ISSN 2816-8089

REVIEW ARTICLE

Natural Language Processing – Finding the Missing Link for Oncologic Data in 2022

Andra Krauze[1] and Kevin Camphausen[1]

[1]*Center for Cancer Research, National Cancer Institute, NIH, Building 10, Bethesda, MD, 20892, United States.*

Abstract

Like most medical specialties, oncology is undergoing a data revolution at the center of which lie vast and growing amounts of clinical data in unstructured, semi-structured and structed formats. Artificial intelligence approaches are widely employed in research endeavors to harness electronic medical records data to advance patient outcomes. As of the end of 2021, the use of clinical oncologic data, although collected on a large scale, particularly with the increased implementation of electronic medical records, remains limited due to missing, incorrect, or manually entered data in registries and the lack of resource allocation to data curation in real-world settings. Natural Language Processing (NLP) may provide an avenue to extract data from electronic medical records and, as a result has grown considerably in medicine to be employed for documentation, outcome analysis, phenotyping, and clinical trial eligibility. Barriers to NLP persist with the inability to aggregate findings across studies due to the use of different methods and significant heterogeneity at all levels with essential parameters such as patient comorbidities and performance status lacking implementation in AI approaches. This review aims to provide an updated overview of natural language processing (NLP) and the current state of its application in oncology for clinicians and researchers that wish to implement NLP to augment registries and/or advance research projects using NLP in 2022 and beyond.

Key Words: *Natural Language Processing (NLP); Artificial Intelligence (AI); Oncology; Electronic Medical Records (EMR); Clinical research*

***Corresponding Author**: *Andra Krauze, Department of radiation Oncologist, Center for Cancer Research, National Cancer Institute, Building 10 - Hatfield CRC, Room B2-3561, Bethesda, MD 20892-1682, USA; Email: andra.krauze@nih.gov*

Received Date: *December 26, 2021*, **Accepted Date**: *January 21, 2022*, **Published Date**: *February 25, 2022*

Citation: *Andra Valentina Krauze and Kevin Camphausen. Natural language processing – finding the missing link for oncologic data in 2022. Int J Bioinfor Intell Comput. 2021;1(1):24-47.*

This open-access article is distributed under the terms of the Creative Commons Attribution Non-Commercial License (CC BY-NC) (http://creativecommons.org/licenses/by-nc/4.0/), which permits reuse, distribution and reproduction of the article, provided that the original work is properly cited, and the reuse is restricted to non-commercial purposes.

Abbreviations

AI: Artificial Intelligence; ML: Machine Learning; DL: Deep Learning; NLP: Natural Language Processing; CAD: Coronary Artery Disease; NYD: Not Yet Diagnosed; NLI: Natural Language Inference; HPV: Human Papilloma Virus; SDoH : Social Determinants of Health; OS: Overall Survival; HER: Electronic Health Records; EMR: Electronic Medical Records; ECOG: Eastern Cooperative Oncology Group; KPS: Karnofsky Performance Status; ICD: International Classification of Diseases coding; NLP-BIRNN: Natural language processing-bidirectional recurrent neural network ; CNN: Convolutional Neural Network; RNN: Recurrent Neural Network ; eCRFs: electronic Case Report Forms; CT: Computer Tomography; LSTM: Long Short-term Memory (recurrent neural network); BERT: Bidirectional Encoder Representations from Transformers; PRO: Patient Reported Outcomes; CTCAE: Common Terminology Criteria for Adverse Events; VAERS: Vaccine Adverse Event Reporting System; PRAPARE: Protocol for Responding to and Assessing Patients' Assets, Risks, and Experiences; SCC: Squamous Cell Carcinoma; SEER: Surveillance, Epidemiology, and End Results; TCGA: The Cancer Genome Atlas; VA: Veterans Affairs Data Warehouse/Registry

1. Introduction

Oncology, like most medical specialties, is amid a data revolution with increasing interest and research that employs artificial intelligence (AI) to address clinical questions [1-6]. Under the AI umbrella, machine learning (ML), deep learning (DL), and hybrids of the two are the subject of a significant rise in publications [7-12]. However, AI approaches, in particular DL, are data-hungry, and it has become increasingly clear that there are significant limitations to both mining existing large-scale data sets as well as robustly acquiring prospective data that can translate into reproducible and generalizable AI mediated analyses and conclusions [13] [14-17]. The use of clinical oncologic data, although collected on a large scale, remains limited due to missing, incorrect or manually entered data and the lack of resource allocation to data curation in real-world settings [18]. As Sanyal et al. astutely put it "free-text clinic notes may offer the greatest nuance and detail about a patient's clinical status, they are largely excluded in previous predictive models due to the increase in processing complexity and need for a complex modeling framework" [18]. The problem may be approached by addressing the individual aspects of unstructured [1,19-24] clinical notes (e.g. history and physical exam) [25,26] , operative reports [3,27], pathology reports [28-32] and imaging reports [33-39] to analyze outcomes: response [21,40-42], toxicity [43-45]and survival [33,46]. Natural Language Processing (NLP) has grown considerably in medicine to be employed for documentation [47,48], outcome prediction [1,27,43,49], phenotyping [50,51], data extraction [17,28,49,52-56] and analysis, clinical trial eligibility [41,57,58], exploration of literature [51,59,60], evaluating impact on workload and recruitment with active parallel growth in the secondary use of expanding clinical data sets [1,58,61]. NLP methodology continues to evolve, with roughly 25% of NLP methods being rule based and up 50% machine learning based [34] with variation on the method based on the data at hand [17]. Significant barriers to wide-ranging implementation remain with an inability to aggregate findings across studies due to different NLP methods, evaluation and reporting, lack of diverse patient samples, heterogeneity across diseases, datasets, data collection methods and applications [16]. Significant parameters known to carry prognostic importance in oncology outcomes, such as performance status, comorbidities, and social history, are still not widely deployed in oncology AI efforts as data sets remain insufficiently robust to allow for their widespread inclusion and the promise of solutions such as NLP has yet to deliver. The goal of this review is to provide an updated overview of natural language processing (NLP) and the current state of its application to oncology documentation for clinicians and researchers that wish to implement NLP to augment registries and/or advance research projects. Attention was lent to the integration of NLP into the clinic to address the missing link for oncologic data showcasing areas where significant advances have been made, and a drive towards transfer of expertise is promising in 2022.

