Digestive Diseases and Sciences (2021) 66:29–40
https://doi.org/10.1007/s10620-020-06156-y

REVIEW

Evolving Role and Future Directions of Natural Language Processing in Gastroenterology

Fredy Nehme[1] · **Keith Feldman**[2,3]

Received: 20 June 2019 / Accepted: 18 February 2020 / Published online: 27 February 2020

Abstract

In line with the current trajectory of healthcare reform, significant emphasis has been placed on improving the utilization of data collected during a clinical encounter. Although the structured fields of electronic health records have provided a convenient foundation on which to begin such efforts, it was well understood that a substantial portion of relevant information is confined in the free-text narratives documenting care. Unfortunately, extracting meaningful information from such narratives is a non-trivial task, traditionally requiring significant manual effort. Today, computational approaches from a field known as Natural Language Processing (NLP) are poised to make a transformational impact in the analysis and utilization of these documents across healthcare practice and research, particularly in procedure-heavy sub-disciplines such as gastroenterology (GI). As such, this manuscript provides a clinically focused review of NLP systems in GI practice. It begins with a detailed synopsis around the state of NLP techniques, presenting state-of-the-art methods and typical use cases in both clinical settings and across other domains. Next, it will present a robust literature review around current applications of NLP within four prominent areas of gastroenterology including endoscopy, inflammatory bowel disease, pancreaticobiliary, and liver diseases. Finally, it concludes with a discussion of open problems and future opportunities of this technology in the field of gastroenterology and health care as a whole.

Keywords Gastroenterology · Artificial intelligence · Natural Language Processing · Health care

Introduction

Fostered by the current trajectory of healthcare reform and widespread adoption of electronic health records (EHR), the availability of large repositories of digital health and wellness data have given rise to lofty expectations around the future of medicine in the age of Big Data. Although notable progress has been made in the areas of predictive analytics and personalized medicine, the complexity of clinical workflows and patient care has presented several notable barriers to the effective utilization of analytical methods in practice. Among the most prominent is the understanding that, to date, many analytical methods operate on an incomplete set of patient data.

Technical constraints of the algorithms underlying many state-of-the-art methods often require that data be discretely coded and represented in a standardized vector, or matrix format. As a result, such methods have relied heavily on clinical billing codes, medication lists, lab procedures, or, even in the most advanced systems, extensive lists of discretely captured clinical events. Yet, from a clinical perspective, it is well understood that much of the defining information regarding a patient's condition is captured not within discrete fields, but in the narrative style "clinical note" [1]. In fact, recent work has found that even for data traditionally viewed as highly distinct, e.g., laboratory and medication records, a significant portion of relevant information may only be available as part of clinical text [2].

Unfortunately, knowledge of where such rich information exists has not translated to the development of more effective analytical tools. As extraction of the necessary structured data from within the unstructured free-form EHR, text

✉ Fredy Nehme
nehme.fredy@gmail.com

1 Department of Gastroenterology and Hepatology, University of Missouri-Kansas City School of Medicine, 5000 Holmes Street, Kansas City, MO 64110, USA

2 Division of Health Services and Outcomes Research, Children's Mercy Kansas City, Kansas City, MO, USA

3 Department of Pediatrics, University of Missouri-Kansas City School of Medicine, Kansas City, MO, USA

embodies a well-established challenge, one in part stemming from extremely specialized terminologies, and high degree of variability across the structure of content of each note and across differing institutions and practitioners [3–6]. While manual annotation by a clinical expert remains a gold standard in the attainment of such data, the approach has become impractical across the ever-growing volumes of available data on which these algorithms operate [7]. Rather, researchers have turned to translational methods, exploiting a sundry of automated computational methodologies from a domain formally known as Natural Language Processing (NLP).

Representing an intersection of linguistics and statistical modeling, NLP techniques offer a means to perform complex low-level language tasks including entity recognition (e.g., identifying medications, procedure names, locations), relation detection (e.g., linking dosages to a specific medication), as well as high-level grouping techniques such as topic modeling to identify themes in text. Consequently, the notion that NLP methods could be used to efficiently process clinical text has existed for almost four decades. However, only recently has the increasing availability of digitalized clinical text documents aligned with improved computational resources and advancing NLP methods to make substantial progress in that goal [8–10].

Today, with demonstrated success in use cases spanning dictated examination notes, to automatic detection of new symptom patterns, NLP is poised to make a transformational impact on the field of health care, particularly in procedure-heavy sub-disciplines such as gastroenterology (GI). As such, this manuscript aims to provide a clinically focused review of NLP systems in GI practice. To do so, it will begin with a detailed overview around the state of NLP techniques, presenting state-of-the-art methods and typical use cases in both clinical settings and across other domains. Next, it will then present a robust literature review around current applications of NLP within four prominent areas of GI practice including endoscopy, inflammatory bowel disease (IBD), pancreaticobiliary, and liver diseases. Finally, it concludes with a discussion of open problems and future opportunities of this technology in the field of gastroenterology and health care as a whole.

Natural Language Processing

Dating back to the 1950s for use in automated language translation, the theory of NLP has since been shown to be highly adaptable, finding utility across numerous domains and applications in modern society [11]. Prominent examples include keyword and phase processing for information retrieval in search engines and databases [12], speech recognition technologies [13], question-and-answer tools such as smart assistants [14], and the automated summarization of lengthy text documents [15]. While the algorithms and end-products across such applications are highly distinct, the underlying NLP methodologies that provide structured data necessary for these tools to operate remain fairly consistent. Often viewed as a pipeline, originating with the raw unstructured text and ending with discrete elements, NLP techniques can be viewed as encompassing a breadth of both syntactic and semantic language tasks.

Syntactic Text Analysis

The process often begins from a syntactic perspective, with a focus on isolating distinct terms, where the complete body of a text document is broken down into unique paragraphs or sentences (a process known as segmentation) and then further into individual words (known as tokenization). Depending on the application, these terms (or *tokens*) are traditionally subjected to various forms of structural and morphological normalization, processes that aim to standardize differences caused by grammatical considerations such as inflection in number, tense, gender, person, and case. Further normalization of capitalization, punctuation, abbreviations, and spelling correction is also common in the effort to standardize the diverse terminology within the original text.

By enumerating the resulting set of "clean" tokens as a single wide vector for each document, where each feature represents a unique token as either raw counts or in more complex weightings such as term frequency and inverse document frequency (tf–idf), it is possible to obtain a basic form of structure on which machine learning tools can operate [12, 16]. This structure provides a direct means to: build predictive models (such as spam classification [17]), machine translation (e.g., English to Spanish [18]), and even develop topical representations of text (e.g., automatically grouping themes of articles published within scientific journals [19]).

