshop on Understanding the City with Urban Informatics - UCUI '15. New York, New York, USA: ACM Press; 2015:27-30.
104. Lampos V, Cristianini N. Tracking the flu pandemic by monitoring the Social Web. In: 2010 2nd International Workshop on Cognitive Information Processing. IEEE; 2010:411-6.
105. Paul MJ, Dredze M, Broniatowski D. Twitter improves influenza Forecasting. PLoS Curr 2014;6:1-13.
106. Lee K, Agrawal A, Choudhary A. Real-time Disease Surveillance using Twitter Data. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '13. New York, New York, USA: ACM Press; 2013:1474..
107. Broniatowski DA, Paul MJ, Dredze M. Twitter: big data opportunities. Science 2014;(345):148.
108. Liu X, Chen H. A Research Framework for Pharmacovigilance in Health Social Media: Identification and Evaluation of Patient Adverse Drug Event Reports. J Biomed Inform 2015;58:268-79.
109. Plachouras V, Leidner JL, Garrow AG. Quantifying Self-Reported Adverse Drug Events on Twitter. In: Proceedings of the 7th 2016 International Conference on Social Media & Society - SMSociety '16. New York, New York, USA: ACM Press; 2016:1-10.
110. Patki A, Sarker A, Pimpalkhute P, Nikfarjam A, Ginn R. Mining Adverse Drug Reaction Signals form Social Media: Going Beyond Extraction. In: Proceedings of BioLinkSig 2014:9-19.
111. Mazzocut M, Truccolo I, Antonini M, Rinaldi F, Omero P, Ferrarin E, et al. Web Conversations About Complementary and Alternative Medicines and Cancer: Content and Sentiment Analysis. J Med Internet Res 2016;18(6):e120.
112. Korkontzelos I, Nikfarjam A, Shardlow M, Sarker A, Ananiadou S, Gonzalez GH. Analysis of the Effect of Sentiment Analysis on Extracting Adverse Drug Reactions from Tweets and Forum Posts. J Biomed Inform 2016;62:148-58.
113. Adrover C, Bodnar T, Huang Z, Telenti A, Salathé M. Identifying Adverse Effects of HIV Drug Treatment and Associated Sentiments Using Twitter. JMIR Public Health Surveill 2015;1(2):e7.
114. Sarker A, Gonzalez G. Portable Automatic Text Classification for Adverse Drug Reaction Detection via Multi-Corpus Training. J Biomed Inform 2014;53:196-207..
115. Dai H-J, Touray M, Jonnagaddala J, Syed-Abdul S. Feature Engineering for Recognizing Adverse Drug Reactions from Twitter Posts. Information 2016;7(2):27.
116. Kendra RL, Karki S, Eickholt JL, Gandy L. Characterizing the Discussion of Antibiotics in the Twittersphere: What is the Bigger Picture? J Med Internet Res 2015;17(6):e154.
117. Sarker A, O'Connor K, Ginn R, Scotch M, Smith K, Malone D, et al. Social Media Mining for Toxicovigilance: Automatic Monitoring of Prescription Medication Abuse from Twitter. Drug Saf 2016 39(3);231-40.
118. Kavuluru R, Sabbir AKM. Toward Automated e-Cigarette Surveillance: Spotting e-Cigarette Proponents on Twitter. J Biomed Inform 2016;61:19-26.
119. Choudhury S, Alani H. Personal Life Event Detection from Social Media; 2014.
120. KıcKıman E, Richardson M. Towards Decision Support and Goal Achievement: Identifying Action-Outcome Relationships from Social Media. Proc 21th ACM SIGKDD 2015.
121. Wen M, Zheng Z, Jang H, Xiang G, Rosé C. Extracting Events with Informal Temporal References in Personal Histories in Online Communities. ACL 2013.
122. Collier N, Son N, Nguyen N. OMG U got flu? Analysis of Shared Health Messages for Bio-surveillance. J Biomed Semantics 2011;2(Suppl 5):S9.
123. Nakov P, Ritter A, Rosenthal S, Sebastiani F, Stoyanov V. SemEval-2016 Task 4: Sentiment Analysis in Twitter. In: International Workshop on Semantic Evaluation Exercises (SemEval); 2016:1-18.
124. Daniulaityte R, Chen L, Lamy FR, Carlson RG, Thirunarayan K, Sheth A. "When 'Bad' is 'Good'": Identifying Personal Communication and Sentiment in Drug-Related Tweets. JMIR Public Health Surveill 2016;2(2):e162.
125. Cobb NK, Mays D, Graham AL. Sentiment Analysis to Determine the Impact of Online Messages on Smokers' Choices to Use Varenicline.J Natl Cancer Inst Monogr 2013;2013(47):224-30.
126. Chan B, Lopez A, Sarkar U. The Canary in the Coal Mine Tweets: Social Media Reveals Public Perceptions of Non-Medical Use of Opioids. PLoS One 2015;10(8):e0135072.