2. Clinical History

The inclusion of clinical information embedded in both the history and the physical exam into AI approaches remains underserved, neglected, or incomplete due to heterogenous and inconsistent capture resulting in a lack of clinician confidence in the inclusion of such data in analyses that explore primary and secondary endpoints and an ongoing lack of large-scale robust data for the computational approaches [19]. Essential areas for data capture to enable AI biomarker identification are wide-ranging but most significant is the capture of patient performance status [66], comorbidities [62],and social history [17,46,67,68], all of which may significantly alter management, treatment response and downstream progression and survival outcomes. Large-scale real-world data is lacking for outcome measures such as treatment response [43], symptom management [63,69,70] and toxicity [43,44] as well as patient-reported outcomes [71], advanced care planning [72,73], and goals of care [47,74,75]. These are significant areas of need in oncology, where NLP approaches are actively evolving. Natural language processing (NLP) represents an intuitive avenue to collect, extract, transform and load unstructured data while avoiding increasing clinician burden in data collection and annotation [19,76]. If effective, NLP can elevate cancer registry data, allow research queries, and identify patients who may benefit from novel treatments or enrollment on clinical trials [58,77]. Much has been written about NLP, and algorithmic advances in AI and NLP methods that have boosted their performance [13,23,35,63,64], and a recent PubMed search revealed 46 systematic reviews examining NLP in medicine or medicine affiliated specialties, while Wang et al. identified 2336 articles on NLP research 1999 to 2018 approximately 100 annually [64]. However, it has become increasingly clear that NLP is both domain-adaptive (e.g., medical specialty specific (emergency room, oncology etc.) and clinical note type specific (e.g., initial consult, follow-up, on treatment note etc.) and task-adaptive (e.g., symptom duration, performance status etc.). This emphasizes the need to build on specific domains and tasks to realize the full potential of NLP in medicine [65,78] (Table 1). Per Davis et al. "History taking is more than information gathering, it affords the opportunity to decipher the patient's body language as the inquiry proceeds. At this stage, no symptom or circumstance should be disregarded. With an understanding of biology and medicine coupled with past experience, the physician tries to connect the salient parts of the patient's story to develop a plausible explanation of the physiologic or pathologic events that lead to illness." [79]. As most health care providers are taught and intuitively understand, history taking is a truly human experience that far exceeds simple acquisition of information. It comprises a complex process of verbal and nonverbal interaction combined with the interpretation of information gathered. This is followed by information capture by clinician, student, or other allied specialist in a handwritten but now increasingly electronic format via telephone dictation or voice recognition powered mechanisms [61]. As has been discussed in recent publications, the framework of history taking is itself evolving to acquire an increased amount of information inherent in various aspects of patient care from symptomatology to treatment related toxicity to outcomes to best inform diagnosis and management [80]. It should also be noted that virtually all aspects of history taking, and physical exam have been impacted by the covid -19 pandemic and increased use of telemedicine [81-84]. As such, there is also an increasing component of telemedicine-generated data and interest in optimizing the collection of information in this context as described by Bragin et al. for example, in neurology [83]. However, it is entirely clear that several factors have affected how data was captured even preceding the pandemic and the growth of telemedicine. Two major aspects have affected how information is captured and therefore how it can be extracted: the implementation of electronic medical records (EMR) and the implementation of voice recognition systems over traditional transcription services [21,47,48,72,74,75,85,86]. Most of the patient history data exists in an unstructured format (Figure 1) and existing research has shown that the use of both structured and unstructured data improves ML approaches, generating more clinically meaningful results [1]. A significant proportion of NLP dedicated literature originates from data that exists in smaller institutional or multi-institutional data sets as opposed to large public databases since these datasets are both richer in clinical information and allow for the capture of data parameters by manual means allowing for annotation (Figure 1). By comparison, large public data sets, are less rich in detailed clinical data but are more accessible to researchers for training and validation of NLP approaches (Figure 1). It should be noted that significant variability persists based on the clinical setting where the data originates (Figure 2). Figure 2 showcases the fictional scenario of a patient who presents with a cough and a neck mass and illustrates typical clinical notes from the first emergency room visit to the first oncology appointment. Each medical encounter captured information from a different point of view, hence lending value to particular aspects and lesser value to others. For example, the emergency room visit is focused on the immediate problem but the diagnosis (malignancy?) or even more specifically a tissue diagnosis is not available, nor are staging investigations, all of which are available to radiation oncologist who sees the patient 2 weeks later, reflecting very different levels of information available and captured. Analogous to this, affiliated specialties that the patient is referred to (anesthesia, medical oncology, nutrition, speech language pathology, dentistry, social work) also will value discipline specific factors that are then captured accordingly. In this fictional scenario, the same information may be captured in different manners in different domains and even within the same domain. For example, the discrepant capture of the smoking history ("20 pack yr hx" vs. "20 pack years") which arguably could be harmonized as one term to be recognized. Achieving commonality in the context of the past medical history is arguably much more comples, as illustrated in this example. This is exemplified using coronary artery disease (CAD) as a feature of the past medical history. It is captured as "not yet diagnosed" (NYD), as "CAD, low blood counts (NYD)", as "CAD" and as "coronary artery disease", "heart attack" and even "heart failure". In this case, the "heart failure" term is actually associated with the patients` brother not the patient himself which would be challenging for NLP to recognize. This example illustrates also how the alcohol intake history ("2-3 drinks per week" (ER note) vs. "denies alcohol use" in subsequent notes) is discrepantly captured posing a challenge for NLP (Figure 2). Explanations for such significant discrepancies may be multifactorial. In the case of the alcohol intake history, the patient may have discontinued alcohol use following the diagnosis or may have forgotten to describe his alcohol use as perhaps evidenced by the "trouble providing a clear history" statement in the note, itself a multifactorial problem or the patient may have not wish to disclose their alcohol use for fear of stigmatization or bias. On the other hand, the provider may have failed to elicit a thorough history or may have done so but forgotten to document it. They may have become aware of the patients` alcohol use at a later date and this may be captured in a later note e.g., after speaking with the patients` daughter and obtaining a collateral history. Therefore, it is comprehensible that the use of NLP for clinical notes is extremely challenging since each note in isolation would likely result in vastly different registry outputs and taking some or all notes into account would result in significant irreconcilable discrepancies (Figure 2). These difficulties are evident in the literature [16,35,61,87,88]. Yan et al. examined 9 studies in sepsis prediction and detection but could not carry out a meta-analysis due to incomparable measurements among the studies [1]. While factors such as age or gender are captured robustly, as are demographics in general and this is generally structured data (Figure 1), unstructured features (Figure 1) are poorly captured in registries, and these include past medical history, family history, social history which can lead to missing important implications as relating to a cancer diagnosis and decisions regarding management. Features such as transportation difficulties, finances, and family-related factors are even more poorly captured potentially leading to bias in AI results that are generated based on such registry data. Staging information pulled from clinical documents is also a challenge since its capture is variable and the addition of staging relevant information over time e.g., HPV status in the fictional scenario (Figure 2) is problematic to extract as is the development of metastatic disease [77] . Several authors have investigated converting and harnessing unstructured data to structured data [17,23,28,29,43,89,90]. Source documents such as pathology reports and some operative reports, which can also include staging information, are generally time consuming to capture manually and require consistent and considerable expertise to do so. A significant number of publications examining NLP as a solution to these data challenges have dealt with breast cancer where the creation of databases is ongoing [18,19,40,43,57,89,91-94] and NLP has been particularly prolific in the breast cancer field representing 23.3 % of all NLP assisted medical research (closely followed by lung cancer with 14.56%) [64]. Many solutions however are not scalable. Percha et al. looked at the feasibility of natural language inference (NLI) as a scalable solution for registry curation using 5 state-of-the-art, pretrained, deep learning based NLI models to clinical, laboratory, and pathology notes employed towards 43 different breast oncology registry fields and evaluated the models against a manually curated, 7439 patient breast oncology research database [19]. Considerable variation in performance was noted both within and across fields due to incorrect inferences through models' tendency to misinterpret historical findings, and confusion based on abbreviations and subtle term variants common in clinical text [19]. Ultimately the authors concluded that NLP methods require specially annotated training sets or constructing a separate model for each registry field. However, they found that a single pretrained NLI model could simultaneously curate dozens of different fields. While NLI methods remain largely unexplored in the clinical domain NLI could increase the efficiency of registry curation, even with no additional training [19]. Bar et al. used NLP to examine the relationship between the method of cancer detection and genomic and clinical risk, and its effect on adjuvant chemotherapy recommendations in breast cancer. They found an association between the method of cancer detection and both genomic and clinical risk noting that symptomatic breast cancer, especially in young women, remains a poor prognostic factor that should be taken into account when evaluating patient prognosis and determining adjuvant treatment plans [91]. However, broad approaches that are more likely to result in reproducible and generalizable results need to be generated. Such studies showcase that clinical history matters but raise questions as to how it is captured. To illustrate this point in the fictional case described in Figure 2, the duration of symptoms as captured in 3 different notes is considerably heterogenous and one can easily imagine how NLP could result in discrepant results (Figure 2). To advance NLP in medicine, in 2022 and beyond, significant efforts will continue to be directed at creating specialized lexicons. Jung et al. describe this most significantly time-consuming step of NLP, the creation of a specialized lexicon, introducing 4243 unique lexicon items for matching patients to clinical trials automatically based on eligibility matching and text mining analysis in the breast cancer domain, which they evaluated by comparing it with the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) [57]. In 2022, we will be looking for increased literature presence of specialized lexicons in various clinical oncologic settings.

Table 1: *Selected systematic reviews on natural language processing (NLP) in medicine published 2017 to present.*

Systematic review	Subject	# of studies included	Reference
Medical specialty			
Sepsis prediction, early detection, and identification using clinical text for machine learning: a systematic review	Sepsis, prediction, early detection	9	Yan et al. J Am Med Inform Assoc. 2021 Dec 13;ocab236. [1]
*Systematic review of current natural language processing methods and applications in cardiology	Cardiology	37	Turchioe et al. Heart. 2021 Oct 28.[2]
Natural Language Processing in Surgery: A Systematic Review and Meta-analysis	Surgery	29	Mellia et al. Ann Surg. 2021 May 1;273(5):900-908.[3]
Natural Language Processing Applications in the Clinical Neurosciences: A Machine Learning Augmented Systematic Review	Clinical Neuroscience	48	Buchlak et al. Acta Neurochir Suppl. 2022;134:277-289.[4]
Machine Learning and Natural Language Processing in Mental Health: Systematic Review	Mental Health	58	Le Glaz et al. J Med Internet Res. 2021 May 4;23(5):e15708.[5]
Application of Artificial Intelligence Methods to Pharmacy Data for Cancer Surveillance and Epidemiology Research: A Systematic Review	Pharmacy Data Oncology	36	Grothen et al. JCO Clin Cancer Inform. 2020 Nov;4:1051-1058. [6]
**The reporting quality of natural language processing studies: systematic review of studies of radiology reports.	Radiology reports (quality)	164	Davidson et al. BMC Med Imaging. 2021 Oct 2;21(1):142[16]
A systematic review of natural language processing applied to radiology reports	Radiology reports	164	Casey et al. BMC Med Inform Decis Mak. 2021 Jun 3;21(1):179. [34]
Deep Learning for Natural Language Processing in Radiology-Fundamentals and a Systematic Review	Radiology (2018-2019)	10	Sorin et al. J Am Coll Radiol. 2020 May;17(5):639-648.[38]
Electronic Health Record (EHR) aspects general and specific			
*Extracting social determinants of health from electronic health records using natural language processing: a systematic review	Social determinants of health from EHR	82	Patra et al. J Am Med Inform Assoc. 2021 Nov 25;28(12):2716-2727.[17]
Natural Language Processing of Clinical Notes on Chronic Diseases: Systematic Review	Chronic Diseases	106	Sheikhalishahi et al. JMIR Med Inform. 2019 Apr 27;7(2):e12239.[62]
*Natural language processing algorithms for mapping clinical text fragments onto ontology concepts: a systematic review and recommendations for future studies	Mapping clinical text to ontology concepts	77	Kersloot et al. J Biomed Semantics. 2020 Nov 16;11(1):14. [13]

Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review	NLP for unstructured clinical information	86	Kreimeyer et al. J Biomed Inform. 2017 Sep;73:14-29.[23]
Impacts of structuring the electronic health record: Results of a systematic literature review from the perspective of secondary use of patient data	Structuring the EHR	85	Vuokko et al. Int J Med Inform. 2017 Jan;97:293-303.[61]
Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review	Free-text narratives in EHR	14	Koleck et al. J Am Med Inform Assoc. 2019 Apr 1;26(4):364-379.[63]
Systematic Evaluation of Research Progress on Natural Language Processing in Medicine Over the Past 20 Years: Bibliometric Study on PubMed	NLP in medicine	2336	Wang et al. J Med Internet Res. 2020 Jan 23;22(1):e16816.[64]
Current approaches to identify sections within clinical narratives from electronic health records: a systematic review	NLP in EHR	39	Pomares-Quimbaya et al. BMC Med Res Methodol. 2019 Jul 18;19(1):155.[65]
Clinical Research			
A systematic review on natural language processing systems for eligibility prescreening in clinical research	Prescreening clinical research	11	Idnay et al. J Am Med Inform Assoc. 2021 Nov 2;ocab228. [58]
Adverse event reporting			
A systematic review of natural language processing for classification tasks in the field of incident reporting and adverse event analysis	incident reporting and adverse event	35	Young et al. Int J Med Inform. 2019 Dec;132:103971.[45]

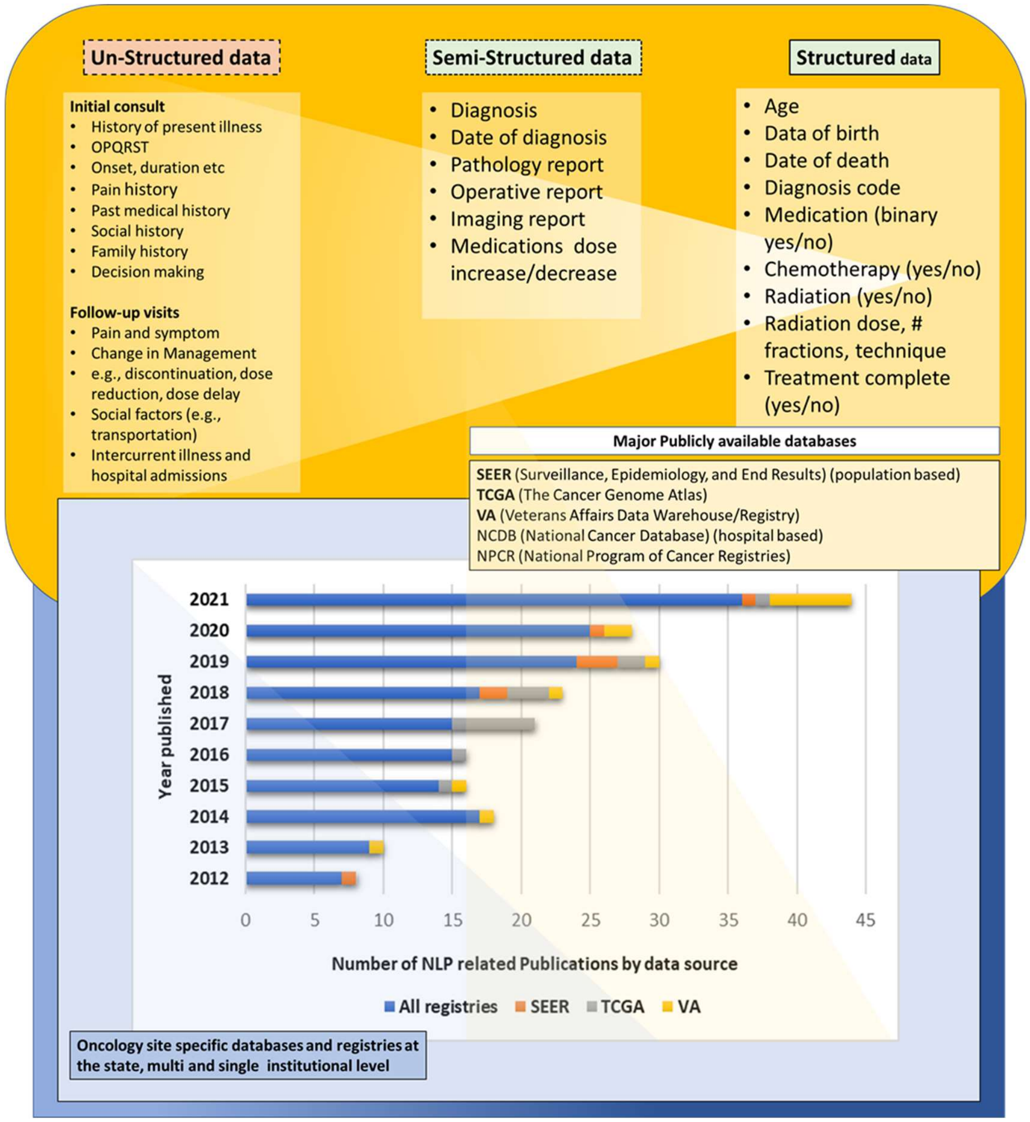

Figure 1: *Top panel: Data components in the spectrum of unstructured, semi-structured and structured data in oncology Middle panel: Major publicly available databases such as SEER (Surveillance, Epidemiology, and End Results), TCGA (The Cancer Genome Atlas) and VA (Veterans Affairs Data Warehouse/Registry), collect structured and semi-structured data. Lower panel: NLP (Natural Language Processing) related publications by data source of origin illustrating all registries (blue), SEER only (orange), TCGA only (grey), VA only (yellow) (PubMed search 1/19/2022), showcasing an emphasis on NLP related publications that employ oncology site specific databases and registries at the state, multi and single institutional level in greater proportion as compared to major publicly available data bases reflecting both lack of un-structured data capture and the role of NLP in converting unstructured data to structured data.*

ISSN 2816-8089

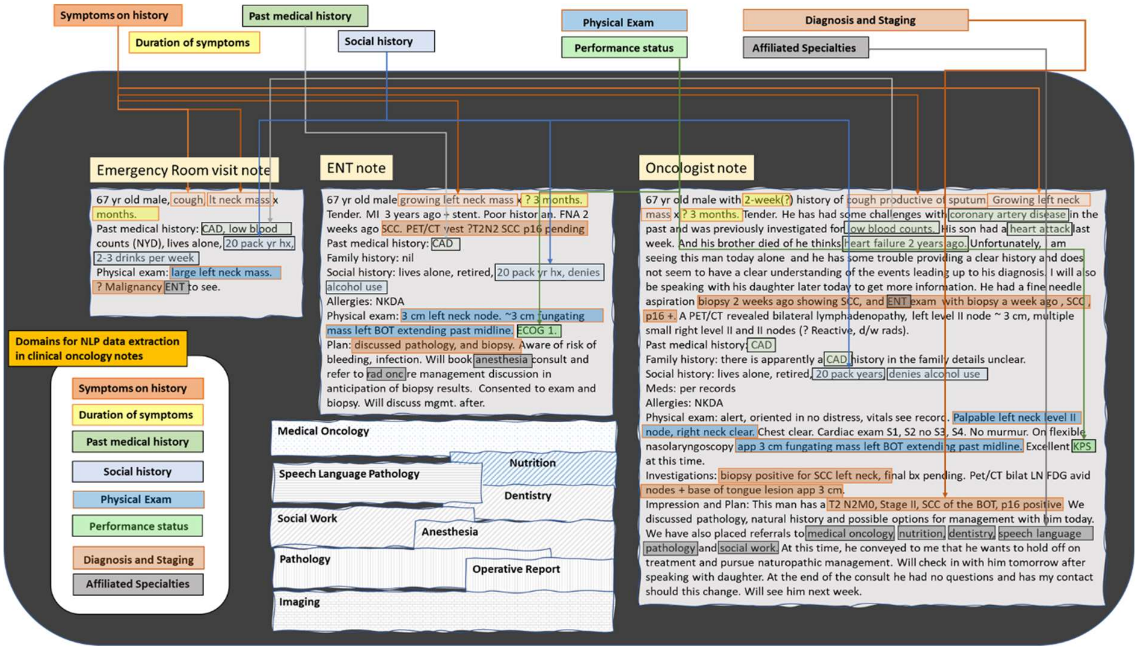

Figure 2: *A fictional clinical scenario of a patient with p16+ squamous cell carcinoma (SCC) of the base of tongue (BOT) and a hypothetical glimpse at clinical notes ranging from the emergency visit note to the first oncology note with emphasis on domains requiring NLP data extraction: clinical history, duration of symptoms, past medical history, social history, physical exam, performance status, diagnosis and staging and referral to affiliated specialties (denoted by a multitude of notes originating within each specialty).*