Semantic Text Analysis

Beyond the enumeration of terms, many state-of-the-art NLP techniques seek to provide a more nuanced form of structure, by extracting meaningful semantic relations from the raw unstructured text. To do so, methods often employ linguistic structure analysis. Again taking "cleaned" tokens, the process begins by going just slightly beyond simple syntactical analysis, utilizing a probabilistic model to label each token with its respective part-of-speech in each sentence. Trained on massive corpuses of annotated text, these models have been shown to be incredibly accurate and exist for a multitude of different languages [20, 21].

Based on the resulting labels, and known grammatical rules of a respective language, parse trees can be constructed for each sentence, together forming a foundation on which several common methodologies act. Most commonly these include a form of *Named Entity Detection*, where specific persons, organizations, products, geographies, and other well-known objects, events, or quantities are recognized within the discrete tokens, e.g., labeling Apple as a company, iPhone as a product, and 2019 as year in the sentence: *Apple released 3 new iPhones in 2019* [22]. Finally, at the conclusion of the modern NLP pipeline we find *Dependency parsing* and *Relation Detection*, where relationships between the named entities are formalized, allowing us to link complex phenomena such as cost of a product to the product name, or time and location [23]. Recently, modern graph-based methods have combined and extended these final two steps to include *entity-linking,* a form of word disambiguation allowing researchers meaningful separate similarly named entities based on context (e.g., recognizing that Apple is a company and not the fruit in the above example) [24].

State-of-the-Art Text Analysis: Embedding Representations

This pipeline has repeatedly demonstrated clear value and has become a standard in many disciplines. Yet, it has become clear the diverse nature of writing has furthered notion that syntax alone cannot be wrangled into a cohesive array of terms, as such a process overlooks the existence of synonyms or other semantically similar entities that are pervasive across text from different sources and individuals. Consequently, state-of-the-art NLP methods have emerged focused on semantic similarity between words in the form of *embeddings*. By learning the relationship between term ordering, these models provide a vector representation for each term such that interchangeable or semantically similar entities are numerically close together. Primarily used for improving syntactically focused models, embeddings techniques have provided a significant boost in modeling tasks, including opinion/sentiment analysis, information retrieval tasks including relevance ranking, and improvements to speech recognition [25–27].

Although outside the scope of this review, the processes used to create these embeddings commonly rely on neural network architectures and require massive repositories of available text (such as the entire body of Wikipedia) to learn accurate relationships. Recently, this concept has been expanded, allowing for singular representations of full sentences, paragraphs, or even entire documents [28]. Conversely, researchers have also begun work to capture relationships at a much more granular level, not only between the series of terms, but between the specific characters that comprise each token. Known as character or byte-pair embeddings, these new methods present a great opportunity for text limited domains, or domains with highly specialized vocabularies that may prove difficult to model [29–31].

Evolving Utility of Natural Language Processing in Health Care

As we look to the multitude of available NLP methods, it is important to remember the efficacy of such techniques has conventionally been predicated on the availability of large text repositories; over which stochastic models (part-of-speech taggers, entity recognition, embedding representations, etc.) could be accurately trained. It is then unsurprising that widespread transition from paper documentation to electronic health records has fostered a great deal of interest in studying the implications of applying these techniques to clinical text problems [32, 33].

Specifically, with workflows that place a significant emphasis on procedural or operational reports, fields such as radiology and oncology have led the way in adoption of NLP techniques. To date, several reviews have highlighted successful NLP applications for diseases or clinical outcomes prediction (pulmonary embolism, tuberculosis, fractures, and several forms of cancer and their respective stages), decision support tools that aid practitioners through reminders or recommendations (e.g., need for additional imaging), QI initiatives to understand division workflows, and even improved cohort identification for use in epidemiologic studies [34–36]. Further, the utility of NLP for healthcare tasks has continued to expand to other areas including bio-surveillance for influenza outbreaks, identification of postoperative complications, and automated code assignment [37–39]. Finally, emergent NLP applications in health care have begun to extend beyond the confines of practitioner-curated data. Most pointedly in the field of mental health, where researchers have looked to less conventional sources of text, such as social media and online forum posts to achieve early detection of depression or suicidal ideation [40, 41].

Clinical Tools for NLP

Notably, challenges in developing these and other NLP methodologies to addressing problems across the clinical domain have stemmed not only from the historically limited availability of data, but also from the complexity of clinical text itself. Rife with abbreviations, nonstandard uses of common terms, multiple phrasing and codings for the same components, and complex relational dependencies between elements, clinical text presents a unique array of challenges to the standard language models used in traditional NLP [42].

Recognizing this, the National Library of Medicine (NLM) has undertaken a significant effort to facilitate the development of tools specifically designed for the unique properties of biomedical text. Known as the Unified Medical Language System (UMLS), the NLM maintains three distinct tools (Metathesaurus®, semantic network, and SPECIALIST NLP Tools) [43]. At a fundamental level, the *SPECIALIST NLP Tools* represents a set of low-level NLP methodologies (part-of-speech taggers, entity recognition and normalization models, etc.) trained on an extensive biomedical lexicon to aid with text analysis tasks. However, these tools are limited in their generalizability, particularly in addressing terms that are syntactically difficult to coalesce but are semantically related (e.g., *Hematochezia*: bright red blood per rectum, or *Dysphagia*: swallowing difficulties).

This form of higher-level standardization often requires curated ontologies to group related terms, instruments found within the two remaining UMLS tools. Drawing on over one million biomedical concepts from nearly 200 source vocabularies, the Metathesaurus represents a means to link alternative names of the same concept (e.g., *Linitis plastica, diffuse stomach cancer, Brinton's disease)* and map between various coding systems and focused ontologies (SNOMED, ICD-9-CM, MeSH, etc.), The semantic network in turn provides a further "Upper-level ontology" categorization of all UMLS concepts, capturing over 100 semantic types (e.g., cell function, neoplastic process, virus) and 54 qualifiers (e.g., interacts with, treats, affects) to allow for improved relationship mapping between concepts [44]. Moreover, in line with trends across the NLP field, semantically focused emergent work has provided state-of-the-art embeddings specifically designed on biomedical vocabularies: Examples include BioWordVec and PMCVec (both trained across the titles and abstracts from all the 27 million documents in PubMed database) [45, 46]. Together with these and other emergency tools, practical NLP applications have been utilized in a myriad of research and practice settings ranging from EHR phenotyping to automated discovery of new medical ontologies [47, 48].

Current Applications of NLP in Digestive Diseases

Together with increased documentation, the tools have provided a means to rapidly advance the uptake of NLP methodologies across several clinical specialties and subspecialties, including that of gastroenterology.

NLP was initially used in the field of gastroenterology in 1984 when a database system was developed by integrating gastrointestinal radiologic, endoscopic, and pathologic studies using a natural language format. Easy retrieval of data on a given patient was made possible allowing intercorrelative studies among radiology, endoscopy, and pathology [49].

