NATURAL LANGUAGE PROCESSING (NLP) APPLICATIONS IN PATIENT CARE: A SYSTEMATIC ANALYSIS

Zixing Shen, New Mexico State University

ABSTRACT

Artificial intelligence (AI) has been gaining fresh momentum with remarkable breakthroughs. AI-powered applications have been developed and deployed rapidly for various business functions across different industries. This paper focuses on natural language processing (NLP) in the context of patient care. The use of NLP in the medical field has been growing fast and drawing more and more attention. Although NLP has been dominantly used in clinical and translation research, the recent technological development and the increasing availability of patient data have provided opportunities for the direct use of NLP to patient care, the core of the medicine. This study assesses the applications of NLP in patient care. Specifically, it conducts a review of publications on how NLP is applied in patient care from 2004 to 2019. The analysis of the literature has provided interesting insights and trends, as well as gaps in the applications of NLP in patient care. This study informs the researchers and practitioners of the status quo of the NLP applications in patient care and helps stimulate research efforts that can lead to more advances in applying NLP to clinical decision support and operation.

Keywords: Natural Language Processing, NLP, NLP applications, patient care, clinical decision support

INTRODUCTION

Artificial intelligence (AI) has been gaining fresh momentum with remarkable breakthroughs. AI-powered applications have been developed and deployed rapidly for various business functions across different industries. This paper focuses on natural language processing (NLP) in the context of patient care. NLP is the automatic analysis and representation of human languages (Joshi, 1991). It is concerned with analyzing, understanding, and generating responses that ultimately make computers to interface with human languages rather than computer languages. NLP can explain a structure or a command to a computer in the natural language, written and spoken, as used by humans, translate it into a format that a computer can understand and process, and generate it back to the human user.

Over the years, NLP has become more and more sophisticated and powerful. It now has various capabilities such as content categorization, topic discovery, sentiment analysis, document summarization, and machine translation. NLP is increasingly used on large amounts of data written in plain text to get insights from texts. In addition, NLP can process languages in voice format. Examples of voice NLP applications include Xbox, Skype, and Apple’s Siri. With its wide and growing usage, NLP is now included in programming languages like Python and R.

NLP has been used in the medical field for decades. Its applications in medicine have been growing fast and drawing more and more attention. Open-source NLP software tailored explicitly to the medical text, such as the clinical text analysis and knowledge extraction system (cTAKES) and the concept extraction-based text analysis system (CETAS), have been developed. NLP methods such as speech information recognition (Xiao et al., 2015), semantic labeling (Zhang et al., 2015), syntactic parsing (Sidorov et al., 2014), and negation detection (Nikfarjam et al., 2015) have been developed for medical information processing. Along with other technologies such as machine learning, NLP is applied in medical research including, but not limited to, disease classification (Reddy & Bhaskar, 2018), disease relation identification (Gu et al., 2019), gene-name recognition (Hakenberg et al., 2005), phenotype discovery (Bai et al., 2018).

Research on NLP applications in the medical field has been reviewed. Some of the reviews focus on the use of NLP in specific areas such as oncology (Yim et al., 2016), radiology (Cai et al., 2016), and chronic diseases (Sheikhalishahi et al., 2019), or in certain types of free-text data like electronic patient-authored text data (Dreisbach et al., 2019). Others synthesize the applications of NLP in biomedical research (Cohen, 2014). and the software tools used in NLP applications (Cohen et al., 2013). There are also analyses of the NLP academic research themes and scientific collaboration patterns (Chen et al., 2018). While these reviews have improved the understanding of NLP applications in the medical field, they have not provided a comprehensive picture of the NLP applications in patient care.

Patient care, at the core of the medical field, consists of testing, diagnosing, and treating patients and involves primarily doctors, nurses, clinical pharmacists, as well as patients. Although NLP has been dominantly applied in clinical and translation research, the recent technological development in NLP and the increasing availability of patient data in the format of text and image have provided opportunities for the direct use of NLP in patient care. Patient data, from various sources, such as Electronic Health Records (EHRs) and Electronic Medical Records (EMRs), as well as patients in social media, are rich in synonymy and semantically similar and related concepts. They are continuously growing in magnitude, particularly with real-time imaging and point of care devices, as well as wearable computing and mobile health technologies. The increasingly massive patient data contain a considerable amount of information valuable to clinicians and patients. However, much of such information comes in an unstructured form. NLP is thus crucial for transforming relevant unstructured information hidden in text (e.g., doctor's notes) and speech (e.g., doctor's dictates) into structured information.

Thanks to technological advances in information processing and data storage, new NLP methods and techniques, such as speech information recognition, semantic labeling, syntactic parsing, word sense disambiguation, negation detection, and temporal analysis, have emerged (Chen et al., 2018). They can meet the need for syntactic and semantic understanding of the languages, helps resolve ambiguity in language, and adds useful numeric structure to the data for many downstream applications, such as speech recognition and text analytics. As such, they enable efficient and effective analyses of vast amounts of patient data. Therefore, NLP applications can be extremely instrumental in improving clinical decision support and advancing patient care.

Given the potentials of NLP applications (Demner-Fushman et al., 2009) as well as the growing interests in AI and big data, it is of great significance to analyze the literature and understand how

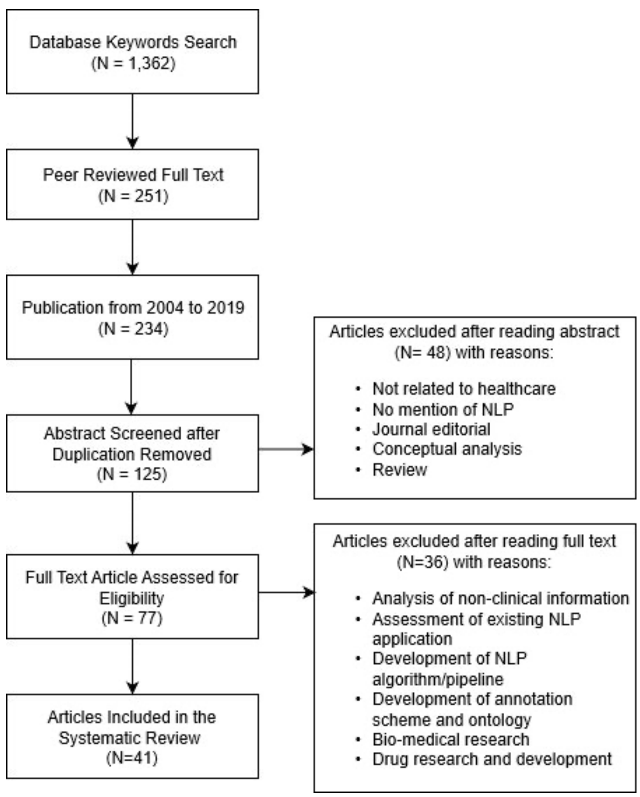
NLP has been applied in patient care. This paper answers this call. It aims to inform researchers and practitioners of the status quo of the NLP applications in patient care. This study further synthesizes and discusses the current trends and gaps across studies and proposes major directions for future research in NLP applications in patient care.

METHOD

Search Strategy and Selection Process

The PRISMA model (Liberati et al., 2009) was followed in searching the literature and selecting papers (See Figure 1). First, the EBSCO database was queried with several keywords – natural language processing, NLP, patient care, healthcare, and medicine. EBSCO is one of the largest database sources for high-quality peer-reviewed journals that publish research related to healthcare and medicine. The keyword search generated 1362 papers. Of them, 251 are peer-reviewed journal papers with full-text available in the database. Then peer-reviewed journal papers published from 2004 to 2019, the most recent one and half decades, were included. This reduced the sample size to 234. After duplicated papers were removed, 125 unique pieces remained.

