

Review

Natural Language Processing of Clinical Notes for Cancer Research and Patient Care Prior to Widespread Adoption of Generative AI: Scoping Review

Alfred B Kayira^{1*}, MSc, MPH, MRES; Hadeel R A Elyazori^{2*}, MSc; Kevin Lybarger², PhD; Fiona M Walter¹, MA, MD; Claude Chelala³, PhD; Garth Funston¹, MB BChir, PhD

¹Centre for Cancer Screening, Prevention, and Early Diagnosis, Wolfson Institute of Population Health, Queen Mary University of London, London, United Kingdom

²Department of Information Sciences and Technology, College of Engineering & Computing, George Mason University, Fairfax, VA, United States

³Barts Cancer Institute, Queen Mary University of London, London, United Kingdom

*these authors contributed equally

Corresponding Author:

Alfred B Kayira, MSc, MPH, MRES
Centre for Cancer Screening, Prevention, and Early Diagnosis
Wolfson Institute of Population Health, Queen Mary University of London
Charterhouse Square
London EC1M 6BQ
United Kingdom
Phone: 44 7415302686
Email: a.b.kayira@qmul.ac.uk

Abstract

Background: Clinical notes are the most abundant data type within electronic health records; however, their highly unstructured format presents significant challenges for supervised natural language processing (NLP) methods. The NLP community is increasingly adapting large language models to analyze clinical notes, achieving strong performance and generalizability with minimal task-specific fine-tuning. We conducted a scoping review of NLP methods applied to clinical notes prior to widespread adoption of generative artificial intelligence (AI) to establish a pre-large language model methodological baseline, showcase potential clinical utility, and highlight key challenges and limitations of extractive, supervised techniques that generative AI approaches may need to overcome.

Objective: This review aimed (1) to characterize the clinical notes used, (2) to identify NLP techniques used to analyze these notes, (3) to determine the clinical applications of NLP in cancer research and patient care, and (4) to highlight challenges and limitations of traditional pregenerative AI methods.

Methods: We systematically searched MEDLINE, Embase, Scopus, and Web of Science for English-language studies published from January 1, 2014, to March 8, 2024. Retrieved references were imported into Covidence, a web-based platform that streamlines management of reviews. Two authors (ABK and HRAE) independently screened studies for eligibility and extracted data using a predefined data extraction template.

Results: A total of 226 studies were included in the review. Research using NLP to derive insights from clinical notes grew significantly, from 4 studies in 2014 to 43 in 2023. NLP methods have evolved from predominantly rule-based and ontology-driven approaches (2014-2017) to hybrid approaches that combine these with deep neural models such as Bidirectional Encoder Representations from Transformers (2018-2024). Most studies (161/226, 71.2%) developed their systems using small, single-institution datasets. Supervised learning approaches with manually annotated corpora were predominant (181/226, 80.1%). Most studies (174/226, 77%) focused on information extraction, with a few applying the extracted data to downstream tasks such as diagnostic and prognostic classification. Clinical domain pretrained models outperformed general domain pretrained models in the majority (11/16, 68.8%) of studies that evaluated multiple model types. In total, 25 studies compared their NLP-based systems with current practice in their respective clinical settings and reported potential benefits, including improved data coverage and completeness, faster information extraction, and improved classification or prediction accuracy. No studies evaluated the utility or impact of their systems in real-world clinical practice. The most common challenges reported by authors were restricted access to clinical notes (n=39) and limited data (n=18).

Conclusions: The application of NLP to clinical notes in oncology has expanded, but most studies focus on information extraction rather than downstream clinical tasks. Oncology NLP has the potential to advance cancer research and patient care, but barriers remain to robust evaluation and clinical deployment of promising tools. Emerging generative AI approaches will need to overcome these challenges to deliver real-world impact.

JMIR AI 2026;5:e73481; doi: [10.2196/73481](https://doi.org/10.2196/73481)

Keywords: natural language processing; clinical notes; electronic health records; clinical NLP challenges; cancer; scoping review

Introduction

Background

Cancer is a major cause of morbidity and mortality globally [1], with 19.3 million new cases and 10 million deaths reported in 2020 [1]. Incidence is projected to rise by 55% by 2040 due to population growth and aging [2]. Research leveraging real-world data is important to support prevention, early detection, and optimized treatment, and ultimately improve patient outcomes, including survival. Electronic health records (EHRs), digital profiles of patient histories created and managed by health care institutions, provide a valuable real-world data resource for cancer research and improve patient care.

While EHR systems have become increasingly available [3], only a small portion consists of structured data (eg, clinical codes, vital signs, clinical and laboratory measurements, and demographics) that can be easily extracted and analyzed using conventional statistical and machine learning methods. Most data (80%) exist in unstructured forms, including clinical notes, diagnostic reports (eg, pathology and radiology), and images [4], limiting usability [5]. Natural language processing (NLP)—a subfield of artificial intelligence (AI) that enables computers to understand, interpret, and generate human language—offers a promising approach to unlock insights from unstructured clinical narratives such as clinical notes and diagnostic reports, enabling their use in research and patient care.

While both diagnostic reports and clinical notes contain valuable information, they differ in complexity for NLP. Diagnostic reports are typically formal and standardized, making them relatively straightforward to process. In contrast, clinical notes are highly diverse due to variations in recording practices across clinicians and health care institutions [6]. They often feature incomplete sentences, poor punctuation, nonstandard abbreviations, shorthand, ambiguous terms, and spelling errors. These characteristics pose significant challenges for NLP processing, even with advanced methodological approaches such as pretrained language models (PLMs), for example, Bidirectional Encoder Representations from Transformers (BERT) [7-9], which dominated the general NLP domain since the introduction of the BERT model in 2018 [10].

However, recent advances in generative AI are reshaping the field of clinical NLP. Large language models (LLMs)—a subset of PLMs designed for generative tasks (eg, OpenAI's GPT [11] and Meta's LLaMA [12])—are

transforming clinical NLP by enabling broader generalization with minimal task-specific fine-tuning. LLMs (GPT-4, Gemma3-27B, and DeepSeek-14B), applied using prompt engineering or task-specific fine-tuning, have demonstrated strong performance in extracting treatment histories [13], social and behavioral determinants of health (employment, housing, marital status, alcohol use, tobacco use, and drug use) [14], and neurofibromatosis type 1–relevant phenotypes [15] from clinical notes. Recent review studies highlight increasing interest in the use of LLMs with prompt-based strategies, including zero-shot and few-shot prompting, for information extraction (IE) [16,17], as well as for tasks such as information summarization, translation, and clinical communication [18].

Given their strong early performance, which has generated considerable interest within the NLP community, LLMs may emerge as a dominant approach, potentially replacing traditional supervised deep learning methods (eg, recurrent neural networks [RNNs], convolutional neural networks [CNNs], and BERT-based models). To better understand the value that LLMs add beyond established NLP approaches, we conducted a scoping review of NLP methods applied to cancer clinical notes prior to the widespread use of generative AI, providing a comprehensive overview of pre-LLM methods, their potential clinical utility, and the limitations and challenges likely to extend to generative AI.

Several reviews have examined the application of NLP to clinical notes before the adoption of LLMs; however, none have specifically focused on clinical notes as the primary text. Prior reviews have included clinical notes only as a subset of broader document categories. Only 35% (43/123), 22% (5/23), and 12% (2/17) of studies included in Wang et al [19], Li et al [20], and Gholipour et al [21], respectively, used clinical notes, often alongside other medical documents (eg, radiology and pathology reports). Sangariyavanich et al [22] included 17 studies but did not specify the proportion or extent of clinical note use. Furthermore, these reviews focused on one NLP task or the other, for example, IE [19, 21], diagnostic classification [20], and prognostic classification [22]. Broader reviews by Wang et al [23], Sim et al [24], and Sheikhalishahi et al [25] covered studies, which included substantial volumes of clinical notes but were not cancer-specific, limiting their utility to the cancer domain. Additionally, these reviews only include studies published up to 2020, predating the widespread adoption of BERT-based PLMs. Notably, in Sheikhalishahi et al [25], only 3 of the 106 studies used deep learning approaches.

Objectives

This review provides a comprehensive synthesis of NLP applications to clinical notes in cancer research prior to widespread experimentation with LLMs. Unlike prior review that included studies based solely on structured diagnostic reports, we restricted inclusion to studies involving clinical notes (exclusively or in combination with diagnostic reports or other documents), so our findings more closely reflect the distinctive challenges—including acquisition—and methodological choices associated with this particularly complex text. We also diverge from earlier reviews by imposing no restrictions on the NLP task, allowing a broader characterization of cancer-related use cases beyond conventional diagnostic or prognostic classification.

By systematically analyzing pregenerative AI methodologies, this review provides important benchmarks for assessing the real “value add” of LLMs, highlights the limitations of extractive, supervised approaches, and anticipates challenges that may need to be overcome. Specifically, our objectives are (1) to characterize the clinical notes used in NLP studies, including their sources and properties; (2) to identify NLP techniques (including annotation methods) used to analyze these notes and examine how these methodologies have evolved over time; (3) to determine the clinical applications of NLP in cancer research and patient care, including reported clinical impact; and (4) to highlight the challenges encountered by researchers in the field.

Methods

This review follows the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) [26].

Working Definitions

We broadly defined NLP as the application of computational techniques to process and analyze unstructured clinical text. This encompasses a diverse range of methods, including domain-specific dictionaries, medical ontologies (eg, Unified Medical Language System [UMLS]), ontology-based tools (eg, MetaMap and Clinical Text Analysis and Knowledge Extraction System), handcrafted rules or search strings, rule-based tools (eg, ConText and NegEx), classical machine learning models (eg, support vector machine), neural networks (eg, RNN), PLMs (eg, BERT), and LLMs (a subset of PLMs distinguished by their larger parameter scale and enhanced capacity for broad generalization with minimal task-specific fine-tuning [eg, GPT]).

Clinical notes were defined as free-text narratives written by health care providers during patient encounters, documenting patient symptoms and signs, investigations, diagnoses, treatment, or treatment plans. They detail a patient’s social and medical history, disease progression, and outcomes. They are distinguished from diagnostic reports, in that they later provide results of diagnostic investigations or imaging studies, often objective and structured. Clinical notes

may, however, contain descriptions and interpretations of diagnostic results from these reports.

Search Strategy and Information Sources

We developed a three-concept search criterion covering (1) NLP, (2) EHR or electronic medical record, and (3) cancer or oncology. Predetermined key terms relating to these concepts were used to search MEDLINE through PubMed. These were further expanded by scanning the titles and abstracts of retrieved records. To avoid missing studies in which clinical notes were only one of several document types and therefore not mentioned in the title or abstract, we intentionally kept the EHR or electronic medical record concept broad. The final search criteria for all 4 databases are provided in [Multimedia Appendix 1](#).

We searched MEDLINE (via PubMed), Embase, Web of Science, and Scopus for primary studies that applied NLP to process and analyze clinical notes to generate actionable information for cancer research or patient care. For PubMed, Embase, and Web of Science, we searched across all available fields. In Scopus, the search was limited to the title, abstract, and keywords fields. We used a mix of MeSH term mappings and exact phrase or term searching to balance the sensitivity and precision of the search. All searches were restricted to English-language publications from January 1, 2014, to March 8, 2024.

Inclusion Criteria

We included peer-reviewed journal papers and conference papers that (1) applied NLP to clinical notes—either exclusively or in combination with other medical documents (eg, pathology, radiology, colonoscopy, or other imaging reports); (2) focused on any part of the cancer care continuum, including screening, diagnosis, staging, treatment, surveillance, outcomes assessment, and risk factor identification or risk stratification; and (3) were conducted in any clinical setting (eg, primary care, outpatient clinics, emergency departments, and hospitals).

Exclusion Criteria

We excluded studies that used non-EHR documents (eg, patient-authored text in online health communities), studies using translated text (eg, from one language to English before applying NLP methods), reviews, editorials, commentaries, abstracts, letters, retracted papers, and veterinary studies.

Study Selection

Study screening (title or abstract and full text) was completed in Covidence (Veritas Health Innovation Ltd), a web-based collaboration software platform that streamlines the production of systematic and other literature reviews. References identified through database searches were imported into Covidence, and duplicates were automatically removed.

Two authors (ABK and HRAE) independently assessed the papers for eligibility based on the title and abstract. Proportionate agreement (the proportion of times that reviewers agree on their assessments) was 96%. Class-specific agreement was 56.2% for the positive (include) class

and 97.9% for the negative (exclude) class. Cohen κ , which measures the agreement between 2 reviewers (ABK and HRAE) adjusting for the possibility of agreement occurring by chance, was 0.54. Full-text papers were retrieved for studies that passed the title-abstract screening, and the same authors assessed the full texts for eligibility. Proportionate agreement was 81.5%. Class-specific agreement was 86.3% for the positive (include) class and 71.8% for the negative (exclude) class. Cohen κ was 0.58. At both stages, discrepancies were discussed and resolved through consensus, with reference to the predefined inclusion or exclusion criteria and the operational definitions of key concepts (NLP, clinical notes, and cancer or oncology). When consensus could not be reached, another author (GF or KL.) adjudicated.

Data Extraction and Analysis

A data extraction template was created in Covidence and refined through several iterations until all authors agreed on the final version. Using this template, we extracted data across 37 predetermined variables, which can be classified into 5 categories: study metadata, clinical note characteristics, methods, applications, and challenges. Two authors

(ABK and HRAE) extracted data from 10% of the papers. The extracted data were compared, and inconsistencies were discussed. Concordance was high and so the remaining papers were extracted by 1 reviewer (ABK). Extracted data were analyzed descriptively, providing counts and percentages.

Study Quality Assessment

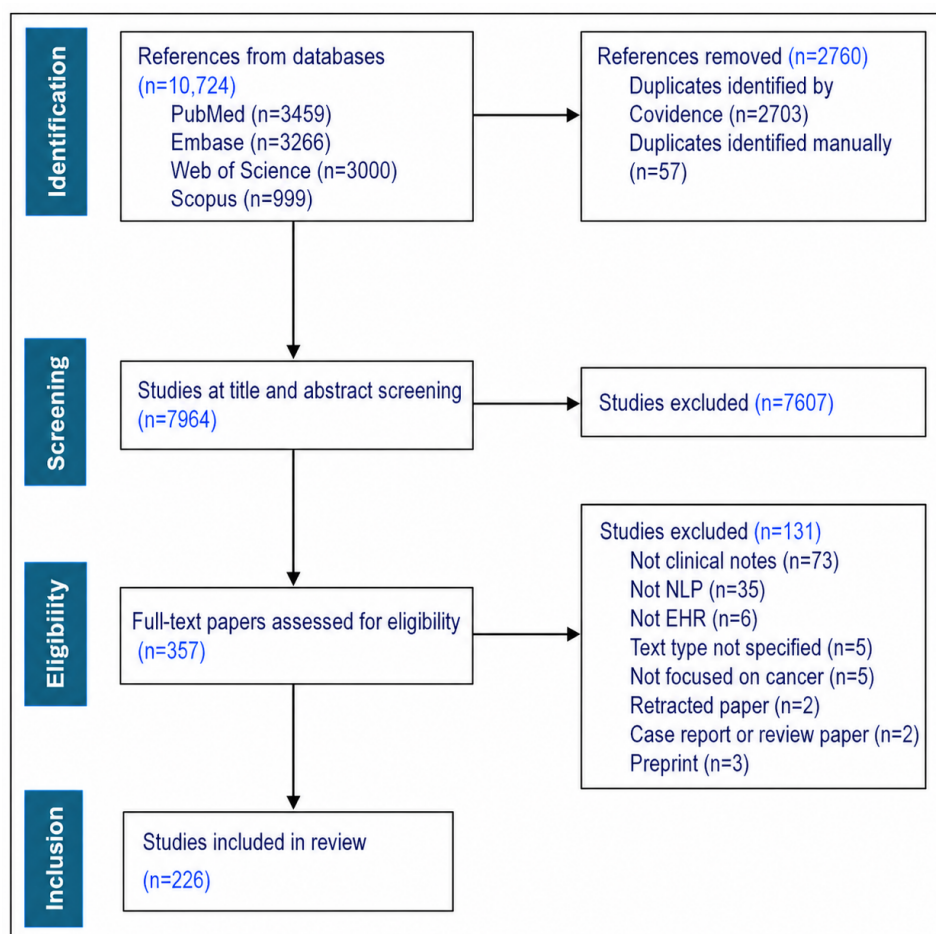
Given the scoping review methodology and our count-based analyses, a risk of bias or quality assessment was not performed [27].