While GI is considered a procedure-heavy domain, gastroenterology specialty societies have emphasized on the importance of maintaining high-quality performance during procedures and routinely assessing certain quality metrics. Such routine measurement has been hampered by the costs and time required to manually review and assess compliance with quality measures. In addition, GI is a rapidly evolving field, and as a result, many procedures, state-of-the-art therapies, and diagnosis lack specific ICD codes making database research challenging and often inaccurate. Therefore, Natural Language Processing appears particularly promising in gastroenterology offering the potential to address the above issues along with several other applications in this field. Multiple studies have utilized NLP covering a wide spectrum of digestives diseases including endoscopy, inflammatory bowel disease (IBD), pancreaticobiliary, and liver diseases.

Review Methodology

Electronic database searches were conducted using Medline, Ovid EMBASE, and Google Scholar through June 2019. The search terms were "Natural Language Processing" and "gastroenterology OR digestive diseases OR endoscopy OR hepatology OR IBD (Inflammatory bowel diseases) OR pancreaticobiliary diseases." Our search yielded 652 results. Abstracts were screened and articles discussing applications of NLP in GI were included. We included articles written in English and excluded articles unrelated to NLP. Our final review included 40 studies utilizing NLP applications in the field of gastroenterology.

Endoscopy

Colonoscopy is a high-volume procedure used routinely for colorectal cancer (CRC) prevention. There is considerable variation in the quality of the procedure among different endoscopists, and lower quality has been linked to increased cancer incidence [50–53]. The American Society of Gastrointestinal Endoscopy (ASGE) published a list of 15 quality indicators to improve safety and performance of colonoscopy [50], and experts have called on providers to routinely report on their colonoscopy performance [54]. However, manually extracting data pertinent to colonoscopy quality is difficult as it is often imbedded as unstructured text within health records leading to very few physicians reporting their quality metrics [55]. In addition, manual data extraction is not free of error. In fact, in a multicenter Veterans Health Administration (VHA) study, the error rate was not statistically significant between the NLP program and certified gastroenterologists (25.4% vs. 21.1%, respectively, $p = 0.07$)

[56]. In addition, currently used endoscopic software can provide data on a limited number of quality measures. Given the time-intensive nature of manual medical record review, several studies have demonstrated the feasibility of NLP in extracting colonoscopy specific quality metrics with > 90% accuracy [56–60]. Quality indicators accurately extracted using NLP include screening indication, family history of CRC, cecal intubation rate, adequacy of bowel preparation, and presence and location of polyps [61]. A common quality measure implemented is the "adenoma detection rate" (ADR) defined by the proportion of screening colonoscopic examinations performed by a physician that detect one or more adenomas. Both Natural Language Processing and manual review produced comparable values for ADRs in multiple studies [57, 58, 62, 63]. By using a NLP program, Marcones et al. reported that the ADR and withdrawal time worsened by the end of the day after analyzing more than 80,000 colonoscopies [64]. Patel et al. [65] used a NLP tool and found that a longer mean withdrawal time of 11 min resulted in statistically significant increase of ADR and proximal polyp–serrated detection rate. Moreover, public reporting of colonoscopy quality was associated with a higher ADR [66]. A NLP system was used to analyze outpatient colonoscopy examinations to determine physician characteristics associated with higher ADR [67]. Hakerma et al. evaluated multiple colonoscopy-related quality measures based on gastroenterology society recommendations by using an NLP engine that extracts 21 variables for 19 quality metrics from free-text pathology and colonoscopy reports. The average accuracy was 0.89, and the average agreement score (Cohen's K) was 0.62 [59]. With further refinement and development, such system can be used on a substantially larger scale for routine quality measurement of several metrics.

A fully automated system using NLP technology and guideline-based clinical decision support (CDS) was accurately used to predict colonoscopy surveillance intervals by analyzing more than 10,000 reports and comparing them to a manually reviewed sample (Cohen's $K = 0.74$ consistent with substantial agreement between manual review and CDS system). Similar to the ADR, a fully automated system would allow tracking of another quality metric that is less likely to be manipulated by the individual provider and compare adherence rates to surveillance guidelines between providers. This would help justify costs and allow for education in proper surveillance intervals.

An NLP tool allowed to analyze more than 100,000 colonoscopy and pathology reports in a multicenter study to determine that gastroenterology specialization, more recent completion of training, and greater procedure volume are associated with serrated–polyp detection [68]. By analyzing a substantial number of reports, NLP increased the strength of the analysis leading to more precise estimates while requiring less manual review. This technology was also used to measure the variation in pathologist's interpretation of colorectal adenomas and serrated polyps by evaluation 85,526 reports [69].

NLP can also be used to generate prediction models in endoscopy. Hong et al. [70] developed a five-item prediction model to determine the risk of advanced colorectal neoplasia at the first screening colonoscopy by means of NLP from the EHR system (AUC 71.6%), while Blumenthal et al. [71] developed a model predicting future non-adherence with outpatient colonoscopy (AUC 70.2%).

Besides colonoscopy, NLP algorithm was highly accurate in identifying dysplasia in Barrett's esophagus with 97.1% accuracy and 93.6% precision [72].

Inflammatory Bowel Diseases

The diagnosis of IBD and other outcome variables is challenging based on diagnosis and procedure billing codes who are often inaccurate or lacking altogether. Therefore, IBD database research has been limited by the absence of administrative codes for key disease-related variables. Natural Language Processing was used to address these issues by identifying data within text reports.

One example is the absence of diagnostic or billing codes to identify patients with non-cancer dysplasia making the follow-up of large cohorts of IBD patients with dysplasia difficult. After creating a validation cohort using a manually reviewed random sample, Hou et al. identified colonic dysplasia in a VHA cohort with IBD using NLP with an accuracy of 97.1% for the detection of low-grade dysplasia. This can help to further study the natural history and outcomes of colonic dysplasia IBD patients [73]. In another study, narrative concepts using NLP were successfully utilized with codified data to enhance case definition of ulcerative colitis and Crohn's disease with an accuracy of 94% and 95%, respectively, compared with 91% and 92% when using codified data alone, allowing for efficient analysis of large IBD cohorts [74].

Medication side effects, such as arthralgia, are not coded by gastroenterologists, while joint pain, if present, is usually mentioned in their notes. NLP was used to identify the notes stating if the patient had joint pain, allowing to compare the prevalence of arthralgia between patients on vedolizumab and those on anti-TNF inhibitors. Performance characteristics were better for NLP than ICD9 codes in identifying patients with arthralgia with a PPV and sensitivity of 90% and 83% for NLP compared to 79% and 52% for ICD9 codes, respectively [75]. Same applies to endoscopy in IBD, where the absence of procedure codes for surveillance and non-surveillance colonoscopy limits the use of available registries in researching surveillance practices and related quality measures. Automatic

Retrieval Console (ARC) is a software that uses NLP pipelines to breakdown documents into structured fragments of text based on parts of speech, negated terms, and a library of medical and non-medical terms. In a VHA study, ARC was able to reliably perform document classification of surveillance and non-surveillance endoscopy with a specificity of 88% and sensitivity of 77% [76].