Figure 1. PRISMA Flow Diagram



Next, the abstract of the 125 papers was checked. Papers that were not related to healthcare and medicine and NLP (e.g., Simon et al., 2014) were excluded. Journal editorials (e.g., Tao et al., 2017), conceptual analyses (e.g., Hope et al., 2012), and reviews (e.g., Kohane, 2011) were also removed. Forty-eight papers were eliminated after reading the abstract. Then the full text of the remaining 77 papers was examined. The following categories of articles were excluded as they did not have any NLP application directly related to patient care: (1) analysis of medical publication text for non-clinical information (e.g., Lerchenmueller and Sorenson, 2016); (2) assessment of the effectiveness of an existing NLP method (e.g., Ferrández et al., 2012); (3) development of an NLP method (e.g., Jackson et al., 2018),; (4) development of an annotation scheme (e.g., Long et al., 2019); (5) discovery research in the biomedical field (Klann et al., 2015); (6) drug research and development (e.g., Peissig et al., 2007). Finally, a total of 41 papers (indicated with * in the reference section) with NLP applications specific to patient care were selected for coding.

Coding Procedure

A coding sheet was first developed. Special attention was paid to NLP utility (that is, what the NLP application was used for in patient care). Lexicon, extraction, and classification information was coded, as well. The software used in the extraction and the algorithm used in the classification, when applicable, were also noted. The second section of the coding sheet collected information on the data used in the NLP application, including text type, data source, and the number of texts. The last section of the coding sheet gathered information on study publication year and country affiliation. Then, each paper was read and analyzed using the coding sheet. Table 1 shows an example of the coding of a paper. The appendix lists the coding of all the 41 papers.

Table 1. Example of the Coding Sheet

Paper: Hong and colleagues, 2017	
NPL Application in Patient Care	
Clinical Utility	Screening advanced colorectal neoplasia (CRN)
Lexicon/dictionary	Custom made
Extraction	
• Method	Matching terms
• Software tool	CETAS (Concept Extraction-based Text Analysis System)
Classification	
• Approach	Statistical analysis
• Algorithm	Logistic regression
Data Used in NLP Application	
Data source	Samsung Medical Center
Text type	colonoscopy reports and pathology reports
Number of texts	49,450
Publication Year and Country	
Publication Year	2017
Country	Korea

RESULTS

Patient Care Application

Utility. A patient care application can be the use of NLP by medical professionals in the observation and treatment of actual patients and by patients in their medical journeys. It can also be the use of NLP in the administration and management of patient care. The utilities of NLP applications in patient care were found in three areas - clinical decision support, patient care administration, and patient decision making and support, as summarized in Table 2.

Table 2. Patient Care Utility

Category	N	%
Clinical Decision Support	27	66%
Diagnose disease	9	
Detect medication/drug use	6	
Predict/identify clinical conditions and outcomes	12	
Patient Care Administration	13	32%
Quality control	3	
Patient care communication	4	
Care cost & utilization	4	
Disease prevention & control	2	
Patient Support & Decision Making	1	2%
Patient behavior and emotions		

The use of NLP for patient support and decision making was scarce (one paper only). De Silva and colleague (2018) applied NLP to investigate the interactions between clinical factors and patient emotions and behaviors in the process of diagnosis, treatment, and recovery of prostate cancer. Clinical decision support was most studied. NLP applications were tasked to provide clinical decision support in diagnosing diseases such as multiple sclerosis (Chase et al., 2017; Xia et al., 2013), influenza (Ye et al., 2017), nontuberculous mycobacterial disease (Jones et al., 2018), dementia (Shibata et al., 2018), staphylococcus aureus (Jones et al., 2012), polycystic ovary syndrome (Castro et al., 2015), depression (Parthipan et al., 2019), and diabetes (Mishra et al., 2012). NLP applications also helped clinicians in detecting medication discrepancy (Li et al., 2015), antipsychotic polypharmacy (Kadra et al., 2015), adverse drug reactions (Iqbal et al., 2015; Li et al., 2013), and identifying aspirin use (Pakhomov et al., 2010) and opioid misuse (Afshar et al., 2019). Additionally, NLP applications were effective in predicting certain medical conditions like suicidal behavior (Carson et al., 2019; Cook et al., 2016; Poulin et al., 2014; Taylor et al., 2016; Zhong et al., 2018), severe injuries after falls (Toyabe, 2012),smoking status (Regan et al., 2016), skeletal site-specific fractures (Wang, Mehrabi, et al., 2019), advanced colorectal neoplasia (Hong et al., 2017), as well as clinical outcome like patient mortality (Beeksma et al., 219; McCoy et al., 2015; Waudby-Smith et al., 2018).

NLP applications were also found useful for patient care administration in several ways. First, they improved quality control because of their ability to detect the discordance between patient self-report and documentation of symptoms in the medical record (Pakhomov et al., 2008) and assess quality in PTSD (post-traumatic stress disorder) care (Shiner et al., 2012) and imaging diagnosis (Zheng et al., 2019). They also streamlined the communication by fostering the shared meaning between clinician and patients (Balyan et al., 2019), identifying communication failures between home health nurses and physicians (Pesko et al., 2018), and providing more accurate and relevant information to patients in their searching care providers (Cook et al., 2019) and to clinicians for more tailored care (Wang, Wang, et al., 2019). Moreover, NLP applications helped with the management of clinical care cost and utilization as they predicted utilization of ADI (advanced diagnostic imaging) (Zhang et al., 2019), identified patients at risk for readmission (Navathe et al., 2018), tracked healthcare disparities (Wieland et al., 2013) and quantified the healthcare costs and utilization for patients with binge-eating disorder (BED) (Bellows et al., 2015). Finally, NLP applications identified information for disease prevention (Workman & Stoddart, 2012) and detected concerned HIV-related messages in online forums to facilitate medical intervention (Sung et al., 2014), and therefore was useful for disease prevention and control.

Lexicon. Central to NLP is the use of standardized terminology for each concept that is of fundamental interest in a particular field. A concept is an intrinsically unique entity with an unambiguous meaning (e.g., a specific disease such as lung cancer or a symptom such as chest pain). The standardized terminologies for unique concepts are collected in lexicons. Summarized in Table 3 is the lexicon information of the reviewed literature.

Table 3. Lexicon

Category	N	%
Not reported	30	73%
Readily available (e.g., UMLS)	3	7%
Custom made	5	13%
Both	3	7%

The vast majority did not provide explicit information on the lexicon. Of the 11 papers that explicitly reported the lexicon, three used readily available lexicons - UMLS (Unified Medical Language System) (Chase et al., 2017; Li et al., 2015), SNOMED CT (Systematized Nomenclature of Medicine – Clinical Terms) (Jones et al., 2018; Li et al., 2015), and RxNorm dictionary (Li et al., 2015). Customized dictionaries were either configured by referring to existing lexicons (Hong et al., 2017) or created by medical specialists (Castro et al., 2015) to meet the specific needs in processing patient data. Few studies (De Silva et al., 2018, Jones et al., 2012; Xia et al.,2013) complemented the readily available lexicons with in-house developed ones as well.