Results

Search Results

Figure 1 shows the study selection process used to arrive at the included studies. A total of 10,724 records were identified from the databases. After removing duplicates, 7964 records were screened. Of these, 7607 were excluded at the title and abstract screening stage. In the full-text screening stage, 357 papers were assessed for eligibility, and 131 were excluded. Ultimately, 226 studies met the inclusion criteria.

Figure 1. PRISMA diagram illustrating the study selection process and reasons for exclusion. EHR: electronic health record; NLP: natural language processing; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

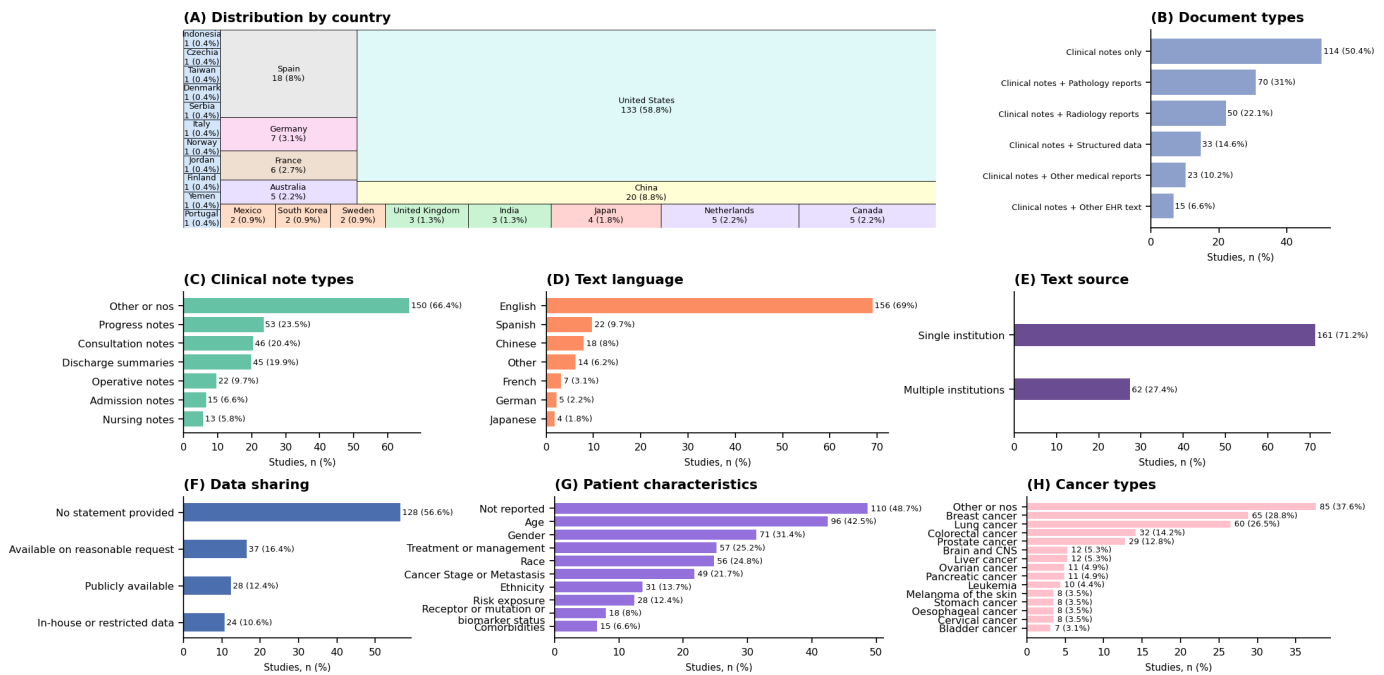


Distribution of Studies by Country

Figure 2A illustrates the distribution of included studies based on the country of institution of affiliation of the major (first

or corresponding) authors. The majority were from the United States (133/226, 58.8%), followed by China (20/226, 8.8%) and Spain (18/226, 8%).

Figure 2. Characterization of included studies. (A) Distribution of studies by country of institution of affiliation of major authors. (B) Document type. (C) Clinical note type. (D) Language of clinical notes and other medical documents. (E) Heterogeneity of the sources of clinical notes and other medical documents. In total, 3 (1.3%) studies had insufficient information to determine the source of clinical notes. (F) Accessibility of data used by the studies (publicly available means that authors used a publicly available corpus [majority] or made their corpus publicly available). (G) Patient characteristics reported in studies. (H) Cancer types targeted by studies. CNS: central nervous system; nos: not otherwise specified.



Characterization of Clinical Notes

Figure 2B shows the document types used across included studies. Out of 226 studies, 114 (50.4%) used clinical notes exclusively, while the remainder used clinical notes and other medical documents, primarily pathology and radiology reports. Progress notes (53/226, 23.5%), consultation notes (46/226, 20.4%), and discharge summaries (45/226, 19.9%) were the most common types of clinical notes used in included studies (Figure 2C). However, in 150 of the 226 (66.4%) studies, authors either used nonspecific terms to describe clinical notes (eg, oncology, urology, and cancer clinic notes) or did not specify the clinical note type. Most of the clinical notes were written in English (156/226, 69%), Spanish (22/226, 9.7%), and Chinese (18/226, 8%; Figure 2D).

Figure 2E illustrates heterogeneity in the sources of clinical notes and other medical documents used in the studies. Most studies (161/226, 71.2%) used documents from a single institution, while 27.4% (62/226) included multi-institution data from the same country. No study used documents from more than 1 country. Regarding data availability, 128 of 226 (56.6%) studies did not provide any

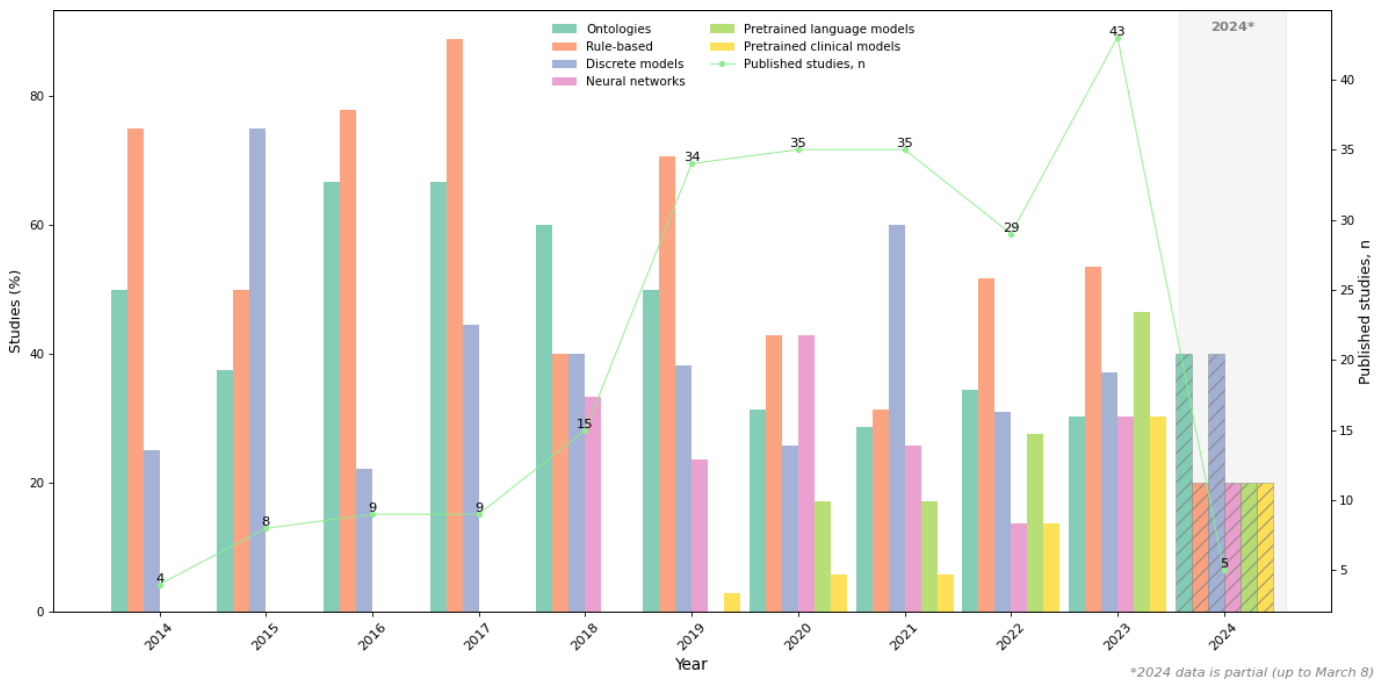
statement on the accessibility of the corpora used. A few studies (37/226, 16.4%) indicated that their corpus could be made available upon reasonable request, and 12.4% (28/226) either used publicly available corpora (majority) or made their corpus publicly accessible (Figure 2F).

Nearly half of the studies (110/226, 48.7%) did not provide any information about the characteristics of the patients associated with the clinical notes they used. When reported, common characteristics included age (96/226, 42.5%), sex or gender (71/226, 31.4%), race (56/226, 24.8%), cancer therapy or management (57/226, 25.2%), and cancer stage or metastasis (49/226, 21.7%; Figure 2G). The most commonly studied cancers were breast (65/226, 28.8%), lung (60/226, 26.5%), colorectal (32/226, 14.2%), and prostate (29/226, 12.8%; Figure 2H).

NLP Publications and Methods Used by Calendar Year

Figure 3 illustrates the number of studies published annually from January 2014 to March 2024, along with the NLP methods applied to clinical notes. The number of publications per year increased from 4 in 2014 to 43 in 2023.

Figure 3. Model architectures used to analyze clinical notes over the years. Percentages are relative to the number of studies published in that year. The line graph depicts the number of published studies per year. *2024 is a partial year; it includes papers published from January 1, 2024, to March 8, 2024. It is common for researchers to use multiple methods from the same class or different classes (either as discrete models or in hybrid architectures), leading to double-counting. “Pretrained language models” refers to general-domain pretrained models (eg, BERT and GPT), while “pretrained clinical models” refers to models with domain-specific pretraining on clinical or biomedical text (eg, BioBERT, ClinicalBERT, and PubMedBERT). These categories are mutually exclusive. BERT: Bidirectional Encoder Representations from Transformers.



NLP methods have evolved over time. Between 2014 and 2017, only ontologies, rule-based approaches, and discrete models were used. Studies using neural networks were first published in 2018, followed by PLMs in 2019 (Figure 3). While neural networks, including PLMs, have gained popularity since their introduction, ontologies, rule-based approaches, and discrete models remained the most prevalent approaches throughout the review period. However, rule-based approaches and ontologies were often used in hybrid workflows, serving specific preprocessing and postprocessing roles, rather than as standalone solutions. Out of 226 studies, only 7 (3.1%) and 27 (11.9%) exclusively used ontologies and rule-based methods, respectively.

Fine-Grained Classification of NLP Methods

Ontologies were used in 87 of 226 (38.5%) studies, with domain-specific or customized dictionaries being the most common approach (42/87, 48.3%), followed by the UMLS at 41.4% (36/87; Table 1). These knowledge resources often supported machine learning and neural models by providing seed terms or domain expertise. Off-the-shelf tools such as MetaMap and Clinical Text Analysis and Knowledge Extraction System, which rely on UMLS mappings to analyze biomedical text, were also used.

Table 1. Breakdown of methods used in included studies (N=226)^a.

Model architecture	Values (N=226), n (%)
Ontologies (n=87)	
Domain-specific dictionary	42 (48.3)
Unified Medical Language System	36 (41.4)
MetaMap	16 (18.4)
cTAKES ^b	10 (11.5)
NCBO ^c BioPortal	7 (8)
MedTagger	3 (3.4)
Other	6 (6.9)
Rule-based (n=112)	
Rules or RegEx ^d	112 (100)
Discrete models (n=87)	
Support vector machine	29 (33.3)
Trees	28 (32.2)

Model architecture	Values (N=226), n (%)
Logistic regression	18 (20.7)
Conditional random fields	16 (18.4)
Clustering	15 (17.2)
Other	11 (12.6)
Naive Bayes classifier	5 (5.7)
K-nearest neighbors classifier	3 (3.4)
Linear regression	2 (2.3)
Neural networks (n=53)	
Recurrent neural network	34 (64.2)
Convolutional neural network	21 (39.6)
Feed forward neural networks	10 (18.9)
Capsule networks	1 (1.9)
Pretrained language models (n=41)	
BERT ^c	39 (95.1)
ChatGPT	1 (2.4)
Google Bard	1 (2.4)
Pretrained clinical models (n=23)	
Clinical BERT	23 (100)

^aNumber of model types per study: 91 (40.3%) studies used 1 model type, 92 (40.7%) studies used 2 model types, 26 (11.5%) studies used 3 model types, and 10 (4.4%) studies used 4 model types. Number of model subtypes per study: 76 (33.6%) studies used 1 model subtype, 72 (31.9%) studies used 2 model subtypes, 39 (17.3%) studies used 3 model subtypes, 25 (11.1%) studies used 4 model subtypes, and 4 (1.8%) studies used 5 model subtypes. Pretrained language models are general-domain pretrained models (eg, BERT and GPT), while pretrained clinical models are models pretrained on clinical or biomedical text (eg, BioBERT, ClinicalBERT, and PubMedBERT).

^bcTAKES: Clinical Text Analysis and Knowledge Extraction System.

^cNCBO: National Center for Biomedical Ontology.

^dRegEx: a rule-based algorithm for negation detection in clinical text.

^eBERT: Bidirectional Encoder Representations from Transformers.

Rule-based methods, including handcrafted rules and off-the-shelf tools such as clinical RegEx and ConText, were used in 112 of 226 (49.6%) studies (Table 1), making them the most prevalent, but rarely used in isolation. Rule-based approaches were used in 53 of 114 (46.5%) studies that analyzed clinical notes exclusively and in 64 of 112 (57.1%) studies that analyzed clinical notes in combination with other medical documents. Although the latter proportion was 10.6 percentage points higher, this difference was not statistically significant (2-proportion z test: $z=-1.60$; $P=.11$).

Discrete models, encompassing classical machine learning and statistical methods, were used in 87 of 226 (38.5%) studies (Table 1). The most common approaches under this category included support vector machines (29/87, 33.3%), tree-based models including random forest (28/87, 32.1%), logistic regression (18/87, 20.7%), conditional random fields (16/87, 18.4%), and clustering algorithms (15/87, 17.2%). Conditional random field was often applied as a classification layer in neural models like long short-term memory and CNN.

Neural networks featured in 53 of 226 (23.5%) studies, with RNN (34/53, 64.2%) and CNN (21/53, 39.6%), being the most popular in this category (Table 1). RNNs were dominated by long short-term memory architectures.

PLMs were used in 41 of 226 (18.1%) studies. These were primarily BERT-based models, with only 2 of the 41 (0.9%) studies [28,29] using LLMs (ChatGPT and Google Bard;

Table 1). Pretrained clinical models—BERT-based models pretrained on clinical or biomedical corpora (eg, Bio_ClinicalBERT)—were used in 23 of 226 (10.2%) studies (Table 1). Among 23 studies that implemented pretrained clinical models, 16 compared clinical domain pretrained models to general domain pretrained models. Clinical domain models outperformed general domain models in 11 of 16 (68.8%) studies, while general domain models performed better in the remaining 5 studies (Multimedia Appendix 2).

Methods for Non-English Corpora

Out of 226 studies, 70 (40%) developed models for non-English clinical notes. Of these, 59 (84.3%) implemented language-specific pipelines built from rules and classical machine learning with engineered features, including some hybrid combinations. Pretrained approaches were present but less common and not mutually exclusive across studies: language-specific pretrained models in 11 of 70 (15.7%) studies, multilingual pretrained models in 7 of 70 (10%) studies, language-specific biomedical or clinical pretrained models in 6 of 70 (8.6%) studies, and language-adapted models in 3 of 70 (4.3%) studies. Language-adapted models typically consisted of models pretrained in English and then further trained on the target language (Multimedia Appendix 3).

