

RESEARCH

Open Access



Extracting cancer concepts from clinical notes using natural language processing: a systematic review

Maryam Gholipour¹, Reza Khajouei², Parastoo Amiri¹, Sadrieh Hajesmaeel Gohari³ and Leila Ahmadian^{2*}

*Correspondence:

lahmadian@kmu.ac.ir;
ahmadianle@yahoo.com

¹ Student Research Committee,
Kerman University of Medical
Sciences, Kerman, Iran

² Department of Health
Information Sciences, Faculty
of Management and Medical
Information Sciences, Kerman
University of Medical Sciences,
Kerman, Iran

³ Medical Informatics Research
Center, Institute for Futures
Studies in Health, Kerman
University of Medical Sciences,
Kerman, Iran

Abstract

Background: Extracting information from free texts using natural language processing (NLP) can save time and reduce the hassle of manually extracting large quantities of data from incredibly complex clinical notes of cancer patients. This study aimed to systematically review studies that used NLP methods to identify cancer concepts from clinical notes automatically.

Methods: PubMed, Scopus, Web of Science, and Embase were searched for English language papers using a combination of the terms concerning “Cancer”, “NLP”, “Coding”, and “Registries” until June 29, 2021. Two reviewers independently assessed the eligibility of papers for inclusion in the review.

Results: Most of the software programs used for concept extraction reported were developed by the researchers ($n = 7$). Rule-based algorithms were the most frequently used algorithms for developing these programs. In most articles, the criteria of accuracy ($n = 14$) and sensitivity ($n = 12$) were used to evaluate the algorithms. In addition, Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT) and Unified Medical Language System (UMLS) were the most commonly used terminologies to identify concepts. Most studies focused on breast cancer ($n = 4$, 19%) and lung cancer ($n = 4$, 19%).

Conclusion: The use of NLP for extracting the concepts and symptoms of cancer has increased in recent years. The rule-based algorithms are well-liked algorithms by developers. Due to these algorithms' high accuracy and sensitivity in identifying and extracting cancer concepts, we suggested that future studies use these algorithms to extract the concepts of other diseases as well.

Keywords: Neoplasms, Natural language processing, NLP, Machine learning, Terminology, Information system, Systematic review

Background

One of the significant public health concerns is cancer. According to the World Health Organization (WHO) report in 2019, this disease is the leading cause of death worldwide [1]. GLOBOCAN (The Global Cancer Observatory) estimated [2] about 10 million deaths from cancer in 2020 (i.e., one in every six patients with cancer) [3]. The global



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

cancer-related deaths are predicted to be around 13 million by 2030 [4]. Due to the growing incidence of cancer, researchers use various methods to combat this disease. Artificial intelligence (AI) is one of the methods that has been used to diagnose cancer [5–9] and predict its risk [10], relapse [11], and symptoms [11–13]. AI can provide a safe, fast, and efficient way to manage such diseases.

Natural language processing (NLP) is a branch of AI that addresses the interpretation and comprehension of texts using a set of algorithms [13–15]. NLP is the key to obtaining structured information from unstructured clinical texts [16]. Today, large amounts of clinical information are recorded and stored as narrative text in electronic systems. Retrieving and using this information can facilitate the diagnosis, treatment, and prediction of diseases. So far, NLP has been widely used in medical and health research, e.g., for identifying care coordination terms in nursing records [17], identifying medical concepts from radiology reports [18], extracting complications from problem lists [19], and determining disease status in discharge summaries [20]. For example, Si et al. [21] proposed a framework-based NLP method for extracting cancer-related information with a two-step strategy including bidirectional long short-term memory and conditional random field. Other studies extracted tumor-related information, such as location and size, using the NLP method [22, 23]. Kehl et al. [24] reported that the neural network-based NLP method could extract significant data from oncologists' notes.

Due to the unique characteristics of clinical texts, such as poor structure, use of specific vocabulary, and abbreviations [24] that make the use of NLP challenging, understanding the new developments of NLP in clinical research is essential. Despite the various studies that have been done on the application of NLP in medicine, there are limited systematic review studies summarizing its application. Previous systematic reviews mostly addressed the extraction of concepts from clinical texts such as radiology, laboratory, pathology, evaluation of postoperative surgical results, assessment of the application of NLP in the clinical practice of mental health, and development and adoption of NLP methods in open-text clinical notes related to chronic diseases [16, 25–29]. Casey et al. [25] investigated the use of NLP algorithms that were used in various studies to analyze radiology reports. In this study, besides determining the NLP algorithms, they focused on the purpose of using these algorithms for analyzing the reports and reported the following main applications: disease information and classification, language discovery and knowledge structure, quality and compliance, and cohort and epidemiology. In their systematic review, Pons et al. [15] also investigated NLP methodologies used on radiology reports and described the application and the purpose of using NLP, the tools used, and the performance results. Concerning the application of NLP in cancer, Santos conducted a study on NLP algorithms and extracted information regarding various models applied in different studies and their performances [24]. Based on the results of a systematic review of the application of NLP models to evaluate postoperative surgical outcomes, the most common outcome was postoperative complications. These complications can be identified more reliably using NLP models compared to traditional non-NLP alternatives. Glaz et al. [27] evaluated studies that used machine learning and NLP techniques in the field of mental health and also the potential application of these methods in mental clinical practice. The main objectives were to extract terms related to symptoms, classify the severity of illness, compare therapy effectiveness, and provide

psychopathological clues. Sheikhalishahi et al. [28] carried out a comprehensive overview of the development and uptake of NLP methods in open-text clinical notes related to chronic diseases, including an examination of the challenges of using NLP methods to extract terms from clinical narratives. The results of this study showed a trend indicating that most studies focused on cardiovascular diseases, while endocrine and metabolic diseases were the least researched topics. This trend may occur because clinical records related to metabolic diseases are more structured than those related to cardiovascular diseases.