Extraction. The overall goal of NLP applications in patient care is to determine which concepts are mentioned in a medical report and in what capacity. Extraction is the first task to achieve this. In extraction, NLP analyzes the text to identify individual concepts and their modification by other terms. When extraction has been completed, each individual concept found in the text is output as a separate item in a structured format. The extraction information of the reviewed literature is summarized in Table 4. The various extraction methods can be put into two broad categories – pattern matching and linguistic analysis. Pattern matching makes use of regular expressions or sequences of characters and special symbols that explicitly define a character pattern to be searched for. For example, a list of key terms in searching the clinical notes was used to identify aspirin use (Pakhomov et al., 2010). Similarly, a regular expression approach was used to extract concepts of interest (Mishra et al., 2012).

Linguistic analyses involve using knowledge, both syntactic and semantic, to infer what concepts are mentioned and how each concept modifies other concepts. The syntactic analysis relies on the rules that control the arrangement of words in a sentence to check the text for meaningfulness. For example, an in-house tokenizer and part-of-speech tagger were developed to tokenize and parse the clinical notes (Li et al., 2015). Similarly, structural features of the clinical notes were analyzed using sentence boundary detector, tokenizer, normalizer, part-of-speech tagger, shallow parser, etc. (Carson et al., 2019). On the other hand, semantic analysis is concerned with the knowledge regarding the different meanings of words in the context of a sentence. For example, topic modeling (Zhang et al., 2019) and neural network (Zheng et al., 2019) were used to capture semantic structures and patterns in clinical texts. Syntactic and semantic analyses were used together in one study (De Silva et al., 2018), where the syntactic analysis was used to retrieve patient demographic information and semantic analysis to extract patient-reported side effects. In addition, natural language is also based on other components such as phonetics and morphology. However, because the use of language in clinical settings is more limited than that in general text, the syntactic and semantic approaches achieve sufficient accuracy for NLP applications. This is why only one paper (Shibata et al., 2018) used morphological analysis.

As shown in Table 4, the linguistic approach was preferred over pattern matching. This may be explained by the fact that linguistic methods offer more information than pattern matching and are therefore more capable of analyzing complex concepts in patient data. The review also indicates that the syntactic analysis was the dominant linguistic approach (24 papers), while the use of semantic analysis was scant (i.e., Zhang et al., 2019; Zheng et al., 2019).

Table 4. Extraction

Category	N	%
Pattern Matching	13	32%
Linguistic Analysis	28	68%
Syntactic	24	
Semantic	2	
Hybrid	1	
Morphological analysis	1	

Presented in Table 5 is the summary of the software tools used in extraction. More than half of the papers used readily available software tools, some of which are specialized in NLP (e.g., cTAKES), while others are general-purpose tools with NLP functions (e.g., Python). The most popular NLP tool was cTAKES, followed by Python, GATE (the general architecture for text engineering), and Medline. Other readily available software tools include CETAS (Hong et al., 2017), MedLEE (Medical Language Extraction and Encoding) (Chase et al., 2017), R (Zhang et al., 2019), NLTK (Natural Language Toolkit) (Parthipan et al., 2019) among others. Customized tools were also built to fit the specific needs of NLP applications. While most tools were developed for the English language, tools for other languages were used in four papers. The software of Text Mining Studio and morphological analyzers such as Juman++ and MeCab were used in processing clinical data written in Japanese (Toyabe, 2012). Chinese segmentation tools like Jieba (Zheng et al., 2019) and CKIP (The Chinese Knowledge Information Processing) (Sung et al., 2014) were used for extraction from Chinese documents.

Table 5. Extraction Tool

Category	N	%
Not reported	5	12%
In-house developed	9	22%
Readily available	26	64%
cTAKES	5	
Python	3	
GATE	2	
Medline	2	
Others (e.g., CETAS)	14	
Both	1	2.%

Classification. Classification is performed after extraction. In classification, NLP analyzes the structured data extracted from textual data to determine whether they contain one or more desired concepts and modifiers that indicate that the textual data possess one or more specified characteristics with a given certainty (e.g., positive for a specific disease). The information on the classification approach is summarized in Table 6.

Table 6. Classification

Category	N	%
Classification not reported	1	2%
Knowledge-based rules	18	44%
Statistically inferred rules	22	54%
Regression models	7	
Bayesian models	2	
Conditional random field	2	
Word2Vec	2	
Latent Dirichlet allocations	2	
Others (e.g., random forests)	7	

Classification information was identified in all papers except one (Taylor et al., 2016). The various classification techniques fell into two broad categories – knowledge-based and statistical approach. The knowledge-based approach utilizes the real-world domain knowledge developed by domain specialists. For example, a four-step process was developed by experts for automated medication discrepancy detection (Li et al., 2015). Similarly, the decision rules were manually constructed for detecting falls (Toyabe, 2012). The statistical approach infers rules and patterns directly from data (i.e., a large corpus, like a book, a collection of sentences). Algorithms used in the statistical approach included regression models (Castro et al., 2015; Hong et al., 2017; McCoy et al., 2015; Navathe et al., 2018; Waudby-Smith et al., 2018; Xia et al., 2013; Zhang et al., 2019), Bayesian models (Chase et al., 2017; Ye et al., 2017), conditional random field (Li et al., 2013; Shiner et al., 2012), Word2Vec (Beeksma et al., 2019; De Silva et al., 2018), latent Dirichlet allocations (Afshar et al., 2019; Wang, Wang, et al., 2019), random forests (Carson et al., 2019), genetic programming (Poulin et al., 2014), genetic algorithm (Sung et al., 2014), and convolutional neural network (Zheng et al., 2019). As all these are related to machine learning, the statistical approaches in the reviewed literature heavily used machine learning to derive insights from human languages. Machine learning can automatically learn rules and patterns but requires a large body of data to make a statistically sound inference.

As revealed in Table 6, more research relied on statistically inferred rules than knowledge-based rules in classification. The increasing popularity of the statistical approach over the knowledge-based approach can be attributed to factors like growing volumes and varieties of available medical data, computational processing that is cheaper and more powerful, and affordable data storages that enable complex mathematical calculations to deliver faster, more accurate results. In addition, compared to the statistical approach, the knowledge-based approach is more costly and time-consuming as it requires considerable manual effort to build the necessary knowledge base. Moreover, medical knowledge continually changes, and updating knowledge may also be challenging.

Data

Data source. Text data were acquired from various sources. These different data sources can be grouped into two categories – publicly available and institutional proprietary, as shown in Table 7. Examples of publicly available data sources include online support group discussion (De Silva et al., 2018), portal (Sung et al., 2014), medical publication database (Workman & Stoddart, 2012), the government agency such as Veteran Affair's Health Administration (Bellows et al., 2015; Jones et al., 2012; Jones et al., 2018; Poulin et al., 2014; Shiner et al., 2012). Intuitional proprietary data were from clinics and hospitals like Mayo Clinic (Pakhomov et al., 2008; Pakhomov et al., 2010; Wang, Mehrabi, et al., 2019; Wang, Wang, et al., 2019; Wieland et al., 2013), Shanghai Tongren Hospital (Zheng et al., 2019). One paper did not report data sources (Shibata et al., 2018). As revealed by Table 8, proprietary data were utilized more frequently than publicly available data. This was largely expected given the sensitivity of as well as the legal and regulatory issues related to clinical data.