3. Social History

Social history is particularly significant in oncology since it can capture important cancer risk factors and features that alter management and outcomes. Smoking is an important feature of any oncologic history [17,95]. Medical providers are trained to not merely capture a binary parameter (smoking yes/no) but rather elicit and capture pack-years, and whether and when the patient quit/restarted and provide counseling. However, smoking histories are complex and heterogeneously captured (Figure 2). Patients may have quit and restarted, cut down and so on and when this is captured, it may be challenging for NLP to "pick up" on nuances. Nonbinary data remains challenging to mine using NLP and using improper data can again lead to bias. Clinicians understand the importance of the smoking history. They may naturally form conclusions from the data they collect from the patient on history hence biasing how they capture information and what aspects they lend importance to e.g., Human Papilloma Virus (HPV) status in the fictional scenario (Figure 2). A patient with a significant smoking and alcohol history may have a different capture of their p16 or HPV status vs. a patient with fewer lifestyle risk factors for head and neck cancer (Figure 2). However, if the patient does have a significant smoking history a different molecular profile in a lung tumor may be discussed or head and neck cancer a different result on p16 analysis of the tumor may be anticipated, impacting how clinicians document and what referrals are made. In a non-oncology example, where smoking history was incorporated, Gomollon et al. performed a multicenter, retrospective study using data from eight Spanish National Healthcare Network hospitals from 2014 to 2018 using NLP to examine the impact of smoking on Crohn's disease relapses. Predictive models were created with ML algorithms, namely, logistic regression, decision trees, and random forests. They identified variables such as smoking as a risk factor, described treatment patterns with biologics and examined relapse prediction noting the importance of patients' age, biochemistry values, and assessment of risk factors for relapse in a clinical setting [95]. Social history on support systems is so unstructured, it is rarely if ever captured in registries but is of significant importance to overall survival [46]. Social Determinants of Health (SDoH) information is often collected however, the lack of standardized data elements, assessment tools, measurable inputs, and data collection practices in clinical notes greatly limits the access to this information. Smoking status (n=27), substance use (n=21), homelessness (n=20), and alcohol use (n=15) are the most frequently studied SDoH categories. Homelessness (n=7) and other less-studied SDoH (e.g., education, financial problems, social isolation and support, family problems) are mostly identified using rule-based approaches. In contrast, machine learning approaches are popular for identifying smoking status (n=13), substance use (n=9), and alcohol use (n=9) [17]. A secondary cross-sectional analysis conducted on data from a prospective longitudinal cohort study of 251 hospitalized patients with aggressive hematologic malignancies used NLP to identify the extent of patients' social support (limited vs. adequate as defined by NLP-aided chart review of the electronic health record). Multivariable regression models were used to examine associations of social support with overall survival (OS), death or readmission within 90 days of discharge from index hospitalization, time to readmission within 90 days, and index hospitalization length of stay [46]. The authors found associations of limited social support with lower OS and a higher likelihood of death or readmission within 90 days of hospital discharge, enforcing the utility of NLP to evaluate the extent of social support and the need for larger studies. Further to this, a recent systematic review [17], identified 6402 publications looking at social determinants of health and NLP, of which, after applying study inclusion criteria, 82 publications were selected. The authors found that smoking status, substance use, homelessness, and alcohol use were the most frequently studied categories. Interestingly, homelessness and other less-studied domains (e.g., education, financial problems, social isolation and support, family problems) were mostly identified using rule-based approaches, while machine learning approaches were popular for smoking status, substance, and alcohol use [17]. New and evolving free text areas include social media where active efforts are ongoing [5,64] although not yet much explored in oncology (Table 1). Ongoing work is needed in the investigation of symptoms and symptom documentation in electronic health records (EHR) free-text narratives with patient characteristics and symptom-related NLP algorithms or pipelines and vocabularies lacking [63]. Over the last several years, with the increased implementation of electronic medical records, digital records can be more effectively mined than handwritten documentation and SDoH categories are increasingly transitioning from unstructured data to structured data paving the way for exciting research and outcome linked publications in 2022.

4. The Physical Exam

The capture of data pertaining to the physical is particularly challenging since there is significant variability between clinician physical exams and yet further variability in how in depth the exam is captured on paper or electronically (Figure 2). The use of abbreviations is variable but accepted and commonplace as evidenced by their use in standard medical licensing exams and in practice [19]. In fact, considering time constraints and need to capture information in electronic medical records (EMR), there has been significant impact in how the physical exam is documented to ensure maximal efficiency especially considering that many clinicians are under pressure to document their findings on electronic terminals in the patient room during the patient interaction. This has led to a decrease in the “richness” of the documented findings and emphasis on capture for future references that also serves medicolegal constraints. One particular area that represents a significant barrier effectively employing large scale data originating in real world settings for oncologic outcomes, is the lack of consistent capture of performance status which has been shown in multiple malignancies to be most associated with overall survival but is also associated with selection of management plans and secondarily progression of disease (Figure 2). Data sets that do carry a robust capture of performance status are usually originating on clinical trials or smaller institutional registries both carrying risk for bias, trials since selected trial patient population and institutional registries the bias in data captured from a select area(s), by select providers. In real world data sets, capture is inconsistent and when present can vary in format (ECOG (Eastern Cooperative Oncology Group) vs KPS (Karnofsky Performance Status)) by specialty (Figure 2) and the evolution of the performance status over the course of the disease is very poorly captured outside of clinical trial data. The physical exam is one area of unstructured data in particular in the "real-world" setting, where we anticipate that progress will require more significant research in the coming years as it represents a facet of oncology that has been impacted most significantly by both the implementation of EMR and the unparalleled growth of telemedicine in the context of the pandemic. In 2022, we will see more research examining the process of data curation in this space.