Given the increasing incidence of cancer, as well as the recent advancements of NLP techniques to assist with the parsing and analysis of cancer-specific medical literature, a new systematic review in this area can help researchers and professionals gain a deeper understanding of this field and identify new techniques in cancer research to support and promote cancer research. To our knowledge, all existing systematic reviews addressed extracting NLP algorithms, and none of them specifically focused on extracting cancer concepts and the terminologies applied to detect the information regarding different types of cancer. Therefore, in this study, we systematically reviewed the studies on extracting cancer concepts to determine which NLP methods have been applied to automatically identify cancer concepts in clinical notes, which terminologies are used to code cancer concepts, and what types of cancers are identified. The results of this study can help researchers identify the existing NLP methods and proper terminological systems in this field.

Method

This systematic review was performed using the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [30]. PRISMA is a guideline that helps researchers to format their reviews and demonstrate the extent of the quality of their reviews. Also, the present study used wordcloud to pinpoint which variables need to be highlighted.

Information resources and searches

The PubMed, Scopus, Web of Science, and Embase databases were searched for relevant literature until June 29, 2021. A list of terms, keywords, and their synonyms were identified and categorized into four groups: "Cancer", "NLP", "Coding", and "Registries". We used the "OR" operator to combine the expressions within groups 1, 2, 3, and 4 and the "AND" operator to combine the results of the four groups (Table 1 shows the keywords for each group).

Inclusion criteria

All articles included in the study were original research articles that sought to retrieve cancer-related terms or concepts in clinical texts. These articles used the NLP technique to retrieve cancer-related concepts.

Exclusion criteria

Articles that used the NLP technique to retrieve concepts related to other diseases were excluded from the study. Studies that used the NLP technique in the field of cancer but

Table 1 Groups of keywords used in the search strategy

Group 1	Automatic mapping OR Automatic coding OR Lexical mapping OR Text parsing OR Text mining OR NLP OR Natural Language processing OR Information extraction OR Text understanding OR Text analysis OR Concept extraction OR Concept mapping OR Narrative parsing OR Automatic annotation
Group 2	Coding OR Terminology OR Thesaurus OR Vocabulary OR Glossary OR Nomenclature OR Classification OR Taxonomy OR Ontology OR Terminological system OR Nosology OR Lexicon
Group 3	Cancer OR Neoplasms OR Neoplasia OR Tumor OR Malignancy OR Anatomical Pathological Condition OR oncology OR Morphology OR Cell differentiation OR Lymphoma OR Carcinoma OR Adenoma OR Sarcoma OR Fibroma OR Thymoma OR Leukemia OR Papilloma OR Pathology
Group 4	Registries OR Information System OR Database OR Surveillance system

*MeSH terms are in bold

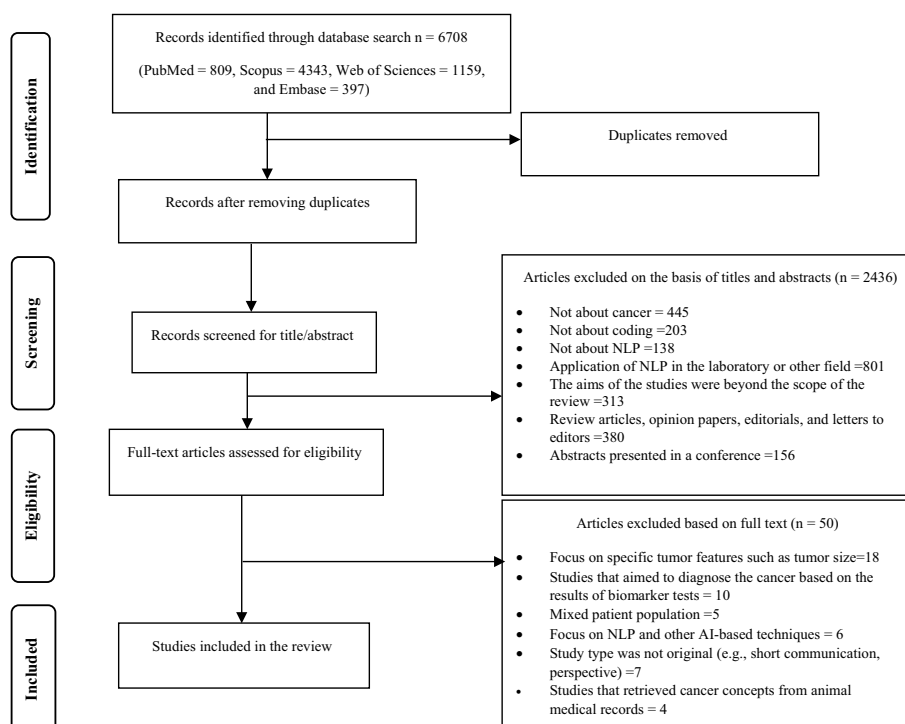


Fig. 1 The PRISMA diagram of study selection

used this technique to extract tumor features, such as tumor size, color, and shape, were also excluded. In addition, articles that used the NLP technique to diagnose cancer based on the patient’s clinical findings were not included in the study. For example, articles that aimed to diagnose cancer based on the results of biomarker tests and measurements in the patient’s body and the symptoms were not eligible for inclusion in the study. Furthermore, all review articles, conferences, and articles that retrieved cancer concepts from animal medical records were also excluded.

