Received: 11 July 2022 | Revised: 23 November 2022 | Accepted: 1 December 2022

DOI: 10.1111/epi.17474

Epilepsia®

CRITICAL REVIEW

Transforming epilepsy research: A systematic review on natural language processing applications

Arister N. J. Yew[1] | Marijn Schraagen[2] | Willem M. Otte[3] | Eric van Diessen[3]

[1]University College Utrecht, Utrecht University, Utrecht, The Netherlands

[2]Department of Information and Computing Sciences, Faculty of Science, Utrecht University, Utrecht, The Netherlands

[3]Department of Child Neurology, Brain Center, University Medical Center Utrecht and Utrecht University, Utrecht, The Netherlands

Correspondence
Eric van Diessen, Department of Child Neurology, UMC Utrecht Brain Center, University Medical Center Utrecht and Utrecht University, Room KG 01.310.0, PO Box 85090, Utrecht 3508 AB, The Netherlands.
Email: e.vandiessen-3@umcutrecht.nl

Funding information
EWUU grant 'AI for Health' (2021)

Abstract

Despite improved ancillary investigations in epilepsy care, patients' narratives remain indispensable for diagnosing and treatment monitoring. This wealth of information is typically stored in electronic health records and accumulated in medical journals in an unstructured manner, thereby restricting complete utilization in clinical decision-making. To this end, clinical researchers increasing apply natural language processing (NLP)—a branch of artificial intelligence—as it removes ambiguity, derives context, and imbues standardized meaning from free-narrative clinical texts. This systematic review presents an overview of the current NLP applications in epilepsy and discusses the opportunities and drawbacks of NLP alongside its future implications. We searched the PubMed and Embase databases with a "natural language processing" and "epilepsy" query (March 4, 2022) and included original research articles describing the application of NLP techniques for textual analysis in epilepsy. Twenty-six studies were included. Fifty-eight percent of these studies used NLP to classify clinical records into predefined categories, improving patient identification and treatment decisions. Other applications of NLP had structured clinical information retrieval from electronic health records, scientific papers, and online posts of patients. Challenges and opportunities of NLP applications for enhancing epilepsy care and research are discussed. The field could further benefit from NLP by replicating successes in other health care domains, such as NLP-aided quality evaluation for clinical decision-making, outcome prediction, and clinical record summarization.

KEYWORDS

clinical epilepsy, machine learning, natural language processing, textual analysis

1 | INTRODUCTION

Epilepsy is one of the most common neurological disorders worldwide, with significant physical, cognitive, and psychiatric comorbidities.[1] Clinical challenges related to diagnosis and treatment options in epilepsy are only partly met by improvements in ancillary investigations, such as electroencephalography (EEG), magnetic resonance imaging, and genetic testing.[2] Taking history and clinical follow-up will remain an indispensable part of diagnosing epilepsy and tailoring treatment options.[3,4] Daily practice, however, tells us that practical and time limitations often restrict the complete utilization of this clinical information, sometimes hidden in patients' narratives. To this end, increasing research efforts are undertaken to facilitate the automatic retrieval of standardized clinical information from electronic health records (EHRs) and patient interviews.[5] The eventual aim here is to assist clinicians and patients in clinical decision-making by augmenting the clinical profile.

An approach often used in automatic retrieval is natural language processing (NLP), a branch of artificial intelligence that enables the automated processing of textual information.[6] Traditionally, NLP involves the conversion of unstructured texts into a structured form that fits with a predefined standardization scheme. The structured textual features can subsequently be processed through a series of computational instructions (i.e., algorithms) to allow the eventual presentation of the desired information (for details, see Box 1). In recent years, NLP research has increasingly focused on end-to-end approaches, using raw text as input for models that can be trained to perform many tasks without additional knowledge modeling, given a sufficient amount of data.[7] Various NLP applications exist, such as extracting phrases from texts, information retrieval, classifying texts into categories or topics, identifying key themes across texts, calculating word frequencies, and determining whether a text has a positive, neutral, or negative connotation.[8] Consequently, the medical community has applied NLP in various contexts, including classifying health records and assessing patient experiences with medications.[9]

NLP is increasingly applied in clinical neurosciences,[9] but a thorough overview for epilepsy is lacking. Here, we present a complete synthesis of the current NLP applications that can support epilepsy clinical research and practice. After systematically reviewing the literature, we present an overview of the various NLP subfields applied in epilepsy and will discuss a selection of the articles in more detail. We elaborate on the opportunities and possible pitfalls of the technique and conclude with possible future directions in this emerging research field.

Key Points

- NLP is a branch of artificial intelligence that allows automated retrieval of data from unstructured texts.
- Studies in epilepsy are increasingly applying NLP for patient identification, risk stratification, and prediction.
- Epilepsy research could benefit from adopting NLP applications in other medical subfields and exploring implementation in clinical settings.