5. Diagnosis, Pathology, Staging and Imaging Reports

Source documents in oncology that are particularly defining for meaningful registry data are diagnosis and staging [36,77,89] often mentioned in clinical notes and defined in the pathology [28,29,31,32,96], operative [3,27] and imaging reports [16,33,34,38,97]. Both pathology and operative reports have increasingly been moving towards synoptic or more structured reporting, but this is still heterogenous and unstructured data is common. On the surface one might surmise that the international classification of diseases (ICD) coding is already embedded EMR system and may provide a robust framework. This is however rarely the case and providers are often aware of the heterogeneity in capture (Figure 2). ICD coding often exists in separate silos e.g., pathology report, operative report and may not be easily pulled from EMR as they represent unstructured data (Figure 1) resulting in variability in capture in clinical notes [3,93,98] (Figure 2). A recent study established an EMR information knowledge system and collected the data of patient medical records and disease diagnostic codes on the front pages of 8 clinical departments (endocrinology, oncology, obstetrics and gynecology, ophthalmology, orthopedics, neurosurgery, cardiovascular medicine) for statistical analysis [98]. Natural language processing-bidirectional recurrent neural network (NLP-BIRNN) algorithm was employed to optimize medical records. This revealed that that the coder unclear about the basic rules of main diagnosis selection and the classification of disease coding and did not code according to the main diagnosis principles. Further, the disease was not coded according to different conditions or specific classification, the code of postoperative complications was inaccurate, the disease diagnosis was incomplete, and the code selection was too general. They concluded that coders and medical personnel should strengthen communication and knowledge training [98]. BIRNN was superior when compared with the convolutional neural network (CNN) and recurrent neural network (RNN) in accuracy, symptom accuracy, and symptom recall, potentially providing a solution to this very challenging problem. As oncology clinicians are aware, in the hematology-oncology setting, pathology reports are particularly complex and the extent of detail captured in the EMR is highly variable. Zaccaria et al. used NLP to transpose unstructured reports into structured, standardized electronic health records by developing an automated tool to recognize information from pathology reports and populate electronic case report forms (eCRFs) [96]. The tool was applied to hemo-lymphopathology reports of diffuse large B-cell, follicular, and mantle cell lymphomas and assessed for accuracy, precision, recall and F1-score on internal and external report series. 326 (98.2%) reports were converted into corresponding eCRFs, and the tool showed high performance in capturing (1) identification report number (all metrics > 90%), (2) biopsy date (all metrics > 90% in both series), (3) specimen type and diagnosis [96]. In a similar study, Abedian et al. developed a NLP pipeline implemented on an open-source framework called Leo using 555,681 surgical pathology reports of 329,076 patients which was then evaluated on subsets of reports from patients with breast, prostate, colorectal, and randomly selected cancer subtypes, and achieved an accuracy of 1.00 for International Classification of Diseases, Tenth Revision codes, 0.89 for T staging, 0.90 for N staging, and 0.97 for M staging with an F1 score of 1.00 for International Classification of Diseases, Tenth Revision codes, 0.88 for T staging, 0.90 for N staging, and 0.24 for M staging [28]. While it was not generalizable to other institutions, the recommendation was made by the authors for other institutions to adopt a similar NLP approach and reuse the code available at GitHub-to support research[28]. In particular, in the metastatic setting the data becomes more unreliable. This aspect was examined by Alba et al. in the setting of metastatic prostate cancer using the Veterans Affairs Corporate Data Warehouse. The authors identified all veterans with prostate cancer and then using an NLP algorithm identified patients with any history or progression of metastatic prostate cancer. They found that out of a total of 1,144,610 Veterans diagnosed with PCa (2000 – 2020), 76,082 (6.6%) were identified by NLP as having metastatic disease at some point during their care with a specificity of 0.979 and sensitivity of 0.919 [77]. Do et al. employed NLP to look at cancer patients' computer tomography (CT) reports and generate a database of metastatic phenotypes with the goal of aggregating this with genomic studies to explore prognostic imaging phenotypes with relevance to treatment planning [37]. An area of significant need is radiologic reporting where outside of prospective protocols structured reporting of specific features can be highly variable. NLP in radiology reporting [16,35,99] has been used to develop tumor spread signatures [16], predict outcomes [33] and real-time screening for clinical trials [41,57]. Kim et al. developed a deep-transfer-learning-based natural language processing model that analyzed serial magnetic resonance imaging reports of rectal cancer patients to predict their overall survival and evaluated the model in a retrospective cohort study of 4,338 rectal cancer patients. The proposed model, utilizing pre-trained clinical linguistic knowledge, could predict the overall survival of patients without any structured information and was superior to the carcinoembryonic antigen in predicting survival [33]. This is a noteworthy example of deep-transfer-learning using free-text radiological reports successfully predicting survival and the authors pointed out significantly increasing utility of unstructured medical big data [33]. Davidson et al. carried out a systematic review of studies applying NLP to radiology reports (164 studies 2015-2019) and noted that the proportion of studies that described their annotated, training, validation, and test set were 67.1%, 63.4%, 45.7%, and 67.7% respectively leading the authors to conclude that suboptimal reporting quality precludes comparison, reproducibility, and replication. They encouraged the development of reporting standards specific to clinical NLP studies [16]. In a non-oncology study looking at 2 datasets consisting of radiologist-annotated reports of both trauma radiographs, chest radiographs and CT studies the authors investigated the impact of class imbalance, variation in dataset size, variation in report complexity, and on NLP performance of four types: a fully connected neural network (Dense), a long short-term memory recurrent neural network (LSTM), a convolutional neural network (CNN), and a Bidirectional Encoder Representations from Transformers (BERT)] of deep learning-based NLP [35]. All four model-architectures demonstrated high performance with metrics up to > 0.90, however the BERT algorithm was superior producing stable results despite variation in training size and prevalence. However, it should be noted that this is an evolving field and despite the promise of BERT, generalizability and the use of multi-language setting NLP are yet the subject of ongoing investigations [29,36,99]. A lack of transparent methodology limits comparison of approaches and reproducibility. Multiple systematic reviews on NLP in radiology have been carried out (Table 1). Oncology (24%) was the most frequent disease area analysed [16]. Most studies had dataset size > 200 (85.4%) but the proportion of studies that described their annotated, training, validation, and test set were 67.1%, 63.4%, 45.7%, and 67.7% respectively. About half of the studies reported precision (48.8%) and recall (53.7%). Few studies reported external validation being performed (10.8%), provided data availability (8.5%) and had code available. Existing reviews support the need for development of reporting standards specific to clinical NLP studies [16]. Diagnosis, pathology, staging and imaging reports have arguably been most often the subject of NLP approaches and in 2022 we will see a greater emphasis on validation and quality assurance.

6. Response, Toxicity, and Patient-reported Outcomes

Outside of prospective clinical trials or natural history protocols, response and toxicity are capture albeit inconsistently in clinical notes. Treatment response is rarely captured quantitatively, and toxicity is unlikely to be captured consistently or according to CTCAE framework [44,45]. The reasons here are multifactorial but often secondary to resource limitations. NLP tools in this space are actively evolving (Table 1). Outcome capture beyond overall survival remains problematic since patient-reported outcomes (PRO) have their own barriers in oncology since patients are ill and PRO tools may result in additional time needed on the part of patient and caregiver. While PRO tools have grown significantly in clinics and many centers have implemented validated tools, information may be incomplete and time points may be missing. Ill and poor performance status patients are less likely to complete PRO questionnaires and data may be biased by larger data sets from better performing patients with longer survivals. The richness of clinical notes remains unmined as often unstructured. Unstructured data free text such as in context of patient centered care and the patient experience feedback [100] is almost exclusively employed in this space (Figure 1). This is of need in children and adolescent long-term survivors of malignancy and rare cancers where data may suffer from additional scarcity. Notably in a child and adolescent survivor cross-sectional study from St. Jude Children's Research Hospital where pain interference and fatigue symptoms were reported through in-depth interviews after transcription, analyzable sentences were semantically labeled by 2 content experts for each attribute (physical, cognitive, social, or unclassified) and two NLP/ML methods were used to extract and validate the semantic features, with the bidirectional encoder representations from transformers (BERT) performed best suggesting that collecting unstructured PROs via interviews or conversations during clinical encounters and applying NLP/ML methods may facilitate PRO assessment in child and adolescent cancer survivors and represent an alternative to using standard PRO surveys [71]. In an interesting study with potential applicability to other areas, Mathew et al. developed NLP-ML models for incident classification in radiation oncology. They integrated these into the incident learning system to generate a drop-down menu such that the model as a semi-automated feature could improve the usability, accuracy and efficiency of the incident reporting system overall [101]. Hong et al. had two independent reviewers identify National Cancer Institute Common Terminology Criteria for Adverse Events (CTCAE) v5.0 symptoms from 100 randomly selected notes for on-treatment visits during radiation therapy with adjudication by a third reviewer. They compared the results to a NLP pipeline based on Apache clinical Text Analysis Knowledge Extraction System used to extract CTCAE terms and found that NLP accurately detected a subset of documented present CTCAE symptoms but was limited for negated symptoms [44]. PRO questionaries have been increasingly implemented in oncology clinics over several years both in the context of clinical trials and real-world settings and in 2022 and beyond. This investment will result in more sophisticated approaches to patient-reported outcomes and treatment-related toxicity analyses using NLP which will likely proliferate rapidly since this data is increasingly available in structured formats.

7. Barriers to Advancement

Heterogeneous approaches to the reporting on the development and evaluation of NLP algorithms that map clinical text to ontology concepts persist (Table 1). Over one-fourth of the identified publications did not perform an evaluation, and over one-fourth of the included studies did not perform validation, and 88% did not perform external validation [13]. Most of the data is English-centric, with 78% English language and most published in last five years[3,64]. There is a scarcity of publicly available data that likely impairs the development of more advanced methods, such as extracting word embeddings from clinical notes [62]. Free text narratives pose a special problem, and more research should be directed at site-specific oncologic stetting to create frameworks for patient characteristics, histories, and symptoms in NLP algorithms and pipelines and vocabularies openly available [63]. Variable reporting and missing metrics (e.g. F-score) are holding back the implementation of NLP in oncology, and transparency is needed regarding data sources and the performance of AI methods [6]. The use of custom dictionaries in most studies also impairs reproducibility and their use in conjunction with the UMLSⓇ meta thesaurus is recommended [65]. It is noteworthy that far less than 25% of papers that involve NLP in medicine meet inclusion criteria for systematic reviews [1,2,5,6,16,34,51,58,62,64]. As NLP is increasingly evolving from traditional methods to deep learning models such as LSTM and BERT [35], the field is gaining a greater understanding of barriers that include dataset size, the lack of ground truth annotation and the impact of prevalence on class imbalance in addition to data variation and the impact of the large size of clinical texts on algorithm performance.