Article selection

Articles retrieved from databases were first entered into EndNote version X10. After eliminating duplicate studies, two authors (M.Gh and P.A) independently reviewed the titles and abstracts of the retrieved articles. Figure 1 shows the PRISMA diagram for the

inclusion and exclusion of articles in the study. After deleting irrelevant articles, the full text of the related articles was independently reviewed by three authors (S.Hg, M.Gh, and P.A). Disagreements among the reviewers were resolved by consensus in a meeting with another author (L.A).

Data collection process

A data extraction form was developed by the researchers. The validity of this form was confirmed by three medical informatics specialists and a health information management specialist. The form included the following headings: Authors, Year of publication, Setting, System, System module, Objective, Cancer type, Outcome, Data standard exchange, Terminological systems, NLP type, and Algorithm.

Results

This study was a systematic review that aimed to review articles that extracted cancer concepts using NLP. In total, 6708 papers were initially retrieved. After removing duplicates, 2503 articles remained for further review. Subsequently, the titles and abstracts of the remaining articles were screened, and inclusion and exclusion criteria were applied. After applying exclusion criteria, a total of 2436 articles were excluded, and 67 studies were deemed relevant. The full texts of these articles were reviewed, and finally, 17 articles were selected, and their information was extracted (Fig. 1).

General characteristics of the included articles

The publication dates of the retrieved articles were between 2012 and 2021 (Fig. 2). Most articles were published between 2016 and 2020. Out of 17 articles included in the present study, 10 were conducted in non-academic settings ($n = 10$, 58%).

Aims of the included articles

The aims of these 17 articles were evaluated after reviewing the full text of the articles (Table 2). The study aims were divided into four general categories: “extraction of cancer concepts” ($n = 12$), “comparison of the results retrieved by NLP algorithms and manual coding” ($n = 3$), “comparison of different NLP algorithms in terms of their ability to extract cancer concepts” ($n = 1$), and “extraction of cancer concepts and coding” ($n = 1$).

Specific characteristics of the included articles

System, Module and Database characteristics of the included articles are shown in Table 3.

System

The data analyzed in the included articles were extracted from various resources such as databases, registers, and health information systems. Data from multiple databases were examined in 10 out of the 17 articles included in the present study. In two of these 10 articles, more than one database was used. In three articles, electronic health record (EHR) data were examined. In these articles, clinical notes, pathology reports, and surgery reports were analyzed. In two articles, the data were retrieved from the electronic medical records (EMR) system, and the reports analyzed in these systems were breast



Fig. 2 General characteristics of included studies: **A** number of academic and non-academic articles; **B** number of articles per year; **C** number of articles per country

Table 2 The aims of the included articles

Aim type	References
Extraction of cancer concepts	[31–42]
Comparison of the retrieval results of NLP algorithms with manual coding	[43–45]
Comparison of different NLP algorithms in terms of their ability to extract cancer concepts	[45, 46]
Extraction of cancer concepts and coding	[47]

Table 3 System, module, and database characteristics of the included articles

Author	System	Module	Database
Hammami et al. [36]	–	Pathology reports	Oracle Data Warehouse of Fondazione IRCCS “Istituto nazionale dei tumori” (Istituto Nazionale dei Tumori)
Ryu et al. [37]	EHR	Pathology reports	–
Oliveira et al. [41]	–	Cervical and anal pathology reports	Clinical pathology laboratory information system
Becker et al. [48]	EHR	Clinical note	–
Wang et al. [40]	Mayo clinic EHR	Clinical notes and pathology reports	–
Kumar et al. [33]	–	Pathology reports	The Dartmouth-Hitchcock Medical Center (DHMC)
Wadia et al. [43]	–	Radiology reports (Ct) Chest CT reports	Clinical text analysis
Bustos et al. [38]	–	Public registry	Free dataset
Faina Linkov et al. [47]	–	UPMC registry	–
Sada et al. [34]	–	Liver pathology reports, abdominal CT, and abdominal MRI reports	Veterans Affairs administrative data
Nguyen et al. [44]	–		
Hoogendoorna et al. [45]	–	–	Primary care dataset originating from a network of general practitioners (GPs) centered around the Utrecht University Medical Center
Löprrich et al. [39]	–	Clinical report	Multiple myeloma research database
Mehrabi et al. [32]	–	–	Indiana University (IU) dataset Mayo Clinic dataset
Sippo et al. [42]	EMR	Breast imaging reports	–
Segagni et al. [31]	HIS and biobank	FSM pathology unit hospital biobank	FSM pathology unit database
Strauss et al. [46]	EMR	Pathology report	–

The Algorithm performance characteristics of the included articles are shown in Table 3. –: Indicates that the information are not reported in the included studies

imaging and pathology reports. In one article, the cancer registry, the Surveillance, Epidemiology, and End Results (SEER) registry data, pathology reports, and radiology reports were examined.

NLP type and algorithm performance

NLP type and Algorithm performance articles are shown in Table 4.

Extensive variations were observed in the software and algorithm evaluation methods used in the articles included in the present study. The reported precision in 14 articles was between 65 and 99% ($n=14$, 82%), sensitivity in 12 articles was between 57 and 100% ($n=12$, 70%), f1-score in 9 articles was between 45 and 99% ($n=9$, 52%), specificity in 3 articles was between 72 and 99% ($n=3$, 17%), and accuracy in 2 articles varied between 98 and 100% ($n=2$, 11%).