2 | MATERIALS AND METHODS

2.1 | Study identification

The review was conducted according to the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) 2020 guidelines.[15] A literature search was performed to identify all relevant publications on NLP applications in the epilepsy domain. The following search strategy was applied to the PubMed and Embase databases on March 4, 2022: ("natural language processing" OR "NLP" OR "text mining" OR "text processing" OR "text analysis" OR "information extraction" OR "language model" OR "entity recognition") AND ("epilepsy" OR "epilepsies" OR "seizure" OR "seizures"). Further studies were identified through the reference lists of included publications and relevant review articles. No time limits or study filters were applied. The detailed search queries can be found in Appendix S1.

2.2 | Study inclusion and exclusion

Studies were included based on abstract and full-text assessments by one reviewer (A.N.J.Y.). The full-text selections were verified by a second reviewer (E.v.D.). Inclusion criteria were: (1) peer-reviewed original research articles; (2) studies involving the development, evaluation, or use of NLP methods for textual analysis; and (3) dedicated to the field of epilepsy. Exclusion criteria were: (1) studies employing NLP for nontextual analysis (such as EEG signal analysis), (2) studies not focusing primarily on epilepsy, and (3) preliminary proof-of-concept studies conducted before NLP model development.

2.3 | Data extraction

One reviewer extracted the study properties (A.N.J.Y.). The information sought from each study was the general characteristics and the details of the NLP systems presented. This information entailed the first author, year of publication, findings related to the study's main objectives, and associated strengths and limitations. Data items extracted regarding the NLP systems included the name of the NLP model, size and type of dataset processed (e.g., 120 discharge summaries), NLP application (e.g., information extraction or sentiment analysis), output of the algorithm

BOX 1 What is NLP?

The broad definition of NLP encompasses two aspects, namely, raw text transformation and information generation from transformed texts.[10] The first step involves using linguistic techniques to break down the complex human language into textual features processible by computers.[6] Some of the common techniques used are described in Table 1. The following step entails the use of computational algorithms to process the textual features for the generation of the desired information, such as a patient's age and condition.[10] The two main computational approaches used are rule-based and machine learning methods. Rule-based approaches involve manually devising a set of rules to recognize textual patterns and perform the designated task, such as text classification.[11] An example of a basic rule is the "if–then" rule.[12] On the other hand, machine learning approaches rely on the ability to self-learn textual patterns from large training datasets.[13] By detecting the associations between the textual inputs and human-annotated data output, the computer can infer the types of information it should produce and automatically assume a learning function. A subfield of machine learning, deep learning, further minimizes the need for human intervention with algorithms.[14] A combination of rule-based and machine learning techniques is often employed in hybrid NLP models to complement and compensate for the weaknesses of each approach.[13] A generic NLP pipeline consisting of the textual input, processing steps, and information output is depicted in Figure 1.

(i.e., what information was identified), and intended purpose and implications (i.e., why was the information identified?). Further technical information obtained was the NLP resources used (i.e., what NLP software packages or program was used), type of algorithm (i.e., rule-based, machine learning, or deep learning), and performance of the model.

2.4 | Evaluation metrics

The performance measures of the NLP models include sensitivity, specificity, positive predictive value, accuracy, area under the receiver operating curve, and *F*-score—a harmonized mean of recall and precision. As *F*-score is one of the most used performance indicators in the field, we report the models' performance using the *F*-measure. We computed the average performance if the included studies indicated separate values for a single model that performed different tasks. When multiple models were used, we presented performances for each model. For studies that did not provide the *F*-score, we indicate the alternative measure provided by the study.

3 | RESULTS

3.1 | Study selection and baseline characteristics

PubMed and Embase queries revealed 124 records after removal of duplicates. Five records were identified additionally from reference lists for further screening. A total of 48 publications were assessed for eligibility, among which 26 articles satisfied the criteria (Figure 2). A summary of the most relevant characteristics of each included study are displayed in Table 2. The earliest work emerged in 2012, followed by an upsurge in publications within the past 2 years. In 58% of the studies ($n = 15$), NLP was applied to examine patients' conditions and perspectives for improving clinical care. The remaining studies explored the potential of NLP in enabling rapid identification of patient cohorts from EHRs and key information from scientific articles to expedite research.

3.2 | Categorization of the studies

To better grasp the variety of NLP studies performed in epilepsy, we grouped the studies according to the research topic and how NLP was applied. We elaborate on the most important findings in each category and provide relevant interrelation between the included studies.

