
Review

Clinical Decision Support and Natural Language Processing in Medicine: Systematic Literature Review

Hans Eguia^{1,2*}, MSc, MD; Carlos Luis Sánchez-Bocanegra^{2*}, PhD; Franco Vinciarelli^{1,3*}, ITE; Fernando Alvarez-Lopez^{4*}, PhD; Francesc Saigí-Rubió^{2*}, PhD

¹SEMERGEN New Technologies Working Group, Madrid, Spain

²Faculty of Health Sciences, Universitat Oberta de Catalunya (UOC), Barcelona, Spain

³Emergency Hospital Clemente Álvarez, Rosario (Santa Fe), Argentina

⁴Faculty of Health Sciences, Universidad de Manizales, Manizales, Colombia

* all authors contributed equally

Corresponding Author:

Francesc Saigí-Rubió, PhD

Faculty of Health Sciences

Universitat Oberta de Catalunya (UOC)

Rambla del Poblenou, 156

Barcelona, 08018

Spain

Phone: 34 933263622

Email: fsaigi@uoc.edu

Abstract

Background: Ensuring access to accurate and verified information is essential for effective patient treatment and diagnosis. Although health workers rely on the internet for clinical data, there is a need for a more streamlined approach.

Objective: This systematic review aims to assess the current state of artificial intelligence (AI) and natural language processing (NLP) techniques in health care to identify their potential use in electronic health records and automated information searches.

Methods: A search was conducted in the PubMed, Embase, ScienceDirect, Scopus, and Web of Science online databases for articles published between January 2000 and April 2023. The only inclusion criteria were (1) original research articles and studies on the application of AI-based medical clinical decision support using NLP techniques and (2) publications in English. A Critical Appraisal Skills Programme tool was used to assess the quality of the studies.

Results: The search yielded 707 articles, from which 26 studies were included (24 original articles and 2 systematic reviews). Of the evaluated articles, 21 (81%) explained the use of NLP as a source of data collection, 18 (69%) used electronic health records as a data source, and a further 8 (31%) were based on clinical data. Only 5 (19%) of the articles showed the use of combined strategies for NLP to obtain clinical data. In total, 16 (62%) articles presented stand-alone data review algorithms. Other studies (n=9, 35%) showed that the clinical decision support system alternative was also a way of displaying the information obtained for immediate clinical use.

Conclusions: The use of NLP engines can effectively improve clinical decision systems' accuracy, while biphasic tools combining AI algorithms and human criteria may optimize clinical diagnosis and treatment flows.

Trial Registration: PROSPERO CRD42022373386; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=373386

(*J Med Internet Res* 2024;26:e55315) doi: [10.2196/55315](https://doi.org/10.2196/55315)

KEYWORDS

artificial intelligence; AI; natural language processing; clinical decision support system; CDSS; health recommender system; clinical information extraction; electronic health record; systematic literature review; patient; treatment; diagnosis; health workers

Introduction

Advancement in medicine continues apace, especially with the emergence of new pathologies such as COVID-19. New treatments are continually being developed to fight not only these diseases but also previous pathologies for which new alternatives are being developed. Consequently, the number of publications in different indexed journals has increased, as shown in search results in various databases such as PubMed. Currently, it is possible to find many articles that mention new treatments or even new diagnostic forms [1,2].

Real and verified information is vital for the treatment and diagnosis of patients and is the cornerstone of medicine. The National Library of Medicine has developed at least 3 major source evaluation systems that provide useful examples for the task at hand: MEDLINE indexing, MedlinePlus indexing, and the Disaster Lit database [3].

Many health workers use the internet to search generally for updated clinical data [4]. However, this method is not the most efficient way to find information, since physicians must determine the type of information they need and then conduct the search themselves in an online medical database. This type of search can not only be time-consuming but also error prone due to not using suitable data. Therefore, automated information recommender systems have been established as a solution that allows medical staff to obtain reliable knowledge very quickly. These types of solutions are known as clinical decision support systems (CDSSs) [5].

CDSSs are composed of multiple platforms that allow the assessment of clinical data and alert clinicians to eventual problems. In addition, decision-making tools can be used to assist clinical staff. For these systems to function properly, they must interact with elements that allow them to obtain updated data for improved development, such as electronic health records (EHRs) [6]. Accordingly, CDSSs are known to focus on 6 specific aspects: data, knowledge, inference, architecture and technology, implementation and integration, and the user [7].

All available technology and tools (eg, artificial intelligence [AI], machine learning, and big data) could be useful for obtaining high-quality, reliable information. Such information could also be obtained by taking a supervised machine learning approach using several natural language processing (NLP) components that are domain independent and related to medical information extraction (text mining) [8]. These resources could include medical sources such as the Unified Medical Language System, different metathesauri, and different medical ontologies.

This study aims to answer the question of whether AI- and NLP-based CDSSs can provide effective results in automated searches that are useful to health care staff. To this end, a systematic review was carried out to assess the current state of these techniques in health care to identify their potential use in EHRs and automated information searches. The results found and conclusions drawn about the research question are subsequently presented.

Methods

Study Design

The protocol for this systematic review was published on November 5, 2022, in PROSPERO (CRD42022373386). This systematic review was performed per the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [9]. A search was conducted in the PubMed, Embase, ScienceDirect, Scopus, and Web of Science online databases for articles published between January 2000 and April 2023 using combinations of the following Medical Subject Headings (MeSH) terms: (((Artificial Intelligence [MeSH Terms]) AND (Natural language processing [MeSH Terms])) AND (Clinical decision support [MeSH Terms])) AND (Electronic health record [MeSH Terms]). The snowballing technique was used to complement the search to find the articles most relevant to the study [10].

Selection Criteria

In total, 2 researchers independently assessed titles and abstracts and analyzed appropriate studies through full-text evaluation. The only inclusion criteria were (1) original research articles and studies on the application of AI-based medical clinical decision support (CDS) using NLP techniques and (2) publications in English. The exclusion criteria were (1) studies describing the use of AI that are not focused on CDS tools; (2) studies related only to NLP; (3) studies related to an algorithm submitted to a challenge; (4) letters to the editor; (5) conference abstracts, books and book reviews; and (6) studies not published in scientific journals (ie, only in science magazines or magazines without a DOI).

Data Extraction and Management

Data were collected as follows: (1) reference, country, and year; (2) objective; (3) study type; (4) research design—intervention; (5) population sample + target (organ); and (6) results and conclusions. Further, 2 researchers independently extracted data. A third investigator resolved discrepancies.

Quality Appraisal of the Studies

The articles were independently assessed by 2 researchers. Disagreements were discussed until a consensus was reached. A Critical Appraisal Skills Programme (CASP) tool for qualitative studies with a 10-item scale (0-10) [11] was used to ensure the quality of the studies, focusing on (1) validity of the study, (2) accuracy of the results, and (3) transferability. A 10-item CASP scale (0-10) was used for systematic reviews, focusing on (1) validity of the study, (2) robustness and relevance of the findings, and (3) applicability and relevance of the results in a local or specific context. Quality appraisal was used to demonstrate the methodological quality of the studies since it would affect the validity of the results and was something that needed to be taken into account when considering the findings of the review.

Ethical Considerations

This study relied on secondary data. No ethics approval or patient consent was therefore required.