In total, 12 studies compared multiple model families. Language-specific biomedical or clinical pretrained models most often yielded the best performance (n=4) [30-33],

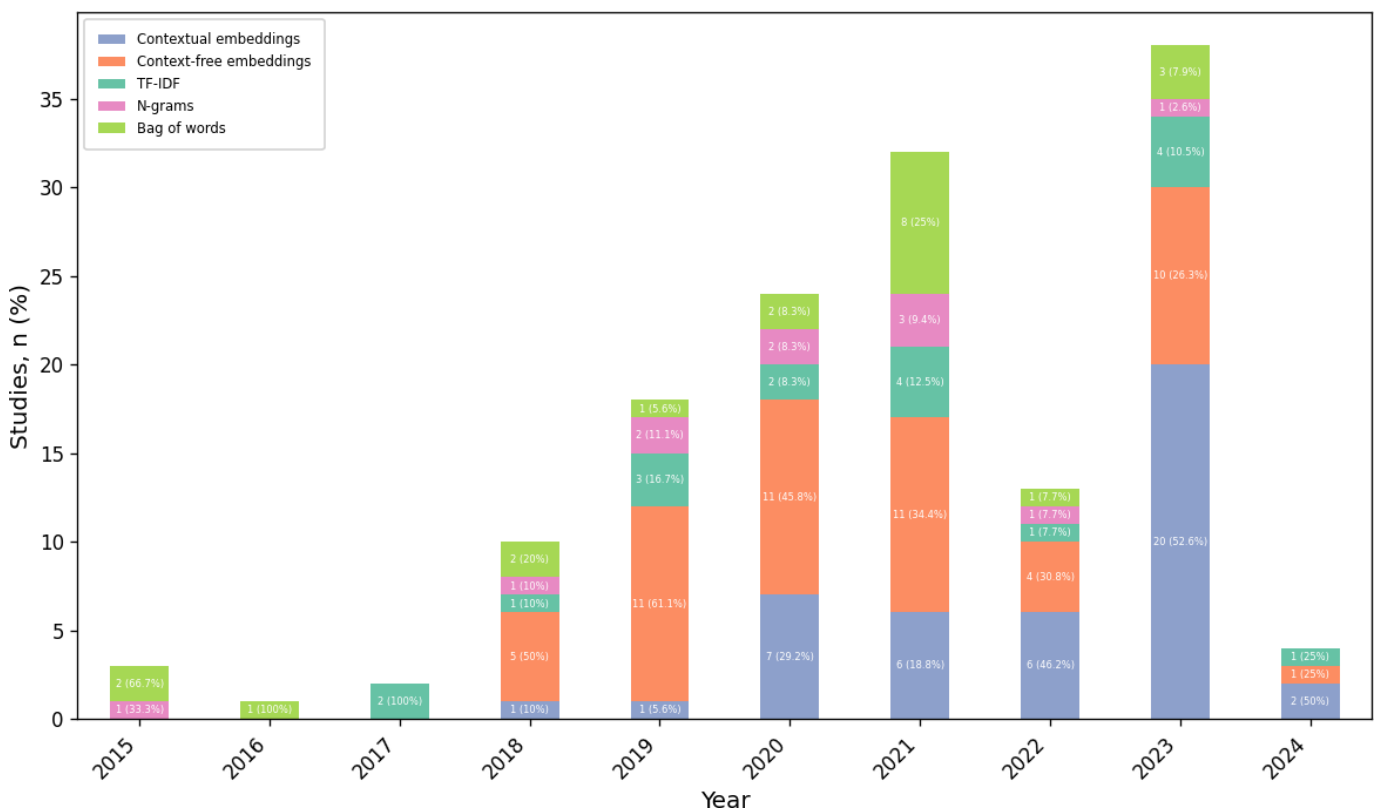
followed by language-specific pretrained models (n=3) [34-36] and language-adapted pretrained models (n=2) [37,38]. In the remaining 3 studies, the best-performing models were a biomedical or clinical pretrained model [39], a language-specific model [40], and a multilingual pretrained model [41].

Text Representation Methods

Figure 4 illustrates the text representation and vectorization methods used in the studies. Out of 226 studies, 120 (53.1%) used at least 1 representation method. From 2015 to 2017,

statistical methods including bag of words, n-grams, and term frequency-inverse document frequency were prevalent. In 2018, context-free embeddings (one fixed vector for each word or token regardless of the context in which it is used, eg, Word2Vec, GloVe, and FastText) and contextual embeddings (a new vector assigned to each word or token depending on the surrounding context, eg, BERT and GPT) were introduced and became the predominant approaches. It was common for studies to test multiple embedding methods to identify the best-performing approaches.

Figure 4. Text representation and embedding methods (n=120). Context-free embeddings include Word2Vec, FastText, and GloVe. N-grams include continuous bag of words, skip-gram, bigrams, and trigrams. Groups are not mutually exclusive—studies may appear in more than 1 category. TF-IDF: term frequency-inverse document frequency.

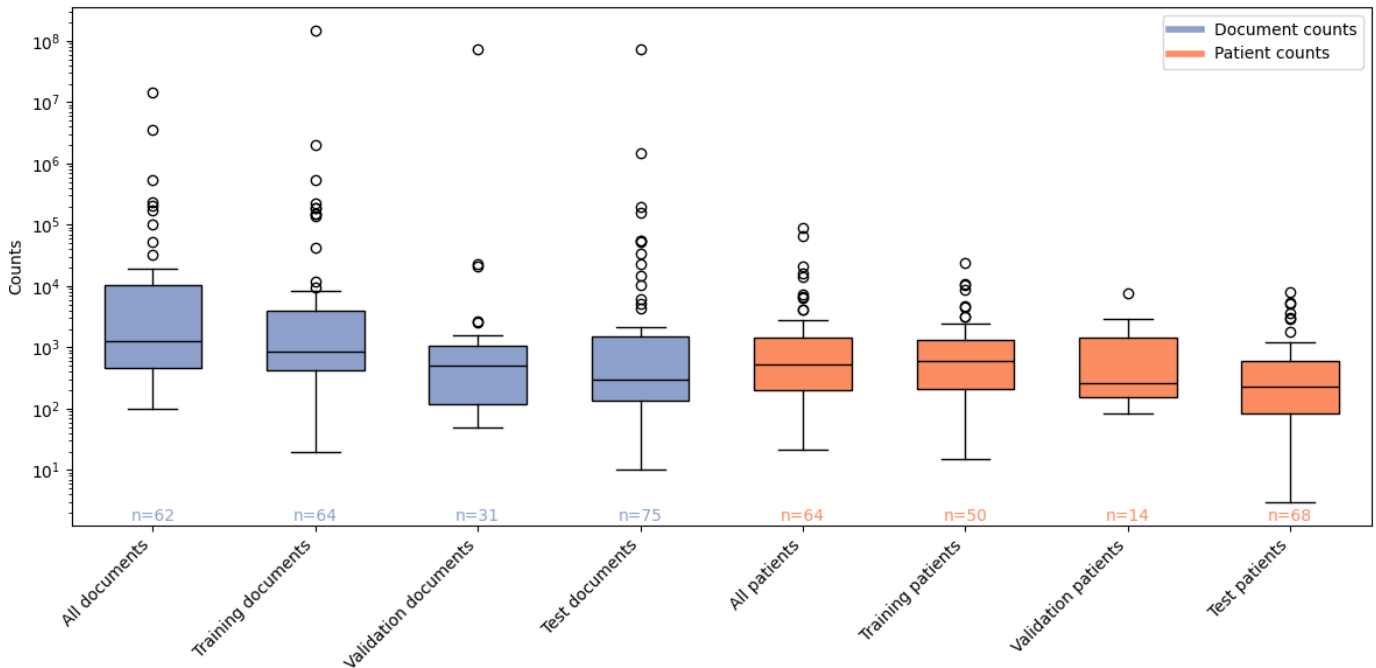


Size of Labeled Data Used to Train and Evaluate NLP Systems

Figure 5 shows the data size (clinical notes, with or without additional medical documents, and patients) used to train and evaluate NLP systems. The median number of documents per partition was fewer than 1000, and the median number of patients associated with these notes was also under 1000. For example, the median number of training documents, test documents, training patients, and test patients was 838 (IQR 439-3905), 300 (IQR 120-1504), 606 (IQR 202-1337), and 231.5 (IQR 86-599), respectively.

Training and test sets were generally created through random splits, except in 3 studies where the test cohort came from a slightly different patient population (prospective palliative radiation cohort vs metastatic cancer retrospective registry-based cohort) [42], a different but overlapping time period with the training cohort [43], a different nonoverlapping time period with the training cohort [44], or where the test cohort had a shorter follow-up time than the training cohort (4 vs 5 years) [45].

Figure 5. Size of the data used in model development and evaluation. Documents refer to entire clinical notes or reports or sentences (a small number of studies reported corpus size in sentences). To cater for instances where train or test split was not specified, we report total data sums (ie, all documents and all patients) as provided by the authors. The number below each boxplot indicates the count of studies reporting data size in that category.



Annotation Methods for Reference Corpus

The majority of studies (181/226, 80.1%) trained and evaluated their systems on corpora that were manually annotated by humans. Few studies (7/226, 3.1%) trained models using weakly supervised labels but evaluated them on human-curated labels. A considerable proportion of studies (38/226, 16.8%) either relied on existing labels within the EHR (eg, *International Classification of Diseases* or ICD codes) or developed unsupervised systems, for which manual annotation was not applicable. A summary of annotation methods is provided in [Multimedia Appendix 4](#).

Implementation Type and Evaluation

[Table 2](#) summarizes model implementation type, evaluation metrics, and whether models were externally evaluated. Most

studies (179/226, 79.2%) developed new models or retrained or fine-tuned an existing one, while 19.5% (44/226) used existing models without retraining. The latter group included studies that used off-the-shelf tools such as MetaMap or repurposed existing models for new extraction tasks.

Evaluation metrics varied by task, with the most commonly reported being recall (155/226, 68.6%), precision (153/226, 67.7%), F_1 -score (136/226, 60.2%), accuracy (44/226, 19.5%), area under the receiver operating characteristic curve (40/226, 17.7%), and specificity (30/226, 13.3%). While metrics such as recall, precision, and F_1 -score were widely used and therefore suitable for summarization, variability in clinical corpora and tasks precluded comparison on NLP methods. Only 21 of 226 (9.3%) studies evaluated their systems on external corpora.

Table 2. Model implementation and evaluation (N=226)^a.

Implementation or evaluation	Values, n (%)
Implementation type	
New model	179 (79.2)
Existing model	44 (19.5)
Reported evaluation metrics	
Recall	155 (68.6)
Precision	153 (67.7)
F_1 -score	136 (60.2)
Accuracy	44 (19.5)
AUC-ROC ^b	40 (17.7)
Specificity	30 (13.3)
Cohen κ	5 (2.2)

Implementation or evaluation	Values, n (%)
Cosine similarity	3 (1.3)
Mean average precision	2 (0.9)
Other	44 (19.5)
External evaluation	
Yes	21 (9.3)
No	158 (69.9)

^aSome studies lacked sufficient information to assess external evaluation; for example, those that used existing tools had their detailed data documented elsewhere.

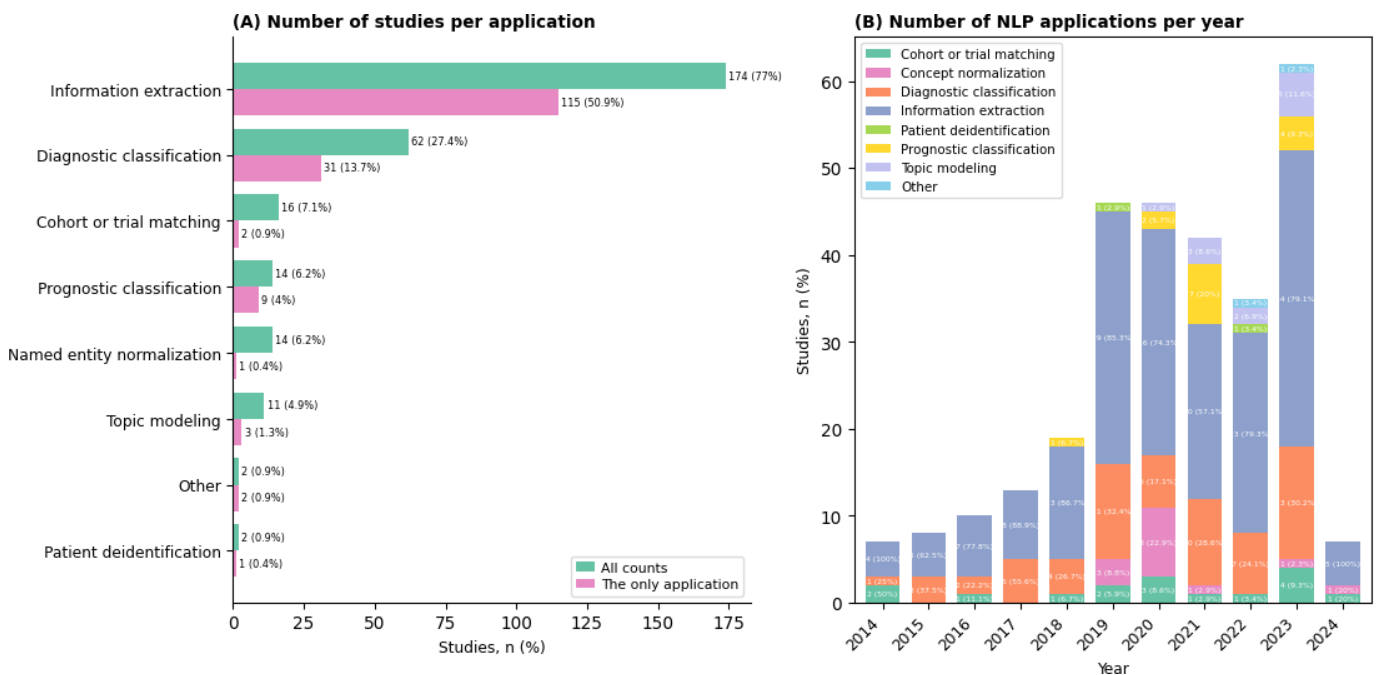
^bAUC-ROC: area under the receiver operating characteristic curve.

Clinical Applications of NLP

Figure 6A summarizes the clinical applications of NLP to clinical notes. IE was the most common task, with 77% (174/226) of the studies. In 50.9% (115/226) of the studies, NLP was exclusively used for IE. Diagnostic classification was performed in 62 of 226 (27.4%) studies, while trials or

cohort matching was the goal in 16 of 226 (7.1%) studies. Other notable applications included prognostic classification (n=14), concept normalization (n=14), and topic modeling (n=11). It was not uncommon, however, for a study to undertake multiple tasks, often with the output of one task feeding into subsequent tasks.

Figure 6. NLP clinical applications with clinical notes. (A) Number of studies per clinical application. (B) Number of clinical applications per year (percentages are relative to the number of papers published in that year). Diagnostic classification refers to document-level or patient-level classification tasks, for example, distinguishing between notes with metastasis and those without metastasis. Prognostic classification refers to predicting that some clinical event of interest will occur within a specified time period in the future, for example, lung cancer recurrence 2 years following lobectomy. NLP: natural language processing.



A subset of studies (n=15) that focused on IE also extracted temporal information. Some studies formulated this task as a document-time relation (DocTimeRel) classification, where events were assigned a temporal relation to the document creation time (before, after, overlap, or before or overlap) [46-48]. Others used an event-date relation classification formulation, classifying event-time pairs as before, after, overlap, or before or overlap [49] or directly linking events to their corresponding dates through contextual pairing [50,51]. One study constructed patient-level temporal timelines by assigning events to coarse temporal bins (way before admission, before admission, admission, after admission, and discharge) and then temporally ordered them

within and across documents [52]. Less complex approaches included proximity- or context-based methods (linking events to nearby date mentions using dependency parsing and rule-based contextual heuristics) [53-58] or simply classifying identified events into broad temporal categories such as current, history, future, or unknown [59,60].