Stoma-related complications from IBD surgeries differ dramatically by stoma type between ileostomies and colostomies. Using solely CPT (Current Procedural Terminology) codes to differentiate between the two types of stoma had poor sensitivity (35% for ileostomy and 75.2% for colostomy). Incorporation of an NLP-based software increased sensitivity to more than 95% [38]. Accurately identifying these outcomes and many more in IBD using NLP may open new opportunities for database research in this field by allowing EHR-based determination of variables.

Pancreaticobiliary Diseases

ERCP remains the highest risk endoscopic procedure widely used in practice [77]. Given the proven efficacy of NLP in quality measures in colonoscopy as described previously and the impact of colonoscopy feedback on ADR, Imler et al. extracted ERCP quality measures from 23,674 procedures over an 8.5-year period using NLP and compared them to society guidelines and across providers. The accuracy of NLP to identify 13 ERCP quality measures ranged from 90 to 100% compared to a manually selected sample with intraprocedure measures having lower values [78]. This may guide quality improvement in advanced endoscopy and identify providers who are not meeting society-endorsed benchmarks. At present, there is no standard nationwide program or system to identify patients with pancreatic cysts and current practices rely on individual practitioners and their use of manually entered patient databases. Intraductal papillary mucinous neoplasms (IPMNs), one of the most common pancreatic cysts with malignant potential, may not always be identified, tracked, and treated optimally. A NLP system used to identify IPMNs from a patient database outperformed a manually maintained surgical registry with sensitivity of 97.5% and positive predictive value of 95.5% [79], while Merhabi et al. [80] efficiently identified pancreatic cysts from medical reports using multiple NLP-based concepts with recall of 97.4% and precision of 98.5%. This could open the door to additional applications of NLP in pancreaticobiliary diseases such as implementing a surveillance program that automatically detects new pancreatic cysts as they are discovered or the ability to recognize concomitant conditions such as patients with pancreatic cysts who develop new or worsening diabetes mellitus.

Hepatology

Kung et al. [81, 82] used a NLP algorithm that included laboratory data, radiology reports, and clinic notes to increase the sensitivity identification of patients with cirrhosis compared to commonly used noninvasive markers. A NAFLD (nonalcoholic fatty liver disease) identification algorithm was also created using NLP and was superior to an algorithm using ICD-9 data alone (AUC of 0.85 vs. 0.75, $p < 0.0001$) [83]. Hepatocellular adenomas were also identified with high accuracy with similar methods [57]. Sada et al. [84] identified patients with hepatocellular carcinoma by using pathology reports from VHA database with significantly increased positive predictive value compared to diagnosis codes alone (97% vs. 68%, respectively).

In a national retrospective VHA database, NLP was used in an effort to identify patients diagnosed with hepatorenal syndrome (HRS), an acute condition that is more challenging to identify with NLP than chronic conditions that typically have much higher data density. Improved phenotyping of HRS was noted using NLP over ICD-9 codes [85].

Imler et al. [86] used NLP as a potential tool to predict patients who will be susceptible to alcohol abuse in order to offer intervention before their disease leads to cirrhosis.

Practice Implications

In brief, studies to date have demonstrated that NLP can be successfully utilized in extracting endoscopy quality metrics and analyzing factors associated with better value leading to quality improvement initiatives. In addition, prediction models for several GI diseases and outcomes can be developed with the assistance of NLP systems and colonoscopy surveillance intervals can be accurately predicted. Narrative concepts could be used to enhance case definition for GI disease-related variables that lack administrative codes opening new opportunities for database research. Moreover, NLP can also be used as a surveillance program for early identification of conditions that may have otherwise gone unnoticed.

Future Applications of Natural Language Processing

The future of NLP consists of routine use of software applications to extract key information from unstructured data for both clinical research and quality improvement [87]. Once the NLP is set up, it can review thousands of records quickly and provide accurate reports. Incorporation of these tools across multiple centers may permit tracking of quality measures through national registries and provide feedback to providers, administrators, and payers in order to demonstrate adherence to national benchmarks and would constitute the basis of quality improvement initiatives. Natural Language Processing potentially can be used for reporting quality metrics to the Centers of Medicare and Medicaid services for national benchmarking. Such programs have the potential to significantly reduce the burden on practitioners in reporting quality metrics in a timely, low-cost, and accurate matter. In fact, NLP linked to clinical decision support software was more accurate in determining colonoscopy surveillance intervals than expert gastroenterologists [88]. This can help in monitoring both overuse and underuse of colonoscopy and may reduce inappropriate referral for open access endoscopy. As many quality metrics are currently only based on billing and administrative data, incorporating NLP tools to efficiently extract information from EHRs may be better able to reflect the quality of care in complex conditions such as IBD. In addition, many studies that were thought to be prohibitively expensive when using a manual data extraction method would be feasible. Another area with potential for future applications is the integration of NLP into clinical decision support systems. CDS requires not only patient-specific information extracted from the EHR but also medical knowledge regarding best practices in diagnosing and treating a range of conditions. Consequently, advanced NLP systems are needed to find and formally present publications, guidelines, and actionable recommendations. The ideal system will have an NLP module that monitors the EHR for insertion of new data into specific fields. For example, when a patient presents to the hospital for upper gastrointestinal bleeding and data are entered into the EHR, the NLP system will extract information and look-up decision rules such as the Glasgow–Blatchford score in order to identify patients who can be safely discharged and others who will likely require early medical intervention. Furthermore, the system can identify patient-specific transfusion thresholds based on comorbidities among other specific interventions required. These advanced NLP systems will be able to process text and document types ranging from informal notes typed to a patient's records from multiple providers and incorporated to highly structured peer-reviewed publications in scientific journals.

An NLP–CDS system can assist on an administrative level, as an integrated NLP system will be able to automatically map endoscopy reports and assign appropriate ICD codes for financial, and analytical purposes. A NLP has also the potential to streamline prior authorizations. "Next-generation" electronic prior authorization will become a decision support tool with the incorporation of evidence-based algorithms and machine learning. The algorithmic programs factor in the therapy policy and restrictions, gather patient-specific data, and outcomes of patient populations with similar characteristics to present the prescriber with therapy recommendations. Selection of one of these recommendations triggers automatic prior authorization approval. This would be particularly of benefit in IBD where prior authorization burden is linked to treatment delays, disease progression, and patient suffering [89].