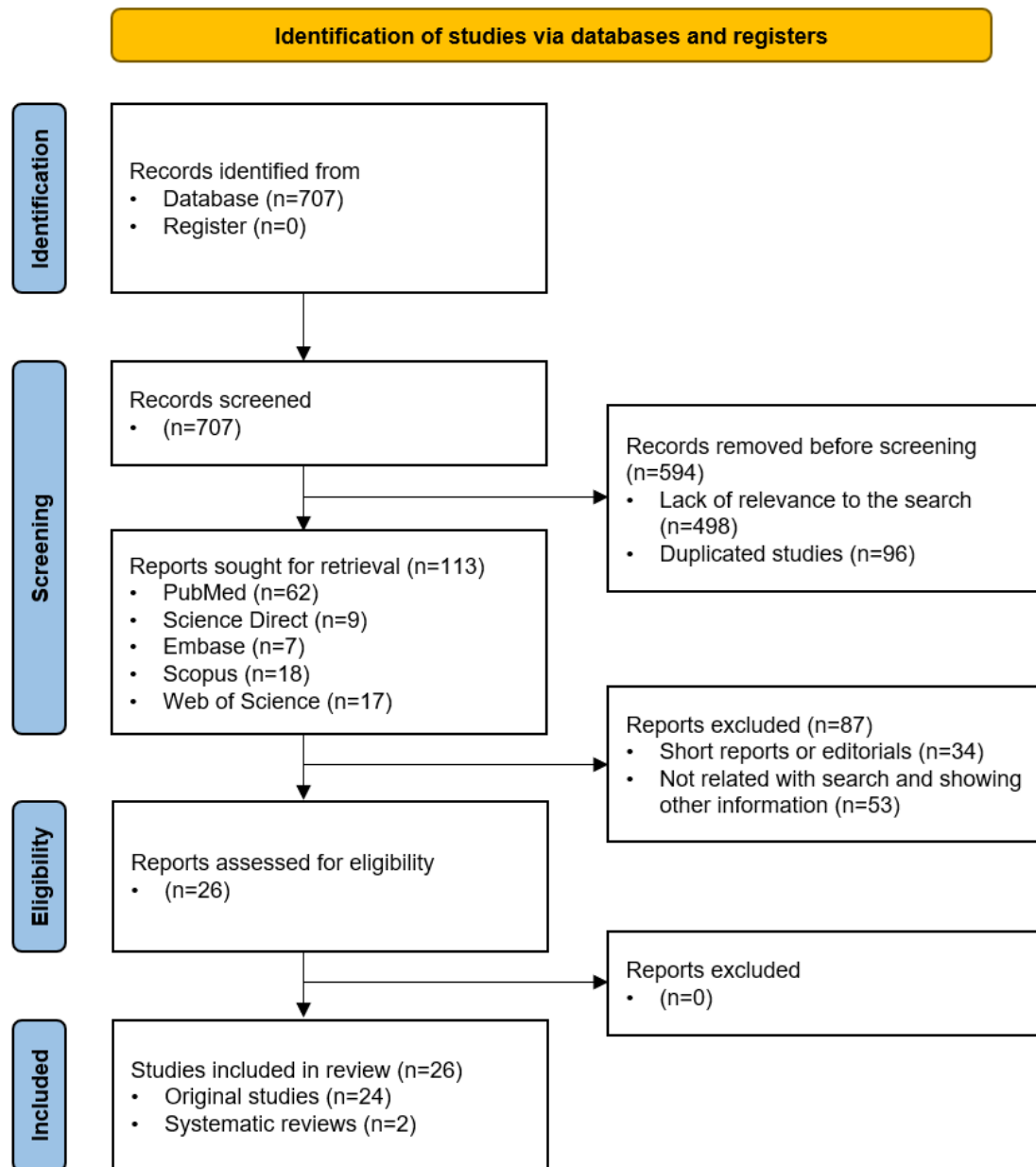
Results

Overview

A systematic review was conducted with the aim of assessing the current state of AI and NLP techniques in health care to identify their potential use in EHRs and automated information searches. In the initial search, 707 articles were retrieved. In title and abstract screening, 594 publications were excluded

either due to their lack of relevance to the search or duplication. After the initial review, 113 articles were chosen for further examination: 62 from PubMed, 9 from ScienceDirect, 7 from Embase, 18 from Scopus, and 17 from Web of Science. Of the remaining studies, 87 were excluded as they showed examples of data mining and algorithms for a challenge, presented nonscientific stories, or gave ultrashort presentations, among others. Therefore, 26 articles were included in the final analysis. The overview flowchart is shown in Figure 1.

Figure 1. PRISMA 2020 flow diagram for new systematic reviews that include searches of databases and registers only. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.



Characteristics of the Included Studies

The characteristics of the included studies are reported in Table 1 and Multimedia Appendix 1. In total, 24 of the evaluated articles were original articles and 2 were systematic reviews

[12,13]. It should be noted that 6 of the original articles reviewed were presentations of studies conducted in relation to the 2010 i2b2/VA Workshop on Natural Language Processing Challenges for Clinical Records [8,14-18].

Table 1. Summary of the studies included in the systematic review (n=26).

Study	Objective	Research design—intervention	Population sample + target (organ)	Results and conclusions
Clark et al [14]	To develop a system for determining the assertion status of clinical reports (extracted from patient records)	<ul style="list-style-type: none"> Determining what kind of clinical assertion was made in a clinical record (problem concept-sorted). 3 assertions: conditional hypothetical and not associated with the patient 	<ul style="list-style-type: none"> 2010 i2b2^a/VA^b: 349 clinical records. Detection of negation, speculation, and conditionals 	<ul style="list-style-type: none"> Linguistic information (negation status and temporal attributes) are features that are not always evaluated. They could give a more accurate classification
Patrick et al [15]	To present a method to deal with different extractions of, and classifications in, clinical data	<ul style="list-style-type: none"> Developing clinical NLP^c that includes a proofreading process with validation and correction 	<ul style="list-style-type: none"> 2010 i2b2/VA: 349 manually annotated notes + 827 raw records: 479 records 	<ul style="list-style-type: none"> NLP in clinical fields needs to address the issues between model complexity and model accuracy
Roberts and Harabagiu [8]	To develop a framework that can optimally identify medical concepts and adequately classify assertions	<ul style="list-style-type: none"> Evaluating and finding the near-optimal subset of features for the 2010 i2b2/VA context task (concept extraction) 	<ul style="list-style-type: none"> 2010 i2b2/VA: 827 summaries 	<ul style="list-style-type: none"> Machine-learning-based supervised methods can be improved with feature-selection techniques
Jiang et al [16]	To design and evaluate a machine-learning algorithm to extract clinical information from hospital discharge summaries	<ul style="list-style-type: none"> Evaluating the effects of different types of features 	<ul style="list-style-type: none"> 2010 i2b2/VA: 349 clinical annotated notes 	<ul style="list-style-type: none"> A system that is useful for concept extraction and assertion
D'Avolio et al [17]	To evaluate if the use of NLP-derived features combined with supervised machine-learning can perform effectively across tasks	<ul style="list-style-type: none"> Creating an engine using cTAKES^d (“noun-phrase bumpers”) for text extraction with NLP combined with machine-learning classification 	<ul style="list-style-type: none"> 2010 i2b2/VA: 349 clinical annotated notes 	<ul style="list-style-type: none"> NLP clinical extraction is well known. Its adoption with acceptable concept-level extraction performance is also known
Garla et al [19]	To develop a classifier extractor using cTAKES to store documents	<ul style="list-style-type: none"> Exploring if the accuracy of rule-based classifiers could be improved with machine-learning algorithms (testing systems after improvement) 	<ul style="list-style-type: none"> 13,000 abdominal radiology reports, ascites, and liver masses of 395 patients 	<ul style="list-style-type: none"> cTAKES extensions simplify feature extraction—it enables a better and more efficient exploration of text including the syntactic structure of documents and the negation context of concepts
Waghlikar et al [20]	To develop a CDSS for cervical cancer screening that can interpret free text Papanicolaou reports	<ul style="list-style-type: none"> Developing a CDSS^e that would be able to identify abnormal Papanicolaou reports as well as interpret other variables accurately 	<ul style="list-style-type: none"> 74 patient cases from cervical cancer screening 	<ul style="list-style-type: none"> EHR^f text can be effectively used through NLP to develop CDS^g tools.
Waghlikar et al [21]	To evaluate a CDSS for cervical cancer screening	<ul style="list-style-type: none"> A previous CDSS was created, care providers' recommendations were requested and compared with the CDSS system 	<ul style="list-style-type: none"> 6053 random patients and 8 reviewers. Cervical cancer screening 	<ul style="list-style-type: none"> CDS and human review can facilitate the evaluation of accuracy and would help to identify decision scenarios that may be missed