Table 7. Data Source

Data Source	N	%
Not Reported	1	2%
Publicly available	16	39%
Intuitional Proprietary	24	59%

Text type. The types of text varied between studies. Table 8 provides a summary. The vast majority used text data from EHRs/EMRs. Some papers reported text type with such general terms as clinical notes and clinical records (Parthipan et al., 2019; Zheng et al., 2019), and others provided specific information on the unstructured data, which included physician note (Navathe et al., 2018), nurse note (Pesko et al., 2018), triage note (Zhang et al., 2019), radiology note (Wang, Mehrabi, et al., 2019), patient chart (Wieland et al., 2013), discharge summary (Mishra et al., 2012), laboratory record (Jones et al., 2012), incident report (Toyabe, 2012), imaging report and pathology report (Hong et al., 2017; Zheng et al., 2019). Text data outside EHRs and EMRs were collected and analyzed, as well. Such data included patient narrative (Shibata et al., 2018), drug label (Li et al., 2013), and the patient provided information (Pakhomov et al., 2008), text message

(Cook et al., 2016), online discussion post (De Silva et al., 2018), state licensure list (Cook et al., 2019), and medical publication (Workman & Stoddart, 2012).

Table 8. Text Type

Text Type	N	%
EHRs & EMRs	33	81%
Other (e.g., online post)	7	17%
Both	1	2%

The number of texts. The number of texts used for NLP applications ranged from 42 patient narratives (Shibata et al., 2018) to 4,795,428 online posts (De Silva et al., 2018). Table 9 summarizes the distribution of numbers of texts. Datasets smaller than 10,000 were analyzed in nearly 60% of the papers. Very few studies (Afshar et al., 2019; Balyan et al., 2019; Beeksma et al., 2019; De Silva et al., 2018; Zhang et al., 2019) worked with datasets larger than 100,000.

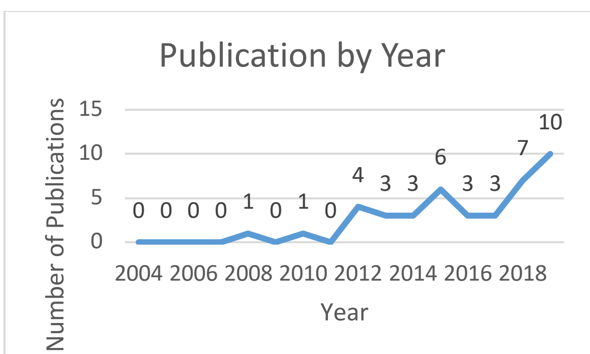
Table 9. Number of Texts

Number of Texts	N	%
N < 1,000	9	21%
1,000 < N < 10,000	15	37%
10,000 < N < 100,000	12	30%
N > 100,000	5	12%

Publication Year and Country

Publication year. As shown in Figure 2, forty studies were published after 2010 and only one (Pakhomov et al., 2008) before 2010. Despite some year to year fluctuation, the interest in NLP applications in patient care has been growing steadily over the past 15 years, especially since 2011. There was a big jump in the last three years, from 2016 to 2019.

Figure 2. Publication by Year

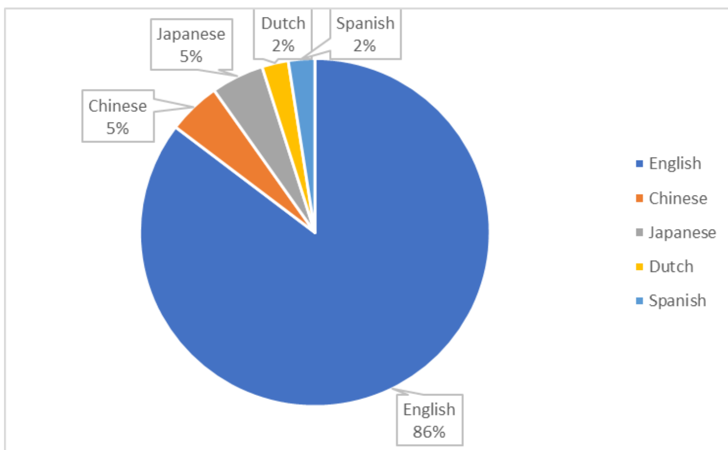


Country. As shown in Table 10, the studies were conducted in nine different countries, with the USA leading the list (29 papers). As shown in Figure 3, English was the most analyzed language (35 papers), and other languages, including Chinese (Sung et al., 2014; Zhang et al., 2019), Japanese (Shibata et al., 2018; Toyabe, 2012), Dutch (Beeksma et al., 2019), and Spanish (Cook et al., 2016), were also studied.

Table 10. Publication by Country

Country	N	%
USA	29	72%
UK	3	8%
China	2	5%
Japan	2	5%
Australia	1	2%
Canada	1	2%
Korea	1	2%
Netherlands	1	2%
Spain	1	2%

Figure 3. Languages in NLP



DISCUSSION

NLP Application

The study has revealed the paucity of NLP applications for patient support and decision-making. The amount of online patient-related texts from social media posts and themed forums is overwhelming and could lead to a significant increase in NLP applications targeting patient support and decision-making. One barrier could be the lack of lexicons for patient-authored text data. Patients use everyday language, not technical, medical terms as clinical professionals. The lexicons for the language used in formal settings (hospitals and clinics) do not work for the language used in informal communities (Twitter and online discussion groups). As such, future research is needed to develop lexicons to help generate patient vocabularies, match them with standard terminologies, and annotate social medial and web communities with medical terms.

Most NLP applications have been used for clinical decision support. Upon further examination of these studies, it is found that NLP was employed for passive clinical decision support that requires input by the user to generate output. Future research can investigate how to move NLP applications from passive to active decision support. Active NLP decision support can push patient-specific information to users. The findings from many studies in the reviewed papers can be capitalized to build active NLP systems. For example, the automatic identification of adverse drug reactions (Li et al., 2013) can become the input of a clinical decision support system that alerts prescribers in real-time to medication problems. Likewise, information on suicidal ideation (Cook et al., 2016) generated from textual data can inform timely clinical interventions to prevent suicide. The use case for patient care administration can go beyond what the existing literature has found. For example, NLP's ability to review medical records more quickly and thoroughly than traditional, manual programs can dramatically increase efficiency and accuracy in patient care documentation. Future NLP research on patient care documentation is needed to help realize the full potential of NLP in transforming patient care operations.

The study has shown the dominance of syntactic representation of text as well as the lack of semantic-based NLP in clinical care. The syntactic analysis, though powerful, can process only information that it can "see" in the text. Such limitation can be overcome in the semantic analysis, which can identify a cascade of concepts related to the words in the text. Therefore, the semantic analysis can support more complex NLP tasks such as word-sense disambiguity, textual entailment, and semantic role labeling (Cambria and White, 2014), all of which are very relevant in understanding the patient data. The semantic analysis warrants more future research to harvest more valuable information from unstructured clinical texts.

Readily available extraction tools, many of which are open-sourced, are widely used in NLP applications in patient care. These general-purpose tools cover a broad set of needs and produce outcomes that are easy to interpret. On the other hand, the phrase structure used by these tools (for example, "colon and lung cancer" may be identified as "colon" and "lung cancer" rather than "colon cancer" and "lung cancer") inherently reduces the possibility to extract complex medical information from text. The in-house developed tools are successful in addressing specific needs, but they have a relatively narrow focus tuned highly to the data at particular institutions, and therefore, low in transferability (Kreimeyer et al., 2017). How to provide versatile NLP flatforms with available pipelines for many specific clinical tasks, simple or complex, is the methodology challenge to be addressed to move NLP systems toward full-scale acceptance and routine use in patient care settings.

Statistical NLP has become more popular than rule-based NLP. A variety of machine learning algorithms were adopted in the reviewed studies. However, only one paper (Zheng et al., 2019) made use of the method of deep learning, the most exciting recent development in machine learning. Given the promise of deep learning for text processing (LeCun et al., 2015), how to apply deep learning to solve clinical care problems is another avenue for future research.