Table 4 NLP type and Algorithm performance characteristics of the included articles

Author	NLP type (algorithm)	Performance
Hammami et al. [36]	Developed by researchers (rule-based)	86.3% < = Sens < = 99.2% 85.9% < = Prec < = 99.2% 84.9% < = F1-score < = 99.2% 98.10% < = Acc < = 99.2%
Ryu et al. [37]	Developed by researchers (rule-based)	Sens = 100% 98.6 < = Prec < = 100 99.3 < = F1-score < = 100
Oliveira et al. [41]	Developed by researchers (rule-based + ML) + CLAMP software	–
Becker et al. [48]	Developed by researchers (semi-automated rule-based system)	Sens = 99.54% Prec = 97.95%
Wang et al. [40]	Developed by researchers (algorithms to generate final normalized concept names for each data element) + ML + Premade (Med Tagger)	0.982 < = Recall < = 1 0.885 < = Prec < = 1
Kumar et al. [33]	Developed by researchers (semantic similarity measures and clustering methods) + cTAKES	P value < = 0.05
Wadia et al. [43]	Premade (cTAKES)	Sens = 77.3 Spec = 72.5 Prec = 88.4 NPV = 54
Bustos et al. [38]	Developed by researchers ((Fast Text. CNN, SVM, KNN))	0.79 < = Sens < = 0.93 0.79 < = Prec < = 0.93 0.79 < = F1-score < = 0.93
Faina Linkov et al. [47]	Premade (TIES)	–
Sada et al. [34]	Premade (ARC = automated retrieval console)	0.94 < = Sens < = 0.96 0.68 < = Spec < = 0.97 0.75 < = PPV < = 0.96
Nguyen et al. [44]	Premade (Medtex) + Developed by researchers (rule-based)	Sens = 0.78 Prec = 0.83 F1-score = 0.80
Hoogendoorna et al. [45]	Developed by researchers (rule-based)	0.870 < = Accu < = 0.831
Löprrich et al. [39]	Developed by researchers (SVM—MEC) + German framed clinical text	0.89 < = F1-score < = 0.92
Mehrabi et al. [32]	Developed by researchers (rule-based)	87.8 < = Prec < = 88.1
Sippo et al. [42]	Premade (BROK)	Sens = 100% 96.6% < = Prec < = 100%
Segagni et al. [31]	Developed by researchers and premade (rule-based system for the onco-i2b2 + researcher-made algorithm)	–
Strauss et al. [46]	Premade (CoPathPlus) + Developed by researchers (Scent = rule-based)	<u>Breast</u> cancer 0.74 > Sens > 1 Spec = 0.99 0. > 90 Prec > 0.94 0.97 < NPV > 1 <u>Prostate</u> cancer 0.71 > Sens > 0.97 0.9 > 8 Spec < 0.99 0.88 > Prec > 0.97 0.95 > NPV > 0.99

Table 4 (continued)

NLP type and Algorithm performance articles are shown in Table 4.

Accu: Accuracy, Sens: Sensitivity, Spec: Specificity, Prec: Precision, SVM: support vector machines, Scent: SAS-based coding, extraction, and nomenclature tool, ML: Machine Learning.

Table 5 Terminological system and data standard exchange and cancer type characteristics of the included articles

Author	Terminological system	Data standard exchange	Cancer type
Hammami et al. [36]	ICD-O-M	–	All
Ryu et al. [37]	LOINC, ICD-O, SNOMED CT-OMOP- Local dictionary	–	Colon
Oliveira et al. [41]	–	–	Cervical and anal
Becker et al. [48]	UMLS	–	Colorectal
Wang et al. [40]	–	–	Lung
Kumar et al. [33]	SNOMED-CT – UMLS	–	Lung
Wadia et al. [43]	UMLS – SNOMED-CT, ICD-9, customized dictionary of additional terms added to SNOMED-CT	–	Lung
Bustos et al. [38]	–	XML	All
Faina Linkov et al. [47]	ICD_O, ICD_9		
ICD 10	NAACCR	Breast cancer	
Sada et al. [34]	ICD-9	–	Hepatocellular
Nguyen et al. [44]			All
Hoogendoorna et al. [45]	UMLS—SNOMED-CT	–	Colorectal
Löprrich et al. [39]	UMLS – open NLP- NP-channker – pos Tagger	–	All
Mehrabi et al. [32]	UMLS	HL7-CDA	Pancreatic
Sippo et al. [42]	–	Java	Breast
Segagni et al. [31]	SNOMED CT, malignant tumors (TNM) classifications	A web service that communicates with other cells through XML messages following REST standard	ALL
Strauss et al. [46]	SNOMED	–	Breast & Prostate

The Terminological system's characteristics of the included articles are shown in Table 5.

LOINC: Logical Observation Identifiers Names and Codes, ICD-O: International Classification of Diseases for Oncology, SNOMED CT: Systemized Nomenclature of Medicine – Clinical Terms, OMOP: Observational Medical Outcomes Partnership, CLAMP software: Clinical Language Annotation Modeling and Processing, UPMC: University of Pittsburgh Medical Center, TIES: Text Information Extraction System, NAACCR: North American Association of Central Cancer Registries, FSM: Fondazione Maugeri hospital, i2b2: Informatics for Integrating Biology and the Bedside, TNM: T describes the size of the tumor and any spread of cancer into nearby tissue; N describes the spread of cancer to nearby lymph nodes; and M describes metastasis, XML: an extensible markup language.

Terminological systems

Terminological System and Data standard exchange and Cancer type characteristics of the included articles in Table 5

The most frequently used terminologies were UMLS and SNOMED-CT. In six of the 17 articles, the data exchange standard was used for data transfer; in two articles, the HL7 standard; in two articles, the XML standard; in two articles, the JAVA standard; and in one article, the CDA standard was employed.

Cancer type

Of all the articles reviewed, 70% focused on a specific type of cancer, with breast cancer ($n=4$, 19%) and lung cancer ($n=4$, 19%) receiving the most attention.