Study	Objective	Research design—intervention	Population sample + target (organ)	Results and conclusions
Sordo et al [22]	To develop a conceptual schema to represent clinical knowledge for decision support	<ul style="list-style-type: none"> • Rule interactions based on semantic web technologies (ontologies). Evaluating the feasibility of the modelling strategy by implementing test scenarios 	<ul style="list-style-type: none"> • 3 applications were analyzed: laboratory results; adverse drug reaction monitoring and immunization protocols 	<ul style="list-style-type: none"> • A richer conceptual representation of production rules will facilitate rule authoring consistency in rule implementation and maintenance. • Interactions are not considered a modelling time but they should be
Mehrabi et al [23]	To develop an NLP system to identify (retrospective) patients with pancreatic cysts	<ul style="list-style-type: none"> • Records were evaluated by the system created. This system double checks negation in the EHR and stratifies the results (low or high risk of cancer) 	<ul style="list-style-type: none"> • Test set of 703 patient records: pancreas and cyst. Training set: 316 control records 	<ul style="list-style-type: none"> • Identification of negation improved the algorithm, increasing recall to 98.12%
Ross et al [12]	To use secondary data and to put data into practice	<ul style="list-style-type: none"> • Paper-based materials containing similar information. Finding examples of BD^h use with EHRs 	<ul style="list-style-type: none"> • 84 papers—EHR research (various organs) 	<ul style="list-style-type: none"> • Greater use of EHRs means more clinical data. There is a need to develop new methods to obtain better data from EHRs
Kotfila and Uzuner [18]	To evaluate if NLP techniques can identify phenotypes in unstructured medical notes	<ul style="list-style-type: none"> • Training support vector machines using individual feature spaces to assess better model performance 	<ul style="list-style-type: none"> • 2 separate i2b2 shared tasks (2008+2014: 790+730 documents) 	<ul style="list-style-type: none"> • A simple feature performed as well as a combination of feature spaces—the addition of training data has weak statistical significance effects
Patterson et al [24]	To determine whether a colonoscopy was performed for screening	<ul style="list-style-type: none"> • Creating a rule-based model and 3 machine-learning models (automated retrieval console) which also uses cTAKES 	<ul style="list-style-type: none"> • A set of 1000 documents with colonoscopies 	<ul style="list-style-type: none"> • A rule-based classification system appears to be more robust than a machine-learning system alone
Divita et al [25]	To use an NLP algorithm to process large corpora of clinical notes to demonstrate a time decrease in the analyses of a large corpus of clinical information	<ul style="list-style-type: none"> • Developing an NLP algorithm to detect urinary catheters in hospitalized patients 	<ul style="list-style-type: none"> • 550,000 notes from patients with urinary catheters from a US Veteran Affairs hospital 	<ul style="list-style-type: none"> • Generalizable and widely applicable techniques can aid BD processing and analysis
Mei et al [26]	To evaluate a decision fusion framework for treatment recommendation systems (combining knowledge-driven and data-driven decisions)	<ul style="list-style-type: none"> • Implementing a fusion engine that obtains inputs from base decision engines 	<ul style="list-style-type: none"> • 3150 records from patients with diabetes in China 	<ul style="list-style-type: none"> • Decision fusion is a promising way toward a more valuable treatment recommendation
Marco-Ruiz et al [27]	To find out if multidisciplinary leverage archetypes and ontologies can model CDSS (better reuse and maintenance)	<ul style="list-style-type: none"> • An archetype was developed and reviewed by 5 practitioners and 5 information architects. • SNOMED for concept simplification 	<ul style="list-style-type: none"> • 126 cases with respirator and symptoms and signs 	<ul style="list-style-type: none"> • Model archetypes can provide information using ontologies. • EHR standards and terminologies can be used by multidisciplinary teams effectively
Danger et al [28]	To explain the methodology for constructing a clinical prediction rule repository	<ul style="list-style-type: none"> • Using data mining algorithms for pattern extraction 	<ul style="list-style-type: none"> • Clinical collection for 126,931 patients from primary care with multiple diagnoses 	<ul style="list-style-type: none"> • Important factors to be included in the rules were defined. With the use of clinical evidence rules, a better diagnostic recommendation can be obtained

Study	Objective	Research design—intervention	Population sample + target (organ)	Results and conclusions
Breischneider et al [29]	To introduce a system for the automated processing of clinical reports of patients with mamma carcinoma to extract relevant textual features and derive therapy suggestions	<ul style="list-style-type: none"> Use of a rule-based grammar approach for information extraction and deriving therapy suggestions based on the extracted variables and clinical guidelines. The accuracy of predicting therapeutic measures is also reported 	<ul style="list-style-type: none"> 8766 clinical text reporting on 2096 patients with mamma carcinoma 	<ul style="list-style-type: none"> Good accuracy of information extraction for different features. The evaluation of therapy suggestions, factors influencing the error between suggested and conducted therapy
Yang et al [30]	To use convolutional neural networks to be able to potentiate extraction without paper construction of rules or knowledge bases	<ul style="list-style-type: none"> Proposing a method using neural networks to extract features from EHRs. A tool was used in medical text showing 98.7% accuracy and 96.02% recall 	<ul style="list-style-type: none"> 18,590 EHRs indicating hypertension, diabetes, COPDⁱ, gout, arrhythmia, asthma, gastritis, and stomach polyps. After validation: a training set of 7000 EHRs and a test set of 400 EHRs 	<ul style="list-style-type: none"> The use of neural networks can increase information extraction. This could assist in a feasible and effective diagnosis
Wissel et al [31]	To validate an NLP application using machine-learning to identify patients for epilepsy surgery	<ul style="list-style-type: none"> Determining the proportion of potential surgical candidates using EHRs, NLP, and a surgical score 	<ul style="list-style-type: none"> 1 app and 4211 patients with an epilepsy diagnosis after surgical treatment of epilepsy and controls 	<ul style="list-style-type: none"> NLP-based app modelled for patient score to identify candidates for epilepsy surgery
Wulff et al [32]	To design a tool to automatically extract important information from medical texts and transform them into standardized data	<ul style="list-style-type: none"> NLP consisting of 5 successive tasks was developed. The tool was implemented to extract fragments from EHRs (97% precision and 94% recall) 	<ul style="list-style-type: none"> 529 medical events from the pediatric cardiology and the intensive care medicine department at Hannover Medical School. Further, 499 events were correctly extracted 	<ul style="list-style-type: none"> NLP-based solution to extract information from medical text. This will benefit health care by analyzing a large amount of information and presenting the most important data
Kulchak et al [33]	To evaluate if modification of a CDS tool using concepts of human-centered design can improve CDS itself	<ul style="list-style-type: none"> Adapting a DDSS^j (design, NLP with cTAKES, and coordination with EHRs). Creating simulated cases using real but deidentified clinical information (varying complexity) 	<ul style="list-style-type: none"> 3 simulated genetic cases of increasing complexity were created, 3 clinicians tested the system and another 3 retested it 	<ul style="list-style-type: none"> Evidence of potential implementation of DDSS using machine-assisted medical record review. The use of mixed technical CDS and human-centered design criteria enhances the real-world clinical workflow
van de Burgt et al [13]	To assess whether data mining can improve the diagnostic and therapeutic processes of CDS	<ul style="list-style-type: none"> Combining data mining and CDSSs may improve diagnostic and therapeutic processes, contributing to increased patient safety 	<ul style="list-style-type: none"> 714 reviewed publications, 39 of which were included 	<ul style="list-style-type: none"> Barriers: complexity of NLP, EHR incompleteness, validation and performance of the tool, lack of input from an expert team, and the adoption rate among professionals
Suh et al [34]	To evaluate the use of clinical NLP to identify elements relevant to preoperative medical history by analyzing clinical notes	<ul style="list-style-type: none"> An NLP pipeline would identify a significant portion of pertinent history captured by a perioperative provider (in an EHR) 	<ul style="list-style-type: none"> 93 patients with 9765 clinical notes were reviewed 	<ul style="list-style-type: none"> The use of NLP identified medical conditions relevant to preanesthetic evaluation from unstructured free-text input. This provides CDS or recommends additional testing or evaluation
Park et al [35]				

Study	Objective	Research design—intervention	Population sample + target (organ)	Results and conclusions
	To evaluate the efficacy of an algorithm of 3 levels of search functionality in supporting information retrieval for clinical users from EHR in a simulated clinical environment	<ul style="list-style-type: none"> Medical tasks were presented to an evaluation system and a panel of medical raters to assess both the speed and accuracy of task completion 	<ul style="list-style-type: none"> Medical records for 3 fictional patients, each with 20 documents, were assessed by 60 clinicians of varying specialties and experience levels 	<ul style="list-style-type: none"> NLP-enhanced search facilitated more accurate task completion than both string search and no search, with similar task speeds between NLP-enhanced search and string search
Afshar et al [36]	To implement a real-time NLP-driven CDS tool for screening opioid misuse in hospitalized adults and assess its effectiveness in providing interventions for substance use disorder treatment	<ul style="list-style-type: none"> A 30-month quasi-experimental pre-post study screened inpatients for 24 months of standard care followed by a 6-month tool implementation phase. Cost-effectiveness analysis and user acceptability evaluations were covered 	<ul style="list-style-type: none"> A total sample size of 12,500 patients (10,000 preintervention and 2500 postintervention) 	<ul style="list-style-type: none"> The study implemented a real-time NLP-driven CDS tool for screening opioid misuse with a sensitivity of 93% and specificity of 92%. The workflow was reproducible and included a shared pseudocode for cloud service implementation

^ai2b2: Informatics for Integrating Biology and the Bedside.