Data

This review has revealed that the availability of public datasets remains limited. The need remains for data availability and accessibility that would increase participation in NLP for patient care. Furthermore, the review has revealed that most existing research has worked on small numbers of texts. As previously mentioned, statistical NLP nowadays relies heavily on machine learning algorithms, which require enormous amounts of data to make valid statistical inferences. It is clear that future research needs to work on more sizable data for more robust models and more powerful outcomes. Finally, all papers in this review analyzed textual data. That is, voice data are not studied at all. The recent breakthroughs in speech recognition make voice NLP valuable to patient care. For example, patient care providers can implement chatbots to automate some of their communications with their patients to improve their operational efficiency. There is also a promise in the use of speech NLP for information entry at the bedside. In short, future research needs to consider the impact of voice data on NLP applications in patient care.

Publication Year and Country

The review has shown a surge of research publications in the last decade, which is aligned with the rapid developments of artificial intelligence technologies in recent years. As AI technologies have been witnessing tremendous and rapid advancements in recent years, NLP applications will become more sophisticated and make greater inroads into patient care. More research will continue to investigate the profound impacts of NLP applications on patient care. It would be interesting to see more research on patient care NLP applications from research groups outside the USA and to develop more NLP tools for languages other than English.

Limitations

The rate of publication in the field of patient care NLP applications is fast increasing, and literature-based reviews like this one may have difficulty in keeping up with new developments. Nevertheless, this review provides a general scope of NLP applications in patient care. The study has limitations as well. First, as only journal papers with full-text available were analyzed, it is possible that the information from the journal papers which were not accessible at the time of the review was not gathered. In addition, this review excluded conference papers. As a result, this review may miss the latest development in the dynamic and fast-paced research field. Though the study is representative of patient NLP applications in the last 15 years, its findings may not capture every aspect of the area. A further review may want to include more databases and more types of papers to get a more comprehensive understanding of the applications of NLP in patient care.

CONCLUSION

As healthcare goes digital, NLP has become indispensable in medicine. Although NLP has been dominantly applied in clinical and translation research, usually through phenotyping, recent developments in NLP methods and techniques provide opportunities to represent patient care knowledge and drive clinical decision support and operation. The direct use of NLP is making inroads into clinical decision support and patient care. This paper presents a systematic analysis of research on NLP applications in patient care in the last 15 years. It has provided encouraging results that NLP is improving the delivery and management of patient care when compared with traditional methods. This paper helps information systems (IS) researchers understand the status quo of NLP in the domain of patient care. Equally important, it highlights important areas of NLP-powered research to move the field forward. NLP applications in patient care are promising, and much more systematic research is needed for the practical or commercial implementation of NLP in patient care.

REFERENCES

Afshar, M., Joyce, C., Dligach, D., Sharma, B., Kania, R., Xie, M., Swope, K., Salisbury-Afshar, E., & Karnik, N. S. (2019). Subtypes in patients with opioid misuse: A prognostic enrichment strategy using Electronic Health Record data in hospitalized patients. *PLoS ONE*, *14*(7), e0219717.

Bai, T., Chanda, A. K., Egleston, B. L., & Vucetic, S. (2018). EHR phenotyping via jointly embedding medical concepts and words into a unified vector space. *BMC Medical Informatics and Decision*, *18*(S 4), article no.123.

Balyan, R., Crossley, S. A., Brown, W., Karter, A. J., McNamara, D. S., Liu, J. Y., Lyles, C. R., & Schillinger, D. (2019). Using natural language processing and machine learning to classify health literacy from secure messages: The ECLIPSE study. *PLOS ONE*, *14*(2), e0212488.

Beeksma, M., Verberne, S., van den Bosch, A., Das, E., Hendrickx, I., & Groenewoud, S. (2019). Predicting life expectancy with a long short-term memory recurrent neural network using Electronic Medical Records. *BMC Medical Informatics and Decision Making*, *19*(1), article no. 36.

Bellows, B. K., DuVall, S. L., Kamauu, A. W. C., Supina, D., Babcock, T., & LaFleur, J. (2015). Healthcare costs and resource utilization of patients with binge-eating disorder and eating disorder not otherwise specified in the Department of Veterans Affairs: Healthcare costs and utilization in binge eating. *International Journal of Eating Disorders, 48*(8), 1082-1091.

Cai, T., T Giannopoulos, A., Yu, S., Kelil, T., Ripley, B., Kumamaru, K., Rybicki, F., & Mitsouras, D. (2016). Natural language processing technologies in radiology research and clinical applications. *Radiographics*, *36*(1), 176-191.

Cambria, E., & White, B. (2014). Jumping NLP curves: A review of natural language processing research. *IEEE Computational Intelligence Magazine*, *9*(2), 48-57.

Carson, N. J., Mullin, B., Sanchez, M. J., Lu, F., Yang, K., Menezes, M., & Cook, B. L. (2019). Identification of suicidal behavior among psychiatrically hospitalized adolescents using natural language processing and machine learning of Electronic Health Records. *PLOS ONE*, *14*(2), e0211116.

Castro, V., Shen, Y., Yu, S., Finan, S., Pau, C. T., Gainer, V., Keefe, C. C., Savova, G., Murphy, S. N., Cai, T., & Welt, C. K. (2015). Identification of subjects with polycystic ovary syndrome using Electronic Health Records. *Reproductive Biology and Endocrinology, 13*(1), article no. 116.

Chase, H. S., Mitrani, L. R., Lu, G. G., & Fulgieri, D. J. (2017). Early recognition of multiple sclerosis using natural language processing of the Electronic Health Record. *BMC Medical Informatics and Decision Making, 17*(1), article no. 24.

Chen, X., Xie, H., Wang, F. L., Liu, Z., Xu, J., & Hao, T. (2018). A bibliometric analysis of natural language processing in medical research. *BMC Medical Informatics and Decision Making*, *18*(s1), article no. 14.

Cohen, K. B. (2014). Chapter 6: Biomedical natural language processing and text mining, in I. N. Sarkar (Ed) *Methods in Biomedical Informatics: A Pragmatic Approach* (pp.141-177). Academic Press.

Cohen, R., Elhadad, M., & Birk, O. (2013). Analysis of free online physician advice services. *PLoS ONE*, *8*(3), e59963.

Cook, B. L., Progovac, A. M., Chen, P., Mullin, B., Hou, S., & Baca-Garcia, E. (2016). Novel use of natural language processing (NLP) to predict suicidal ideation and psychiatric symptoms in a text-based mental health intervention in Madrid. *Computational and Mathematical Methods in Medicine*, 1-8.

Cook, M. J., Yao, L., & Wang, X. (2019). Facilitating accurate health provider directories using natural language processing. *BMC Medical Informatics and Decision Making*, *19*(S3), article no. 80.

De Silva, D., Ranasinghe, W., Bandaragoda, T., Adikari, A., Mills, N., Iddamalgoda, L., Alahakoon, D., Lawrentschuk, N., Persad, R., Osipov, E., Gray, R., & Bolton, D. (2018). machine learning to support social media empowered patients in cancer care and cancer treatment decisions. *PLOS ONE, 13*(10), e0205855.

Demner-Fushman, D., Chapman, W.W., & McDonald C. J. (2009). What can natural language processing do for clinical decision support? *Journal of Biomedical Informatics*, *42*(5), 760-772.

Dreisbach, C, Koleck, TA, Bourne, P.E., & Bakken, S. (2019). A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data. *International Journal of Medical Informatics*, *125*, 37-46.

Ferrández, O., South, B. R., Shen, S., Friedlin, F. J., Samore, M. H., & Meystre, S. M. (2012). Evaluating current automatic de-identification methods with veteran's health administration clinical documents. *BMC Medical Research Methodology, 12*(1), article no. 109.