Results from wordcloud analysis

The wordclouds of three variables (cancer types, algorithms, terminologies) are presented in Fig. 3. The wordclouds represents the most common terms used in the included articles. The more frequent a word, the bigger and more central its representation in the cloud.

Discussion

This study aimed to review and synthesize the results of the articles focusing on concept retrieval concerning cancer using NLP software. The most commonly used terminologies in the articles included in this study were SNOMED, SNOMED-CT, and UMLS. Studies that evaluated only one or more specific types of cancer provided data

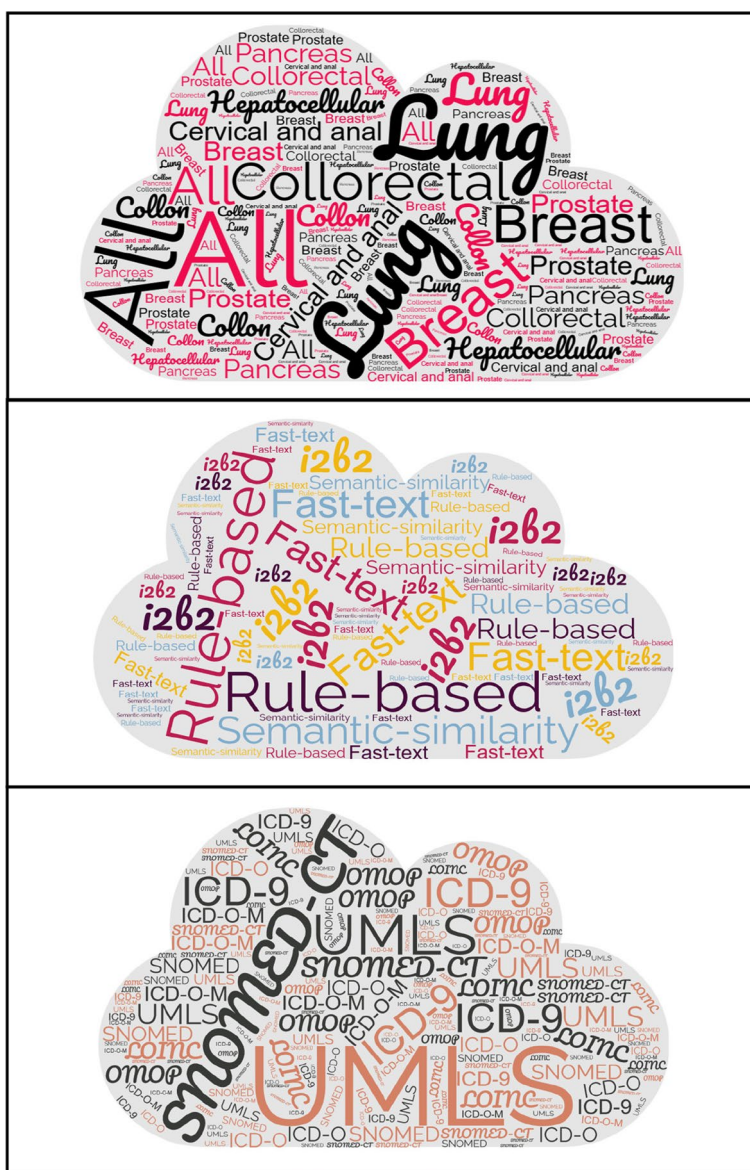


Fig. 3 A wordcloud view of extracted three variables (cancer types, algorithms, terminologies)

on high-prevalence cancers such as breast, colon, and lung cancer. Moreover, the most frequently used algorithm in the software development of these studies was the rule-based algorithm. In recent years, the number of studies that used NLP to retrieve and extract concepts and words has increased ($n=70\%$), which confirms a growth in the use of NLP in medicine [1, [49]. With the development of health information systems, electronic information registration, and electronic preparation of medical reports, the volume of textual data recorded in these systems has increased. The rise in the diversity and volume of data prompted researchers to use various techniques to retrieve these texts.

NLP applications provide a significant advantage via automation. They effectively reduce or even eliminate the need for manual narrative reviews, which makes it possible to assess vast amounts of data quickly. As a consequence, previously impractical tasks can be achieved. Furthermore, NLP can enhance clinical workflows by continuously monitoring and providing advice to healthcare professionals concerning reporting. The implementation of various NLP techniques varies among applications. Tokenization is a common feature of all systems, and stemming is common in most systems. A segmentation step is crucial in many systems, with almost half incorporating this step. However, limited performance improvement has been observed in studies incorporating syntactic analysis [50–52]. Instead, systems frequently enhance their performance through the utilization of attributes originating from semantic analysis. This approach usually involves a specialized lexicon to detect relevant terms and their synonyms. These lexicons are typically crafted manually by experts in a particular field, but they can also be integrated with pre-existing lexicons [53–58].

The results of our study showed that to retrieve concepts from electronic texts recorded in the field of cancer, researchers have employed several methods and algorithms. The rule-based algorithm was the most frequently used algorithm in the included studies. However, deep learning has been used more frequently in healthcare [30, 59]; in certain studies that have compared rule-based and machine learning algorithms, it has been observed that both rule-based algorithms and machine learning classifiers can demonstrate comparable performance when evaluated using the same dataset [60, 61]. In recent years, the popularity of machine learning algorithms has increased considerably, most likely due to their improved scalability and user-friendliness [62]. Despite the widespread adaption of deep learning methods, this study showed that both rule-based and traditional algorithms are still popular. A likely reason for this may be that these algorithms are simple and easier to implement and understand, as well as more interpretable compared to deep learning methods [63]. Interpretation of deep learning can be challenging because the steps that are taken to arrive at the final analytical output are not always as clear as those used in more traditional methods [63–65]. In addition, rule-based and traditional algorithms are more useful for smaller datasets with few features as these algorithms do not require massive amounts of data that are necessary for the development and successful implementation of machine learning. Furthermore, ML techniques can lead to a phenomenon known as overfitting, in which the developed model is too close to the underlying data set, which can limit the generalizability of the model to different data sets and making accurate predictions in other situations. However, this does not mean that using traditional algorithms is always a better approach

than using deep learning since some situations may require more flexible and complex techniques [63].