^bVA: US Department of Veterans Affairs.

^cNLP: natural language processing.

^dcTAKES: Clinical Text Analysis and Knowledge Extraction System.

^eCDSS: clinical decision support system.

^fEHR: electronic health record.

^gCDS: clinical decision support.

^hBD: big data.

ⁱCOPD: chronic obstructive pulmonary disease.

^jDDSS: diagnostic decision support system.

Of the 26 articles reviewed, 18 (69%) corresponded to authors from the United States [8,12-14,16-25,31,33,34,36]; China [26,30], Germany [29,32], the United Kingdom [28,35] each had 2 (8) articles; and the rest (n=2, 8%) were from Norway [27] and Australia [15].

Of the evaluated articles, 21 (81%) explained the use of NLP as a source for data collection, and 18 (69%) articles used EHRs as a data source; meanwhile, a further 8 (31%) articles were based on clinical data [8,12,14-16,18-21,23-25,29-33,36]. Only 5 (19%) articles showed the use of an NLP tool called Apache cTAKES (Clinical Text Analysis and Knowledge Extraction System) as a set of combined strategies for NLP to obtain clinical data [14,16,17,19,36].

A total of 16 (62%) articles presented stand-alone data review algorithms [20-33,35,36]. Other studies (n=9, 35%) showed that the CDSS alternative was also a way of displaying the information obtained for immediate clinical use [12,20,22,26-28,30,33,36].

Some of the articles focused on specific pathologies such as epilepsy [31]; genomics [33]; pancreatic cysts [23]; radiographic images [19]; diabetes, obesity, hypertension, dyslipidemia, and cardiovascular diseases [18]; colonoscopies [24]; urinary problems or catheters [25]; posttraumatic stress disorder [17]; preanesthetic evaluation [34]; breast cancer [29]; opioid misuse [36]; and cervical cancer assessment [20,21], thus emphasizing the use that can be made of these types of system in almost all medical specialties. In addition, it should be noted that the vast

majority of studies reviewed did not consist of just a few cases. This is demonstrated by the fact that 74 patients and their EHRs were reviewed in 1 study [20]; 349 clinical cases were reviewed in several studies [8,14-18]; and more than 126,000 clinical cases were analyzed by information collection systems in another [28].

Quality Appraisal Results

In the CASP checklist for the 24 qualitative studies (Multimedia Appendix 2), all had a clear statement of the aims of the research, an adequate qualitative study design, and clearly defined outcomes. Data collection and analysis were sufficiently rigorous in the 24 studies, all were adequately designed to achieve the research aims, and the results obtained were readily transferrable to other settings. However, only 4 of the evaluated studies [20,21,27,31] indicated interaction with the participants (patients, in this instance), which involved informing them about the study and the use of the data obtained from it. A further 12 studies [18,19,23-25,28-30,32-34,36] also used patient data or health records, but there was no mention of patients being informed. Importantly, while the remaining 7 studies used data for their research, 5 [8,15-17,37] obtained the necessary information from a database for a challenge (i2b2/VA), 1 study generated clinical data specifically for the research [35], and the other 2 [22,26] did not mention the type of data they used or where they obtained it from. Nevertheless, none of this influenced or affected the results of the research. All the studies were analyzed using standard means of content analysis and provided sufficient information on the design to replicate the

study. This was sufficient to demonstrate the credibility of the studies and that the data analysis was sufficiently rigorous.

In the CASP checklist for 2 systematic reviews ([Multimedia Appendix 2](#)), both studies had a clear statement of research objectives, an appropriate study design, clearly defined outcomes, and sufficiently rigorous data collection and analysis. Both studies were adequately designed to achieve the research aims, and the results obtained were readily transferable to other settings.

Discussion

Principal Findings

This systematic review produced a synthesis of the current state of AI and NLP techniques in health care to identify their potential use in EHRs and automated information searches. Most of the studies showed good internal validity and decent quality. What stands out from our study is the use of NLP as a source for data collection, and while most of the included studies used EHRs, some were based on clinical data. Only 5 of the articles indicated the use of combined NLP strategies to obtain clinical data. While more than half (16/26, 62%) of the articles presented stand-alone algorithms for data review, others (9/26, 35%) indicated that CDSSs also served to present the information obtained for immediate clinical use.

NLP, as a data mining technique, is considered one of the most appropriate tools to find useful information in the data contained in large databases [12,16]. This is because it is an instrument that enables large amounts of information to be clinically analyzed, showing only the parts with the greatest interest or importance to health professionals [32]. While its use has significantly advanced in extracting concepts from clinical data [17], it faces challenges when dealing with the unstructured format of EHRs, which can impede accurate responses to queries submitted to NLP [15].

To overcome these challenges, various techniques have been proposed. One approach involves the combined use of clinical scores that serve as a guide for obtaining results [31], which could be very useful in improving health systems. Another technique to enhance data collection could be the use of neural networks to increase information extraction (and thus achieve more effective diagnoses) [30]. An alternative option offered by NLP includes a sentiment-based model that goes beyond the traditional collaborative filtering approach. This model uses machine learning algorithms to analyze human language text. The metrics used in sentiment analysis aim to determine whether the overall tone of a text is positive, negative, or neutral [38].

Algorithms are the basis of NLP, which consist of any well-defined computational procedure that takes a value or set of values as input and produces a value or set of values as output [39]. However, despite algorithms being versatile tools used in programming and software development, and predominantly acknowledged for their pivotal role in data mining and AI [40], the process of algorithm development is not always straightforward and can sometimes become complicated. For example, many algorithms have difficulties with “negation,” as it can be interpreted as a positive part of a patient’s clinical

history (thus “does not smoke” can be understood as “smoker”). This is a linguistic problem with features that are not always valued, which can lead to inaccurate classification [14,23,41]. It is for this reason that solutions such as NegEx, an algorithm developed in 2001, have been created to try to correct the problem with negation [42]. It should be noted that the use of rules (heuristics) in the search for clinical evidence can generate a better diagnostic recommendation [28], and this is probably because classification systems using rules present more robust machine learning models [24]. The use of tools such as cTAKES is also an alternative, as they are more efficiently and accurately able to scan texts and even the syntactic structure of documents, including negation [19].

The results show the potential of using NLP not only in reviewing clinical notes but also with algorithms that can help find specific information in large volumes of medical information [25]. This may explain its widespread use in epidemiology, public health, and disease surveillance [43,44]. The data obtained could be used to prevent new outbreaks of different diseases worldwide and to identify the main characteristics of pathologies to guide diagnoses even before the disease develops to chronic levels. Health care professionals could benefit from integrating NLP with AI in CDSS to improve medical consultations, streamline tasks such as data analysis, document clinical information in an automated and structured way, and refine treatment strategies and diagnostic processes by automated identification and extraction of key data from medical records [29]. Providing accurate information in real time could improve medical decision-making that better suits each patient’s individual needs, which could translate into better medical outcomes.

When grouping the results by their findings, several conclusions could be drawn, such as that NLP is effective in a CDSS, very accurate, and faster than manual search, especially when accompanied by a human review to facilitate the evaluation of the results and check their accuracy. However, it is necessary to consider the fact that more clinical data from EHRs may complicate its use and that new methods would have to be developed to better obtain large amounts of data. Another striking aspect is that all the reviewed articles focus on the detection of clinical data in EHRs in closed environments. That is, the information obtained was used to account for specific pathologies or diagnostic procedures, and the accuracy was assessed by someone able to understand EHRs. However, none of the articles reviewed referred to the use of external data (medical databases); they all use the data found—using NLP—in the EHRs only. By using external data sources, more appropriate or updated diagnostic aids and treatments could be obtained.