8. Ongoing Efforts and Future Directions

The increased need to leverage existing data to identify clinically relevant biomarkers has given rise to Meta Map, a natural language processing tool developed and funded by the National Library of Medicine to map biomedical text to the Unified Medical Language System Meta thesaurus by applying specific tags to clinically relevant terms [87]. With Meta map the goal is to use terminology-driven semantic tags, incorporate them into a semantic frame that is task-specific to add context. This was used by Holmes et al. to select 6,713 relevant reports containing standard-of-care biomarkers for metastatic breast cancer. The name, type, numeric quantifiers, non-numeric qualifiers, and time frame for several features (breast cancer gene 1 and 2, estrogen receptor, progesterone receptor, human epidermal growth factor receptor 2, and programmed death-ligand 1) were extracted and then tested on pathology reports from the internal pathology laboratory at Henry Ford Health System. This was compared to a certified tumor registrar who reviewed 400 tests and resulted in 95% accuracy. In 2016 the FDA launched the NLP Web Service for Structuring and Standardizing Unstructured Clinical Information in anticipation of addressing growing unstructured data sets in medicine [23] and in 2018 this effort resulted in creating a corpus of 1000 Vaccine Adverse Event Reporting System (VAERS) reports annotated for 36,726 clinical features, 13,365 temporal features, and 22,395 clinical-temporal links [102]. A similar reference standard in oncologic specialties is evolving. Concerning social determinants of health, attempts to improve standardization at the national level have been made by the Protocol for Responding to and Assessing Patients' Assets, Risks, and Experiences (PRAPARE) [17]. Moving forward, lexicons are actively being developed but are not available outside of oncologic sites such as breast and lung. There is currently no standardized approach to clinical notes and the implementation of EMR and voice recognition has placed additional burden on health care providers, driving a trend towards a less rich documentation to allow for time management in the clinic. NLP of existing documentation and its change over time can and should direct what aspects of clinical notes are potentially vital for oncologic biomarker analysis but are currently captured poorly. This understanding will allow for EMR to be optimized to allow rapid, effective, and consistently captured features e.g., dropdown menus. Efforts are underway but EMR implementation is often suffering from lack of direct provider input and the secondary use of data is often not the priority of implementation of EMR systems. Lexicons are now increasingly being generated in several domains and oncologic sites addressing a crucial limiting step to harnessing NLP in converting unstructured clinical data to structured data that can alter care and improve outcomes. Multiple reviews support the development of reporting standards specific to clinical NLP studies [16] (Table 1). Kersloot et al. generated a list of sixteen recommendations regarding using NLP systems and algorithms, usage of data, evaluation and validation, presentation of results, and generalizability of results that may be employed [13]. Depending on the data and the specific oncologic setting, it will also be crucial to determine whether NLP confirms clinical hypotheses rather than developing entirely new information [5]. One of the most critical obstacles identified to machine learning approaches in clinical NLP is often listed as data annotation. However, as illustrated in Figure 2, this is likely to remain a barrier due to the workload and expertise involved. Active learning and distant supervision have been explored, and future research in this field in 2022 will benefit from data augmentation and transfer learning, or unsupervised learning, which do not require data annotation [103]. Automated data extraction pipelines may advance this effort but standardizing NLP methodology and accuracy reporting for greater generalizability will be needed as well as the use of crowdsourcing competitions to spur innovative NLP pipelines would further [104].

9. Conclusions

There has been increased focus on imaging and genomic data in AI-driven methods implemented in oncology. This is in part secondary to the lack of robust clinical annotated data sets and large-scale registries that possess detailed clinical data for DL methods. While data is collected on a large-scale in all oncology settings, the lack of consistency in large scale clinical data outside of clinical trials has limited its use. Crucial parameters such as patient comorbidities, family or social history and performance status are lacking. Staging investigations, pathology, operative and imaging reports are often unstructured or semi-structured. NLP may overcome existing clinical resource constraints to allow the capture of information from these sources, augment growing radio genomic analyses and draw meaningful clinically applicable conclusions that allow for improvement in patient outcomes. However, it is increasingly clear that several variables impact the approaches employed and the results of NLP, most importantly the documentation itself which is location, setting and population dependent. There has been increased growth in NLP methods however, quality assurance, reporting, and lexicon generation is evolving as is the addition of data from social media and wearables, all of which we will see increasingly employed and the subject of NLP approaches in 2022. The goal remains to identify and grow data that can be mined with practical NLP methods to allow for robust data acquisition in oncology that enables superior computational analysis of large-scale oncologic data sets. At this time most areas of oncologic need have benefited from research activity in NLP but generalizability is lacking. Open-source code may be used to optimize NLP, reproduce findings, and allow for increasingly effective algorithms. This is an area of ongoing development where clinician engagement will be paramount in moving the needle to incorporate the richness of human gathered information into AI learning. In 2022, two important questions will need to be asked when EMR is implemented towards clinical data: does the data captured have the potential to be employed in secondary research? And can clinical knowledge transfer be produced as a result? In oncology, the focus should be placed on specific niche areas where biomarkers are likely to play the most significant role and solving a clinical need should be paramount. As of the end of 2021, we are still in the process of finding the missing link for big oncologic data and several avenues discussed here are likely to produce increasingly pivotal results in the coming years as data matures and NLP approaches become increasingly robust.

Consent for publication

All authors consented to publication of the material.

Funding

This work was supported by NCI NIH Radiation Oncology Branch funding.

Competing Interests

The authors declare that they have no competing interests.

References

1. Yan MY, Gustad LT, Nytro O. Sepsis prediction, early detection, and identification using clinical text for machine learning: a systematic review. J Am Med Inform Assoc. 2021;236:1-17.

2. Reading Turchioe M, Volodarskiy A, Pathak J, et al. Systematic review of current natural language processing methods and applications in cardiology. Heart. 2021.

3. Mellia JA, Basta MN, Toyoda Y, et al. Natural language processing in surgery: A systematic review and meta-analysis. Ann Surg. 2021;273:900-8.

4. Buchlak QD, Esmaili N, Bennett C, et al. Natural language processing applications in the clinical neurosciences: A machine learning augmented systematic review. Acta Neurochir Suppl. 2022;134:277-89.

5. Le Glaz A, Haralambous Y, Kim-Dufor DH et al. Machine learning and natural language processing in mental health: systematic review. J Med Internet Res. 2021;23:1-95.

6. Grothe AE, Tennant B, Wang C, et al. Application of artificial intelligence methods to pharmacy data for Cancer Surveillance and Epidemiology Research: A Systematic Review. JCO Clin Cancer Inform 2020;4:1051-8.

7. Streiner DL, Saboury B, Zukotynski KA. Evidence-based artificial intelligence in medical imaging. PET Clin 2022;17:51-5.

8. Hasani N, Farhadi F, Morris MA, et al. Artificial intelligence in medical imaging and its impact on the rare disease community: Threats, challenges and opportunities. PET Clin. 2022;17:13-29.

9. Echle A, Rindtorff NT, Brinker TJ, et al. Deep learning in cancer pathology: a new generation of clinical biomarkers. Br J Cancer. 2021;124:686-96.

10. Johnson KB, Wei WQ, Weeraratne D, et al. Precision medicine, AI, and the future of personalized health Care. Clin Transl Sci. 2021;14:86-93.

11. Shui L, Ren H, Yang X, et al. The era of radiogenomics in precision medicine: An emerging approach to support diagnosis, treatment decisions, and prognostication in oncology. Front Oncol. 2020;10:1-21.

12. Shimizu H, Nakayama KI. Artificial intelligence in oncology. Cancer Sci. 2020;111:1452-60.

13. Kersloot MG, Van Putten FJP, Abu-Hanna A, et al. Natural language processing algorithms for mapping clinical text fragments onto ontology concepts: a systematic review and recommendations for future studies. J Biomed Semantics. 2020;11:1-21.

14. Weber GM, Mandl KD, Kohane, I.S. Finding the missing link for big biomedical data. JAMA. 2014;311: 2479-80.

15. Kim E, Rubinstein SM, Nead KT, et al. The evolving use of electronic health records (EHR) for research. Semin Radiat Oncol. 2019;29:354-61.

16. Davidson EM, Poon MTC, Casey A, et al. The reporting quality of natural language processing studies: systematic review of studies of radiology reports. BMC Med Imaging. 2021;21:1-13.