Another advantage of NLP is that it may help providers in using natural language in patient care documentation, as structured note systems force providers to create unnatural and overly structured notes which take extra time and are hard to read [90, 91]. Future NLP systems will decrease the workload on healthcare providers through automatic text summarization. This consists of automatically summarizing multiple records to produce a concise and fluent summary. This is generated through extractive methods that work by identifying important sections of the text and generating them verbatim, and more importantly abstractive methods that generate a new shorter text that conveys the most critical information [92]. For example, when a patient with extensive history of Crohn's disease presents for a new clinic visit, automatic text summarization can effectively condense the high-yield information required for a treatment plan including previous therapies, endoscopy reports, and immunizations records. Another technology that will further NLP is continuous voice recognition. Integrating voice recognition software with a NLP system will substantially enhance its functionality and enable physicians to dictate their report while the natural processor translates the report into a structured encoded form [93]. Ambient virtual scribes can interpret the physician's narrative, use NLP to parse the information, detect structured data, assign the necessary ICD-10 code, and prepare orders and electronic prescriptions for sign-off.

In the research domain, NLP may have a role in enhancing the efficiency of physician decision-making in clinical trial enrollment by reducing the pool of potential candidates for screening and increasing the efficacy of trial–patient matching. This has the potential to significantly reduce the effort to conduct clinical research and expand both access to trials and number of trials available [94]. The pharmaceutical industry may also benefit from NLP applications through computational phenotyping and biomarker discovery. With the sheer numbers of research publications accumulating in public and proprietary repositories, no human team, however specialized, can maintain an up-to-date overview. While current information extraction software consists of text mining of named entity types with explicitly known relations, the future of NLP lies in finding new pieces of information that are not explicitly stated in available documents and have to be discovered by associative, semantically unspecified relationships. An autonomous, self-organized semantic engine was able to discover potential novel biomarkers and phenotypes for diabetes and obesity by self-organized text mining of 120,000 PubMed abstracts, public clinical trial summaries, and internal research documents. This shows promise and has the potential to impact pharmaceutical research, for example by shortening time to market of novel drugs, or speed up early recognition of dead ends and adverse events [95].

On a larger scale, NLP may have an important role in the future for population surveillance. For example, it can assist in identifying health disparities in an ethnic or racial group in certain chronic conditions. In addition, NLP may have a role in predictive analytics in health care. A recent study showed that a NLP system accurately predicted suicide attempts by monitoring social media activity [96]. With the high prevalence of chronic diseases in gastroenterology, NLP may be used to predict patients at higher risk of developing psychosocial conditions and medication non-adherence and would therefore benefit from early interventions.

Limitations and Open Problems

As with many new technologies, the effective adoption and integration of NLP and NLP-related tools into clinical practice remain an open challenge. This section highlights several of the most salient and pressing concerns as they relate to limitations of NLP in GI and other sub-specialty practices, as well as presents opportunities for their advancement.

Generalizability of NLP Application Tools

Although the application and validation of NLP methodologies in single-site settings are an essential component of the development pipeline, if such tools are to be used widely they must be adapted *across* EHR systems and reporting styles [96]. Work by Carrell et al. has illustrated a number of important challenges encountered in translating an NLP system measuring colonoscopy quality developed in one academic medical center to four diverse healthcare systems where notable challenges included the diverse language in endoscopy and pathology reports used between different sites, heterogeneous report structure, and a lack of metadata establishing linkages between pathology and colonoscopy reports in comprehensive EHRs [6].

However, challenges in generalizing methods span well-beyond changes and updates in EHR systems, or variation of policies and practices among different institutions. Variability in the availability of free text itself (vs. scanned documents and images saved into the EHR) has a profound impact on application usability, a concern that has become increasingly relevant as NLP systems are now beginning to rely on the merger of endoscopy images and reports. This integration of sources presents several challenges in GI as, for example, current NLP tools aiming to evaluate ADR and other quality measures in endoscopy may only focus on free text provided in the procedure and pathology reports. Yet, the indication may, however, be on a clinic progress note, and the withdrawal time may be calculated by comparing the time stamps on the photographs of the appendiceal orifice and retroflexion in the rectum. Moreover, while a large number of NLP-based tools have been applied on colonoscopy reports, expanding to outpatient reports and discharge summaries has illuminated similar issues given the diversity of systems used to record patient data.

First steps in addressing such diversity have utilized large multicenter systems that utilize standardized EHR systems such as those found in the VHA [97]. Nonetheless, these controlled approaches have yet to be tested with registries such as The GI Quality Improvement Consortium. For such high stakes applications for public reporting of quality and pay-for-performance tools, gastroenterologists may not accept the level of accuracy provided by currently used NLP tools.

Report Language and Structure

Similar to other NLP tasks, there have been challenges to the standardization of clinical text found in GI reports as a result of spellings and alternative phrasings discussed earlier. Moreover, ambiguity in abbreviations is also common: "ANC," for example, may stand for "acute necrotic collection" or "absolute neutrophil count," and "IM" may stand for "intestinal metaplasia," "intramucosal," or "intramuscular," making any form of semantic and phrase analysis difficult. Moreover, as a product of how these reports are generated, similar terms can take on extremely different meanings based on the section being completed. For example, the term "bleeding" has different meanings whether it is in the indications, findings, or recommendations section of a colonoscopy report. These complexities are further compounded, as it is known that a high degree of variability exists in the description of pathology specimens provided during procedures such as endoscopy both within and across sites (e.g., A, B, C; 1, 2, 3; part 1, part 2, part 3). Finally, reliance on NLP techniques can be problematic in both clinical and research settings when evaluating subjective GI symptoms in such as dyspepsia or bloating that may need to be methodologically evaluated to delineate the clinical phenotype. Depending on software to quantitatively evaluate these symptoms may erroneously affect research results and clinical decisions.

Direct adaptation of existing methods to address these barriers represents a challenging endeavor, as the specific characteristics of each section can be revised and updated continuously. Thus, doing so would necessitate dedicating substantial time and effort to software modification, frequent development cycles, and the creation of novel NLP techniques. Broadly, current efforts have focused on the linguistic elements. On the one side, expanding the current view of embeddings as a single representation of a term, with the creation of context-specific representations [98, 99], while from another prospective seeking to provide meaningful translation between subjective symptom descriptions and clinical presentation [100, 101].

Practitioner Utilization: General Acceptability, Training, and Incentive Structure

At a high level, NLP tools encounter many of the well-documented challenges of translating decision support systems into practice without disrupting the patient–provider workflow [4, 102, 103], including poor usability of models in the constraints of daily practice, and concerns over situations in which provider expertise conflicts with the computerized recommendations, especially as outcomes are increasingly tied to compensation or reimbursement. These concerns have only been magnified by a lack of familiarity of NLP by practitioners and study coordinators [6]. Additionally, widespread concern has been seen around acceptability of such technology by the provider and patient alike. Driven by a lack of interpretability of NLP tools, there may be limited satisfaction with computer-based recommendations for therapy or colonoscopy intervals, particularly from those systems using embedding representations and neural network architectures common in recent NLP research.