There are also some barriers preventing the development and improvement of NLP systems. One such barrier is the lack of data or incomplete data in EHRs. Another is associated with the lack of use or knowledge of NLP by health professionals. The latter significant issue is the lack of multidisciplinary working practices (health care and computer specialists), which hinders adequate progress concerning NLP algorithms. Establishing a multidisciplinary team involving physicians and information systems professionals would be the most effective approach, as demonstrated in various health care environments

[45-47]. In total, 6 of the reviewed articles described the results of a challenge to find an algorithm that best uses NLP in clinical notes, underscoring the efficacy of such initiatives in catalyzing technological advancement, thereby enhancing the performance of algorithms applicable in AI and big data domains.

Incorporating several medical ontologies to increase the coverage of medical entities may enhance results [48]. A semantic term, representing a single clinical concept, serves as a starting point for ontologies. The combination of these concepts defines a set of properties, allowing interconnections (mapping) between them. This process generates semantic ontologies, characterized by controlled terminology and formal semantic relationships in a particular area of interest using a particular modelling language and terminology [22], such as the terms in EHRs [27]. The incorporation of machine learning techniques into EHRs not only produces better results but also plays a key role in the development of predictive rules. Through the use of ontologies, diagnoses are standardized with a unified vocabulary, facilitating seamless exchange and validation across diverse populations [18,28].

The weight given to ontologies in the studies reviewed varies. While some of them define their use very well [22,27], others only mention ontologies as an important part of information extraction [8,12,15,17-19,23,28,29,33,34] or not at all [13,14,16,21,24-26,30-32,35,36]. Further, 1 study mentions the word “ontology” in the keywords but not in the text. This is surprising, as health ontologies are a fundamental part of clinical data extraction projects, and even more so considering the emergence of new ontologies with almost every new study. The levels of understanding of ontology concepts where the knowledge domains of medicine and computer science intersect could be reviewed as a future line of research.

Regarding the use of ontologies, their inclusion with the use of the semantic web, along with medical NLP, will lead to a better assessment of annotation tasks [49,50]. The use of ontologies is extremely important to overcome the barriers that may arise with the use of NLP. For example, to overcome them, some proposals could be adopted, such as the use of (1) AI assistants (special fusion engines) combining knowledge-based engines and data-based engines; (2) biphasic tools (adding human intervention) with the addition of a human reviewer, which would improve search results and identify potentially lost data; and (3) semantic graphs (sentiment analysis), where ontology-based AI tools would allow relevant information about pathologies in clinical data to be found.

The use of appropriate ontologies in NLP systems would serve to facilitate the real-time extraction of information that could be used for the development of real-time clinical decision tools [51,52]. Ontologies can also be useful for avoiding the ambiguity and inconsistencies found in some health care documents such as EHRs [53]. This is very important because these clinical documents could be converted into more understandable semantic structures by the NLP algorithm, allowing the most important information to be extracted [54]. Thus, a CDSS with incorporated NLP could provide physicians with contextual information, meaning that better clinical decisions could be made to the benefit of patients [55]. Such improvements could

take the form of system-generated alerts when alterations in vital sign monitoring or interactions between prescribed drugs are detected [56-58].

Although CDSSs have great potential for use by health care staff to increase adherence to clinical guidelines and to assist in the correct diagnosis, treatment, follow-up, and prevention of various pathologies, with the consequent better maintenance of the population's health [7], some studies suggest that they may disrupt physicians' workflow or alter or be inconsistent with the initial clinical decisions, and may also require technical maintenance with additional costs [59]. Thus, depending on the algorithm and validation, it may present incorrect or low-quality data. Furthermore, as a different program from the EHR one, there may not be adequate interoperability between the two.

With this information, the advantages of using NLP and CDSSs are obvious. However, it is noteworthy that all the studies in our review have a “closed behavior,” meaning that only specific information is searched for in the data present in clinical notes, without searching for further information in the large medical databases available. If the latter were to be carried out, it would allow a new line of research to be developed, in which NLP-based algorithms combined with keyword searches in clinical databases such as PubMed could potentially enable better and faster diagnoses to be made, and also updated treatments to be offered, all based on EHR data and in real time.

The ongoing evolution of generative AI, namely large language models (LLMs), represents a type of AI that is capable of generating text through a process of training on large data sets in multiple languages. These models demonstrate the ability to produce “human-like” responses [60]. A well-known example is ChatGPT, whose architecture uses a neural network to process natural language, thus generating responses based on the context of the input text [61]. It is essential to recognize that the synergistic use of these techniques presents a significant opportunity. The integration of tools based on LLM, medical ontologies, and NLP has the potential to offer a substantial positive influence on the health care process [62].

These results support the need to conduct research aimed not only at improving algorithms and generating new knowledge but also at suggesting new research directions for the development of AI tools. This includes the integration of NLP, medical ontologies, and LLM for enhanced search capabilities in EHRs and other external sources. A promising research path could be to develop algorithms whose architecture is based on web systems and contrasted medical databases, supported by AI with NLP, and that gather information about semantic terms from health care ontologies such as those in the National Library of Medicine. Such developments of AI-based tools may have a positive impact on research into their use in certain areas, such as health care [63]. In addition, the development of AI-based skills also enhances the development of further algorithms and research, as evidenced by the publications resulting from the challenges mentioned above. However, it is imperative to acknowledge the potential ethical implications inherent in this field, which require thorough assessment and subsequent integration into clinical practice [64]. While the potential benefits are substantial, it is paramount to rigorously address

ethical considerations and data privacy concerns, emphasizing cybersecurity and privacy requirements to effectively protect patients' sensitive data and ensure their confidentiality [36,65,66].

Limitations

Despite conducting an exhaustive search across 5 databases, which specifically targeted studies on the application of AI-based medical CDS using NLP techniques, a total of 113 studies were initially identified for screening. However, upon thorough review, only 26 studies were deemed to meet the stringent inclusion and exclusion criteria established for this review. Consequently, the representativeness of our findings may be questioned given the number of records primarily identified and the possible paucity of research on this particular study topic. A significant number of articles were excluded from our review due to their failure to establish a clear connection between NLP, AI, medical records, and their integration with CDSS. Despite delving into NLP and AI within the context of medical records, these articles lacked sufficient exploration of their relationship with CDSS [67]. The sources of information were peer-reviewed publications, so relevant information from other sources (eg, gray literature) was omitted. CASP-based quality scores [11] may have reflected incomplete reporting, since the vast majority of studies did not compare their results to those of other studies along similar lines (eg, the 2010 i2b2/VA challenge), or had short lists of references (between 8 and 20) in which nonscientific ones were included [14,22,23,25-28]. Nevertheless, all the articles were very robust in terms of the presentation of their results, which could be

extrapolated to different local communities without losing their essence.

Conclusions

The use of NLP engines can effectively obtain results that guide the development of more accurate clinical decision systems. The implementation of decision systems using AI assistants is a potential use of this type of tool. Furthermore, the use of biphasic tools using AI criteria as algorithms combined with human criteria may improve the flow of clinical diagnosis or treatment. Human review can improve the accuracy of the search results as well as identify scenarios that might have been missed. The implementation of a special fusion engine (combining knowledge-driven and data-driven engines) is a promising technique that has shown results in terms of more relevant (or improved) recommendations.

Most CDSSs are designed to recommend text based on keywords. However, this leads to problems regarding the effectiveness of the method using NLP. Some proposals, such as the use of semantic graphs, have been put forward to solve this problem. Some controversy has arisen over the fact that CDSSs endure problems related to a certain coldness in their responses, as well as a paucity of data. A sentiment analysis technique to evaluate user preferences may help to overcome this.

The results found allow us to establish new lines of research for the development of AI tools based on NLP with the use of medical ontologies for information searching in both EHRs and external sources (clinical databases) to obtain better results and extra information that could be used to the benefit of patients.

Acknowledgments

This research received no external funding.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Main characteristics of the studies included in the systematic review.

[\[DOCX File, 134 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Quality appraisal—CASP checklist for qualitative studies and for systematic review. CASP: Critical Appraisal Skills Programme.