Despite considerable variety among the evaluation methods when using NLP algorithms that have been reported in previous studies and published articles [66, 67], most of the retrieved articles in our study used the recall (R), f1-score (F1), and precision (P) metrics, to evaluate the findings of the algorithms being investigated. The recall ranged from 0.71 to 1.0, the precision ranged from 0.75 to 1.0, and the f1-score ranged from 0.79 to 0.93. The present study included articles that used pre-developed software or software developed by researchers to interpret the text and extract the cancer concepts. Pons et al. [13] systematically reviewed articles that used image processing software to automatically encode radiology reports. Similar to our study, this review extracted concepts identified by included studies, the NLP methodology and tools used, and their application purpose and performance results.

The most commonly used terminologies in the articles were UMLS and SNOMED-CT, among which UMLS was utilized more frequently [30]. A study in 2020 showed that 42% of UMLS users were researchers, and 28% of terminology users were programmers and software developers. Both groups acknowledged that terminologies were used to find concepts in the texts and the relationship between terms [68]. In this study, the articles concerning the use of UMLS were divided into six categories, with more than half of the articles (about 78%) falling under the NLP category [68].

The use of SNOMED-CT terminology in implementations has increased in recent years, while its use in theoretical discussions has recently been reduced [69]. The results of our study also indicated the practical use of this terminology to retrieve concepts from medical texts or documents.

In 2013, a review paper [70] on the application of SNOMED-CT in 2001 and 2012 categorized the included articles into five groups: unknown, theoretical, development and design, implementation, and evaluation. In this review, the number of studies related to implementation was 44 out of 488 relevant articles, which was a small number compared to the total number of articles. However, in the study by Change et al. [69], 124 articles out of 622 addressed this topic, which shows the importance of this field and the attention it has received in recent years. Most of these articles focused on the classification or coding of free-text clinical notes/narratives and radiology reports.

Despite the importance of content coverage as a metric in the evaluation of terminological systems, most of the articles included in our review did not include this information in their results, and only five articles reported this information. The reason can be that the focus of the included studies has been more on the extraction of the concepts from the narrative and identification of the best algorithms rather than the evaluation of applied terminological systems. Usually, studies that have been conducted to evaluate terminological systems focused on their content coverage [71, 72].

Implication

The results of this study will help researchers to identify the most common techniques used to process cancer-related texts. This study also identified the terminologies that were mainly used to retrieve the concepts concerning cancer. The findings of this study

will assist software developers in identifying the most beneficial algorithms and terminologies to retrieve the concepts from narrative text.

Strengths and limitations of the study

In this article, in addition to examining NLP algorithms, we also reviewed the coding systems used for identifying concepts. This study had some limitations. We only searched for articles that were related to cancer-specific concepts. Studies that used the NLP technique in the field of cancer but extracted tumor features, such as tumor size, color, and shape, were excluded from the study. In addition, articles that used the results of tests and clinical examinations to diagnose cancer were also excluded. Articles that used AI and ML methods were also excluded from the study. One of the other limitations of this study was that due to the insufficiency of information concerning datasets used in the included studies, it was impossible to categorize studies based on the public and non-public nature of the datasets. Our contact with the authors of the articles did not reach any specific results. We suggest that future studies consider these limitations.

Conclusions

This systematic review was the first comprehensive evaluation of NLP algorithms applied to cancer concept extraction. Information extraction from narrative text and coding the concepts using NLP is a new field in biomedical, medical, and clinical fields. The results of this study showed UMLS and SNOMED-CT systems are the most used terminologies in the field of NLP for extracting cancer concepts. We have also reviewed NLP algorithms that help researchers retrieve cancer concepts and found that rule-based methods were the most frequently used techniques in this field. Considering that limited studies applied ML and deep learning algorithms to extract concepts from the narrative text, it is recommended that researchers focus on the application of these methods in information extraction and synthesize the results of these types of studies. In addition, in the future, researchers can compare the results of natural language processing software to extract the concepts of various diseases from clinical documents such as radiology or laboratory reports. Moreover, as most of the included studies had not reported the content coverage of the applied terminological systems, future studies should address this type of results as it can help developers of the systems to choose the right terminological system with proper coverage.

Abbreviations

NLP Natural language processing
M Machine learning

Acknowledgements

Not applicable.

Author contributions

All authors took part in the entire study and approved the final manuscript. MGh and LA contributed to the study design. PA, MGh, and SH conducted data extraction. PA and MGh drafted the manuscript. RKh, LA, and SH critically revised the manuscript for important intellectual content.

Funding

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

All methods were performed in accordance with the relevant guidelines and regulations.

Consent for publication

Not applicable.

Competing interests

None of the authors of this study have personal, professional, or financial conflicts of interest to declare.