Figure 6B shows the evolution of clinical NLP applications over time. IE remained the predominant task throughout the years, followed by diagnostic classification. Newer applications introduced after 2018 include concept normalization, prognostic classification, and topic modeling. Task chaining, where the output of one task is used as the input for

downstream tasks, was common in studies that went beyond IE. For example, in 2014, there were 4 publications of NLP applied to clinical notes. All 4 (100%) studies used NLP to extract information of some kind, 2 (50%) studies used the extracted information to match patients to clinical trials, 1 (25%) study used the extracted information for diagnostic classification, and 1 (25%) study had IE as the end point. Discrete models were almost exclusively used for these downstream tasks.

System Deployment Stage and Clinical Impact

Of the 226 reviewed studies, 224 (99.1%) developed proof-of-concept systems that were evaluated only in research settings rather than deployed in routine clinical practice. One study piloted their system in clinical practice [61], while another described the use of an NLP-based system that had just been integrated in clinical use [62].

Since most studies were evaluated only as research implementations (ie, no real-world deployments), clinical impact was not evaluated. However, 25 of the 226 (11.1%) studies compared their systems with current practice in their research implementation. These studies reported benefits

such as improved data coverage (identified more patients with the relevant attribute than from structured data alone) and completeness (curated further variables not available as structured data) [63-75], taking less time to extract relevant information [29,61,76-82], fewer clinician man-hours for certain tasks (eg, fewer clinicians needed to complete clinical audits) [29], and higher classification or prediction accuracy compared with human experts or existing methods [76,83,84]. One study that described an IE system in routine use [62] focused on characterizing use patterns, including which clinical specialties used the system and for what purposes.

Challenges and Limitations Reported by the Authors

Table 3 details challenges and limitations faced by researchers applying different NLP techniques to clinical notes. Common challenges were single-institution corpora (39/226, 17.3%), limited data (18/226, 8%), incomplete EHR data (14/226, 6.2%), label imbalance (12/226, 5.3%), rules or dictionary not comprehensive or generalizable (9/226, 4%), and word sense and abbreviation disambiguation (6/226, 2.7%). Overall, authors reported a range of challenges, some unique to the task, corpora, or methodological approach.

Table 3. Challenges and limitations reported in studies^a.

Challenge or limitation	Values, n (%)
Single institution corpus [37,42,44,61,70,72,73,75,77,79,82,84-111]	39 (17.3)
Limited data [32,50,57,61,62,65,79,90,92,104,111-117]	18 (8)
Incomplete recording in the EHR ^b [42,57,74,78,81,94,98,103,109,118-122]	14 (6.2)
Label imbalance [31,38,44,73,82,98,102,104,123-126]	12 (5.3)
Negation detection and resolution [41,74,97,119,126-131]	10 (4.4)
Dictionary or rules not comprehensive or generalizable [65,66,92,119,120,132-135]	9 (4)
Word sense or abbreviation disambiguation [48,130,131,136-138]	6 (2.7)
Variability in terminology used to describe the same concept [78,120,136,139]	4 (1.8)
Spelling errors or typos [90,130,137,140]	4 (1.8)
Imbalanced data [57,102,105,141]	4 (1.8)
Use of speculative language [117,128,136]	3 (1.3)
Use of nonstandard terminology [90,128,142]	3 (1.3)
Rarity of concepts of interest [41,45,143]	3 (1.3)
Institutional differences in documentation style or note structure [42,81,117]	3 (1.3)
Quality of human annotations [51,80]	2 (0.9)
Multilingualism in text [79,128]	2 (0.9)
Temporal reasoning (current vs historical events) [129,138]	2 (0.9)
Short notes or sentences (insufficient context for context-dependent models) [72,144]	2 (0.9)
Model computationally expensive [38]	1 (0.4)
Distant (intersentence) relations [124]	1 (0.4)
Frequency of co-occurrence of unrelated concepts [143]	1 (0.4)
Long execute-response time [145]	1 (0.4)
Very long documents (>512 token limit for BERT ^c -based models) [125]	1 (0.4)
Significant n-gram method insensitive to evolution of patient's notes over time and between patients [146]	1 (0.4)
Resolution of patient and nonpatient references [97]	1 (0.4)
Nonstandard date formats [57]	1 (0.4)

^aNegation detection and resolution includes detecting the negation itself, distant negations, and resolving the scope of the negation. Limited data encompass the following: small corpus, only a small number of patients associated with those notes, and small annotated or labeled notes for model

development and evaluation. Imbalanced data refer to instances where notes are overrepresented by text from one patient group (eg, private insurance vs noninsured). Label imbalance is when one label of interest (eg, a certain biomarker) is more prevalent in the notes, hence, easily learned by the model at the expense of other labels (biomarkers). Quality of human annotations is where human annotated corpora for model training and evaluation are erroneous.

^bEHR: electronic health record.

^cBERT: Bidirectional Encoder Representations from Transformers.

Discussion

Summary of Main Findings

Research applying NLP to clinical notes in the cancer domain grew substantially during the review period, rising from 4 publications in 2014 to 43 in 2023, likely driven by the increasing availability of digital records and advances in scalable NLP methods. However, most studies relied on English language (156/226, 69%) and single-institution (161/226, 71.2) datasets. The majority of studies originated from the United States (133/226, 58.8%), which aligns with trends in clinical NLP publishing in which the United States dominates [147]. Almost half of the studies (110/226, 48.7%) provided no information on the characteristics of patients whose clinical notes were used, while 56.6% (128/226) did not provide a statement on data sharing, limiting interpretability and reproducibility. The most commonly studied cancers (breast, lung, colorectal, and prostate) likely reflect their prevalence in the United States and hence dedicated EHR systems, which in turn increases the availability of clinical notes.

NLP methods for processing clinical notes evolved from exclusively ontology-based, rule-based, and discrete models (2014-2017) to hybrid approaches incorporating neural networks and PLMs such as BERT (2018-2024). Only a few studies applied LLMs, with publications starting from October 2023. Contextual embeddings have become increasingly prevalent, reflecting the wider adoption of pretrained models. Most studies used small single-institution datasets (<1000 documents or <1000 patients), likely due to challenges in accessing clinical notes. Annotation methods were mostly manual. A subanalysis of non-English corpora studies showed that the majority (59/70, 84.3%) implemented language-specific, nonpretrained models. Domain-specific pretrained clinical models were superior to other model types in the majority (11/16, 68.8%) of studies across both English and non-English corpora. Only 9.3% (21/226) of studies evaluated their systems on external datasets.

Most studies (174/226, 77%) focused on IE. A subset of these used the extracted information in downstream tasks, but the majority (115/226, 50.9%) focused solely on IE. In total, 15 studies extracted temporal information from clinical notes using various approaches, including DocTimeRel classification, event-time relation classification, and proximity- or context-based methods. No studies evaluated clinical impact following implementation, but several studies compared their systems to current practice in their respective settings (eg, manual review of notes in clinical audits) and demonstrated potential clinical utility. The most common challenge in

clinical NLP was restricted access to sufficient clinical notes, reported by 17.3% (39/226) of studies.

Evolution of NLP Methods for Clinical Notes

NLP methods for clinical notes have become more diverse over time. While new deep learning-based techniques have gained popularity, they have largely complemented rather than replaced traditional methods such as rules and ontologies, resulting in widespread adoption of hybrid architectures. Prior reviews that included substantial volumes of clinical notes reported similar findings, namely, the predominance of rule-based methods alongside increasing use of hybrid architectures that combine rules with machine learning or neural networks [24,25]. However, a review of NLP applied to diagnostic (radiology) reports reported slightly different findings, with rule-based and classical machine learning methods being prevalent but often used as baselines against which deep learning approaches were compared [148].

The continued use of rule-based approaches for clinical notes likely reflects the unique challenges posed by these documents, which often require substantial preprocessing before neural models can be applied, as well as postprocessing to structure model outputs into clinically meaningful formats. The overall prevalence of rule-based methods may also partly reflect the inclusion of semistructured diagnostic reports, which—owing to their templated design and restricted, domain-specific vocabulary—are generally more amenable to rule-based processing [149]. Combining knowledge resources with deep neural models, on the other hand, may reflect authors' efforts to enhance the explainability of predictions made by these complex networks, given the importance of explainability in health care AI [150,151] and evidence from prior work that integrating knowledge into deep learning may improve explainability [152].

Text representation methods have evolved alongside machine learning models. Earlier NLP approaches commonly relied on discrete word representations, such as term frequency-inverse document frequency and n-grams [153]. Our review shows that context-free word embeddings (eg, Word2Vec, GloVe, and FastText) were the most widely used, typically with classical machine learning models. The results also suggest that these approaches are increasingly being complemented or replaced by contextual embeddings derived from transformer-based models, which represent words as vectors that capture richer semantic and syntactic relationships.

Trends in NLP Clinical Applications

NLP applications to clinical notes focused predominantly on IE, accounting for over three-quarters of included studies, with comparatively limited use in downstream clinical decision-making tasks. This emphasis reflects both the pragmatic advantages and the perceived safety of IE. By structuring free-text data into clinically meaningful variables, IE enables expert oversight, produces interpretable intermediate outputs, and supports a broad range of secondary applications, including diagnostic or prognostic modeling, cohort identification, and decision support [154]. In contrast, approaches that predict outcomes directly from unstructured text without an explicit IE step are often less transparent, constrain the incorporation of domain knowledge, and are typically optimized for a single task [155].

Potential Clinical Impact of NLP

Although none of the included studies evaluated the direct clinical impact of NLP systems on patient care following research implementation, several studies compared their systems with current clinical practice as part of their evaluation. These comparisons demonstrated the potential of NLP to support tasks such as IE, clinical auditing, and diagnostic or prognostic classification. However, most studies (161/226, 71.2%) relied on small, single-institution datasets, raising concerns about generalizability, as such models often perform less well when applied to more representative or external datasets due to differences in both population characteristics and data structure. Without extensive evaluation across diverse datasets, there remains limited evidence of real-world effectiveness, thereby impeding adoption into routine clinical use.

Beyond technical performance, the application of NLP systems to high-risk tasks, such as cancer diagnosis or risk prediction, is subject to stringent regulatory oversight as medical devices [156,157]. These regulatory requirements, together with challenges in integrating NLP systems into existing clinical workflows [158], further hinder translation into routine clinical care and help explain the limited real-world impact observed across studies.

Challenges and Opportunities in Advancing Clinical NLP

Our findings indicate that restricted access to clinical data remains the dominant barrier in oncology NLP. Access to clinical corpora is complicated by multiple barriers, including national data protection regulations governing privacy and confidentiality (eg, the General Data Protection Regulation [159] in the European Union and the Health Insurance Portability and Accountability Act [160] in the United States), additional institutional governance restrictions imposed to mitigate disclosure risk and legal liability [161], and technical obstacles such as EHR interoperability [161]. This is compounded by limited data sharing practices, with many studies providing no clear data availability statement or listing data as “available on reasonable request,” a practice that often creates substantial practical barriers, including low response rates and protracted negotiations that effectively limit access.

As a result, researchers have to rely on small, single-institution datasets, resulting in proof-of-concept systems with limited generalizability.

Limited data accessibility undermines reproducibility, hinders meaningful comparison across studies, prevents the establishment of standardized benchmarks for performance evaluation, and reinforces reliance on small, single-institution datasets. Collectively, these challenges derail real-world deployment of clinical NLP systems.

Several methodological approaches have attempted to mitigate data scarcity, each with notable limitations. Transfer learning through clinical PLMs (eg, ClinicalBERT) is constrained by training on relatively small and institutionally narrow corpora, reflecting the same access limitations they aim to overcome, which can result in suboptimal performance on downstream tasks [162,163]. Publicly available deidentified datasets curated for clinical NLP shared tasks (eg, Cancer Text Mining Shared Task [164]) face similar limitations, being small and single-center.

More recently, LLMs have shown promise in mitigating data scarcity by enabling zero-shot or few-shot learning, thereby reducing dependence on large, manually annotated corpora [165,166]. However, LLMs introduce additional challenges, including the propagation of embedded biases [167], privacy breaches [168], model obsolescence and drift [168], hallucination and confidently stated falsehoods [169,170], and substantial computational and environmental costs. These shortcomings can be detrimental to clinical practice, for example, by systematically underrecommending investigations, procedures, or treatments for underrepresented patient groups. Therefore, research on LLMs should also focus on addressing these ethical concerns in addition to technical performance and generalizability.

Model-centric privacy-preserving approaches, such as federated learning, where models are trained locally and aggregated without sharing raw data [171], offer a potential pathway toward multi-institutional collaboration without direct data transfer. However, practical deployment remains challenging, requiring compatible infrastructure, sustained institutional partnerships, and strategies to manage data heterogeneity and site imbalance, which can bias global models toward dominant contributors and degrade performance for underrepresented populations [172]. Related techniques, such as differential privacy, may further reduce reidentification risk but introduce trade-offs between privacy protection and model utility that must be carefully managed [173].

Beyond algorithmic solutions, structural and policy-level interventions are likely to be critical. National initiatives, such as those implemented in Denmark, where clinical notes are rigorously deidentified and made accessible within secure research environments [69], demonstrate the feasibility of balancing privacy protection with research utility. Broader adoption of such frameworks, alongside clearer institutional agreements that permit sharing of rigorously deidentified clinical text and accompanying code, could substantially improve reproducibility and accelerate progress in oncology

NLP. Furthermore, to support open science in oncology, NLP future studies should adopt more transparent reporting of data access conditions, and where feasible, publicly release the code, alongside clear governance mechanisms to balance reproducibility with patient privacy.

Limitations of the Review

This review has several limitations. First, approximately half of the included studies analyzed clinical notes alongside more structured medical documents such as pathology or radiology reports. These document types differ substantially in linguistic complexity, with diagnostic reports often being more templated and semistructured compared to free-text clinical notes such as progress notes or discharge summaries. As a result, the NLP methods and challenges reported in such studies may not be fully representative of those encountered when analyzing highly unstructured clinical narratives.

Second, we were unable to determine the proportion of clinical notes versus other document types in each study, as this was rarely reported. While we distinguished document types where possible, inconsistent reporting limited further quantification of these documents. Consequently, our findings reflect the broader landscape of clinical text processing in oncology rather than exclusively characterizing NLP applied to highly unstructured clinical notes. Nonetheless, we provide a more faithful representation of pregenerative AI methodological choices and challenges associated with clinical notes, as all included studies incorporated clinical notes. In addition, we could not systematically compare model performance across studies due to substantial heterogeneity in corpora and NLP tasks.

Third, the predominance of studies authored by researchers from the United States (133/226, 58.8%), primarily using local datasets, may have introduced some geographical and system-level bias. Our findings are therefore more reflective of the US health care context, including workflows,

documentation styles, clinical note structures, and data access provisions.

Finally, Cohen κ for title or abstract screening (0.54) and full-text screening (0.58) indicated moderate interrater agreement. This primarily reflects challenges in operationalizing eligibility criteria. In particular, disagreement frequently arose from ambiguity in how studies described their textual data sources, as some authors used the term “clinical notes” broadly to refer to any textual medical document, including diagnostic reports. This was exacerbated by limited methodological detail in the abstract, making it difficult to determine whether clinical notes were included. Despite this, class-specific agreement for exclusions at title or abstract screening was high (97.9%), while agreement for included studies improved substantially at full-text screening (86.3%) once detailed information was available. The moderate κ values could therefore be partly attributed to class imbalance inherent to evidence synthesis, as most records are excluded at the title or abstract, and κ adjusts for agreement expected by chance.

Conclusions

This review establishes a comprehensive pregenerative AI baseline for NLP applied to clinical notes in oncology. Over the past decade, research volume increased substantially, and methods evolved from rule-based approaches to hybrid architectures incorporating rules and neural networks, including PLMs. However, most studies focused on IE rather than diagnosis or prognostication, relied on small single-institution datasets, and lacked external validation. While several systems demonstrated superior performance compared to current practice in research settings, significant barriers to clinical deployment remain, including limited generalizability, poor reproducibility, and restricted data access. Emerging generative AI approaches will need to address these barriers, as well as broader ethical challenges, to enable the translation of NLP systems into clinical settings for real-world impact.

Acknowledgments

The authors thank Paula Funnell (Academic Skills and Liaison Librarian, Faculty of Medicine and Dentistry, Queen Mary University of London, Whitechapel Campus) for her assistance in developing the search strategy. The authors used ChatGPT (a generative artificial intelligence tool developed by OpenAI) to refine the Python code used to plot figures and to edit selected sections of the manuscript to improve grammar, sentence structure, and brevity. All outputs (code and text) were checked and, where necessary, revised by the authors.