[\[DOCX File, 143 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

PRISMA 2020 checklist. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

[\[PDF File \(Adobe PDF File\), 85 KB-Multimedia Appendix 3\]](#)

References

1. Bradley-Ridout G, Nikolaichuk E, Jamieson T, Jones C, Morson N, Chuang R, et al. UpToDate versus DynaMed: a cross-sectional study comparing the speed and accuracy of two point-of-care information tools. *J Med Libr Assoc*. 2021;109(3):382-387. [\[FREE Full text\]](#) [doi: [10.5195/jmla.2021.1176](https://doi.org/10.5195/jmla.2021.1176)] [Medline: [34629966](https://pubmed.ncbi.nlm.nih.gov/34629966/)]
2. Abdill R, Blekhman R. Tracking the popularity and outcomes of all bioRxiv preprints. *Elife*. 2019;8:e45133. [\[FREE Full text\]](#) [doi: [10.7554/eLife.45133](https://doi.org/10.7554/eLife.45133)] [Medline: [31017570](https://pubmed.ncbi.nlm.nih.gov/31017570/)]

3. Kington RS, Arnesen S, Chou WS, Curry SJ, Lazer D, Villarruel AM. Identifying credible sources of health information in social media: principles and attributes. *NAM Perspect*. 2021;2021:10.31478/202107a. [[FREE Full text](#)] [doi: [10.31478/202107a](https://doi.org/10.31478/202107a)] [Medline: [34611600](https://pubmed.ncbi.nlm.nih.gov/34611600/)]
4. Bocanegra CLS, Ramos JLS, Rizo C, Civit A, Fernandez-Luque L. HealthRecSys: a semantic content-based recommender system to complement health videos. *BMC Med Inform Decis Mak*. 2017;17(1):63. [[FREE Full text](#)] [doi: [10.1186/s12911-017-0431-7](https://doi.org/10.1186/s12911-017-0431-7)] [Medline: [28506225](https://pubmed.ncbi.nlm.nih.gov/28506225/)]
5. Muhiyaddin R, Abd-Alrazaq AA, Househ M, Alam T, Shah Z. The impact of clinical decision support systems (CDSS) on physicians: a scoping review. *Stud Health Technol Inform*. 2020;272:470-473. [doi: [10.3233/SHTI200597](https://doi.org/10.3233/SHTI200597)] [Medline: [32604704](https://pubmed.ncbi.nlm.nih.gov/32604704/)]
6. Marcos M, Maldonado JA, Martínez-Salvador B, Boscá D, Robles M. Interoperability of clinical decision-support systems and electronic health records using archetypes: a case study in clinical trial eligibility. *J Biomed Inform*. 2013;46(4):676-689. [[FREE Full text](#)] [doi: [10.1016/j.jbi.2013.05.004](https://doi.org/10.1016/j.jbi.2013.05.004)] [Medline: [23707417](https://pubmed.ncbi.nlm.nih.gov/23707417/)]
7. Middleton B, Sittig DF, Wright A. Clinical decision support: a 25 year retrospective and a 25 year vision. *Yearb Med Inform*. 2016;25(Suppl 1):S103-S116. [[FREE Full text](#)] [doi: [10.15265/IYS-2016-s034](https://doi.org/10.15265/IYS-2016-s034)] [Medline: [27488402](https://pubmed.ncbi.nlm.nih.gov/27488402/)]
8. Roberts K, Harabagiu SM. A flexible framework for deriving assertions from electronic medical records. *J Am Med Inform Assoc*. 2011;18(5):568-573. [[FREE Full text](#)] [doi: [10.1136/amiajnl-2011-000152](https://doi.org/10.1136/amiajnl-2011-000152)] [Medline: [21724741](https://pubmed.ncbi.nlm.nih.gov/21724741/)]
9. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *PLoS Med*. 2021;18(3):e1003583. [[FREE Full text](#)] [doi: [10.1371/journal.pmed.1003583](https://doi.org/10.1371/journal.pmed.1003583)] [Medline: [33780438](https://pubmed.ncbi.nlm.nih.gov/33780438/)]
10. Greenhalgh T, Peacock R. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. *BMJ*. 2005;331(7524):1064-1065. [[FREE Full text](#)] [doi: [10.1136/bmj.38636.593461.68](https://doi.org/10.1136/bmj.38636.593461.68)] [Medline: [16230312](https://pubmed.ncbi.nlm.nih.gov/16230312/)]
11. CASP (qualitative studies) checklist. Critical Appraisal Skills Programme. 2018. URL: https://casp-uk.net/images/checklist/documents/CASP-Qualitative-Studies-Checklist/CASP-Qualitative-Checklist-2018_fillable_form.pdf [accessed 2020-12-22]
12. Ross MK, Wei W, Ohno-Machado L. "Big data" and the electronic health record. *Yearb Med Inform*. 2014;23(1):97-104. [[FREE Full text](#)] [doi: [10.15265/IY-2014-0003](https://doi.org/10.15265/IY-2014-0003)] [Medline: [25123728](https://pubmed.ncbi.nlm.nih.gov/25123728/)]
13. van de Burt BWM, Wasylewicz A, Dullemond B, Grouls R, Egberts T, Bouwman A, et al. Combining text mining with clinical decision support in clinical practice: a scoping review. *J Am Med Inform Assoc*. 2023;30(3):588-603. [[FREE Full text](#)] [doi: [10.1093/jamia/ocac240](https://doi.org/10.1093/jamia/ocac240)] [Medline: [36512578](https://pubmed.ncbi.nlm.nih.gov/36512578/)]
14. Clark C, Aberdeen J, Coarr M, Tresner-Kirsch D, Wellner B, Yeh A, et al. MITRE system for clinical assertion status classification. *J Am Med Inform Assoc*. 2011;18(5):563-567. [[FREE Full text](#)] [doi: [10.1136/amiajnl-2011-000164](https://doi.org/10.1136/amiajnl-2011-000164)] [Medline: [21515542](https://pubmed.ncbi.nlm.nih.gov/21515542/)]
15. Patrick JD, Nguyen DHM, Wang Y, Li M. A knowledge discovery and reuse pipeline for information extraction in clinical notes. *J Am Med Inform Assoc*. 2011;18(5):574-579. [[FREE Full text](#)] [doi: [10.1136/amiajnl-2011-000302](https://doi.org/10.1136/amiajnl-2011-000302)] [Medline: [21737844](https://pubmed.ncbi.nlm.nih.gov/21737844/)]
16. Jiang M, Chen Y, Liu M, Rosenbloom ST, Mani S, Denny JC, et al. A study of machine-learning-based approaches to extract clinical entities and their assertions from discharge summaries. *J Am Med Inform Assoc*. 2011;18(5):601-606. [[FREE Full text](#)] [doi: [10.1136/amiajnl-2011-000163](https://doi.org/10.1136/amiajnl-2011-000163)] [Medline: [21508414](https://pubmed.ncbi.nlm.nih.gov/21508414/)]
17. D'Avolio LW, Nguyen TM, Goryachev S, Fiore LD. Automated concept-level information extraction to reduce the need for custom software and rules development. *J Am Med Inform Assoc*. 2011;18(5):607-613. [[FREE Full text](#)] [doi: [10.1136/amiajnl-2011-000183](https://doi.org/10.1136/amiajnl-2011-000183)] [Medline: [21697292](https://pubmed.ncbi.nlm.nih.gov/21697292/)]
18. Kotfila C, Uzuner Ö. A systematic comparison of feature space effects on disease classifier performance for phenotype identification of five diseases. *J Biomed Inform*. 2015;58 Suppl(Suppl):S92-S102. [[FREE Full text](#)] [doi: [10.1016/j.jbi.2015.07.016](https://doi.org/10.1016/j.jbi.2015.07.016)] [Medline: [26241355](https://pubmed.ncbi.nlm.nih.gov/26241355/)]
19. Garla V, Lo Re V, Dorey-Stein Z, Kidwai F, Scotch M, Womack J, et al. The Yale cTAKES extensions for document classification: architecture and application. *J Am Med Inform Assoc*. 2011;18(5):614-620. [[FREE Full text](#)] [doi: [10.1136/amiajnl-2011-000093](https://doi.org/10.1136/amiajnl-2011-000093)] [Medline: [21622934](https://pubmed.ncbi.nlm.nih.gov/21622934/)]
20. Waghlikar KB, MacLaughlin KL, Henry MR, Greenes RA, Hankey RA, Liu H, et al. Clinical decision support with automated text processing for cervical cancer screening. *J Am Med Inform Assoc*. 2012;19(5):833-839. [[FREE Full text](#)] [doi: [10.1136/amiajnl-2012-000820](https://doi.org/10.1136/amiajnl-2012-000820)] [Medline: [22542812](https://pubmed.ncbi.nlm.nih.gov/22542812/)]
21. Waghlikar KB, MacLaughlin KL, Kastner TM, Casey PM, Henry M, Greenes RA, et al. Formative evaluation of the accuracy of a clinical decision support system for cervical cancer screening. *J Am Med Inform Assoc*. 2013;20(4):749-757. [[FREE Full text](#)] [doi: [10.1136/amiajnl-2013-001613](https://doi.org/10.1136/amiajnl-2013-001613)] [Medline: [23564631](https://pubmed.ncbi.nlm.nih.gov/23564631/)]
22. Sordo M, Rocha BH, Morales AA, Maviglia SM, Oglio EDO, Fairbanks A, et al. Modeling decision support rule interactions in a clinical setting. *Stud Health Technol Inform*. 2013;192:908-912. [Medline: [23920690](https://pubmed.ncbi.nlm.nih.gov/23920690/)]
23. Mehrabi S, Schmidt CM, Waters J, Beesley C, Krishnan A, Kesterson J, et al. An efficient pancreatic cyst identification methodology using natural language processing. *Stud Health Technol Inform*. 2013;192:822-826. [Medline: [23920672](https://pubmed.ncbi.nlm.nih.gov/23920672/)]