127. Lei Y, Pereira JA, Quach S, et al. Examining Perceptions about Mandatory Influenza Vaccination of Healthcare Workers through Online Comments on News Stories. PLoS One 2015;10(6):e0129993.
128. Ramagopalan S, Wasiak R, Cox AP. Using Twitter to Investigate Opinions about Multiple Sclerosis Treatments: a Descriptive, Exploratory Study. F1000Res 2014;3:216.
129. Mollema L, Harmsen IA, Broekhuizen E, Clijnk R, De Melker H, Paulussen T, et al. Disease Detection or Public Opinion Reflection? Content Analysis of Tweets, Other Social Media, and Online Newspapers During the Measles Outbreak in the Netherlands in 2013. J Med Internet Res 2015;17(5):e128.
130. Shutler L, Nelson LS, Portelli I, Blachford C, Perrone J. Drug Use in the Twittersphere: A Qualitative Contextual Analysis of Tweets About Prescription Drugs. J Addict Dis 2015;34(4):303-10.
131. Aramaki E, Maskawa S, Morita M. Twitter Catches the flu: Detecting influenza Epidemics using Twitter. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics; 2011. p. 1568-76.
132. Kralj Novak P, Smailović J, Sluban B, Mozetič I. Sentiment of Emojis. PLoS One 2015;10(12):e0144296.
133. Aronson AR, Lang F-M. An Overview of MetaMap: Historical Perspective and Recent Advances. J Am Med Inform Assoc 2010;17(3):229-36.
134. Li X, Li J, Wu Y. A Global Optimization Approach to Multi-polarity Sentiment Analysis. PLoS One 2015;10(4):e0124672.
135. Agarwal V, Zhang L, Zhu J, Fang S, Cheng T, Hong C, et al. Impact of Predicting Health Care Utilization Via Web Search Behavior: A Data-Driven Analysis. J Med Internet Res 2016;18(9):e251.
136. Peng Y, Moh M, Moh T-S. Efficient Adverse Arug Event Extraction using Twitter Sentiment Analysis. In: 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE; 2016:1011-8.
137. Huynh T, He Y, Willis A, Uger S. Adverse Drug Reaction Classification With Deep Neural Networks. In: COLING. Osaka; 2016. p. 877-87.
138. Paul MJ, Dredze M. Discovering Health Topics in Social Media Using Topic Models. PLoS One 2014;9(8):e103408.
139. Wang S, Paul MJ, Dredze M. Exploring Health Topics in Chinese Social Media: An Analysis of Sina Weibo. In: Workshops at the Twenty-Eighth AAAI Conference on Artificial Intelligence; 2014. p. 20-3.
140. Kumar M, Dredze M, Coppersmith G, De Choudhury M. Detecting Changes in Suicide Content Manifested in Social Media Following Celebrity Suicides. In: Proceedings of the 26th ACM Conference on Hypertext & Social Media - HT '15. New York. New York, USA: ACM Press; 2015. p. 85-94.
141. Surian D, Nguyen DQ, Kennedy G, Johnson M, Coiera E, Dunn AG. Characterizing Twitter Discussions About HPV Vaccines Using Topic Modeling and Community Detection. J Med Internet Res 2016;18(8):e232.
142. Li J, Ritter A, Cardie C, Hovy E. Major Life Event Extraction from Twitter based on Congratulations/Condolences Speech Acts. EMNLP. 2014.
143. Li J, Cardie C. Timeline Generation. In: Proceedings of the 23rd International Conference on World Wide Web - WWW '14. New York, New York, USA: ACM Press; 2014. p. 643-52.
144. Sullivan R, Sarker A, O'Connor K, Goodin A, Karlsrud M, Gonzalez G. Finding Potentially Unsafe Nutritional Supplements from User Reviews with Topic Modeling. Pac Symp Biocomput 2016;21:528-39.
145. Wang T, Huang Z, Gan C. On Mining Latent Topics from Healthcare Chat Logs. J Biomed Inform 2016;61:247-59.
146. Zou B, Lampos V, Gorton R, Cox IJ. On Infectious Intestinal Disease Surveillance using Social Media Content. In: Proceedings of the 6th International Conference on Digital Health Conference - DH '16. New York, New York, USA: ACM Press; 2016. p. 157-61.
147. Leaman R, Wojtulewicz L, Sullivan R, Skariah A, Yang J, Gonzalez G. Towards Internet-Age Pharmacovigilance: Extracting Adverse Drug Reactions from User Posts to Health-Related Social Networks. In: Workshop on Biomedical Natural Language Processing. Uppsala, Sweden; 2010. p. 117-25.
148. De Choudhury M, Counts S, Horvitz E. Major Life Changes and Behavioral Markers in Social Media: Case of Childbirth. In: Social Networks

During Major Transitions; 2013. p. 1431-42.
149. Wiley MT, Jin C, Hristidis V, Esterling KM. Pharmaceutical drugs chatter on Online Social Networks. J Biomed Inform 2014;49:245-54.