Gu, J, Sun, F, Qian, L, & Zhou, G. (2019). Chemical-induced disease relation extraction via attention-based distant supervision. *BMC Bioinformatics*, *20*, article no. 403.

Hakenberg, J, Bickel, S, Plake, C, Brefeld, U., Zahn, H., Faulstich, L., Leser, U., & Scheffer, T. (2005). Systematic feature evaluation for gene name recognition. *BMC Bioinformatics*, *6*, article no. S9.

Hong, S. N., Son, H. J., Choi, S. K., Chang, D. K., Kim, Y.-H., Jung, S.-H., & Rhee, P.-L. (2017). A Prediction model for advanced colorectal neoplasia in an asymptomatic screening population. *PLOS ONE, 12*(8), e0181040.

Hope, C. J., Garvin, J. H., & Sauer, B. C. (2012). Information extraction from narrative data. *American Journal of Health-System Pharmacy*, *69*(6), 455-461.

Iqbal, E., Mallah, R., Jackson, R. G., Ball, M., Ibrahim, Z. M., Broadbent, M., Dzahini, O., Stewart, R., Johnston, C., & Dobson, R. J. B. (2015). identification of adverse drug events from free-text electronic patient records and information in a large mental health case register. *PLOS ONE, 10*(8), e0134208.

Jackson, R., Kartoglu, I., Stringer, C., Gorrell, G., Roberts, A., Song, X., Wu, H., Agrawal, A., Lui, K., Groza, T., Lewsley, D., Northwood, D., Folarin, A., Stewart, R., & Dobson, R. (2018). CogStack - Experiences of deploying integrated information retrieval and extraction services in a large national health service foundation trust hospital. *BMC Medical Informatics and Decision Making, 18*(1), article no. 47.

Jones, M. M., DuVall, S. L., Spuhl, J., Samore, M. H., Nielson, C., & Rubin, M. (2012). Identification of methicillin-resistant staphylococcus aureus within the nation's Veterans Affairs medical centers using natural language processing. *BMC Medical Informatics and Decision Making, 12*(1), article no. 34.

Jones, M. M., Winthrop, K. L., Nelson, S. D., Duvall, S. L., Patterson, O. V., Nechodom, K. E., Findley, K. E., Radonovich, L. J., Samore, M. H., & Fennelly, K. P. (2018). Epidemiology of nontuberculous mycobacterial infections in the U.S. Veterans Health Administration. *PLOS ONE, 13*(6), e0197976.

Joshi, A. (1991). Natural Language Processing. *Science*, 253(5025), 1242-1249.

Kadra, G., Stewart, R., Shetty, H., Jackson, R. G., Greenwood, M. A., Roberts, A., Chang, C.-K., MacCabe, J. H., & Hayes, R. D. (2015). Extracting antipsychotic polypharmacy data from Electronic Health Records: Developing and evaluating a novel process. *BMC Psychiatry, 15*(1), article no. 166.

Klann, J. G., Phillips, L. C., Turchin, A., Weiler, S., Mandl, K. D., & Murphy, S. N. (2015). A numerical similarity approach for using retired current procedural terminology (CPT) codes for electronic phenotyping in the scalable collaborative infrastructure for a learning health system (SCILHS). *BMC Medical Informatics and Decision Making, 15*(1), article no. 104.

Kohane, I. S. (2011). Using Electronic Health Records to Drive Discovery in Disease Genomics. *Nature Reviews Genetics*, *12*(6), 417-428.

Kreimeyer, K., Foster, M., Pandey, A., Arya, N., Halford, G., Jones, S. F., Forshee, R., Walderhaug, M., & Botsis, T. (2017). Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review. *Journal of Biomedical Informatics*, *73*, 14-29.

LeCun, Y., Bengio, Y., and Hinton, G. (2015, May). *Nature*, 521, 436-444.

Lerchenmueller, M. J., & Sorenson, O. (2016). Author disambiguation in PubMed: Evidence on the precision and recall of authority among NIH-funded scientists. *PLOS ONE*, *11*(7), e0158731.

Li, Q., Deleger, L., Lingren, T., Zhai, H., Kaiser, M., Stoutenborough, L., Jegga, A. G., Cohen, K. B., & Solti, I. (2013). Mining FDA drug labels for medical conditions. *BMC Medical Informatics and Decision Making, 13*(1), article no. 53.

Li, Q., Spooner, S. A., Kaiser, M., Lingren, N., Robbins, J., Lingren, T., Tang, H., Solti, I., & Ni, Y. (2015). An end-to-end hybrid algorithm for automated medication discrepancy detection. *BMC Medical Informatics and Decision Making, 15*(1), article no. 37.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *Annals of Internal Medicine*, 339: b2700.

Long, H., Zhu Y., Jia L., Gao B., Liu J., Liu L., & Herre H. (2019). An ontological framework for the formalization, organization, and usage of TCM-knowledge. *BMC Medical Informatics and Decision Making*, *19*(S2), article no. 53.

McCoy, T. H., Castro, V. M., Cagan, A., Roberson, A. M., Kohane, I. S., & Perlis, R. H. (2015). Sentiment measured in hospital discharge notes is associated with readmission and mortality risk: An Electronic Health Record study. *PLOS ONE, 10*(8), e0136341.

Mishra, N. K., Son, R. Y., & Arnzen, J. J. (2012). Towards automatic diabetes case detection and ABCS protocol compliance assessment. *Clinical Medicine & Research*, *10*(3), 106-121.

Navathe, A. S., Zong, F., Lie, V. J., Chang, F. Y., Sordo, M., Topaz, M., Navathe, S. B., & Rocha, R. A. (2018). Hospital readmission and social risk factors identified from physician notes. *Health Services Research*, *53*(2), 1110-1136.

Névéol A, & Zweigenbaum P. (2016). Clinical natural language processing in 2015: Leveraging the variety of texts of clinical interest. *Yearbook of Medical Informatics*, *10*(1), 234-239.

Nikfarjam, A., Sarker, A., O'Connor, K., Ginn, R., & Gonzalez, G. (2015). Pharmacovigilance from social media: Mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *Journal of American Medical Informatics Association*, *22*(3), 671-81.

Pakhomov, S. V., Jacobsen, S. J., Chute, C. G., & Roger, V. L. (2008). Agreement between patient-reported symptoms and their documentation in the medical record. *The American Journal of Managed Care, 14*(8), 530-539.

Pakhomov, S., Shah, N., Hanson, P., Balasubramaniam, S., & Smith, S. (2010). Automated processing of Electronic Medical Records is a reliable method of determining aspirin use in populations at risk for cardiovascular events. *Journal of Innovation in Health Informatics, 18*(2), 125-133.

Parthipan, A., Banerjee, I., Humphreys, K., Asch, S. M., Curtin, C., Carroll, I., & Hernandez-Boussard, T. (2019). Predicting inadequate postoperative pain management in depressed patients: A machine learning approach. *PLOS ONE, 14*(2), e0210575.

Peissig, P., Sirohi, E., Berg, R.L., Brown-Switzer, C., Ghebranious, N., McCarty, C. A., & Wilke, R.A. (2007). Construction of atorvastatin dose-response relationships using data from a large population-based DNA Biobank. *Basic & Clinical Pharmacology & Toxicology*, *100*, 286-288.

Pesko, M. F., Gerber, L. M., Peng, T. R., and Press, M. J. (2018). Home health care: Nurse – physician communication, patient severity, and hospital readmission. *Health Services Research*, *53*(2):1008-1024.