Received: 13 December 2022 Accepted: 13 September 2023

Published online: 29 October 2023

References:

1. Organization WH. Global Health Estimates 2019: deaths by cause, age, sex, by country and by region, 2000–2019. Genf, Geneva: World Health Organization; 2020.
2. Torre LA, et al. Global cancer statistics. *CA*. 2015;65(2):87–108.
3. Sung H, et al., Global cancer statistics GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA* 2021; 71(3): p. 209–249.
4. Allemani C, et al. Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *The Lancet*. 2018;391(10125):1023–75.
5. Acs B, Rantalainen M, Hartman J. Artificial intelligence as the next step towards precision pathology. *J Intern Med*. 2020;288(1):62–81.
6. Crowley RJ, Tan YJ, Ioannidis JP. Empirical assessment of bias in machine learning diagnostic test accuracy studies. *J Am Med Inform Assoc*. 2020;27(7):1092–101.
7. Saberi-Movahed F, et al. Dual regularized unsupervised feature selection based on matrix factorization and minimum redundancy with application in gene selection. *Knowl-Based Syst*. 2022;256: 109884.
8. Azadifar S, et al. Graph-based relevancy-redundancy gene selection method for cancer diagnosis. *Comput Biol Med*. 2022;147: 105766.
9. Afolayan JO, et al. Breast cancer detection using particle swarm optimization and decision tree machine learning technique. In: *Intelligent Healthcare: Infrastructure, Algorithms and Management*. Springer; 2022. p. 61–83.
10. Izci H, et al. A systematic review of estimating breast cancer recurrence at the population level with administrative data. *JNCI*. 2020;112(10):979–88.
11. Yang L, et al. Prediction model of the response to neoadjuvant chemotherapy in breast cancers by a Naive Bayes algorithm. *Comput Methods Programs Biomed*. 2020;192: 105458.
12. Takada M, et al. Prediction of postoperative disease-free survival and brain metastasis for HER2-positive breast cancer patients treated with neoadjuvant chemotherapy plus trastuzumab using a machine learning algorithm. *Breast Cancer Res Treat*. 2018;172(3):611–8.
13. Kourou K, et al. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J*. 2015;13:8–17.
14. Yim W-W, et al. Natural language processing in oncology: a review. *JAMA Oncol*. 2016;2(6):797–804.
15. Hirschberg J, Manning CD. Advances in natural language processing. *Science*. 2015;349(6245):261–6.
16. Pons E, et al. Natural language processing in radiology: a systematic review. *Radiology*. 2016;279(2):329–43.
17. Popejoy LL, et al. Quantifying care coordination using natural language processing and domain-specific ontology. *J Am Med Inform Assoc*. 2015;22(e1):e93–103.
18. Flynn RW, et al. Automated data capture from free-text radiology reports to enhance accuracy of hospital inpatient stroke codes. *Pharmacoepidemiol Drug Saf*. 2010;19(8):843–7.
19. Kung R, et al., Mo1043 a natural language processing Algorithm for identification of patients with cirrhosis from electronic medical records. *Gastroenterology*, 2015. 148(4): S-1071-S-1072.
20. Yang H, et al. A text mining approach to the prediction of disease status from clinical discharge summaries. *J Am Med Inform Assoc*. 2009;16(4):596–600.
21. Si Y, Roberts K. A frame-based NLP system for cancer-related information extraction. in *AMIA annual symposium proceedings*. 2018. American Medical Informatics Association.
22. Savova GK, et al. DeepPhe: a natural language processing system for extracting cancer phenotypes from clinical records. *Can Res*. 2017;77(21):e115–8.
23. Ping X-O, et al. Information extraction for tracking liver cancer patients' statuses: from mixture of clinical narrative report types. *TELEMEDICINE and e-HEALTH*. 2013;19(9):704–10.
24. Kehl KL, et al. Natural language processing to ascertain cancer outcomes from medical oncologist notes. *JCO Clin Cancer Informat*. 2020;4:680–90.
25. Casey A, et al. A systematic review of natural language processing applied to radiology reports. *BMC Med Inform Decis Mak*. 2021;21(1):179.
26. Santos T, et al. Automatic classification of cancer pathology reports: a systematic review. *J Pathol Informat*. 2022;13: 100003.
27. Mellia JA, et al. Natural language processing in surgery: a systematic review and meta-analysis. *Ann Surg*. 2021;273(5):900–8.
28. Le Glaz A, et al. Machine learning and natural language processing in mental health: systematic review. *J Med Internet Res*. 2021;23(5): e15708.