Funding

This study was conducted without any funding. However, the first author (ABK) completed this study as part of his PhD funded by the Wellcome Trust through the Health Data in Practice Doctoral Training Programme at Queen Mary University of London (grant 218584/Z/19/Z).

Data Availability

All data generated and analyzed during this study are included in this published paper as [Multimedia Appendix 2](#).

Authors' Contributions

ABK conceptualized and designed the study under the supervision of GF. KL, HRAE, FMW, and CC reviewed the study methodology. ABK performed the database searches and reference retrieval. ABK and HRAE completed the title or abstract screening, full-text screening, and data extraction, and analyzed and interpreted the data. ABK drafted the manuscript, and

GF, HRAE, KL, FMW, and CC reviewed the draft. ABK revised the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search criteria.

[DOCX File (Microsoft Word File), 17 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Studies included in the review and variables extracted.

[XLSX File (Microsoft Excel File), 69 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Models for non-English corpora.

[PNG File (Portable Network Graphics File), 125 KB-Multimedia Appendix 3]

Multimedia Appendix 4

Annotation methods for reference corpus. Annotation granularity ranged from the entity or concept level to the patient level, including sentence, document section, and document levels. No information: no description of annotation methods (studies that used existing tools, detailed methods described elsewhere).

[PNG File (Portable Network Graphics File), 76 KB-Multimedia Appendix 4]

Checklist 1

PRISMA-ScR checklist.

[PDF File (Adobe File), 185 KB-Checklist 1]

References

1. GLOBOCAN 2020: new global cancer data. UICC. URL: <https://www.uicc.org/news/globocan-2020-new-global-cancer-data> [Accessed 2023-12-19]
2. Worldwide cancer incidence statistics. Cancer Research UK. URL: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/worldwide-cancer/incidence#heading-One> [Accessed 2023-12-19]
3. Kim E, Rubinstein SM, Nead KT, Wojcieszynski AP, Gabriel PE, Warner JL. The evolving use of electronic health records (EHR) for research. *Semin Radiat Oncol*. Oct 2019;29(4):354-361. [doi: [10.1016/j.semradonc.2019.05.010](https://doi.org/10.1016/j.semradonc.2019.05.010)] [Medline: [31472738](https://pubmed.ncbi.nlm.nih.gov/31472738/)]
4. Structured vs unstructured data in healthcare. HealthTech. URL: <https://healthtechmagazine.net/article/2023/05/structured-vs-unstructured-data-in-healthcare-perfcon#:~:text=Bring%20order%20to%20unstructured%20data,two%20dozen%20ICD-10%20codes> [Accessed 2025-01-07]
5. Tayefi M, Ngo P, Chomutare T, et al. Challenges and opportunities beyond structured data in analysis of electronic health records. *WIREs Comput Stats*. Nov 2021;13(6). [doi: [10.1002/wics.1549](https://doi.org/10.1002/wics.1549)]
6. Meystre SM, Savova GK, Kipper-Schuler KC, Hurdle JF. Extracting information from textual documents in the electronic health record: a review of recent research. *Yearb Med Inform*. 2008;17:128-144. [doi: [10.1055/s-0038-1638592](https://doi.org/10.1055/s-0038-1638592)] [Medline: [18660887](https://pubmed.ncbi.nlm.nih.gov/18660887/)]
7. Perera S, Sheth A, Thirunarayan K, et al. Challenges in understanding clinical notes: why NLP engines fall short and where background knowledge can help. Presented at: International Conference on Information and Knowledge Management, Proceedings; Nov 3-7, 2013; Shanghai, China. [doi: [10.1145/2512410.2512427](https://doi.org/10.1145/2512410.2512427)]
8. Madan S, Lentzen M, Brandt J, Rueckert D, Hofmann-Apitius M, Fröhlich H. Transformer models in biomedicine. *BMC Med Inform Decis Mak*. Jul 29, 2024;24(1):214. [doi: [10.1186/s12911-024-02600-5](https://doi.org/10.1186/s12911-024-02600-5)] [Medline: [39075407](https://pubmed.ncbi.nlm.nih.gov/39075407/)]
9. Klotzman V, et al. The difficulties of clinical NLP. In: Kunze H, Torre D, Riccoboni A, editors. *Engineering Mathematics and Artificial Intelligence: Foundations, Methods, and Applications*. CRC Press; 2023:413-423. [doi: [10.1201/9781003283980-17](https://doi.org/10.1201/9781003283980-17)] ISBN: 9781032255675
10. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. Presented at: NAACL HLT 2019—2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies—Proceedings of the Conference; Jun 2-7, 2019; Minneapolis, MN, United States. [doi: [10.18653/v1/N19-1423](https://doi.org/10.18653/v1/N19-1423)]

11. Brown TB, Mann B, Ryder N, et al. Language models are few-shot learners. Presented at: 34th Conference on Neural Information Processing Systems (NeurIPS 2020); Dec 8-10, 2020. URL: <https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfc4967418bfb8ac142f64a-Paper.pdf> [Accessed 2026-04-24]
12. Touvron H, Lavril T, Izacard G, et al. LLaMA: open and efficient foundation language models. Preprint posted online on Feb 27, 2023. [doi: [10.48550/arXiv.2302.13971](https://doi.org/10.48550/arXiv.2302.13971)]
13. Tariq A, Sikha M, Kurian AW, et al. Open-source hybrid large language model integrated system for extraction of breast cancer treatment pathway from free-text clinical notes. *JCO Clin Cancer Inform*. Jun 2025;9(9):e2500002. [doi: [10.1200/CCI-25-00002](https://doi.org/10.1200/CCI-25-00002)] [Medline: [40577660](https://pubmed.ncbi.nlm.nih.gov/40577660/)]
14. Gu Z, He L, Naeem A, et al. SBDH-Reader: a large language model-powered method for extracting social and behavioral determinants of health from clinical notes. *J Am Med Inform Assoc*. Oct 1, 2025;32(10):1570-1580. [doi: [10.1093/jamia/ocaf124](https://doi.org/10.1093/jamia/ocaf124)]
15. Kaster L, Hillis E, Oh IY, et al. Comparison of rule- and large language model-based phenotype extraction from clinical notes for neurofibromatosis type 1. *J Am Med Inform Assoc*. Nov 1, 2025;32(11):1663-1673. [doi: [10.1093/jamia/ocaf155](https://doi.org/10.1093/jamia/ocaf155)]
16. Chen D, Alnassar SA, Avison KE, Huang RS, Raman S. Large language model applications for health information extraction in oncology: scoping review. *JMIR Cancer*. Mar 28, 2025;11:e65984. [doi: [10.2196/65984](https://doi.org/10.2196/65984)] [Medline: [40153782](https://pubmed.ncbi.nlm.nih.gov/40153782/)]
17. Zhong R, Chen S, Li Z, et al. Large language models in lung cancer: systematic review. *J Med Internet Res*. Sep 30, 2025;27:e74177. [doi: [10.2196/74177](https://doi.org/10.2196/74177)] [Medline: [41026980](https://pubmed.ncbi.nlm.nih.gov/41026980/)]
18. Hao Y, Qiu Z, Holmes J, et al. Large language model integrations in cancer decision-making: a systematic review and meta-analysis. *NPJ Digit Med*. Jul 17, 2025;8(1):450. [doi: [10.1038/s41746-025-01824-7](https://doi.org/10.1038/s41746-025-01824-7)] [Medline: [40676129](https://pubmed.ncbi.nlm.nih.gov/40676129/)]
19. Wang L, Fu S, Wen A, et al. Assessment of electronic health record for cancer research and patient care through a scoping review of cancer natural language processing. *JCO Clin Cancer Inform*. Jul 2022;6:e2200006. [doi: [10.1200/CCI.22.00006](https://doi.org/10.1200/CCI.22.00006)] [Medline: [35917480](https://pubmed.ncbi.nlm.nih.gov/35917480/)]
20. Li C, Zhang Y, Weng Y, Wang B, Li Z. Natural language processing applications for computer-aided diagnosis in oncology. *Diagnostics (Basel)*. Jan 12, 2023;13(2):286. [doi: [10.3390/diagnostics13020286](https://doi.org/10.3390/diagnostics13020286)] [Medline: [36673096](https://pubmed.ncbi.nlm.nih.gov/36673096/)]
21. Gholipour M, Khajouei R, Amiri P, Hajesmael Gohari S, Ahmadian L. Extracting cancer concepts from clinical notes using natural language processing: a systematic review. *BMC Bioinformatics*. Oct 29, 2023;24(1):405. [doi: [10.1186/s12859-023-05480-0](https://doi.org/10.1186/s12859-023-05480-0)] [Medline: [37898795](https://pubmed.ncbi.nlm.nih.gov/37898795/)]
22. Sangariyavanich E, Ponthongmak W, Tansawet A, et al. Systematic review of natural language processing for recurrent cancer detection from electronic medical records. *Inform Med Unlocked*. 2023;41:101326. [doi: [10.1016/j.imu.2023.101326](https://doi.org/10.1016/j.imu.2023.101326)]
23. Wang Y, Wang L, Rastegar-Mojarad M, et al. Clinical information extraction applications: a literature review. *J Biomed Inform*. Jan 2018;77:34-49. [doi: [10.1016/j.jbi.2017.11.011](https://doi.org/10.1016/j.jbi.2017.11.011)] [Medline: [29162496](https://pubmed.ncbi.nlm.nih.gov/29162496/)]
24. Sim JA, Huang X, Horan MR, et al. Natural language processing with machine learning methods to analyze unstructured patient-reported outcomes derived from electronic health records: a systematic review. *Artif Intell Med*. Dec 2023;146:102701. [doi: [10.1016/j.artmed.2023.102701](https://doi.org/10.1016/j.artmed.2023.102701)] [Medline: [38042599](https://pubmed.ncbi.nlm.nih.gov/38042599/)]
25. Sheikhalishahi S, Miotto R, Dudley JT, Lavelli A, Rinaldi F, Osmani V. Natural language processing of clinical notes on chronic diseases: systematic review. *JMIR Med Inform*. Apr 27, 2019;7(2):e12239. [doi: [10.2196/12239](https://doi.org/10.2196/12239)] [Medline: [31066697](https://pubmed.ncbi.nlm.nih.gov/31066697/)]
26. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med*. Oct 2, 2018;169(7):467-473. [doi: [10.7326/M18-0850](https://doi.org/10.7326/M18-0850)] [Medline: [30178033](https://pubmed.ncbi.nlm.nih.gov/30178033/)]
27. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med Res Methodol*. Nov 19, 2018;18(1):143. [doi: [10.1186/s12874-018-0611-x](https://doi.org/10.1186/s12874-018-0611-x)] [Medline: [30453902](https://pubmed.ncbi.nlm.nih.gov/30453902/)]
28. Sultan I, Al-Abdallat H, Alnajjar Z, et al. Using ChatGPT to predict cancer predisposition genes: a promising tool for pediatric oncologists. *Cureus*. Oct 2023;15(10):e47594. [doi: [10.7759/cureus.47594](https://doi.org/10.7759/cureus.47594)] [Medline: [38021917](https://pubmed.ncbi.nlm.nih.gov/38021917/)]
29. McGowan M, Correia Martins F, Keen JL, et al. Can natural language processing be effectively applied for audit data analysis in gynaecological oncology at a UK cancer centre? *Int J Med Inform*. Feb 2024;182:105306. [doi: [10.1016/j.ijmedinf.2023.105306](https://doi.org/10.1016/j.ijmedinf.2023.105306)] [Medline: [38065003](https://pubmed.ncbi.nlm.nih.gov/38065003/)]
30. Solarte-Pabon O, Blazquez-Herranz A, Torrente M, Rodriguez-Gonzalez A, Provencio M, Menasalvas E. Extracting cancer treatments from clinical text written in Spanish: a deep learning approach. Presented at: 2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA); Oct 6-9, 2021; Porto, Portugal. 2021. [doi: [10.1109/DSAA53316.2021.9564137](https://doi.org/10.1109/DSAA53316.2021.9564137)]

31. Paolo D, Bria A, Greco C, et al. Named entity recognition in Italian lung cancer clinical reports using transformers. Presented at: 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); Dec 5-8, 2023:2023; Istanbul, Turkiye. 2023. [doi: [10.1109/BIBM58861.2023.10385778](https://doi.org/10.1109/BIBM58861.2023.10385778)]
32. Zhang X, Zhang Y, Zhang Q, et al. Extracting comprehensive clinical information for breast cancer using deep learning methods. *Int J Med Inform.* Dec 2019;132:103985. [doi: [10.1016/j.ijmedinf.2019.103985](https://doi.org/10.1016/j.ijmedinf.2019.103985)] [Medline: [31627032](https://pubmed.ncbi.nlm.nih.gov/31627032/)]
33. Rivera-Zavala R, Martinez P. Deep neural model with contextualized-word embeddings for named entity recognition in Spanish clinical text. Presented at: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020) CEUR Workshop Proceedings (CEUR-WS.org); Sep 22-24, 2024; Málaga, Spain.
34. Zelina P, Halámková J, Nováček V. Extraction, labeling, clustering, and semantic mapping of segments from clinical notes. *IEEE Transon Nanobioscience.* 2023;22(4):781-788. [doi: [10.1109/TNB.2023.3275195](https://doi.org/10.1109/TNB.2023.3275195)]
35. Araki K, Matsumoto N, Togo K, et al. Developing artificial intelligence models for extracting oncologic outcomes from Japanese electronic health records. *Adv Ther.* Mar 2023;40(3):934-950. [doi: [10.1007/s12325-022-02397-7](https://doi.org/10.1007/s12325-022-02397-7)] [Medline: [36547809](https://pubmed.ncbi.nlm.nih.gov/36547809/)]
36. García-Pablos A, Perez N. Vicomtech at CANTEMIST 2020. Presented at: Proceedings of the Iberian Languages Evaluation Forum (IberLEF2020) CEUR Workshop Proceedings (CEUR-WS.org); Sep 22-24, 2020; Málaga, Spain.
37. Karlsson A, Ellonen A, Irjala H, et al. Impact of deep learning-determined smoking status on mortality of cancer patients: never too late to quit. *ESMO Open.* Jun 2021;6(3):100175. [doi: [10.1016/j.esmooop.2021.100175](https://doi.org/10.1016/j.esmooop.2021.100175)] [Medline: [34091262](https://pubmed.ncbi.nlm.nih.gov/34091262/)]
38. Chapman K, Neumann G. Automatic ICD code classification with label description attention mechanism. Presented at: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020) CEUR Workshop Proceedings (CEUR-WS.org); Sep 22-24, 2020; Málaga, Spain.
39. Solarte-Pabón O, Montenegro O, García-Barragán A, et al. Transformers for extracting breast cancer information from Spanish clinical narratives. *Artif Intell Med.* Sep 2023;143:102625. [doi: [10.1016/j.artmed.2023.102625](https://doi.org/10.1016/j.artmed.2023.102625)] [Medline: [37673566](https://pubmed.ncbi.nlm.nih.gov/37673566/)]
40. Osborne JD, O'leary T, Del MJ, et al. Identification of cancer entities in clinical text combining transformers with dictionary features. Presented at: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020) CEUR Workshop Proceedings (CEUR-WS.org); Sep 22-24, 2020; Málaga, Spain.
41. Solarte Pabón O, Montenegro O, Torrente M, Rodríguez González A, Provencio M, Menasalvas E. Negation and uncertainty detection in clinical texts written in Spanish: a deep learning-based approach. *PeerJ Comput Sci.* 2022;8:e913. [doi: [10.7717/peerj-cs.913](https://doi.org/10.7717/peerj-cs.913)]
42. Banerjee I, Gensheimer MF, Wood DJ, et al. Probabilistic prognostic estimates of survival in metastatic cancer patients (PPES-Met) utilizing free-text clinical narratives. *Sci Rep.* Jul 3, 2018;8(1):10037. [doi: [10.1038/s41598-018-27946-5](https://doi.org/10.1038/s41598-018-27946-5)] [Medline: [29968730](https://pubmed.ncbi.nlm.nih.gov/29968730/)]
43. Gray SW, Ottesen RA, Currey M, et al. Leveraging an informatics approach to identify an unmet clinical need for *BRCA1/2* testing among patients with ovarian cancer. *JCO Clin Cancer Inform.* Sep 2022;6:e2200034. [doi: [10.1200/CCI.22.00034](https://doi.org/10.1200/CCI.22.00034)] [Medline: [36049148](https://pubmed.ncbi.nlm.nih.gov/36049148/)]
44. Kaka H, Michalopoulos G, Subendran S, et al. Pretrained neural networks accurately identify cancer recurrence in medical record. *Stud Health Technol Inform.* May 25, 2022;294:93-97. [doi: [10.3233/SHTI220403](https://doi.org/10.3233/SHTI220403)] [Medline: [35612023](https://pubmed.ncbi.nlm.nih.gov/35612023/)]
45. Banerjee I, Bozkurt S, Caswell-Jin JL, Kurian AW, Rubin DL. Natural language processing approaches to detect the timeline of metastatic recurrence of breast cancer. *JCO Clin Cancer Inform.* Oct 2019;3:1-12. [doi: [10.1200/CCI.19.00034](https://doi.org/10.1200/CCI.19.00034)] [Medline: [31584836](https://pubmed.ncbi.nlm.nih.gov/31584836/)]
46. Velupillai S, Mowery DL, Abdelrahman S, Christensen L, Chapman W. BluLab: temporal information extraction for the 2015 clinical tempeval challenge. Presented at: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015); Jun 4-5, 2015; Denver, CO, United States. [doi: [10.18653/v1/S15-2137](https://doi.org/10.18653/v1/S15-2137)]
47. Miller T, Laparra E, Bethard S, et al. Domain adaptation in practice: lessons from a real-world information extraction pipeline. In: Ben-David E, Cohen S, McDonald R, editors. Presented at: Proceedings of the Second Workshop on Domain Adaptation for NLP; Aug 1, 2019; Florence, Italy.
48. Hong J, Davoudi A, Yu S, Mowery DL. Annotation and extraction of age and temporally-related events from clinical histories. *BMC Med Inform Decis Mak.* Dec 30, 2020;20(Suppl 11):338. [doi: [10.1186/s12911-020-01333-5](https://doi.org/10.1186/s12911-020-01333-5)] [Medline: [33380319](https://pubmed.ncbi.nlm.nih.gov/33380319/)]
49. Li Z, Li C, Long Y, Wang X. A system for automatically extracting clinical events with temporal information. *BMC Med Inform Decis Mak.* Dec 2020;20(1):1-13. [doi: [10.1186/s12911-020-01208-9](https://doi.org/10.1186/s12911-020-01208-9)]
50. Bitterman DS, Goldner E, Finan S, et al. An end-to-end natural language processing system for automatically extracting radiation therapy events from clinical texts. *Int J Radiat Oncol Biol Phys.* Sep 1, 2023;117(1):262-273. [doi: [10.1016/j.ijrobp.2023.03.055](https://doi.org/10.1016/j.ijrobp.2023.03.055)] [Medline: [36990288](https://pubmed.ncbi.nlm.nih.gov/36990288/)]