Poulin, C., Shiner, B., Thompson, P., Vepstas, L., Young-Xu, Y., Goertzel, B., Watts, B., Flashman, L., & McAllister, T. (2014). Predicting the risk of suicide by analyzing the text of clinical notes. *PLoS ONE, 9*(1), e85733.

Reddy, E. M., & Bhaskar, P. (2018). Able machine learning method for classifying disease-treatment semantic relations from bio-medical sentences. *International Journal of Recent Research Aspects*, *5*(1), 223-226.

Regan, S., Meigs, J. B., Grinspoon, S. K., & Triant, V. A. (2016). Determinants of smoking and quitting in HIV-infected individuals. *PLOS ONE, 11*(4), e0153103.

Sheikhalishahi, S., Miotto, R., Dudley, J. T., Lavelli, A., Rinaldi, F., & Osmani, V. (2019). Natural language processing of clinical notes on chronic diseases: Systematic review. *JMIR Medical Informatics*, *7*(2), e12239.

Shibata, D., Ito, K., Nagai, H., Okahisa, T., Kinoshita, A., & Aramaki, E. (2018). Idea density in Japanese for the early detection of dementia based on narrative speech. *PLOS ONE, 13*(12), e0208418.

Shiner, B., D'Avolio, L. W., Nguyen, T. M., Zayed, M. H., Watts, B. V., & Fiore, L. (2012). Automated classification of psychotherapy note text: Implications for quality assessment in PTSD care: NLP of PTSD psychotherapy notes. *Journal of Evaluation in Clinical Practice, 18*(3), 698-701.

Sidorov, G., Velasquez, F., Stamatatos, E., Gelbukh, A., & Chanona-Hernández, L. (2014). Syntactic n-grams as machine learning features for natural language processing. *Expert Systems with Applications*, *41*(3), 853-860.

Simon, T., Goldberg, A., Aharonson-Daniel, L., Leykin, D., & Adini, B. (2014). Twitter in the crossfire - The use of social media in the Westgate Mall terror attack in Kenya. *PLoS ONE*, *9*(8), e104136.

Sung, R.-J., Chiu, C., Chiu, N.-H., & Hsiao, C.-H. (2014). Online detection of concerned HIV-Related messages in web forums. *AIDS Care*, *26*(3), 337-342.

Tao, C., Gong, Y., Xu, H., & Zhao, Z. (2017). Introduction: The International Conference on Intelligent Biology and Medicine (ICIBM) 2016: Special focus on medical informatics and big data. *BMC Medical Informatics and Decision Making*, *17*(S2), article no. 77.

Taylor, C. L., van Ravesteyn, L. M., van denBerg, M. P. L., Stewart, R. J., & Howard, L. M. (2016). The prevalence and correlates of self-harm in pregnant women with psychotic disorder and bipolar disorder. *Archives of Women's Mental Health, 19*(5), 909-915.

Toyabe, S. (2012). Detecting inpatient falls by using natural language processing of Electronic Medical Records. *BMC Health Services Research, 12*(1), article no. 448.

Wang, L., Wang, Y., Shen, F., Rastegar-Mojarad, M., & Liu, H. (2019). Discovering associations between problem list and practice setting. *BMC Medical Informatics and Decision Making*, *19*(S3), article no. 69.

Wang, Y., Mehrabi, S., Sohn, S., Atkinson, E. J., Amin, S., & Liu, H. (2019). Natural language processing of radiology reports for identification of skeletal site-specific fractures. *BMC Medical Informatics and Decision Making*, *19*(S3), article no. 73.

Waudby-Smith, I. E. R., Tran, N., Dubin, J. A., & Lee, J. (2018). Sentiment in nursing notes as an indicator of out-of-hospital mortality in intensive care patients. *PLOS ONE, 13*(6), e0198687.

Wieland, M. L., Wu, S. T., Kaggal, V. C., & Yawn, B. P. (2013). Tracking health disparities through natural-language processing. *American Journal of Public Health, 103*(3), 448-449.

Workman, T. E., & Stoddart, J. M. (2012). Rethinking information delivery: Using a natural language processing application for point-of-care data discovery. *Journal of the Medical Library Association: JMLA, 100*(2), 113-120.

Xia, Z., Secor, E., Chibnik, L. B., Bove, R. M., Cheng, S., Chitnis, T., Cagan, A., Gainer, V. S., Chen, P. J., Liao, K. P., Shaw, S. Y., Ananthakrishnan, A. N., Szolovits, P., Weiner, H. L., Karlson, E. W., Murphy, S. N., Savova, G. K., Cai, T., Churchill, S. E., Plenge, R. M., Kohane, I. S., & De Jager, P. L. (2013). Modeling disease severity in multiple sclerosis using Electronic Health Records. *PLoS ONE, 8*(11), e78927.

Xiao, B, Imel, Z. E., Georgiou, P. G., Atkins, D. C., Narayanan, S. S. (2015). Rate my therapist: Automated detection of empathy in drug and alcohol counseling via speech and language processing. *PLoS One*, *10*(12), e0143055.

Ye, Y., Wagner, M. M., Cooper, G. F., Ferraro, J. P., Su, H., Gesteland, P. H., Haug, P. J., Millett, N. E., Aronis, J. M., Nowalk, A. J., Ruiz, V. M., López Pineda, A., Shi, L., Van Bree, R., Ginter, T., & Tsui, F. (2017). A study of the transferability of influenza case detection systems between two large healthcare systems. *PLOS ONE, 12*(4), e0174970.

Yim, W, Yetisgen M, Harris W.P., & Kwan S.W. (2016). Natural language processing in oncology: A review. JAMA Oncology, *2*, 1-8.

Zhang, X., Kim, J., Patzer, R. E., Pitts, S. R., Chokshi, F. H., & Schrager, J. D. (2019). Advanced diagnostic imaging utilization during emergency department visits in the United States: A predictive modeling study for emergency department triage. *PLOS ONE*, *14*(4), e0214905.

Zhang, Y. Y., Tang, B. Z., Jiang, M., Wang, J. Q., & Xu, H. (2015). Domain adaptation for semantic role labeling of clinical text. *Journal of American Medical Informatics Association*, *22*(5), 967-79.

Zheng, T., Gao, Y., Wang, F., Fan, C., Fu, X., Li, M., Zhang, Y., Zhang, S., & Ma, H. (2019). Detection of medical text semantic similarity based on convolutional neural network. *BMC Medical Informatics and Decision Making*, *19*(1), article no. 156.

Zhong, Q.-Y., Karlson, E. W., Gelaye, B., Finan, S., Avillach, P., Smoller, J. W., Cai, T., & Williams, M. A. (2018). Screening pregnant women for suicidal behavior in Electronic Medical Records: Diagnostic codes vs. clinical notes processed by natural language processing. *BMC Medical Informatics and Decision Making, 18*(1), article no. 30.