17. Patra BG, Sharma MM, Vekaria V et al. Extracting social determinants of health from electronic health records using natural language processing: a systematic review. J Am Med Inform Assoc. 2021;28: 2716-27.

18. Sanyal J, Tariq A, Kurian AW, et al. Weakly supervised temporal model for prediction of breast cancer distant recurrence. Sci Rep. 2021;11:1-12.

19. Percha B, Pisapati K, Gao C, et al. Natural language inference for curation of structured clinical registries from unstructured text. J Am Med Inform Assoc. 2021;29:97-108.

20. Ritzwoller DP, Hassett MJ, Uno H. Regarding the utility of unstructured data and natural language processing for identification of breast cancer recurrence. JCO Clin Cancer Inform. 2021;5:1024-5.

21. Kehl KL, Xu W, Lepisto E, et al. Natural language processing to ascertain cancer outcomes from medical oncologist notes. JCO Clin Cancer Inform. 2020;4:680-90.

22. Venkataraman GR, Pineda AL, Bear Don't Walk Iv, et al. FasTag: Automatic text classification of unstructured medical narratives. PLoS One. 2020;15:1-18.

23. Kreimeyer K, Foster M, Pandey A, et al. Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review. J Biomed Inform. 2017;73:14-29.

24. Warner JL, Anick P, Hong P, et al. Natural language processing and the oncologic history: is there a match? J Oncol Pract. 2011;7:1-5.

25. Johnson DA. Value of the lost art of a good history and physical exam. Clin Transl Gastroenterol. 2016;7:1-2.

26. Dumont-Driscoll MC. Foreword: too little, too late, too much, too long, just right? Reinforcing the importance of a thorough history and physical exam for correct diagnosis and ongoing patient management. Curr Probl Pediatr Adolesc Health Care. 2015;45:1-2.

27. Barber EL, Garg R, Persenaire C, et al. Natural language processing with machine learning to predict outcomes after ovarian cancer surgery. Gynecol Oncol. 2021;160:182-6.

28. Abedian S, Sholle ET, Adekkanattu PM, et al. Automated extraction of tumor staging and diagnosis information from surgical pathology reports. JCO Clin Cancer Inform. 2021;5:1054-61.

29. Hammami L, Paglialonga A, Pruneri G, et al. Automated classification of cancer morphology from Italian pathology reports using natural language processing techniques: A rule-based approach. J Biomed Inform. 2021;116:103712.

30. Malke JC, Jin S, Camp SP, et al. Enhancing case capture, quality and completeness of primary melanoma pathology records via natural language processing. JCO Clin Cancer Inform. 2019;3:1-11.

31. Leyh-Bannurah SR, Tian Z, Karakiewicz PI, et al. Deep learning for natural language processing in urology: State-of-the-art automated extraction of detailed pathologic prostate cancer data from narratively written electronic health records. JCO Clin Cancer Inform. 2018;2:1-9.

32. Schroeck FR, Patterson OV, Alba PR, et al. Development of a natural language processing engine to generate bladder cancer pathology data for health services research. Urology. 2017;110:84-91.

33. Kim S, Lee CK, Choi Y, et al. Deep-learning-based natural language processing of serial free-text radiological reports for predicting rectal cancer patient survival. Front Oncol. 2021;11:1-9.

34. Casey A, Davidson E, Poon M, et al. A systematic review of natural language processing applied to radiology reports. BMC Med Inform Decis Mak. 2021;21:179.

35. Olthof AW, Van Ooijen PMA, Cornelissen LJ. Deep learning-based natural language processing in radiology: The impact of report complexity, disease prevalence, dataset size and algorithm type on model performance. J Med Syst. 2021;45:1-16.

36. Nobel JM, Puts S, Weiss J, et al. T-staging pulmonary oncology from radiological reports using natural language processing: translating into a multi-language setting. Insights Imaging. 2021;12:1-7.

37. Do RKG, Lupton K, Causa Andrieu PI, et al. Patterns of metastatic disease in patients with cancer derived from natural language processing of structured CT radiology reports over a 10-year period. Radiology. 2021;301:115-22.

38. Sorin V, Barash Y, Konen E, et al. Deep learning for natural language processing in radiology-fundamentals and a systematic review. J Am Coll Radiol. 2020;17:639-48.

39. Senders JT, Karhade AV, Cote DJ, et al. Natural language processing for automated quantification of brain metastases reported in free-text radiology reports. JCO Clin Cancer Inform. 2019;3:1-9.

40. Ribelles N, Jerez JM, Rodriguez-Brazzarola P, et al. Machine learning and natural language processing (NLP) approach to predict early progression to first-line treatment in real-world hormone receptor-positive (HR+)/HER2-negative advanced breast cancer patients. Eur J Cancer. 2021;144:224-31.

41. Kehl KL, Groha S, Lepisto EM, et al. Clinical inflection point detection on the basis of EHR data to identify clinical trial-ready patients with cancer. JCO Clin Cancer Inform. 2021;5:622-30.

42. Savova GK, Danciu I, Alamudun F, et al. Use of natural language processing to extract clinical cancer phenotypes from electronic medical records. Cancer Res. 2019;79:5463-70.

43. Alkaitis MS, Agrawal MN, Riely GJ, et al. Automated NLP extraction of clinical rationale for treatment discontinuation in breast cancer. JCO Clin Cancer Inform. 2021;5:550-60.

44. Hong JC, Fairchild AT, Tanksley JP, et al. Natural language processing for abstraction of cancer treatment toxicities: accuracy versus human experts. JAMIA Open. 2020;3:513-7.

45. Young IJB, Luz S, Lone N. A systematic review of natural language processing for classification tasks in the field of incident reporting and adverse event analysis. Int J Med Inform. 2019;132:103971.

46. Johnson PC, Markovitz NH, Gray TF, et al. Association of social support with overall survival and healthcare utilization in patients with aggressive hematologic malignancies. J Natl Compr Canc Netw. 2021;1-7.

47. Voytovich L, Greenberg C. Natural language processing: practical applications in medicine and investigation of contextual autocomplete. Acta Neurochir Suppl. 2022;134:207-14.

48. Nguyen OT, Turner K, Apathy NC, et al. Primary care physicians' electronic health record proficiency and efficiency behaviors and time interacting with electronic health records: a quantile regression analysis. J Am Med Inform Assoc. 2021;29:461-71.

49. Hernandez-Boussard T, Kourdis PD, Seto T, et al. Mining electronic health records to extract patient-centered outcomes following prostate cancer treatment. AMIA Annu Symp Proc. 2017;2017:876-82.

50. Coquet J, Bozkurt S, Kan KM, et al. Comparison of orthogonal NLP methods for clinical phenotyping and assessment of bone scan utilization among prostate cancer patients. J Biomed Inform. 2019;94:1-9.

51. Buchlak QD, Esmaili N, Leveque JC, et al. Machine learning applications to neuroimaging for glioma detection and classification: An artificial intelligence augmented systematic review. J Clin Neurosci. 2021;89:177-98.

52. Rybinski M, Dai X, Singh S, et al. Extracting family history information from electronic health records: natural language processing analysis. JMIR Med Inform. 2021;9:1-21.

53. Van Laar SA, Gombert-Handoko KB, Guchelaar HJ, et al. An electronic health record text mining tool to collect real-world drug treatment outcomes: A validation sudy in patients with metastatic renal cell carcinoma. Clin Pharmacol Ther. 2020;108:644-52.

54. Li Y, Luo YH, Wampfler JA, et al. Efficient and accurate extracting of unstructured EHRs on cancer therapy responses for the development of RECIST natural language processing tools: part I, the corpus. JCO Clin Cancer Inform. 2020;4:383-91.

55. Doan S, Yang EW, Tilak SS, et al. Extracting health-related causality from twitter messages using natural language processing. BMC Med Inform Decis Mak. 2019;19:1-7.

56. Nguyen AN, Moore J, O'Dwyer J, et al. Assessing the utility of automatic cancer registry notifications data extraction from free-text pathology reports. AMIA Annu Symp Proc. 2015;2015:953-62.

57. Jung E, Jain H, Sinha AP, et al. Building a specialized lexicon for breast cancer clinical trial subject eligibility analysis. Health Informatics J. 2021;27:1-15.

58. Idnay B, Dreisbach C, Weng C, et al. A systematic review on natural language processing systems for eligibility prescreening in clinical research. J Am Med Inform Assoc. 2021;29:197-206.

59. Scaccia JP, Scott VC. 5335 days of Implementation science: using natural language processing to examine publication trends and topics. Implement Sci. 2021;16:1-12.