Widespread efforts have been undertaken to improve acceptability of analytic tools into clinical practice at both the national and organization levels; however, significant work remains to be done. For example, Internet-based clinical decision support systems for appropriateness of colonoscopy have seen limited success but still require further testing and removal of certain organizational and cultural obstacles prior to widespread adoption. These include the development of rigorous evaluation metrics and more consistent reporting standards [4], improved provider buy-in through improved training and integration of providers and administrators into the design of such tools, and finally interdisciplinary collaborations with practitioners and human–computer interaction (HCI) experts remain needed to improve the usability of the tools themselves [104].

Conclusion

In summary, NLP facilitates the access and retrieval of valuable and meaningful healthcare information for multiple uses in the field of digestive diseases with the potential to reduce medical costs and errors. This is particularly beneficial with the availability of large databases and the current healthcare reform. However, NLP systems remain underutilized and still have many challenges to overcome. The development of NLP applications requires close collaboration between NLP experts and clinicians who can provide the necessary domain knowledge [8].

Future integration of this system into EHR may allow for more direct clinician interaction with the resulted data and effective quality improvement. The potential applications for NLP are so complex and numerous that this area of research will succeed only as a result of coordinated community-wide effort.

Acknowledgments We would like to thank An-Lin Cheng, PhD from the department of Bioinformatics at the University of Missouri-Kansas City for her assistance and expertise in developing this manuscript.

Funding None.

Compliance with Ethical Standards

Conflict of interest This statement is to certify that all authors have no conflict of interest related to the current manuscript.

References

1. 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;17:128–144.
2. Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? *J Biomed Inform*. 2009;42:760–772.
3. Leaman R, Khare R, Lu Z. Challenges in clinical natural language processing for automated disorder normalization. *J Biomed Inform*. 2015;57:28–37.
4. Velupillai S, Suominen H, Liakata M, et al. Using clinical Natural Language Processing for health outcomes research: overview and actionable suggestions for future advances. *J Biomed Inform*. 2018;88:11–19.
5. Névéol A, Dalianis H, Velupillai S, Savova G, Zweigenbaum P. Clinical natural language processing in languages other than english: opportunities and challenges. *J Biomed Semant*. 2018;9:12.
6. Carrell DS, Schoen RE, Leffler DA, et al. Challenges in adapting existing clinical natural language processing systems to multiple, diverse health care settings. *J Am Med Inform Assoc JAMIA*. 2017;24:986–991.
7. Zozus MN, Pieper C, Johnson CM, et al. Factors affecting accuracy of data abstracted from medical records. *PloS One*. 2015;10:e0138649.
8. Wang Y, Wang L, Rastegar-Mojarad M, et al. Clinical information extraction applications: a literature review. *J Biomed Inform*. 2018;77:34–49.
9. Luque C, Luna JM, Luque M, Ventura S. An advanced review on text mining in medicine. *WIREs Data Min Knowl Discov*. 2019;9:e1302.
10. Simpson MS, Demner-Fushman D. Biomedical text mining: a survey of recent progress. In: Aggarwal CC, Zhai C, eds. *Mining Text Data*. Boston: Springer; 2012:465–517.
11. Belskaja IK. Machine translation of languages. *Research*. 1957;10:383–389.
12. Manning CD, Raghavan P, Schütze H. *Introduction to Information Retrieval*. Cambridge: Cambridge University Press; 2008.
13. Jurafsky D, Martin JH. *Speech and language processing: an introduction to natural language processing, computational*