51. Adamson B, Waskom M, Blarre A, et al. Approach to machine learning for extraction of real-world data variables from electronic health records. *Front Pharmacol*. 2023;14:1180962. [doi: [10.3389/fphar.2023.1180962](https://doi.org/10.3389/fphar.2023.1180962)] [Medline: [37781703](https://pubmed.ncbi.nlm.nih.gov/37781703/)]
52. Raghavan P, Chen JL, Fosler-Lussier E, Lai AM. How essential are unstructured clinical narratives and information fusion to clinical trial recruitment? *AMIA Jt Summits Transl Sci Proc*. 2014;2014(218):218-223. [Medline: [25717416](https://pubmed.ncbi.nlm.nih.gov/25717416/)]
53. Solarte Pabón O, Torrente M, Provencio M, Rodríguez-Gonzalez A, Menasalvas E. Integrating speculation detection and deep learning to extract lung cancer diagnosis from clinical notes. *Appl Sci (Basel)*. 2021;11(2):865. [doi: [10.3390/app11020865](https://doi.org/10.3390/app11020865)]
54. Guin S, Jun T, Patel VG, et al. Extraction of treatment information from electronic health records and evaluation of testosterone recovery in patients with prostate cancer. *JCO Clin Cancer Inform*. Jun 2022;6:e2200010. [doi: [10.1200/CCI.22.00010](https://doi.org/10.1200/CCI.22.00010)] [Medline: [35696627](https://pubmed.ncbi.nlm.nih.gov/35696627/)]
55. Najafabadipour M, Zanin M, Rodríguez-González A, et al. Reconstructing the patient's natural history from electronic health records. *Artif Intell Med*. May 2020;105:101860. [doi: [10.1016/j.artmed.2020.101860](https://doi.org/10.1016/j.artmed.2020.101860)] [Medline: [32505419](https://pubmed.ncbi.nlm.nih.gov/32505419/)]
56. Wang L, Wampfler J, Dispenzieri A, Xu H, Yang P, Liu H. Achievability to extract specific date information for cancer research. *AMIA Annu Symp Proc*. 2019;2019(893):893-902. [Medline: [32308886](https://pubmed.ncbi.nlm.nih.gov/32308886/)]
57. Fu JT, Sholle E, Krichevsky S, Scandura J, Campion TR. Extracting and classifying diagnosis dates from clinical notes: a case study. *J Biomed Inform*. Oct 2020;110:103569. [doi: [10.1016/j.jbi.2020.103569](https://doi.org/10.1016/j.jbi.2020.103569)] [Medline: [32949781](https://pubmed.ncbi.nlm.nih.gov/32949781/)]
58. Solarte-Pabon O, Torrente M, Rodriguez-Gonzalez A, Provencio M, Menasalvas E, Tunas JM. Lung cancer diagnosis extraction from clinical notes written in Spanish. Presented at: 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS); Jul 28-30, 2020; Rochester, MN, United States. [doi: [10.1109/CBMS49503.2020.00099](https://doi.org/10.1109/CBMS49503.2020.00099)]
59. Rumeng L, Abhyuday N J, Hong Y. A hybrid neural network model for joint prediction of presence and period assertions of medical events in clinical notes. *AMIA Annu Symp Proc*. 2017;2017(1149):1149-1158. [Medline: [29854183](https://pubmed.ncbi.nlm.nih.gov/29854183/)]
60. Palmer EL, Hassanpour S, Higgins J, Doherty JA, Onega T. Building a tobacco user registry by extracting multiple smoking behaviors from clinical notes. *BMC Med Inform Decis Mak*. Jul 25, 2019;19(1):141. [doi: [10.1186/s12911-019-0863-3](https://doi.org/10.1186/s12911-019-0863-3)] [Medline: [31340796](https://pubmed.ncbi.nlm.nih.gov/31340796/)]
61. Yu S, Le A, Feld E, et al. A natural language processing-assisted extraction system for Gleason scores: development and usability study. *JMIR Cancer*. Jul 2, 2021;7(3):e27970. [doi: [10.2196/27970](https://doi.org/10.2196/27970)] [Medline: [34255641](https://pubmed.ncbi.nlm.nih.gov/34255641/)]
62. Biron P, Metzger MH, Pezet C, Sebban C, Barthuet E, Durand T. An information retrieval system for computerized patient records in the context of a daily hospital practice: the example of the Léon Bérard Cancer Center (France). *Appl Clin Inform*. 2014;5(1):191-205. [doi: [10.4338/ACI-2013-08-CR-0065](https://doi.org/10.4338/ACI-2013-08-CR-0065)] [Medline: [24734133](https://pubmed.ncbi.nlm.nih.gov/24734133/)]
63. Zhu W, Teh JB, Li H, Armenian SH. Knowledge extraction of long-term complications from clinical narratives of blood cancer patients with HCT treatments. Presented at: BCB '18: Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics; Aug 29 to Sep 1, 2018; Washington, DC, United States. Aug 15, 2018.[doi: [10.1145/3233547.3233635](https://doi.org/10.1145/3233547.3233635)]
64. Osborne JD, Wyatt M, Westfall AO, Willig J, Bethard S, Gordon G. Efficient identification of nationally mandated reportable cancer cases using natural language processing and machine learning. *J Am Med Inform Assoc*. Nov 2016;23(6):1077-1084. [doi: [10.1093/jamia/ocw006](https://doi.org/10.1093/jamia/ocw006)] [Medline: [27026618](https://pubmed.ncbi.nlm.nih.gov/27026618/)]
65. Cohen AB, Rosic A, Harrison K, et al. A natural language processing algorithm to improve completeness of ECOG performance status in real-world data. *Appl Sci (Basel)*. 2023;13(10):6209. [doi: [10.3390/app13106209](https://doi.org/10.3390/app13106209)]
66. Tamang S, Patel MI, Blayney DW, et al. Detecting unplanned care from clinician notes in electronic health records. *J Oncol Pract*. May 2015;11(3):e313-9. [doi: [10.1200/JOP.2014.002741](https://doi.org/10.1200/JOP.2014.002741)] [Medline: [25980019](https://pubmed.ncbi.nlm.nih.gov/25980019/)]
67. 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*. Apr 2021;5(5):469-478. [doi: [10.1200/CCI.20.00165](https://doi.org/10.1200/CCI.20.00165)] [Medline: [33929889](https://pubmed.ncbi.nlm.nih.gov/33929889/)]
68. Hernandez-Boussard T, Tamang S, Blayney D, Brooks J, Shah N. New paradigms for patient-centered outcomes research in electronic medical records: an example of detecting urinary incontinence following prostatectomy. *EGEMS (Wash DC)*. 2016;4(3):1231. [doi: [10.13063/2327-9214.1231](https://doi.org/10.13063/2327-9214.1231)] [Medline: [27347492](https://pubmed.ncbi.nlm.nih.gov/27347492/)]
69. Hjaltelin JX, Novitski SI, Jørgensen IF, et al. Pancreatic cancer symptom trajectories from Danish registry data and free text in electronic health records. *Elife*. Nov 21, 2023;12:e84919. [doi: [10.7554/eLife.84919](https://doi.org/10.7554/eLife.84919)] [Medline: [37988407](https://pubmed.ncbi.nlm.nih.gov/37988407/)]
70. Wang L, Ruan X, Yang P, Liu H. Comparison of three information sources for smoking information in electronic health records. *Cancer Inform*. 2016;15:237-242. [doi: [10.4137/CIN.S40604](https://doi.org/10.4137/CIN.S40604)] [Medline: [27980387](https://pubmed.ncbi.nlm.nih.gov/27980387/)]
71. Prado MG, Kessler LG, Au MA, et al. Symptoms and signs of lung cancer prior to diagnosis: case-control study using electronic health records from ambulatory care within a large US-based tertiary care centre. *BMJ Open*. Apr 20, 2023;13(4):e068832. [doi: [10.1136/bmjopen-2022-068832](https://doi.org/10.1136/bmjopen-2022-068832)] [Medline: [37080616](https://pubmed.ncbi.nlm.nih.gov/37080616/)]

72. Shi J, Morgan KL, Bradshaw RL, et al. Identifying patients who meet criteria for genetic testing of hereditary cancers based on structured and unstructured family health history data in the electronic health record: natural language processing approach. *JMIR Med Inform.* Aug 11, 2022;10(8):e37842. [doi: [10.2196/37842](https://doi.org/10.2196/37842)] [Medline: [35969459](https://pubmed.ncbi.nlm.nih.gov/35969459/)]
73. Bozkurt S, Magnani CJ, Seneviratne MG, Brooks JD, Hernandez-Boussard T. Expanding the secondary use of prostate cancer real world data: automated classifiers for clinical and pathological stage. *Front Digit Health.* 2022;4:793316. [doi: [10.3389/fdgth.2022.793316](https://doi.org/10.3389/fdgth.2022.793316)] [Medline: [35721793](https://pubmed.ncbi.nlm.nih.gov/35721793/)]
74. Breitenstein MK, Liu H, Maxwell KN, Pathak J, Zhang R. Electronic health record phenotypes for precision medicine: perspectives and caveats from treatment of breast cancer at a single institution. *Clin Transl Sci.* Jan 2018;11(1):85-92. [doi: [10.1111/cts.12514](https://doi.org/10.1111/cts.12514)] [Medline: [29084368](https://pubmed.ncbi.nlm.nih.gov/29084368/)]
75. Liu S, McCoy AB, Aldrich MC, et al. Leveraging natural language processing to identify eligible lung cancer screening patients with the electronic health record. *Int J Med Inform.* Sep 2023;177:105136. [doi: [10.1016/j.ijmedinf.2023.105136](https://doi.org/10.1016/j.ijmedinf.2023.105136)] [Medline: [37392712](https://pubmed.ncbi.nlm.nih.gov/37392712/)]
76. Schiappa R, Contu S, Culie D, et al. Validation of RUBY for breast cancer knowledge extraction from a large French electronic medical record system. *JCO Clin Cancer Inform.* May 2023;7:e2200130. [doi: [10.1200/CCI.22.00130](https://doi.org/10.1200/CCI.22.00130)] [Medline: [37235837](https://pubmed.ncbi.nlm.nih.gov/37235837/)]
77. Beck JT, Rammage M, Jackson GP, et al. Artificial intelligence tool for optimizing eligibility screening for clinical trials in a large community cancer center. *JCO Clin Cancer Inform.* Jan 2020;4:50-59. [doi: [10.1200/CCI.19.00079](https://doi.org/10.1200/CCI.19.00079)] [Medline: [31977254](https://pubmed.ncbi.nlm.nih.gov/31977254/)]
78. Lindvall C, Lilley EJ, Zupanc SN, et al. Natural language processing to assess end-of-life quality indicators in cancer patients receiving palliative surgery. *J Palliat Med.* Feb 2019;22(2):183-187. [doi: [10.1089/jpm.2018.0326](https://doi.org/10.1089/jpm.2018.0326)] [Medline: [30328764](https://pubmed.ncbi.nlm.nih.gov/30328764/)]
79. Gauthier MP, Law JH, Le LW, et al. Automating access to real-world evidence. *JTO Clin Res Rep.* Jun 2022;3(6):100340. [doi: [10.1016/j.jto.2022.100340](https://doi.org/10.1016/j.jto.2022.100340)] [Medline: [35719866](https://pubmed.ncbi.nlm.nih.gov/35719866/)]
80. Lin E, Zwolinski R, Wu JTY, et al. Machine learning-based natural language processing to extract PD-L1 expression levels from clinical notes. *Health Informatics J.* 2023;29(3):14604582231198021. [doi: [10.1177/14604582231198021](https://doi.org/10.1177/14604582231198021)] [Medline: [37635280](https://pubmed.ncbi.nlm.nih.gov/37635280/)]
81. 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.* Jan 2022;63(1):e29-e36. [doi: [10.1016/j.jpainsymman.2021.06.025](https://doi.org/10.1016/j.jpainsymman.2021.06.025)] [Medline: [34271146](https://pubmed.ncbi.nlm.nih.gov/34271146/)]
82. Wang K, Cui H, Zhu Y, et al. Evaluation of an artificial intelligence-based clinical trial matching system in Chinese patients with hepatocellular carcinoma: a retrospective study. *BMC Cancer.* 2024;24(1):1-7. [doi: [10.1186/s12885-024-11959-7](https://doi.org/10.1186/s12885-024-11959-7)]
83. Lin FPY, Salih OSM, Scott N, Jameson MB, Epstein RJ. Development and validation of a machine learning approach leveraging real-world clinical narratives as a predictor of survival in advanced cancer. *JCO Clin Cancer Inform.* Oct 2022;PMID(6):36265112. [doi: [10.1200/CCI.22.00064](https://doi.org/10.1200/CCI.22.00064)]
84. Gensheimer MF, Aggarwal S, Benson KRK, et al. Automated model versus treating physician for predicting survival time of patients with metastatic cancer. *J Am Med Inform Assoc.* Jun 12, 2021;28(6):1108-1116. [doi: [10.1093/jamia/ocaa290](https://doi.org/10.1093/jamia/ocaa290)]
85. Moseley ET, Wu JT, Welt J, et al. A corpus for detecting high-context medical conditions in intensive care patient notes focusing on frequently readmitted patients. Preprint posted online on Mar 6, 2020. [doi: [10.48550/arXiv.2003.03044](https://doi.org/10.48550/arXiv.2003.03044)]
86. Xu Y, Li N, Lu M, et al. Development and validation of method for defining conditions using Chinese electronic medical record. *BMC Med Inform Decis Mak.* Aug 20, 2016;16:110. [doi: [10.1186/s12911-016-0348-6](https://doi.org/10.1186/s12911-016-0348-6)] [Medline: [27542973](https://pubmed.ncbi.nlm.nih.gov/27542973/)]
87. Poort H, Zupanc SN, Leiter RE, Wright AA, Lindvall C. 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.* Feb 2020;9(1):100-104. [doi: [10.1089/jayao.2019.0040](https://doi.org/10.1089/jayao.2019.0040)] [Medline: [31411524](https://pubmed.ncbi.nlm.nih.gov/31411524/)]
88. Ernecoff NC, Wessell KL, Hanson LC, et al. Electronic health record phenotypes for identifying patients with late-stage disease: a method for research and clinical application. *J Gen Intern Med.* Dec 2019;34(12):2818-2823. [doi: [10.1007/s11606-019-05219-9](https://doi.org/10.1007/s11606-019-05219-9)] [Medline: [31396813](https://pubmed.ncbi.nlm.nih.gov/31396813/)]
89. Warner JL, Levy MA, Neuss MN, Warner JL, Levy MA, Neuss MN. ReCAP: feasibility and accuracy of extracting cancer stage information from narrative electronic health record data. *J Oncol Pract.* Feb 2016;12(2):157-158. [doi: [10.1200/JOP.2015.004622](https://doi.org/10.1200/JOP.2015.004622)] [Medline: [26306621](https://pubmed.ncbi.nlm.nih.gov/26306621/)]
90. Chen L, Song L, Shao Y, Li D, Ding K. Using natural language processing to extract clinically useful information from Chinese electronic medical records. *Int J Med Inform.* Apr 2019;124:6-12. [doi: [10.1016/j.ijmedinf.2019.01.004](https://doi.org/10.1016/j.ijmedinf.2019.01.004)] [Medline: [30784428](https://pubmed.ncbi.nlm.nih.gov/30784428/)]