60. Qin X, Liu J, Wang Y, et al. Natural language processing was effective in assisting rapid title and abstract screening when updating systematic reviews. J Clin Epidemiol. 2021;133:121-9.

61. Vuokko R, Makela-Bengs P, Hypponen H, et al. Impacts of structuring the electronic health record: Results of a systematic literature review from the perspective of secondary use of patient data. Int J Med Inform. 2017;97:293-303.

62. Sheikhalishahi S, Miotto R, Dudley JT, et al. Natural language processing of clinical notes on chronic diseases: systematic review. JMIR Med Inform. 2019;7:1-298.

63. Koleck TA, Dreisbach C, Bourne PE, et al. Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. J Am Med Inform Assoc. 2019;26: 364-79.

64. Wang J, Deng H, Liu B, et al. Systematic evaluation of research progress on natural language processing in medicine over the past 20 years: bibliometric study on PubMed. J Med Internet Res. 2020;22:1-46.

65. Pomares-Quimbaya A, Kreuzthaler M, Schulz S. Current approaches to identify sections within clinical narratives from electronic health records: a systematic review. BMC Med Res Methodol. 2019;19:1-20.

66. Agaronnik ND, Lindvall C, Iezzoni LI. Natural language processing to assess frequency of functional status documentation for patients newly diagnosed with colorectal cancer-reply. JAMA Oncol. 2021;7: 463.

67. Badal VD, Nebeker C, Shinkawa K, et al. Do words matter? Detecting social isolation and loneliness in older adults using natural language processing. Front Psychiatry. 2021;12:1-12.

68. Zhu VJ, Lenert LA, Bunnell BE, et al. Automatically identifying social isolation from clinical narratives for patients with prostate Cancer. BMC Med Inform Decis Mak. 2019;19:1-10.

69. Leiter RE, Santus E, Jin Z, et al. Deep natural language processing to identify symptom documentation in clinical notes for patients with heart failure undergoing cardiac resynchronization therapy. J Pain Symptom Manage. 2020;60:948-58.

70. Naseri H, Kafi K, Skamene S, et al. Development of a generalizable natural language processing pipeline to extract physician-reported pain from clinical reports: Generated using publicly-available datasets and tested on institutional clinical reports for cancer patients with bone metastases. J Biomed Inform. 2021;120:103864.

71. Lu Z, Sim JA, Wang JX, et al. Natural language processing and machine learning methods to characterize unstructured patient-reported outcomes: validation study. J Med Internet Res. 2021;23:1-16.

72. Bharadwaj P, Dooley R, Dabagh S, et al. Using machine learning to optimize appropriate advance care planning documents in electronic health records. J Palliat Med. 2021;24:1754.

73. Lindvall C, Deng CY, Moseley E, et al. Natural language processing to identify advance care planning documentation in a multisite pragmatic clinical trial. J Pain Symptom Manage. 2021;63:29-36.

74. Lee RY, Brumback LC, Lober WB, et al. Identifying goals of care conversations in the electronic health record using natural language processing and machine learning. J Pain Symptom Manage. 2021;61:136-42.

75. Poort H, Zupanc SN, Leiter RE, et al. Documentation of palliative and end-of-life Care process measures among young adults who died of cancer: A natural language processing approach. J Adolesc Young Adult Oncol. 2020;9:100-4.

76. Yousefirizi F, Pierre D, Amyar A, et al. AI-based detection, classification and prediction/prognosis in medical imaging: towards radiophenomics. PET Clin. 2022;17:183-212.

77. Alba PR, Gao A, Lee KM, et al. Ascertainment of veterans with metastatic prostate cancer in electronic health records: demonstrating the case for natural language processing. JCO Clin Cancer Inform. 2021;5:1005-14.

78. Bitterman DS, Miller TA, Mak RH, et al. Clinical natural language processing for radiation oncology: A review and practical primer. Int J Radiat Oncol Biol Phys. 2021;110:641-55.

79. Davis JL, Murray JF. History and physical examination. Elsevier public health emergency collection: Online, 2015;263-77.

80. Williams BC, Ward DA, Chick DA, et al. Using a six-domain framework to include biopsychosocial information in the standard medical history. Teach Learn Med. 2019;31:87-98.

81. Mbunge E, Muchemwa B, Batani J. Are we there yet? Unbundling the potential adoption and integration of telemedicine to improve virtual healthcare services in African health systems. Sens Int. 2022:3:1-9.

82. Meghiref Y, Parnot C, Duverger C, et al. Implementation of telemedicine in cancer clinical trials: connect patient to doctor study. JMIR Cancer. 2021;8:1-11.

83. Bragin I, Cohen DT. Certified examination assistants in the age of telemedicine: A blueprint through eurology. JMIR Med Educ. 2021;7:1-6.

84. Alpert JM, Campbell-Salome G, Gao C, et al. Secure messaging and COVID-19: A content analysis of patient-clinician communication during the pandemic. Telemed J E Health. 2021;1-7.

85. Zeng J, Banerjee I, Henry AS, et al. Natural language processing to identify cancer treatments with electronic medical records. JCO Clin Cancer Inform. 2021;5:379-93.

86. Alpert JM, Morris BB, Thomson MD, et al. Open notes in oncology: oncologists' perceptions and a baseline of the content and style of their clinician notes. Transl Behav Med. 2019;9:347-56.

87. Holmes B, Chitale D, Loving J, et al. Customizable natural language processing biomarker extraction tool. JCO Clin Cancer Inform. 2021;5:833-41.

88. Najafabadipour M, Zanin M, Rodriguez-Gonzalez A, et al. Reconstructing the patient's natural history from electronic health records. Artif Intell Med. 2020;105:101860.

89. Deshmukh PR, Phalnikar R. Information extraction for prognostic stage prediction from breast cancer medical records using NLP and ML. Med Biol Eng Comput. 2021;59:1751-72.

90. Zhao Y, Weroha SJ, Goode EL, et al. Generating real-world evidence from unstructured clinical notes to examine clinical utility of genetic tests: use case in BRCAness. BMC Med Inform Decis Mak. 2021;21:1-13.

91. Bar Y, Bar K, Itzhak I, et al. The impact of tumor detection method on genomic and clinical risk and chemotherapy recommendation in early hormone receptor positive breast cancer. Breast. 2021;60:78-85.

92. Jin Y, Junren W, Jingwen J, et al. Research on the construction and application of breast cancer-specific database system based on full data lifecycle. Front Public Health. 2021;9:1-11.

93. Karimi YH, Blayney DW, Kurian AW, et al. Development and use of natural language processing for identification of distant cancer recurrence and sites of distant recurrence using unstructured electronic health record data. JCO Clin Cancer Inform. 2021;5:469-78.

94. Levine MN, Alexander G, Sathiyapalan A, et al. Learning health system for breast cancer: pilot project experience. JCO Clin Cancer Inform. 2019;3:1-11.

95. Gomollon F, Gisbert JP, Guerra I, et al. Clinical characteristics and prognostic factors for Crohn's disease relapses using natural language processing and machine learning: a pilot study. Eur J Gastroenterol Hepatol. 2021;1-9.

96. Zaccaria GM, Colella V, Colucci S, et al. Electronic case report forms generation from pathology reports by ARGO, automatic record generator for onco-hematology. Sci Rep. 2021;11:1-11.

97. Verma AA, Masoom H, Pou-Prom C, et al. Developing and validating natural language processing algorithms for radiology reports compared to ICD-10 codes for identifying venous thromboembolism in hospitalized medical patients. Thromb Res. 2021;209:51-8.

98. Wang C, Yao C, Chen P, et al. Artificial intelligence algorithm with ICD coding technology guided by the embedded electronic medical record system in medical record information management. J Healthc Eng. 2021;2021:1-9.

99. Khambete MP, Su W, Garcia JC, et al. Quantification of BERT diagnosis generalizability across medical specialties using semantic dataset distance. AMIA Jt Summits Transl Sci Proc. 2021;2021:345-54.

100. Khanbhai M WL, Symons J, Flott K, et al. Using natural language processing to understand, facilitate and maintain continuity in patient experience across transitions of care. International journal of medical informatics. 2022;157:1-7.

101. Mathew F, Wang H, Montgomery L, et al. Natural language processing and machine learning to assist radiation oncology incident learning. J Appl Clin Med Phys. 2021;22:172-84.

102. Foster M, Pandey A, Kreimeyer K, et al. Generation of an annotated reference standard for vaccine adverse event reports. Vaccine. 2018;36:4325-30.

103. Spasic I, Nenadic G. Clinical text data in machine learning: systematic review. JMIR Med Inform. 2020; 8:1-19.

104. Decker BM, Hill CE, Baldassano SN, et al. Can antiepileptic efficacy and epilepsy variables be studied from electronic health records? A review of current approaches. Seizure. 2021;85:138-44.