24. Patterson OV, Forbush TB, Saini SD, Moser SE, DuVall SL. Classifying the indication for colonoscopy procedures: a comparison of NLP approaches in a diverse national healthcare system. *Stud Health Technol Inform.* 2015;216:614-618. [Medline: [26262124](#)]
25. Divita G, Carter M, Redd A, Zeng Q, Gupta K, Trautner B, et al. Scaling-up NLP pipelines to process large corpora of clinical notes. *Methods Inf Med.* 2015;54(6):548-552. [doi: [10.3414/ME14-02-0018](#)] [Medline: [26534722](#)]
26. Mei J, Liu H, Li X, Xie G, Yu Y. A decision fusion framework for treatment recommendation systems. *Stud Health Technol Inform.* 2015;216:300-304. [Medline: [26262059](#)]
27. Marco-Ruiz L, Maldonado JA, Karlsen R, Bellika JG. Multidisciplinary modelling of symptoms and signs with archetypes and SNOMED-CT for clinical decision support. *Stud Health Technol Inform.* 2015;210:125-129. [Medline: [25991115](#)]
28. Danger R, Corrigan D, Soler J, Kazienko P, Kajdanowicz T, Majeed A, et al. A methodology for mining clinical data: experiences from TRANSFoRm project. *Stud Health Technol Inform.* 2015;210:85-89. [Medline: [25991107](#)]
29. Breischneider C, Zillner S, Hammon M, Gass P, Sonntag D. Automatic extraction of breast cancer information from clinical reports. 2017. Presented at: 2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS); June 22-24, 2017:213-218; Thessaloniki, Greece. [doi: [10.1109/CBMS.2017.138](#)]
30. Yang Z, Huang Y, Jiang Y, Sun Y, Zhang Y, Luo P. Clinical assistant diagnosis for electronic medical record based on convolutional neural network. *Sci Rep.* 2018;8(1):6329. [FREE Full text] [doi: [10.1038/s41598-018-24389-w](#)] [Medline: [29679019](#)]
31. Wissel BD, Greiner HM, Glauser TA, Holland-Bouley KD, Mangano FT, Santel D, et al. Prospective validation of a machine learning model that uses provider notes to identify candidates for resective epilepsy surgery. *Epilepsia.* 2020;61(1):39-48. [FREE Full text] [doi: [10.1111/epi.16398](#)] [Medline: [31784992](#)]
32. Wulff A, Mast M, Hassler M, Montag S, Marschollek M, Jack T. Designing an openEHR-Based pipeline for extracting and standardizing unstructured clinical data using natural language processing. *Methods Inf Med.* 2020;59(S 02):e64-e78. [FREE Full text] [doi: [10.1055/s-0040-1716403](#)] [Medline: [33058101](#)]
33. Kulchak Rahm A, Walton NA, Feldman LK, Jenkins C, Jenkins T, Person TN, et al. User testing of a diagnostic decision support system with machine-assisted chart review to facilitate clinical genomic diagnosis. *BMJ Health Care Inform.* 2021;28(1):e100331. [FREE Full text] [doi: [10.1136/bmjhci-2021-100331](#)] [Medline: [33962988](#)]
34. Suh HS, Tully JL, Meineke MN, Waterman RS, Gabriel RA. Identification of preanesthetic history elements by a natural language processing engine. *Anesth Analg.* 2022;135(6):1162-1171. [FREE Full text] [doi: [10.1213/ANE.0000000000006152](#)] [Medline: [35841317](#)]
35. Park EH, Watson HI, Mehendale FV, O'Neil AQ, Clinical Evaluators. Evaluating the impact on clinical task efficiency of a natural language processing algorithm for searching medical documents: prospective crossover study. *JMIR Med Inform.* 2022;10(10):e39616. [FREE Full text] [doi: [10.2196/39616](#)] [Medline: [36287591](#)]
36. Afshar M, Adelaine S, Resnik F, Mundt MP, Long J, Leaf M, et al. Deployment of real-time natural language processing and deep learning clinical decision support in the electronic health record: pipeline implementation for an opioid misuse screener in hospitalized adults. *JMIR Med Inform.* 2023;11:e44977. [FREE Full text] [doi: [10.2196/44977](#)] [Medline: [37079367](#)]
37. Clark R, Moloney G. Facebook and older adults: fulfilling psychological needs? *J Aging Stud.* 2020;55:100897. [doi: [10.1016/j.jaging.2020.100897](#)] [Medline: [33272457](#)]
38. Osman NA, Mohd Noah SA, Darwich M, Mohd M. Integrating contextual sentiment analysis in collaborative recommender systems. *PLoS One.* 2021;16(3):e0248695. [FREE Full text] [doi: [10.1371/journal.pone.0248695](#)] [Medline: [33750957](#)]
39. Yanofsky NS. Towards a definition of an algorithm. *J Log Comput.* 2010;21(2):253-286. [doi: [10.1093/logcom/exq016](#)]
40. Ding S, Zhao H, Zhang Y, Xu X, Nie R. Extreme learning machine: algorithm, theory and applications. *Artif Intell Rev.* 2013;44:103-115. [doi: [10.1007/s10462-013-9405-z](#)]
41. Guo W, Kraines SB. Semantic content-based recommendations using semantic graphs. *Adv Exp Med Biol.* 2010;680:653-659. [doi: [10.1007/978-1-4419-5913-3_72](#)] [Medline: [20865551](#)]
42. Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. *J Biomed Inform.* 2001;34(5):301-310. [FREE Full text] [doi: [10.1006/jbin.2001.1029](#)] [Medline: [12123149](#)]
43. Kostkova P, Saigí-Rubió F, Eguia H, Borbolla D, Verschuuren M, Hamilton C, et al. Data and digital solutions to support surveillance strategies in the context of the COVID-19 pandemic. *Front Digit Health.* 2021;3:707902. [FREE Full text] [doi: [10.3389/fdgh.2021.707902](#)] [Medline: [34713179](#)]
44. Schoene AM, Basinas I, van Tongeren M, Ananiadou S. A narrative literature review of natural language processing applied to the occupational exposome. *Int J Environ Res Public Health.* 2022;19(14):8544. [FREE Full text] [doi: [10.3390/ijerph19148544](#)] [Medline: [35886395](#)]
45. Martín-Noguerol T, Paulano-Godino F, López-Ortega R, Górriz JM, Riascos R, Luna A. Artificial intelligence in radiology: relevance of collaborative work between radiologists and engineers for building a multidisciplinary team. *Clin Radiol.* 2021;76(5):317-324. [doi: [10.1016/j.crad.2020.11.113](#)] [Medline: [33358195](#)]