linguistics, and speech recognition. Upper Saddle River: Pearson Prentice Hall; 2009.
14. Calijorne Soares MA, Parreiras FS. A literature review on question answering techniques, paradigms and systems. *J King Saud Univ Comput Inf Sci*. 2018. https://doi.org/10.1016/j.jksuci.2018.08.005.
15. Lloret E. Palomar M Text summarisation in progress: a literature review. *Artif Intell Rev*. 2012;37:1–41.
16. Gentzkow M, Kelly B, Taddy M. Text as data. *J Econ Lit.*. 2019;57:535–574.
17. Crawford M, Khoshgoftaar TM, Prusa JD, Richter AN, Al Najada H. Survey of review spam detection using machine learning techniques. *J Big Data*. 2015;2:23.
18. Liddy ED. Natural language processing. In: Drake MA, ed. *Encyclopedia of Library and Information Science*, 2nd Ed. New York: Marcel Decker, Inc.; 2001.
19. Blei DM. Probabilistic topic models. *Commun ACM*. 2012;55:77–84.
20. Manning CD. Part-of-speech tagging from 97% to 100%: Is it time for some linguistics? In: Gelbukh AF, ed. *Computational Linguistics and Intelligent Text Processing*. Berlin: Springer; 2011:171–189.
21. Ferraro JP, Daumé H, DuVall SL, Chapman WW, Harkema H, Haug PJ. Improving performance of natural language processing part-of-speech tagging on clinical narratives through domain adaptation. *J Am Med Inform Assoc JAMIA*. 2013;20:931–939.
22. Goyal A, Gupta V, Kumar M. Recent Named Entity Recognition and Classification techniques: a systematic review. *Comput Sci Rev*. 2018;29:21–43.
23. Kübler S, McDonald R, Nivre J. Dependency parsing. *Synth Lect Hum Lang Technol*. 2008;2:1–127.
24. Shen W, Wang J, Han J. Entity linking with a knowledge base: issues, techniques, and solutions. *IEEE Trans Knowl Data Eng*. 2015;27:443–460.
25. Giatsoglou M, Vozalis MG, Diamantaras K, Vakali A, Sarigiannidis G, Chatzisavvas KCh. Sentiment analysis leveraging emotions and word embeddings. *Expert Syst Appl*. 2017;69:214–224.
26. Mitra B, Craswell N. Neural text embeddings for information retrieval. In: *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. Association for Computing Machinery, New York, NY, USA; 2017:813–814.
27. Bengio S, Heigold G. Word embeddings for speech recognition. In: *Interspeech*; 2014:1053–1057.
28. Le Q, Mikolov T. Distributed representations of sentences and documents. In: *Proceedings of the 31st International Conference on International Conference on Machine Learning*; volume 32; 2014.
29. Sennrich R, Haddow B, Birch A. Neural machine translation of rare words with subword units. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers). Berlin, Germany: Association for Computational Linguistics; 2016:1715–1725.
30. Kim Y, Jernite Y, Sontag D, Rush AM. Character-aware neural language models. *CoRR*, 2015.
31. Wieting J, Bansal M, Gimpel K, Livescu K. Charagram: embedding words and sentences via character n-grams. In: *EMNLP*. 2016.
32. Assale M, Dui LG, Cina A, Seveso A, Cabitza F. The revival of the notes field: leveraging the unstructured content in electronic health records. *Front Med*. 2019. https://doi.org/10.3389/fmed.2019.00066.
33. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. *J Am Med Inform Assoc JAMIA*. 2011;18:544–551.
34. Cai T, Giannopoulos AA, Yu S, et al. Natural language processing technologies in radiology research and clinical applications. *Radiogr Rev Publ Radiol Soc N Am Inc*. 2016;36:176–191.
35. Yim W-W, Yetisgen M, Harris WP, Kwan SW. Natural language processing in oncology: a review. *JAMA Oncol*. 2016;2:797–804.
36. Pons E, Braun LMM, Hunink MGM, Kors JA. Natural language processing in radiology: a systematic review. *Radiology*. 2016;279:329–343.
37. Murff HJ, FitzHenry F, Matheny ME, et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. *JAMA*. 2011;306:848–855.
38. Vo E, Davila JA, Hou J, et al. Differentiation of ileostomy from colostomy procedures: assessing the accuracy of current procedural terminology codes and the utility of natural language processing. *Surgery*. 2013;154:411–417.
39. Elkin PL, Froehling DA, Wahner-Roedler DL, Brown SH, Bailey KR. Comparison of natural language processing biosurveillance methods for identifying influenza from encounter notes. *Ann Intern Med*. 2012;156:11–18.
40. Calvo RA, Milne DN, Hussain MS, Christensen H. Natural language processing in mental health applications using non-clinical texts. *Nat Lang Eng*. 2017;23:649–685.
41. Conway M, O'Connor D. Social media, big data, and mental health: current advances and ethical implications. *Curr Opin Psychol*. 2016;9:77–82.
42. Sun W, Cai Z, Li Y, Liu, F, Fang S, Wang G. Data processing and text mining technologies on electronic medical records: a review. *J Healthc Eng*. 2018;2018:4302425.
43. Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic Acids Res*. 2004;32:267–270.
44. McCray AT. An upper-level ontology for the biomedical domain. *Int J Genom*. 2003;4:80–84.
45. Zhang Y, Chen Q, Yang Z, Lin H, Lu Z. BioWordVec, improving biomedical word embeddings with subword information and MeSH. *Sci Data*. 2019. https://doi.org/10.1038/s41597-019-0055-0.
46. Gero Z, Ho J. PMCVec: distributed phrase representation for biomedical text processing. *J Biomed Inform*. 2019;3:100047.
47. Bai T, Chanda AK, Egleston BL, Vucetic S. EHR phenotyping via jointly embedding medical concepts and words into a unified vector space. *BMC Med Inform Decis Mak*. 2018;18:123.
48. Liu K, Hogan WR, Crowley RS. Natural Language Processing methods and systems for biomedical ontology learning. *J Biomed Inform*. 2011;44:163–179.
49. Feczko PJ, Ackerman LV, Halpert RD, Simms SM. A computer-based gastrointestinal information management system. *Radiology*. 1984;152:297–300.
50. Rex DK, Schoenfeld PS, Cohen J, et al. Quality indicators for colonoscopy. *Am J Gastroenterol*. 2015;110:72–90.
51. Corley DA, Jensen CD, Marks AR, et al. Adenoma detection rate and risk of colorectal cancer and death. *N Engl J Med*. 2014;370:1298–1306.
52. Barclay RL, Vicari JJ, Doughty AS, Johanson JF, Greenlaw RL. Colonoscopic withdrawal times and adenoma detection during screening colonoscopy. *N Engl J Med*. 2006;355:2533–2541.
53. Adler A, Wegscheider K, Lieberman D, et al. Factors determining the quality of screening colonoscopy: a prospective study on adenoma detection rates, from 12,134 examinations (Berlin colonoscopy project 3, BECOP-3). *Gut*. 2013;62:236–241.
54. Rex DK. Who is the best colonoscopist? *Gastrointest Endosc*. 2007;65:145–150.
55. Pike IM. Quality improvement in gastroenterology: a US perspective. *Nat Clin Pract Gastroenterol Hepatol*. 2008;5:550–551.