Appendix: Summary of coding

	Application in Patient Care					Data			Publication	
Paper	Clinical Utility	Lexicon	Extraction	Extraction Tool	Classification /Algorithm	Source	Type	No. of Texts	Year	Country
Afshar et al. (2019)	detect opioid misuse	not reported	syntactic analysis	cTAKES	statistical - latent Dirichlet allocation	an urban tertiary academic center	clinical note	422,147	2019	USA
Balyan et al. (2019)	foster shared meaning	not reported	syntactic analysis	TAALES, TAACO, etc.	statistical - support vector machines, etc.	KPNC Diabetes Registry	secure message	283,216	2019	USA
Beeksma et al. (2019)	predict patient mortality	not reported	pattern matching	in-house developed	statistical - Word2Vec	FaMe-net repository	Consultation document	149,314	2019	Nether-land
Bellows et al. (2015)	quantify costs & utilization for binge-eating disorder patients	not reported	pattern matching	not reported	knowledge-based	A VA clinic	clinical note	1,487	2015	USA
Carson et al. (2019)	predict suicidal behavior	not reported	syntactic analysis	Invenio	statistical - random forest	a community health system	clinical note	9,415	2019	USA
Castro et al. (2015)	diagnose polycystic ovary syndrome	custom made	syntactic analysis	cTAKES	statistical - regression	Partners Healthcare Research Patients Data Registry	clinic note	640	2015	USA
Chase et al. (2017)	diagnose multiple sclerosis	UMLS	syntactic analysis	MedLEE	statistical - Bayesian	Columbia University Medical Center	clinical note	2,999	2017	USA
Cook et al. (2016)	predict suicidal behavior	not reported	syntactic analysis	in-house developed	statistical - LIBLINEAR	a hospital system in Madrid, Spain	text message	1,453	2016	Spain
Cook et al. (2019)	provide provider information to patients	not reported	pattern matching	in-house developed	knowledge-based	Connecticut eLicensing website, the National Plan Provider Enumeration System	state licensure list	7,408	2019	USA
De Silva et al. (2018)[0]	understand patient behavior and emotion	both	hybrid	Python	statistical - Word2Vec	online support group	online discussion post	4,795,428	2018	Australia
Hong et al. (2017)	predict advanced colorectal neoplasia	custom made	pattern matching	CETAS	statistical - regression	Samsung Medical Center	Colonoscopy, pathology report	49,450	2017	Korea
Iqbal et al. (2015)	detect adverse drug reactions	not reported	pattern matching	in-house developed	statistical - Java Annotation Patterns Engine	South London and Maudsley (SLaM)	clinical record	15,908	2015	UK
Jones et al. (2012)	diagnose staphylococcus aureus	both	pattern matching	in-house developed	knowledge-based	VA network of databases	microbiology record	68,427	2012	USA
Jones et al. (2018)	diagnose Nontuberculous mycobacterial disease	SNOMED CT	pattern matching	not reported	knowledge-based	VA Corporate Data Warehouse	laboratory data	6,031	2018	USA
Kadra et al. (2015)	detect antipsychotic polypharmacy	not reported	syntactic analysis	in-house developed	knowledge-based	South London and Maudsley (SLaM)	diagnostic data	7,201	2015	UK
Li et al. (2013)	detect adverse drug reactions	not reported	syntactic analysis	cTAKES	statistical - conditional random field	Cincinnati Children's Hospital	drug label	38,071	2013	USA
Li et al. (2015)	detect medication discrepancy	UMLS, SNOMED CT, RxNorm	syntactic analysis	both	knowledge-based	Cincinnati Children's Hospital	clinical note, discharge prescription list	975	2015	USA
McCoy et al. (2015)	predict patient mortality	not reported	syntactic analysis	Python	statistical - regression	a large New England health system	discharge note	23,343	2015	USA
Mishra et al. (2012)	diagnose diabetes	not reported	pattern matching	ConText	knowledge-based	i2b2 data set	discharge summary	889	2012	USA
Navathe et al. (2018)	identify readmission risk	custom made	syntactic analysis	MTERMS	statistical - regression	Partners Healthcare System	physician note	93,606	2018	USA
Pakhomov et al. (2008)	detect the discordance between self-report and documented symptoms	not reported	pattern matching	in-house developed	knowledge-based	Mayo Clinic	clinical note, patient form	1,119	2008	USA
Pakhomov et al. (2010)	detect aspirin use	not reported	pattern matching	Perl	knowledge-based	Mayo Clinic	clinical note	499	2010	USA

Parthipan et al. (2019)	diagnose depression	not reported	syntactic analysis	NLTK	knowledge-based	Stanford University Hospitals and Clinical	clinical note	41,713	2019	USA
Pesko et al. (2018)	identify communication failures	custom made	syntactic analysis	not reported	knowledge-based	Medicare Provider Analysis and Review file, VNSNY electronic medical records	nurse note	2,680	2018	USA
Poulin et al. (2014)	predict suicidal behavior	not reported	syntactic analysis	not reported	statistical - genetic programming	VA medical records	clinical note	210	2014	USA
Regan et al. (2016)	identify smoking status	not reported	syntactic analysis	GATE	knowledge-based	Partners Healthcare System	clinic note	3,487	2016	USA
Shibata et al. (2018)	diagnose dementia	not reported	morpho-logical	Juman++, MeCab	knowledge-based	Not reported	patient narrative	42	2018	Japan
Shiner et al. (2012)	assess quality in PTSD care	not reported	syntactic analysis	in-house developed	statistical - conditional random field	A VA clinic	mental health note	221	2012	USA
Sung et al. (2014)	facilitate intervention for HIV	not reported	syntactic analysis	CKIP	statistical - genetic algorithm	Yahoo! Chinese-based portal	online post	2,083	2014	China
Taylor et al. (2016)	predict suicidal behavior	not reported	syntactic analysis	GATE	not reported	South London and Maudsley (SLaM)	clinical record	420	2016	UK
Toyabe (2012)	predict injuries after falls	not reported	syntactic analysis	Text Mining Studio	knowledge-based	Niigata University Hospital, Japan	incident report	2,590	2012	Japan
Wang, Mehrabi et al. (2019)	identify skeletal site-specific fractures	not reported	syntactic analysis	Medline	statistical - latent Dirichlet allocation	Mayo clinic	radiology note	2,356	2019	USA
Wang, Wang et al. (2019)	provide more information to clinicians	not reported	pattern matching	Med-Tagger	knowledge-based	Mayo clinic	clinical note	64,250	2019	USA
Waudby-Smith et al. (2018)	predict patient mortality	not reported	syntactic analysis	Python	statistical - regression	MIMIC-III	nurse note	27,477	2018	Canada
Wieland et al. (2013)	tracked healthcare disparities	not reported	pattern matching	not reported	knowledge-based	Mayo Clinic	patient chart	5,782	2013	USA
Workman & Stoddart (2012)	identified information for disease prevention	not reported	syntactic analysis	Medline	knowledge-based	PubMed	PubMed citation	3,276	2012	USA
Xia et al. (2018)	diagnose multiple sclerosis	hybrid	syntactic analysis	cTAKES	statistical - regression	Partners Healthcare System	clinical note	595	2013	USA
Ye et al. (2017)	diagnose influenza	not reported	pattern matching	in-house developed	statistical - Bayesian	University of Pittsburgh Medical Center, Intermountain Healthcare in Utah	clinical note	88,693	2017	USA
Zhang et al. (2019)	predict the utilization of advanced diagnostic imaging	not reported	semantic analysis	R	statistical - regression	US National Hospital Ambulatory Medical Care Survey data	triage note	139,150	2019	USA
Zheng et al. (2019)	assess imaging diagnosis quality	not reported	semantic analysis	Jieba	statistical - convolutional neural network	Shanghai Tongren Hospital, China	imaging and pathology report	16,354	2019	China
Zhong et al. (2019)	predict suicidal behavior	custom made	syntactic analysis	cTAKES	knowledge-based	Partners Healthcare System	clinical note	1,120	2018	USA

QRBD

Quarterly Review of Business Disciplines

November 2020

Volume 7
Number 3

A Journal of International Academy of Business Disciplines
Sponsored by University of North Florida
ISSN 2334-0169 (print)
ISSN 2329-5163 (online)