46. Lavdas I, Glocker B, Rueckert D, Taylor S, Aboagye E, Rockall A. Machine learning in whole-body MRI: experiences and challenges from an applied study using multicentre data. *Clin Radiol*. 2019;74(5):346-356. [doi: [10.1016/j.crad.2019.01.012](https://doi.org/10.1016/j.crad.2019.01.012)] [Medline: [30803815](https://pubmed.ncbi.nlm.nih.gov/30803815/)]
47. Ponomariov V, Chirila L, Apipie F, Abate R, Rusu M, Wu Z, et al. Artificial intelligence versus doctors' intelligence: a glance on machine learning benefaction in electrocardiography. *Discoveries (Craiova)*. 2017;5(3):e76. [FREE Full text] [doi: [10.15190/d.2017.6](https://doi.org/10.15190/d.2017.6)] [Medline: [32309594](https://pubmed.ncbi.nlm.nih.gov/32309594/)]
48. Rink B, Harabagiu S, Roberts K. Automatic extraction of relations between medical concepts in clinical texts. *J Am Med Inform Assoc*. 2011;18(5):594-600. [FREE Full text] [doi: [10.1136/amiajnl-2011-000153](https://doi.org/10.1136/amiajnl-2011-000153)] [Medline: [21846787](https://pubmed.ncbi.nlm.nih.gov/21846787/)]
49. Madkour M, Benhaddou D, Tao C. Temporal data representation, normalization, extraction, and reasoning: a review from clinical domain. *Comput Methods Programs Biomed*. 2016;128:52-68. [FREE Full text] [doi: [10.1016/j.cmpb.2016.02.007](https://doi.org/10.1016/j.cmpb.2016.02.007)] [Medline: [27040831](https://pubmed.ncbi.nlm.nih.gov/27040831/)]
50. Ceusters W, Bona J. Ontological foundations for tracking data quality through the internet of things. *Stud Health Technol Inform*. 2016;221:74-78. [Medline: [27071880](https://pubmed.ncbi.nlm.nih.gov/27071880/)]
51. Li F, Du J, He Y, Song HY, Madkour M, Rao G, et al. Time event ontology (TEO): to support semantic representation and reasoning of complex temporal relations of clinical events. *J Am Med Inform Assoc*. 2020;27(7):1046-1056. [FREE Full text] [doi: [10.1093/jamia/ocaa058](https://doi.org/10.1093/jamia/ocaa058)] [Medline: [32626903](https://pubmed.ncbi.nlm.nih.gov/32626903/)]
52. Denny JC, Peterson JF, Choma NN, Xu H, Miller RA, Bastarache L, et al. Extracting timing and status descriptors for colonoscopy testing from electronic medical records. *J Am Med Inform Assoc*. 2010;17(4):383-388. [FREE Full text] [doi: [10.1136/jamia.2010.004804](https://doi.org/10.1136/jamia.2010.004804)] [Medline: [20595304](https://pubmed.ncbi.nlm.nih.gov/20595304/)]
53. Ceusters W, Smith B. Biomarkers in the ontology for general medical science. *Stud Health Technol Inform*. 2015;210:155-159. [Medline: [25991121](https://pubmed.ncbi.nlm.nih.gov/25991121/)]
54. Gaebel J, Kolter T, Arlt F, Denecke K. Extraction of adverse events from clinical documents to support decision making using semantic preprocessing. *Stud Health Technol Inform*. 2015;216:1030. [Medline: [26262330](https://pubmed.ncbi.nlm.nih.gov/26262330/)]
55. Garcia-Jimenez A, Moreno-Conde A, Martínez-García A, Marín-León I, Medrano-Ortega F, Parra-Calderón CL. Clinical decision support using a terminology server to improve patient safety. *Stud Health Technol Inform*. 2015;210:150-154. [Medline: [25991120](https://pubmed.ncbi.nlm.nih.gov/25991120/)]
56. Jafarpour B, Abidi SR, Ahmad AM, Abidi SSR. INITIATE: an intelligent adaptive alert environment. *Stud Health Technol Inform*. 2015;216:285-289. [Medline: [26262056](https://pubmed.ncbi.nlm.nih.gov/26262056/)]
57. Rosier A, Mabo P, Temal L, van Hille P, Dameron O, Deleger L, et al. Remote monitoring of cardiac implantable devices: ontology driven classification of the alerts. *Stud Health Technol Inform*. 2016;221:59-63. [Medline: [27071877](https://pubmed.ncbi.nlm.nih.gov/27071877/)]
58. Lardon J, Asfari H, Souvignet J, Trombert-Paviot B, Bousquet C. Improvement of diagnosis coding by analysing EHR and using rule engine: application to the chronic kidney disease. *Stud Health Technol Inform*. 2015;210:120-124. [Medline: [25991114](https://pubmed.ncbi.nlm.nih.gov/25991114/)]
59. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ Digit Med*. 2020;3:17. [FREE Full text] [doi: [10.1038/s41746-020-0221-y](https://doi.org/10.1038/s41746-020-0221-y)] [Medline: [32047862](https://pubmed.ncbi.nlm.nih.gov/32047862/)]
60. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare (Basel)*. 2023;11(6):887. [FREE Full text] [doi: [10.3390/healthcare11060887](https://doi.org/10.3390/healthcare11060887)] [Medline: [36981544](https://pubmed.ncbi.nlm.nih.gov/36981544/)]
61. Deng J, Lin Y. The benefits and challenges of ChatGPT: an overview. *Front Comput Intell Sys*. 2023;2(2):81-83.
62. Kim JK, Chua M, Rickard M, Lorenzo A. ChatGPT and large language model (LLM) chatbots: the current state of acceptability and a proposal for guidelines on utilization in academic medicine. *J Pediatr Urol*. 2023;19(5):598-604. [doi: [10.1016/j.jpuro.2023.05.018](https://doi.org/10.1016/j.jpuro.2023.05.018)] [Medline: [37328321](https://pubmed.ncbi.nlm.nih.gov/37328321/)]
63. Panch T, Pearson-Stuttard J, Greaves F, Atun R. Artificial intelligence: opportunities and risks for public health. *Lancet Digit Health*. 2019;1(1):e13-e14. [FREE Full text] [doi: [10.1016/S2589-7500\(19\)30002-0](https://doi.org/10.1016/S2589-7500(19)30002-0)] [Medline: [33323236](https://pubmed.ncbi.nlm.nih.gov/33323236/)]
64. Morley J, Machado CC, Burr C, Cows J, Joshi I, Taddeo M, et al. The ethics of AI in health care: a mapping review. *Soc Sci Med*. 2020;260:113172. [doi: [10.1016/j.socscimed.2020.113172](https://doi.org/10.1016/j.socscimed.2020.113172)] [Medline: [32702587](https://pubmed.ncbi.nlm.nih.gov/32702587/)]
65. Bear Don't Walk OJ, Reyes Nieva H, Lee SSJ, Elhadad N. A scoping review of ethics considerations in clinical natural language processing. *JAMIA Open*. 2022;5(2):ooac039. [FREE Full text] [doi: [10.1093/jamiaopen/ooac039](https://doi.org/10.1093/jamiaopen/ooac039)] [Medline: [35663112](https://pubmed.ncbi.nlm.nih.gov/35663112/)]
66. Fu S, Wang L, Moon S, Zong N, He H, Pejaver V, et al. Recommended practices and ethical considerations for natural language processing-assisted observational research: a scoping review. *Clin Transl Sci*. 2023;16(3):398-411. [FREE Full text] [doi: [10.1111/cts.13463](https://doi.org/10.1111/cts.13463)] [Medline: [36478394](https://pubmed.ncbi.nlm.nih.gov/36478394/)]
67. Gauthier MP, Law JH, Le LW, Li JJ, Zahir S, Nirmalakumar S, et al. Automating access to real-world evidence. *JTO Clin Res Rep*. 2022;3(6):100340. [FREE Full text] [doi: [10.1016/j.jtocrr.2022.100340](https://doi.org/10.1016/j.jtocrr.2022.100340)] [Medline: [35719866](https://pubmed.ncbi.nlm.nih.gov/35719866/)]

Abbreviations

AI: artificial intelligence

CASP: Critical Appraisal Skills Programme
CDS: clinical decision support
CDSS: clinical decision support system
cTAKES: Clinical Text Analysis and Knowledge Extraction System
EHR: electronic health record
LLM: large language model
MeSH: Medical Subject Headings
NLP: natural language processing
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by A Mavragani; submitted 08.12.23; peer-reviewed by D Singh, L Zhu; comments to author 29.02.24; revised version received 20.04.24; accepted 24.07.24; published 30.09.24

Please cite as:

*Eguia H, Sánchez-Bocanegra CL, Vinciarelli F, Alvarez-Lopez F, Saigí-Rubió F
Clinical Decision Support and Natural Language Processing in Medicine: Systematic Literature Review
J Med Internet Res 2024;26:e55315
URL: <https://www.jmir.org/2024/1/e55315>
doi: [10.2196/55315](https://doi.org/10.2196/55315)
PMID:*

©Hans Eguia, Carlos Luis Sánchez-Bocanegra, Franco Vinciarelli, Fernando Alvarez-Lopez, Francesc Saigí-Rubió. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 30.09.2024. 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 the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.