56. Imler TD, Morea J, Kahi C, et al. Multi-center colonoscopy quality measurement utilizing natural language processing. *Am J Gastroenterol*. 2015;110:543–552.
57. Imler TD, Morea J, Kahi C, Imperiale TF. Natural language processing accurately categorizes findings from colonoscopy and pathology reports. *Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc*. 2013;11:689–694.
58. Mehrotra A, Dellon ES, Schoen RE, et al. Applying a natural language processing tool to electronic health records to assess performance on colonoscopy quality measures. *Gastrointest Endosc*. 2012;75:1233–1239.
59. Harkema H, Chapman WW, Saul M, Dellon ES, Schoen RE, Mehrotra A. Developing a natural language processing application for measuring the quality of colonoscopy procedures. *J Am Med Inform Assoc JAMIA*. 2011;18:i150–i156.
60. Gawron AJ, Thompson WK, Keswani RN, Rasmussen LV, Kho AN. Anatomic and advanced adenoma detection rates as quality metrics determined via natural language processing. *Am J Gastroenterol*. 2014;109:1844–1849.
61. Lee JK, Jensen CD, Levin TR, et al. Accurate identification of colonoscopy quality and polyp findings using natural language processing. *J Clin Gastroenterol*. 2019;53:e25–e30.
62. Raju GS, Lum PJ, Slack RS, et al. Natural language processing as an alternative to manual reporting of colonoscopy quality metrics. *Gastrointest Endosc*. 2015;82:512–519.
63. Deutsch JC. Colonoscopy quality, quality measures, and a natural language processing tool for electronic health records. *Gastrointest Endosc*. 2012;75:1240–1242.
64. Marcondes FO, Gourevitch RA, Schoen RE, Crockett SD, Morris M, Mehrotra A. Adenoma detection rate falls at the end of the day in a large multi-site sample. *Dig Dis Sci*. 2018;63:856–859. https://doi.org/10.1007/s10620-018-4947-1.
65. Patel VD, Thompson WK, Lapin BR, Goldstein JL, Yen EF. Screening colonoscopy withdrawal time threshold for adequate proximal serrated polyp detection rate. *Dig Dis Sci*. 2018;63:3084–3090. https://doi.org/10.1007/s10620-018-5187-0.
66. Abdul-Baki H, Schoen RE, Dean K, et al. Public reporting of colonoscopy quality is associated with an increase in endoscopist adenoma detection rate. *Gastrointest Endosc*. 2015;82:676–682.
67. Mehrotra A, Morris M, Gourevitch RA, et al. Physician characteristics associated with higher adenoma detection rate. *Gastrointest Endosc*. 2018;87:778–786.
68. Crockett SD, Gourevitch RA, Morris M, et al. Endoscopist factors that influence serrated polyp detection: a multicenter study. *Endoscopy*. 2018;50:984–992.
69. Gourevitch RA, Rose S, Crockett SD, et al. Variation in pathologist classification of colorectal adenomas and serrated polyps. *Am J Gastroenterol*. 2018;113:431–439.
70. Hong SN, Son HJ, Choi SK, et al. A prediction model for advanced colorectal neoplasia in an asymptomatic screening population. *PloS One*. 2017;12:e0181040.
71. Blumenthal DM, Singal G, Mangla SS, Macklin EA, Chung DC. Predicting non-adherence with outpatient colonoscopy using a novel electronic tool that measures prior non-adherence. *J Gen Intern Med*. 2015;30:724–731.
72. Thrift AP, Natarajan Y, Mansour NM, et al. Tu1117—using natural language processing to accurately identify dysplasia in pathology reports for patients with Barrett's Esophagus. *Gastroenterology*. 2018;154:896–897.
73. Hou JK, Soysal E, Moon S, et al. Su1815 natural language processing accurately identifies colorectal dysplasia in a national cohort of veterans with inflammatory bowel disease. *Gastroenterology*. 2016;150:S560–S561.
74. Ananthakrishnan AN, Cai T, Cheng S-C, et al. Tu1276 improving case definition of Crohn's disease and ulcerative colitis in electronic medical records using natural language processing—a novel informatics approach. *Gastroenterology*. 2012;142:S-791.
75. Cai T, Lin T-C, Bond A, et al. The association between arthralgia and Vedolizumab using natural language processing. *Inflamm Bowel Dis*. 2018;24:2242–2246.
76. Hou JK, Chang M, Nguyen T, et al. Automated identification of surveillance colonoscopy in inflammatory bowel disease using natural language processing. *Dig Dis Sci*. 2013;58:936–941. https://doi.org/10.1007/s10620-012-2433-8.
77. ASGE Standards of Practice Committee, Anderson MA, Fisher L, et al. Complications of ERCP. *Gastrointest Endosc*. 2012;75:467–473.
78. Imler TD, Sherman S, Imperiale TF, et al. Provider-specific quality measurement for ERCP using natural language processing. *Gastrointest Endosc*. 2018;87:164–173.
79. Al-Haddad MA, Friedlin J, Kesterson J, Waters JA, Aguilar-Saavedra JR, Schmidt CM. Natural language processing for the development of a clinical registry: a validation study in intraductal papillary mucinous neoplasms. *HPB*. 2010;12:688–695.
80. Mehrabi S, Schmidt CM, Waters JA, et al. An efficient pancreatic cyst identification methodology using natural language processing. *Stud Health Technol Inform*. 2013;192:822–826.
81. Chang EK, Yu CY, Clarke R, et al. Mo1050 using an automated diagnostic algorithm that utilizes electronic health records and natural language processing to define a population with cirrhosis. *Gastroenterology.*. 2015;148:S-1074.
82. Kung R, Ma A, Dever JB, et al. Mo1043 a natural language processing algorithm for identification of patients with cirrhosis from electronic medical records. *Gastroenterology*. 2015;148:1071–1072.
83. Corey KE, Kartoun U, Zheng H, Shaw SY. Development and validation of an algorithm to identify nonalcoholic fatty liver disease in the electronic medical record. *Dig Dis Sci*. 2016;61:913–919. https://doi.org/10.1007/s10620-015-3952-x.
84. Sada Y, Hou J, Richardson P, El-Serag H, Davila J. Validation of case finding algorithms for hepatocellular cancer from administrative data and electronic health records using natural language processing. *Med Care*. 2016;54:e9–e14.
85. Koola JD, Davis SE, Al-Nimri O, et al. Development of an automated phenotyping algorithm for hepatorenal syndrome. *J Biomed Inform*. 2018;80:87–95.
86. Imler TD, Ring N, Crabb DW. Su1403 medical, social, and legal risks to predict alcoholic liver disease using natural language processing and advanced analytics. *Gastroenterology*. 2015;148:S-499.
87. Jha AK. The promise of electronic records: around the corner or down the road? *JAMA*. 2011;306:880–881.
88. Imler TD, Morea J, Imperiale TF. Clinical decision support with natural language processing facilitates determination of colonoscopy surveillance intervals. *Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc*. 2014;12:1130–1136.
89. Bhat S, Zahorian T, Robert R, Farraye FA. Advocating for patients with inflammatory bowel disease: how to navigate the prior authorization process. *Inflamm Bowel Dis*. 2019;25:1621–1628.
90. Walsh SH. The clinician's perspective on electronic health records and how they can affect patient care. *BMJ*. 2004;328:1184–1187.
91. Coiera E. When conversation is better than computation. *J Am Med Inform Assoc JAMIA*. 2000;7:277–286.
92. Maybury M. *Advances in Automatic Text Summarization*. New York: MIT Press; 1999.
93. Friedman C, Hripcsak G. Natural language processing and its future in medicine. *Acad Med*. 1999;74:890–895.
94. Ni Y, Wright J, Perentesis J, et al. Increasing the efficiency of trial-patient matching: automated clinical trial eligibility

Pre-screening for pediatric oncology patients. *BMC Med Inform Decis Mak*. 2015. https://doi.org/10.1186/s12911-015-0149-3.
95. Trugenberger CA, Wälti C, Peregrim D, Sharp ME, Bureeva S. Discovery of novel biomarkers and phenotypes by semantic technologies. *BMC Bioinform*. 2013;14:51.
96. Coppersmith G, Leary R, Crutchley P, Fine A. Natural language processing of social media as screening for suicide risk. *Biomed Inform Insights*. 2018. https://doi.org/10.1177/1178222618792860.
97. McCoy AB, Wright A, Eysenbach G, et al. State of the art in clinical informatics: evidence and examples. *Yearb Med Inform*. 2013;8:13–19.
98. Zheng X, Feng J, Chen Y, Peng H, Zhang W. Learning context—specific word/character embeddings. In: *Proceedings of AAAI Conference on Artificial Intelligence*; 2017:3393–3399.
99. Li P, Bing L, Lam W, Li H, Liao Y. Reader-aware multi-document summarization via sparse coding. In: *IJCAI*; 2015.
100. Forbush TB, Gundlapalli AV, Palmer MN, et al. Sitting on Pins and needles: characterization of symptom descriptions in clinical notes. *AMIA Summits Transl Sci Proc*. 2013;2013:67–71.
101. Koleck TA, Dreisbach C, Bourne PE, Bakken S. Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. *J Am Med Inform Assoc*. 2019;26:364–379.
102. Chapman WW. Closing the gap between NLP research and clinical practice. *Methods Inf Med*. 2010;49:317–319.
103. Liberati EG, Ruggiero F, Galuppo L, et al. What hinders the uptake of computerized decision support systems in hospitals? A qualitative study and framework for implementation. *Implement Sci*. 2017;12:113.
104. Horsky J, Schiff GD, Johnston D, Mercincavage L, Bell D, Middleton B. Interface design principles for usable decision support: a targeted review of best practices for clinical prescribing interventions. *J Biomed Inform*. 2012;45:1202–1216.