91. Kondratieff KE, Brown JT, Barron M, Warner JL, Yin Z. Mining medication use patterns from clinical notes for breast cancer patients through a two-stage topic modeling approach. *AMIA Jt Summits Transl Sci Proc.* 2022;2022(303):303-312. [Medline: [35854740](#)]
92. Hong JC, Fairchild AT, Tanksley JP, Palta M, Tenenbaum JD. Natural language processing for abstraction of cancer treatment toxicities: accuracy versus human experts. *JAMIA Open.* Feb 15, 2021;3(4):513-517. [doi: [10.1093/jamiaopen/ooaa064](#)]
93. Gregg JR, Lang M, Wang LL, et al. Automating the determination of prostate cancer risk strata from electronic medical records. *JCO Clin Cancer Inform.* 2017;1(1):1-8. [doi: [10.1200/CCI.16.00045](#)] [Medline: [29541700](#)]
94. Li K, Banerjee I, Magnani CJ, Blayne DW, Brooks JD, Hernandez-Boussard T. Clinical documentation to predict factors associated with urinary incontinence following prostatectomy for prostate cancer. *Res Rep Urol.* 2020;12:7-14. [doi: [10.2147/RRU.S234178](#)] [Medline: [32158720](#)]
95. Bozkurt S, Kan KM, Ferrari MK, et al. Is it possible to automatically assess pretreatment digital rectal examination documentation using natural language processing? A single-centre retrospective study. *BMJ Open.* Jul 18, 2019;9(7):e027182. [doi: [10.1136/bmjopen-2018-027182](#)] [Medline: [31324681](#)]
96. Laios A, Kalampokis E, Mamalis ME, et al. RoBERTa-assisted outcome prediction in ovarian cancer cytoreductive surgery using operative notes. *Cancer Control.* 2023;30:10732748231209892. [doi: [10.1177/10732748231209892](#)] [Medline: [37915208](#)]
97. Joffe E, Pettigrew EJ, Herskovic JR, Bearden CF, Bernstam EV. Expert guided natural language processing using one-class classification. *J Am Med Inform Assoc.* Sep 2015;22(5):962-966. [doi: [10.1093/jamia/ocv010](#)] [Medline: [26063744](#)]
98. 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.* Jun 2019;94:103184. [doi: [10.1016/j.jbi.2019.103184](#)] [Medline: [31014980](#)]
99. Bozkurt S, Park JI, Kan KM, et al. An automated feature engineering for digital rectal examination documentation using natural language processing. *AMIA Annu Symp Proc.* 2018;2018(288):288-294. [Medline: [30815067](#)]
100. Sanyal J, Tariq A, Kurian AW, Rubin D, Banerjee I. Weakly supervised temporal model for prediction of breast cancer distant recurrence. *Sci Rep.* May 4, 2021;11(1):9461. [doi: [10.1038/s41598-021-89033-6](#)] [Medline: [33947927](#)]
101. Kehl KL, Xu W, Gusev A, et al. Artificial intelligence-aided clinical annotation of a large multi-cancer genomic dataset. *Nat Commun.* Dec 15, 2021;12(1):7304. [doi: [10.1038/s41467-021-27358-6](#)] [Medline: [34911934](#)]
102. Chen S, Guevara M, Ramirez N, et al. Natural language processing to automatically extract the presence and severity of esophagitis in notes of patients undergoing radiotherapy. *JCO Clin Cancer Inform.* Jul 2023;7:e2300048. [doi: [10.1200/CCI.23.00048](#)] [Medline: [37506330](#)]
103. Lindvall C, Deng CY, Agaronnik ND, et al. Deep learning for cancer symptoms monitoring on the basis of electronic health record unstructured clinical notes. *JCO Clin Cancer Inform.* Jun 2022;6:e2100136. [doi: [10.1200/CCI.21.00136](#)] [Medline: [35714301](#)]
104. Yim WW, Kwan SW, Johnson G, Yetisgen M. Classification of hepatocellular carcinoma stages from free-text clinical and radiology reports. *AMIA Annu Symp Proc.* 2017;2017(1858):1858-1867. [Medline: [29854257](#)]
105. Derton A, Guevara M, Chen S, et al. Natural language processing methods to empirically explore social contexts and needs in cancer patient notes. *JCO Clin Cancer Inform.* May 2023;7:e2200196. [doi: [10.1200/CCI.22.00196](#)] [Medline: [37235847](#)]
106. Khor RC, Nguyen A, O'Dwyer J, et al. Extracting tumour prognostic factors from a diverse electronic record dataset in genito-urinary oncology. *Int J Med Inform.* Jan 2019;121:53-57. [doi: [10.1016/j.ijmedinf.2018.10.008](#)] [Medline: [30545489](#)]
107. Delorme J, Charvet V, Wartelle M, et al. Natural language processing for patient selection in Phase I or II oncology clinical trials. *JCO Clin Cancer Inform.* Jun 2021;5:709-718. [doi: [10.1200/CCI.21.00003](#)] [Medline: [34197179](#)]
108. Kehl KL, Xu W, Lepisto E, et al. Natural language processing to ascertain cancer outcomes from medical oncologist notes. *JCO Clin Cancer Inform.* Aug 2020;4:680-690. [doi: [10.1200/CCI.20.00020](#)] [Medline: [32755459](#)]
109. DiMartino L, Miano T, Wessell K, Bohac B, Hanson LC. Identification of uncontrolled symptoms in cancer patients using natural language processing. *J Pain Symptom Manage.* Apr 2022;63(4):610-617. [doi: [10.1016/j.jpainsymman.2021.10.014](#)] [Medline: [34743011](#)]
110. Zeng J, Banerjee I, Henry AS, et al. Natural language processing to identify cancer treatments with electronic medical records. *JCO Clin Cancer Inform.* Apr 2021;5:379-393. [doi: [10.1200/CCI.20.00173](#)] [Medline: [33822653](#)]
111. Bozkurt S, Paul R, Coquet J, et al. Phenotyping severity of patient-centered outcomes using clinical notes: a prostate cancer use case. *Learn Health Syst.* Oct 2020;4(4):e10237. [doi: [10.1002/lrh2.10237](#)] [Medline: [33083539](#)]
112. Meystre SM, Heider PM, Cates A, et al. Piloting an automated clinical trial eligibility surveillance and provider alert system based on artificial intelligence and standard data models. *BMC Med Res Methodol.* Apr 11, 2023;23(1):88. [doi: [10.1186/s12874-023-01916-6](#)] [Medline: [37041475](#)]

113. Araki K, Matsumoto N, Togo K, et al. Real-world treatment response in Japanese patients with cancer using unstructured data from electronic health records. *Health Technol*. Mar 2023;13(2):253-262. [doi: [10.1007/s12553-023-00739-1](https://doi.org/10.1007/s12553-023-00739-1)]
114. Guan M, Cho S, Petro R, Zhang W, Pasche B, Topaloglu U. Natural language processing and recurrent network models for identifying genomic mutation-associated cancer treatment change from patient progress notes. *JAMIA Open*. Apr 2019;2(1):139-149. [doi: [10.1093/jamiaopen/ooy061](https://doi.org/10.1093/jamiaopen/ooy061)] [Medline: [30944913](https://pubmed.ncbi.nlm.nih.gov/30944913/)]
115. Li F, Yu H. An investigation of single-domain and multidomain medication and adverse drug event relation extraction from electronic health record notes using advanced deep learning models. *J Am Med Inform Assoc*. Jul 1, 2019;26(7):646-654. [doi: [10.1093/jamia/ocz018](https://doi.org/10.1093/jamia/ocz018)]
116. Dai HJ, Wang FD, Chen CW, Su CH, Wu CS, Jonnagaddala J. Cohort selection for clinical trials using multiple instance learning. *J Biomed Inform*. Jul 2020;107:103438. [doi: [10.1016/j.jbi.2020.103438](https://doi.org/10.1016/j.jbi.2020.103438)] [Medline: [32360937](https://pubmed.ncbi.nlm.nih.gov/32360937/)]
117. Forsyth AW, Barzilay R, Hughes KS, et al. Machine learning methods to extract documentation of breast cancer symptoms from electronic health records. *J Pain Symptom Manage*. Jun 2018;55(6):1492-1499. [doi: [10.1016/j.jpainsymman.2018.02.016](https://doi.org/10.1016/j.jpainsymman.2018.02.016)] [Medline: [29496537](https://pubmed.ncbi.nlm.nih.gov/29496537/)]
118. Yuan Q, Cai T, Hong C, et al. Performance of a machine learning algorithm using electronic health record data to identify and estimate survival in a longitudinal cohort of patients with lung cancer. *JAMA Netw Open*. Jul 1, 2021;4(7):e2114723. [doi: [10.1001/jamanetworkopen.2021.14723](https://doi.org/10.1001/jamanetworkopen.2021.14723)] [Medline: [34232304](https://pubmed.ncbi.nlm.nih.gov/34232304/)]
119. Banerjee I, Li K, Seneviratne M, et al. Weakly supervised natural language processing for assessing patient-centered outcome following prostate cancer treatment. *JAMIA Open*. Apr 2019;2(1):150-159. [doi: [10.1093/jamiaopen/ooy057](https://doi.org/10.1093/jamiaopen/ooy057)] [Medline: [31032481](https://pubmed.ncbi.nlm.nih.gov/31032481/)]
120. Agaronnik ND, Lindvall C, El-Jawahri A, He W, Iezzoni LI. Challenges of developing a natural language processing method with electronic health records to identify persons with chronic mobility disability. *Arch Phys Med Rehabil*. Oct 2020;101(10):1739-1746. [doi: [10.1016/j.apmr.2020.04.024](https://doi.org/10.1016/j.apmr.2020.04.024)] [Medline: [32446905](https://pubmed.ncbi.nlm.nih.gov/32446905/)]
121. Leis A, Casadevall D, Albanell J, et al. Exploring the association of cancer and depression in electronic health records: combining encoded diagnosis and mining free-text clinical notes. *JMIR Cancer*. Jul 11, 2022;8(3):e39003. [doi: [10.2196/39003](https://doi.org/10.2196/39003)] [Medline: [35816382](https://pubmed.ncbi.nlm.nih.gov/35816382/)]
122. Lin FPY, Pokorny A, Teng C, Epstein RJ. TEPAPA: a novel in silico feature learning pipeline for mining prognostic and associative factors from text-based electronic medical records. *Sci Rep*. Jul 31, 2017;7(1):6918. [doi: [10.1038/s41598-017-07111-0](https://doi.org/10.1038/s41598-017-07111-0)] [Medline: [28761061](https://pubmed.ncbi.nlm.nih.gov/28761061/)]
123. Redd DF, Shao Y, Zeng-Treitler Q, et al. Identification of colorectal cancer using structured and free text clinical data. *Health Informatics J*. 2022;28(4):14604582221134406. [doi: [10.1177/14604582221134406](https://doi.org/10.1177/14604582221134406)] [Medline: [36300566](https://pubmed.ncbi.nlm.nih.gov/36300566/)]
124. Liu F, Pradhan R, Druhl E, et al. Learning to detect and understand drug discontinuation events from clinical narratives. *J Am Med Inform Assoc*. Oct 1, 2019;26(10):943-951. [doi: [10.1093/jamia/ocz048](https://doi.org/10.1093/jamia/ocz048)] [Medline: [31034028](https://pubmed.ncbi.nlm.nih.gov/31034028/)]
125. Liu K, Kulkarni O, Witteveen-Lane M, Chen B, Chesla D. MetBERT: a generalizable and pre-trained deep learning model for the prediction of metastatic cancer from clinical notes. *AMIA Jt Summits Transl Sci Proc*. 2022;2022(331):331-338. [Medline: [35854741](https://pubmed.ncbi.nlm.nih.gov/35854741/)]
126. Koleck TA, Topaz M, Tatonetti NP, et al. Characterizing shared and distinct symptom clusters in common chronic conditions through natural language processing of nursing notes. *Res Nurs Health*. Dec 2021;44(6):906-919. [doi: [10.1002/nur.22190](https://doi.org/10.1002/nur.22190)] [Medline: [34637147](https://pubmed.ncbi.nlm.nih.gov/34637147/)]
127. Ehrentraut C, Sundström K, Dalianis H. Exploration of known and unknown early symptoms of cervical cancer and development of a symptom spectrum—outline of a data and text mining based approach. Presented at: Proceeding from CAiSE 2015 Industry Track CEUR Workshop Proc; Jun 8-12, 2015:34-44; Stockholm, Sweden.
128. Lazic I, Jakovljevic N, Boban J, Nosek I, Loncar-Turukalo T. Information extraction from clinical records: an example for breast cancer. Presented at: 2022 IEEE 21st Mediterranean Electrotechnical Conference (MELECON); Jun 14-16, 2022; Palermo, Italy. [doi: [10.1109/MELECON53508.2022.9842995](https://doi.org/10.1109/MELECON53508.2022.9842995)]
129. Stevens M, Kennedy G, Churches T. Applying and improving a publicly available medication NER pipeline in a clinical cancer EMR. *Stud Health Technol Inform*. Jan 25, 2024;310:679-684. [doi: [10.3233/SHTI231051](https://doi.org/10.3233/SHTI231051)] [Medline: [38269895](https://pubmed.ncbi.nlm.nih.gov/38269895/)]
130. Luo X, Gandhi P, Storey S, Zhang Z, Han Z, Huang K. A computational framework to analyze the associations between symptoms and cancer patient attributes post chemotherapy using EHR data. *IEEE J Biomed Health Inform*. Nov 2021;25(11):4098-4109. [doi: [10.1109/JBHI.2021.3117238](https://doi.org/10.1109/JBHI.2021.3117238)] [Medline: [34613922](https://pubmed.ncbi.nlm.nih.gov/34613922/)]
131. Luo X, Storey S, Gandhi P, Zhang Z, Metzger M, Huang K. Analyzing the symptoms in colorectal and breast cancer patients with or without type 2 diabetes using EHR data. *Health Informatics J*. 2021;27(1):14604582211000785. [doi: [10.1177/14604582211000785](https://doi.org/10.1177/14604582211000785)] [Medline: [33726552](https://pubmed.ncbi.nlm.nih.gov/33726552/)]
132. Schiappa R, Contu S, Culie D, et al. RUBY: natural language processing of French electronic medical records for breast cancer research. *JCO Clin Cancer Inform*. Jul 2022;6(6):e2100199. [doi: [10.1200/CCI.21.00199](https://doi.org/10.1200/CCI.21.00199)] [Medline: [35960900](https://pubmed.ncbi.nlm.nih.gov/35960900/)]
133. Tan HJ, Clarke R, Chamie K, et al. Development and validation of an automated method to identify patients undergoing radical cystectomy for bladder cancer using natural language processing. *Urol Pract*. Sep 2017;4(5):365-372. [doi: [10.1016/j.urpr.2016.09.011](https://doi.org/10.1016/j.urpr.2016.09.011)] [Medline: [37592698](https://pubmed.ncbi.nlm.nih.gov/37592698/)]