29. Sheikhalishahi S, et al. Natural language processing of clinical notes on chronic diseases: systematic review. *JMIR Med Inform.* 2019;7(2): e12239.
30. Casey A, et al. A systematic review of natural language processing applied to radiology reports. *BMC Med Inform Decis Mak.* 2021;21(1):1–18.
31. Segagni D, et al. An ICT infrastructure to integrate clinical and molecular data in oncology research. *BMC Bioinformatics.* 2012;13(4):1–8.
32. Mehrabi S, et al. Identification of patients with family history of pancreatic cancer-Investigation of an NLP System Portability. *Stud Health Technol Informat.* 2015;216:604.
33. Kumar N, et al. Identifying associations between somatic mutations and clinicopathologic findings in lung cancer pathology reports. *Methods Inf Med.* 2018;57(01/02):63–73.
34. Sada Y, et al. Validation of case finding algorithms for hepatocellular cancer from administrative data and electronic health records using natural language processing. *Med Care.* 2016;54(2): e9.
35. Becker M, et al. Natural language processing of German clinical colorectal cancer notes for guideline-based treatment evaluation. *Int J Med Informatics.* 2019;127:141–6.
36. Hammami L, et al. Automated classification of cancer morphology from Italian pathology reports using natural language processing techniques: A rule-based approach. *J Biomed Inform.* 2021;116: 103712.
37. Ryu B, et al. Transformation of pathology reports into the common data model with oncology module: use case for colon cancer. *J Med Internet Res.* 2020;22(12): e18526.
38. Bustos A, Pertusa A. Learning eligibility in cancer clinical trials using deep neural networks. *Appl Sci.* 2018;8(7):1206.
39. Löpprich M, et al. Automated classification of selected data elements from free-text diagnostic reports for clinical research. *Methods Inf Med.* 2016;55(04):373–80.
40. Wang L, et al. Natural language processing for populating lung cancer clinical research data. *BMC Med Inform Decis Mak.* 2019;19(5):1–10.
41. Oliveira CR, et al. Natural language processing for surveillance of cervical and anal cancer and precancer: algorithm development and split-validation study. *JMIR Med Inform.* 2020;8(11): e20826.
42. Sippo DA, et al. Automated extraction of BI-RADS final assessment categories from radiology reports with natural language processing. *J Digit Imaging.* 2013;26(5):989–94.
43. Wadia R, et al. Comparison of natural language processing and manual coding for the identification of cross-sectional imaging reports suspicious for lung cancer. *JCO Clin Cancer Informat.* 2018;2:1–7.
44. Nguyen AN, et al. Assessing the utility of automatic cancer registry notifications data extraction from free-text pathology reports. in *AMIA annual symposium proceedings.* 2015. American Medical Informatics Association.
45. Hoogendoorn M, et al. Utilizing uncoded consultation notes from electronic medical records for predictive modeling of colorectal cancer. *Artif Intell Med.* 2016;69:53–61.
46. Strauss JA, et al. Identifying primary and recurrent cancers using a SAS-based natural language processing algorithm. *J Am Med Inform Assoc.* 2013;20(2):349–55.
47. Linkov F, et al. Integration of cancer registry data into the text information extraction system: leveraging the structured data import tool. *J Pathol Informat.* 2018;9(1):47.
48. Mamlin BW, Heinze DT, McDonald CJ. Automated extraction and normalization of findings from cancer-related free-text radiology reports. In: *AMIA Annual Symposium Proceedings.* 2003. American Medical Informatics Association.
49. Locke S, et al. Natural language processing in medicine: a review. *Trends Anaesthesia Crit Care.* 2021;38:4–9.
50. Yetisgen-Yildiz M, et al. A text processing pipeline to extract recommendations from radiology reports. *J Biomed Inform.* 2013;46(2):354–62.
51. Yetisgen-Yildiz M, et al. Automatic identification of critical follow-up recommendation sentences in radiology reports. In: *AMIA Annual Symposium Proceedings.* 2011. American Medical Informatics Association.
52. Garla V, et al. The Yale cTAKES extensions for document classification: architecture and application. *J Am Med Inform Assoc.* 2011;18(5):614–20.
53. Pham A-D, et al. Natural language processing of radiology reports for the detection of thromboembolic diseases and clinically relevant incidental findings. *BMC Bioinformatics.* 2014;15(1):1–10.
54. Percha B, et al. Automatic classification of mammography reports by BI-RADS breast tissue composition class. *J Am Med Inform Assoc.* 2012;19(5):913–6.
55. Sohn S, et al. Identifying abdominal aortic aneurysm cases and controls using natural language processing of radiology reports. *AMIA Summit Transl Sci Proc.* 2013;2013:249.
56. Rubin D, et al. Natural language processing for lines and devices in portable chest x-rays. In: *AMIA Annual Symposium Proceedings.* 2010. American Medical Informatics Association.
57. Gerstmair A, et al. Intelligent image retrieval based on radiology reports. *Eur Radiol.* 2012;22:2750–8.
58. Do BH, et al. Automatic retrieval of bone fracture knowledge using natural language processing. *J Digit Imaging.* 2013;26:709–13.
59. Shickel B, et al. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE J Biomed Health Inform.* 2017;22(5):1589–604.
60. Solti I, et al. Automated classification of radiology reports for acute lung injury: comparison of keyword and machine learning based natural language processing approaches. In: *2009 IEEE international conference on bioinformatics and biomedicine workshop.* 2009. IEEE.
61. Chapman WW, et al. A comparison of classification algorithms to automatically identify chest X-ray reports that support pneumonia. *J Biomed Inform.* 2001;34(1):4–14.
62. van Rijsbergen CJ, Lalmas M. Information calculus for information retrieval. *J Am Soc Inf Sci.* 1996;47(5):385–98.
63. Rajula HSR, et al. Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. *Medicina.* 2020;56(9):455.
64. Rudin C, et al. Interpretable machine learning: Fundamental principles and 10 grand challenges. *Stat Surv.* 2022;16:1–85.
65. Con D, van Langenberg DR, Vasudevan A. Deep learning vs conventional learning algorithms for clinical prediction in Crohn's disease: a proof-of-concept study. *World J Gastroenterol.* 2021;27(38):6476.

66. Dash TK, et al. Gradient boosting machine and efficient combination of features for speech-based detection of COVID-19. *IEEE J Biomed Health Inform.* 2022;26(11):5364–71.
67. Dash TK, et al., Mitigating information interruptions by COVID-19 face masks: a three-stage speech enhancement scheme. *IEEE Trans Comput Soc Syst*, 2022.
68. Amos L, et al. UMLS users and uses: a current overview. *J Am Med Inform Assoc.* 2020;27(10):1606–11.
69. Chang E, Mostafa J. The use of SNOMED CT, 2013–2020: a literature review. *J Am Med Inform Assoc.* 2021;28(9):2017–26.
70. Lee D, et al. Literature review of SNOMED CT use. *J Am Med Inform Assoc.* 2014;21(e1):e11–9.
71. Goss FR, et al. Evaluating standard terminologies for encoding allergy information. *J Am Med Inform Assoc.* 2013;20(5):969–79.
72. Elkin PL, et al. Evaluation of the content coverage of SNOMED CT: ability of SNOMED clinical terms to represent clinical problem lists. in *Mayo Clinic Proceedings*. 2006. Elsevier.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