150. Denecke K. Information Extraction from Medical Social Media. In: Health Web Science. Cham: Springer International Publishing; 2015. p. 61-73.
151. Yates A, Goharian N, Frieder O. Extracting Adverse Drug Reactions from Social Media. In: Proceedings of the National Conference on Artificial Intelligence. Vol 3; 2015. p. 2460-7.
152. Morlane-Hon F, Grouin C, Zweigenbaum P. Identification of Drug-Related Medical Conditions in Social Media. In: LREC; 2016. p. 2022-8.
153. Mikolov T, Chen K, Corrado G, Dean J. Distributed Representations of Words and Phrases and their Compositionality. Nips; 2013. p. 1-9.
154. Han B, Cook P, Baldwin T. Lexical Normalization for Social Media Text. ACM Trans Intell Syst Technol 2013;4(1):1-27.
155. Choudhury M, Saraf R, Jain V, Mukherjee A, Sarkar S, Basu A. Investigation and Modeling of the Structure of Texting Language. Int J Doc Anal Recognit 2007;10(3-4):157-74.
156. Cook P, Stevenson S. An Unsupervised Model for Text Message Normalization. In: Proceedings of the Workshop on Computational Approaches to Linguistic Creativity. Association for Computational Linguistics; 2009. p. 71-8.
157. Xue Z, Yin D, Davison B. Normalizing Microtext. Anal Microtext 2011;(September):74-9.
158. Liu F, Weng F, Jiang X. A Broad-Coverage Normalization System for Social Media Language. Proc 50th Annu Meet Assoc Comput Linguist Vol 1 Long Pap; 2012;(July):1035-44.
159. Karimi S, Metke-Jimenez A, Kemp M, Wang C. Cadec: a Corpus of Adverse Drug Event Annotations. J Biomed Inform 2015;55:73-81.
160. O'Connor K, Pimpalkhute P, Nikfarjam A, Ginn R, Smith KL, Gonzalez G. Pharmacovigilance on Twitter? Mining Tweets for Adverse Drug Reactions. AMIA Annu Symp Proc 2014;2014:924-33.
161. Limsopatham N, Collier N. Adapting Phrase-based Machine Translation to Normalise Medical Terms in Social Media Messages. In: Conference Proceedings - EMNLP 2015: Conference on Empirical Methods in Natural Language Processing; 2015. p. 1675-80.
162. Limsopatham N, Collier N. Learning Orthographic Features in Bi-directional LSTM for Biomedical Named Entity Recognition. In: Fifth Workshop on Building and Evaluating Resources for Biomedical Text Mining. Osaka; 2016. p. 10-9.
163. Chandrashekar PB, Magge A, Sarker A, Gonzalez G. Social Media Mining for Identification and Exploration of Health-related Information from Pregnant Women. In: Proceedings of the First Workshop on Mining Online Health Reports. Cambridge, UK; 2017. p. 1-9.
164. Milne DN, Pink G, Hachey B, Calvo RA. CLPsych 2016 Shared Task: Triaging Content in Online Peer-support Forums. In: 3rd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. San Diego, California: Association for Computational Linguistics; 2016. p. 118-27.
165. Kim S Mac, Wang Y, Wan S, Paris C. Data61-CSIRO Systems at the CLPsych 2016 Shared Task. In: 3rd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. San Diego, California; 2016. p. 128-32.
166. Qntfy GC, Dredze M, Harman C, Hollingshead Ihmc K, Mitchell M. CLPsych 2015 Shared Task: Depression and PTSD on Twitter. In: E 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. Denver, Colorado; 2015. p. 31-9.
167. Resnik P, Armstrong W, Claudino L, Nguyen T. The University of Maryland CLPsych 2015 Shared Task System. In: 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. Denver, Colorado; 2015. p. 54-60.
168. Rastegar-Mojarad M, Elayavilli RK, Yu Y, Liu H. Detecting Signals in Noisy Data - Can Ensemble Classifiers Help Identify Adverse Drug Reaction in Tweets? In: Social Media Mining Shared Task Workshop. Hawaii; 2016.
169. Wang W. MiningAdverse Drug Reaction Mentions in Twitter with Word Embeddings. In: Social Media Mining Shared Task Workshop. Hawaii; 2016.
170. Natarajan S, Bangera V, Khot T, Picado J, Wazalwar A, Santos Costa V, et al. Markov Logic Networks for Adverse Drug Event Extraction from Text. Knowl Inf Syst 2016. p. 1-23.
171. Segura-Bedmar I, Martínez P, Revert R, Moreno-Schneider J. Exploring Spanish Health Social Media for Detecting Drug Effects. BMC Med Inform Decis Mak 2015;14(2):s6.

Correspondence to:
Dr. Graciela Gonzalez Hernandez
Department of Biostatistics, Epidemiology and Informatics
University of Pennsylvania
Perelman School of Medicine
13212 East Shea Boulevard
Scottsdale, AZ 85259
USA
E-mail: gragon@upenn.edu