134. Agaronnik N, Lindvall C, El-Jawahri A, He W, Iezzoni L. Use of natural language processing to assess frequency of functional status documentation for patients newly diagnosed with colorectal cancer. *JAMA Oncol.* Oct 1, 2020;6(10):1628-1630. [doi: [10.1001/jamaoncol.2020.2708](https://doi.org/10.1001/jamaoncol.2020.2708)] [Medline: [32880603](https://pubmed.ncbi.nlm.nih.gov/32880603/)]
135. Afzal M, Hussain M, Ali Khan W, et al. Comprehensible knowledge model creation for cancer treatment decision making. *Comput Biol Med.* Mar 2017;82:119-129. [doi: [10.1016/j.combiomed.2017.01.010](https://doi.org/10.1016/j.combiomed.2017.01.010)]
136. Loda S, Krebs J, Danhof S, et al. Exploration of artificial intelligence use with ARIES in multiple myeloma research. *J Clin Med.* Jul 9, 2019;8(7):999. [doi: [10.3390/jcm8070999](https://doi.org/10.3390/jcm8070999)] [Medline: [31324026](https://pubmed.ncbi.nlm.nih.gov/31324026/)]
137. Tahabi FM, Storey S, Luo X. SymptomGraph: identifying symptom clusters from narrative clinical notes using graph clustering. Presented at: SAC '23: Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing; Mar 27-31, 2023; Tallinn Estonia. [doi: [10.1145/3555776.3577685](https://doi.org/10.1145/3555776.3577685)]
138. Percha B, Pisapati K, Gao C, Schmidt H. Natural language inference for curation of structured clinical registries from unstructured text. *J Am Med Inform Assoc.* Dec 28, 2021;29(1):97-108. [doi: [10.1093/jamia/ocab243](https://doi.org/10.1093/jamia/ocab243)] [Medline: [34791282](https://pubmed.ncbi.nlm.nih.gov/34791282/)]
139. 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.* Sep 2021;5:1005-1014. [doi: [10.1200/CCL.21.00030](https://doi.org/10.1200/CCL.21.00030)] [Medline: [34570630](https://pubmed.ncbi.nlm.nih.gov/34570630/)]
140. Kersloot MG, Lau F, Abu-Hanna A, Arts DL, Cornet R. Automated SNOMED CT concept and attribute relationship detection through a web-based implementation of cTAKES. *J Biomed Semantics.* Sep 18, 2019;10(1):14. [doi: [10.1186/s13326-019-0207-3](https://doi.org/10.1186/s13326-019-0207-3)] [Medline: [31533810](https://pubmed.ncbi.nlm.nih.gov/31533810/)]
141. Ahmad PN, Liu Y, Khan K, Jiang T, Burhan U. BIR: biomedical information retrieval system for cancer treatment in electronic health record using transformers. *Sensors (Basel).* Nov 23, 2023;23(23):9355. [doi: [10.3390/s23239355](https://doi.org/10.3390/s23239355)] [Medline: [38067736](https://pubmed.ncbi.nlm.nih.gov/38067736/)]
142. Jamaluddin M, Wibawa AD. Patient diagnosis classification based on electronic medical record using text mining and support vector machine. Presented at: 2021 International Seminar on Application for Technology of Information and Communication (iSemantic); Sep 18-19, 2021; Semarangin, Indonesia. [doi: [10.1109/iSemantic52711.2021.9573178](https://doi.org/10.1109/iSemantic52711.2021.9573178)]
143. Shah S, Luo X, Kanakasabai S, Tuason R, Klopper G. Neural networks for mining the associations between diseases and symptoms in clinical notes. *Health Inf Sci Syst.* Dec 2019;7(1):1. [doi: [10.1007/s13755-018-0062-0](https://doi.org/10.1007/s13755-018-0062-0)] [Medline: [30588291](https://pubmed.ncbi.nlm.nih.gov/30588291/)]
144. Rohanian O, Jauncey H, Nouriborji M, et al. Using bottleneck adapters to identify cancer in clinical notes under low-resource constraints. Presented at: The 22nd Workshop on Biomedical Natural Language Processing and BioNLP Shared Tasks; Jun 13, 2023; Toronto, Canada. [doi: [10.18653/v1/2023.bionlp-1.5](https://doi.org/10.18653/v1/2023.bionlp-1.5)]
145. Bouvry C, Tvardik N, Kergourlay I, et al. The SYNODOS Project: system for the normalization and organization of textual medical data for observation in healthcare. *IRBM.* Apr 2016;37(2):109-115. [doi: [10.1016/j.irbm.2016.03.002](https://doi.org/10.1016/j.irbm.2016.03.002)]
146. Rahimian M, Warner JL, Jain SK, Davis RB, Zerillo JA, Joyce RM. Significant and distinctive n-grams in oncology notes: a text-mining method to analyze the effect of OpenNotes on clinical documentation. *JCO Clin Cancer Inform.* Jun 2019;3:1-9. [doi: [10.1200/CCL.19.00012](https://doi.org/10.1200/CCL.19.00012)] [Medline: [31184919](https://pubmed.ncbi.nlm.nih.gov/31184919/)]
147. Chen X, Xie H, Wang FL, Liu Z, Xu J, Hao T. A bibliometric analysis of natural language processing in medical research. *BMC Med Inform Decis Mak.* Mar 22, 2018;18(Suppl 1):14. [doi: [10.1186/s12911-018-0594-x](https://doi.org/10.1186/s12911-018-0594-x)] [Medline: [29589569](https://pubmed.ncbi.nlm.nih.gov/29589569/)]
148. Casey A, Davidson E, Poon M, et al. A systematic review of natural language processing applied to radiology reports. *BMC Med Inform Decis Mak.* Jun 3, 2021;21(1):179. [doi: [10.1186/s12911-021-01533-7](https://doi.org/10.1186/s12911-021-01533-7)] [Medline: [34082729](https://pubmed.ncbi.nlm.nih.gov/34082729/)]
149. Goff DJ, Loehfelm TW. Automated radiology report summarization using an open-source natural language processing pipeline. *J Digit Imaging.* Apr 2018;31(2):185-192. [doi: [10.1007/s10278-017-0030-2](https://doi.org/10.1007/s10278-017-0030-2)] [Medline: [29086081](https://pubmed.ncbi.nlm.nih.gov/29086081/)]
150. Dong H, Suárez-Paniagua V, Whiteley W, Wu H. Explainable automated coding of clinical notes using hierarchical label-wise attention networks and label embedding initialisation. *J Biomed Inform.* Apr 2021;116:103728. [doi: [10.1016/j.jbi.2021.103728](https://doi.org/10.1016/j.jbi.2021.103728)] [Medline: [33711543](https://pubmed.ncbi.nlm.nih.gov/33711543/)]
151. Payrovnaziri SN, Chen Z, Rengifo-Moreno P, et al. Explainable artificial intelligence models using real-world electronic health record data: a systematic scoping review. *J Am Med Inform Assoc.* Jul 1, 2020;27(7):1173-1185. [doi: [10.1093/jamia/ocaa053](https://doi.org/10.1093/jamia/ocaa053)] [Medline: [32417928](https://pubmed.ncbi.nlm.nih.gov/32417928/)]
152. Wu S, Roberts K, Datta S, et al. Deep learning in clinical natural language processing: a methodical review. *J Am Med Inform Assoc.* Mar 1, 2020;27(3):457-470. [doi: [10.1093/jamia/ocz200](https://doi.org/10.1093/jamia/ocz200)]
153. Manning CD, Raghavan P, Schütze H. *Introduction to Information Retrieval.* Cambridge University Press; 2008. ISBN: 9780511809071
154. Shivade C, Raghavan P, Fosler-Lussier E, et al. A review of approaches to identifying patient phenotype cohorts using electronic health records. *J Am Med Inform Assoc.* Mar 2014;21(2):221-230. [doi: [10.1136/amiajnl-2013-001935](https://doi.org/10.1136/amiajnl-2013-001935)]
155. Zhu M, Lin H, Jiang J, et al. Large language model trained on clinical oncology data predicts cancer progression. *NPJ Digit Med.* 2025;8(1):1-15. [doi: [10.1038/s41746-025-01780-2](https://doi.org/10.1038/s41746-025-01780-2)]
156. Gottlieb S. New FDA policies could limit the full value of AI in medicine. *JAMA Health Forum.* Feb 7, 2025;6(2):e250289. [doi: [10.1001/jamahealthforum.2025.0289](https://doi.org/10.1001/jamahealthforum.2025.0289)] [Medline: [39913129](https://pubmed.ncbi.nlm.nih.gov/39913129/)]

157. Van Laere S, Muylle KM, Cornu P. Clinical decision support and new regulatory frameworks for medical devices: are we ready for it?—A viewpoint paper. *Int J Health Policy Manag.* Dec 19, 2022;11(12):3159-3163. [doi: [10.34172/ijhpm.2021.144](https://doi.org/10.34172/ijhpm.2021.144)] [Medline: [34814678](https://pubmed.ncbi.nlm.nih.gov/34814678/)]
158. Artsi Y, Sorin V, Glicksberg BS, et al. Challenges of implementing LLMs in clinical practice: perspectives. *J Clin Med.* Sep 1, 2025;14(17):6169. [doi: [10.3390/jcm14176169](https://doi.org/10.3390/jcm14176169)] [Medline: [40943929](https://pubmed.ncbi.nlm.nih.gov/40943929/)]
159. General Data Protection Regulation (GDPR). Intersoft Consulting. URL: <https://gdpr-info.eu/> [Accessed 2025-01-05]
160. Health Insurance Portability and Accountability Act (HIPAA). US Department of Health and Human Services. URL: <https://www.hhs.gov/hipaa/index.html> [Accessed 2025-01-05]
161. Moorthie S, Hayat S, Zhang Y, et al. Rapid systematic review to identify key barriers to access, linkage, and use of local authority administrative data for population health research, practice, and policy in the United Kingdom. *BMC Public Health.* Jun 28, 2022;22(1):1263. [doi: [10.1186/s12889-022-13187-9](https://doi.org/10.1186/s12889-022-13187-9)] [Medline: [35764951](https://pubmed.ncbi.nlm.nih.gov/35764951/)]
162. Alsentzer E. Publicly available clinical BERT embeddings. Presented at: Proceedings of the 2nd Clinical Natural Language Processing Workshop ACL Anthology 2019; Jun 7, 2019; Minneapolis, MN, United States. [doi: [10.18653/v1/W19-1909](https://doi.org/10.18653/v1/W19-1909)]
163. Peng Y, Yan S, Lu Z. Transfer learning in biomedical natural language processing: an evaluation of BERT and ELMO on ten benchmarking datasets. Presented at: BioNLP 2019—SIGBioMed Workshop on Biomedical Natural Language Processing, Proceedings of the 18th BioNLP Workshop and Shared Task; Aug 1, 2019;58-65; Florence, Italy. 2019.[doi: [10.18653/v1/W19-5006](https://doi.org/10.18653/v1/W19-5006)]
164. Miranda-Escalada A, Farré E, Krallinger M. Named entity recognition, concept normalization and clinical coding: overview of the cantemist track for cancer text mining in spanish, corpus, guidelines, methods and results published online first. Presented at: Iberian Languages Evaluation Forum 2020; Sep 22-24, 2024; Málaga, Spain. [doi: [10.5281/zenodo.3773228](https://doi.org/10.5281/zenodo.3773228)]
165. Agrawal M, Heggelmann S, Lang H, Kim Y, Sontag D. Large language models are few-shot clinical information extractors. Presented at: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing; Jun 7-11, 2022; Abu Dhabi, United Arab Emirates. [doi: [10.18653/v1/2022.emnlp-main.130](https://doi.org/10.18653/v1/2022.emnlp-main.130)]
166. Keloth VK, Selek S, Chen Q, et al. Social determinants of health extraction from clinical notes across institutions using large language models. *NPJ Digit Med.* May 17, 2025;8(1):287. [doi: [10.1038/s41746-025-01645-8](https://doi.org/10.1038/s41746-025-01645-8)] [Medline: [40379919](https://pubmed.ncbi.nlm.nih.gov/40379919/)]
167. Ferrara E. Should ChatGPT be biased? Challenges and risks of bias in large language models. *FM.* 2023. [doi: [10.5210/fm.v28i11.13346](https://doi.org/10.5210/fm.v28i11.13346)]
168. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med.* Aug 2023;29(8):1930-1940. [doi: [10.1038/s41591-023-02448-8](https://doi.org/10.1038/s41591-023-02448-8)] [Medline: [37460753](https://pubmed.ncbi.nlm.nih.gov/37460753/)]
169. Thirunavukarasu AJ, Hassan R, Mahmood S, et al. Trialling a large language model (ChatGPT) in general practice with the applied knowledge test: observational study demonstrating opportunities and limitations in primary care. *JMIR Med Educ.* Apr 21, 2023;9:e46599. [doi: [10.2196/46599](https://doi.org/10.2196/46599)] [Medline: [37083633](https://pubmed.ncbi.nlm.nih.gov/37083633/)]
170. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLOS Digit Health.* Feb 2023;2(2):e0000198. [doi: [10.1371/journal.pdig.0000198](https://doi.org/10.1371/journal.pdig.0000198)] [Medline: [36812645](https://pubmed.ncbi.nlm.nih.gov/36812645/)]
171. Lin BY, He C, Ze Z, et al. FedNLP: benchmarking federated learning methods for natural language processing tasks. Presented at: Findings of the Association for Computational Linguistics; Jul 10-15, 2021; Seattle, United States. [doi: [10.18653/v1/2022.findings-naacl.13](https://doi.org/10.18653/v1/2022.findings-naacl.13)]
172. Mohan K. A study on performance limitations in federated learning. Preprint posted online on Jan 7, 2025. [doi: [10.48550/arXiv.2501.03477](https://doi.org/10.48550/arXiv.2501.03477)]
173. Xu C, Qu Y, Xiang Y, Gao L. Asynchronous federated learning on heterogeneous devices: a survey. *Comput Sci Rev.* Nov 2023;50:100595. [doi: [10.1016/j.cosrev.2023.100595](https://doi.org/10.1016/j.cosrev.2023.100595)]

Abbreviations

- AI:** artificial intelligence
- BERT:** Bidirectional Encoder Representations from Transformers
- CNN:** convolutional neural network
- EHR:** electronic health record
- ICD:** International Classification of Diseases
- IE:** information extraction
- LLM:** large language model
- NLP:** natural language processing
- PLM:** pretrained language model
- PRISMA-ScR:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
- RNN:** recurrent neural network

UMLS: Unified Medical Language System

Edited by Andrew Coristine; peer-reviewed by Álvaro García-Barragán, Dillon Chrimes, Kola Adegoke; submitted 05.Mar.2025; final revised version received 15.Feb.2026; accepted 16.Feb.2026; published 14.May.2026

Please cite as:

Kayira AB, Elyazori HRA, Lybarger K, Walter FM, Chelala C, Funston G

Natural Language Processing of Clinical Notes for Cancer Research and Patient Care Prior to Widespread Adoption of Generative AI: Scoping Review

JMIR AI 2026;5:e73481

URL: <https://ai.jmir.org/2026/1/e73481>

doi: [10.2196/73481](https://doi.org/10.2196/73481)

© Alfred B Kayira, Hadeel R A Elyazori, Kevin Lybarger, Fiona M Walter, Claude Chelala, Garth Funston. Originally published in JMIR AI (<https://ai.jmir.org>), 14.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR AI, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.ai.jmir.org/>, as well as this copyright and license information must be included.