3.2.1 | Patient identification

Most of the included studies exploited NLP techniques to identify—and discriminate between—patients based on their documented clinical history and conditions.[16–30] The process often entailed classifying patient reports or interviews into predefined categories based on the prominent differences in textual features. For instance, NLP could differentiate between EEG reports that contained textual features indicative of epileptiform discharges and reports with descriptions that ruled out the presence of epileptiform discharges. Before model training, techniques such as

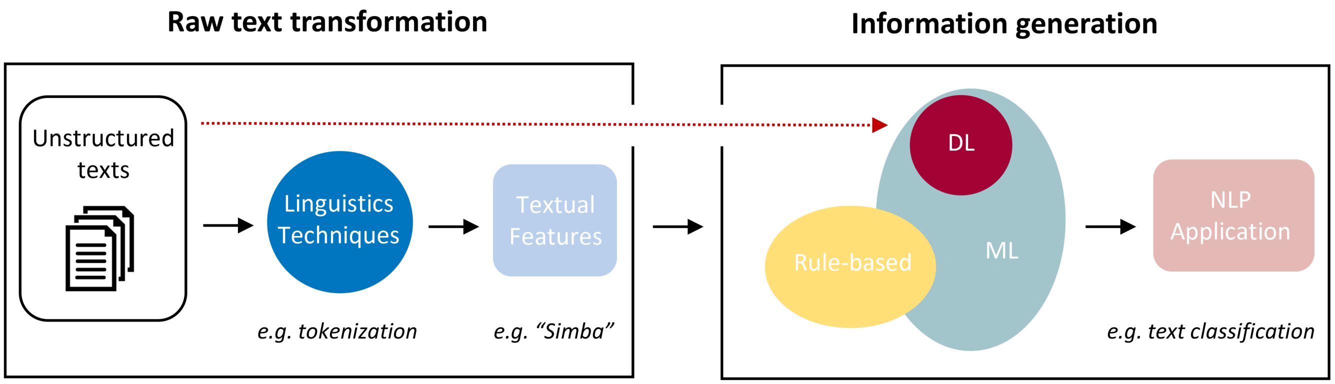

FIGURE 1 A generic natural language processing (NLP) pipeline. The first component of the pipeline transforms raw texts into structured textual features through a combination of linguistics techniques. Next, textual features using computational algorithms are processed to generate the desired output. Many studies apply both a rule-based and machine learning (ML) approach—so-called "hybrid models." A subclass of ML studies uses deep learning (DL) techniques. DL techniques usually skip the linguistic techniques and especially the textual features step and go directly from unstructured texts to information generation (dashed line)

tokenization and stemming were employed to facilitate detection of seizure-related and epileptiform discharge-related keywords in the reports.[20] In doing so, NLP-aided patient identification based on clinical features present in EEG reports can improve diagnostic accuracy and treatment decisions alongside preventing false patient selections.

In contrast to epileptic seizures,[31] no systematic method exists for labeling psychogenic nonepileptic seizures (PNES). Hamid and colleagues showed that it was feasible to distinguish patients with epilepsy from those with PNES. An NLP tool built on a modular pipeline of annotators was used to identify syntactic constructs and named entities from EHRs. The algorithm detected concepts indicative of PNES and those negating PNES.[16] As a different approach, Pevy and colleagues used written transcripts of spoken conversations between patients and doctors to differentiate epileptic seizures from PNES. Features of formulation efforts (i.e., hesitations, reformulations, syntactic repairs) were selected from the transcripts using an NLP toolkit to train a machine learning classifier. The algorithm discriminated between patients based on their differences in the efforts needed to formulate spoken sentences.[29] Connolly and colleagues further showed the feasibility of using NLP to differentiate between partial epilepsy, generalized epilepsy, and unclassified epilepsy patients. Based on textual features from EHRs, the algorithm allowed classification in the respective epilepsy subtypes with a good performance.[17] A similar approach was taken by Glauser and colleagues, who revealed that different classes of psychiatric comorbidities could be attributed to patients using unstructured interviews.[26] Only two studies focused on a pediatric population and showed that prominent differences in textual features between EHRs permit the differentiation of patients with Dravet and West syndromes from those without these syndromes.[19,28]

A series of research extended the applicability of NLP to identifying potential patient candidates for epilepsy surgery.[21–24] Wissel and colleagues validated the ability of a machine learning-based NLP model to recognize various types of tokenized clinical features from EHRs, such as features describing seizure types and drug resistance. An added value of the model compared to standard classifiers was the model's ability to generate surgical candidacy scores rather than clear-cut classifications. The model assigned scores based on the textual features in EHRs, where the presence of such terms as "generalized epilepsy" and "under excellent control" resulted in low candidacy scores and weighed against patients' surgical suitability. This supplementary aspect diversified the clinical applicability of the model by enabling the monitoring of patient candidacy scores over time and clinicians to adjust the minimum score needed for a patient to be recognized as a surgical candidate.

Two studies explored the role of NLP in identifying patients in the context of sudden unexplained death in epilepsy (SUDEP).[25,27] Barbour and colleagues developed rule-based NLP algorithms that could distinguish patients at risk of SUDEP based on SUDEP risk variables in physician notes, including generalized tonic–clonic seizure (GTCS), refractory epilepsy, and potential or previous epilepsy surgery candidacy. The hand-crafted rules built on mathematical notations and SUDEP risk-related terms supported several NLP techniques, including concept and negation detection (see Table 1 for clarification), to determine the presence or absence of risk variables in the texts.[25] For instance, GTCS was indicated by concepts such as "generalized convulsive" and "grand mal seizure," whereas "nonepileptic seizure" pointed to the opposite. Keller and colleagues demonstrated the ability of an NLP algorithm to identify autopsy reports of patients with a

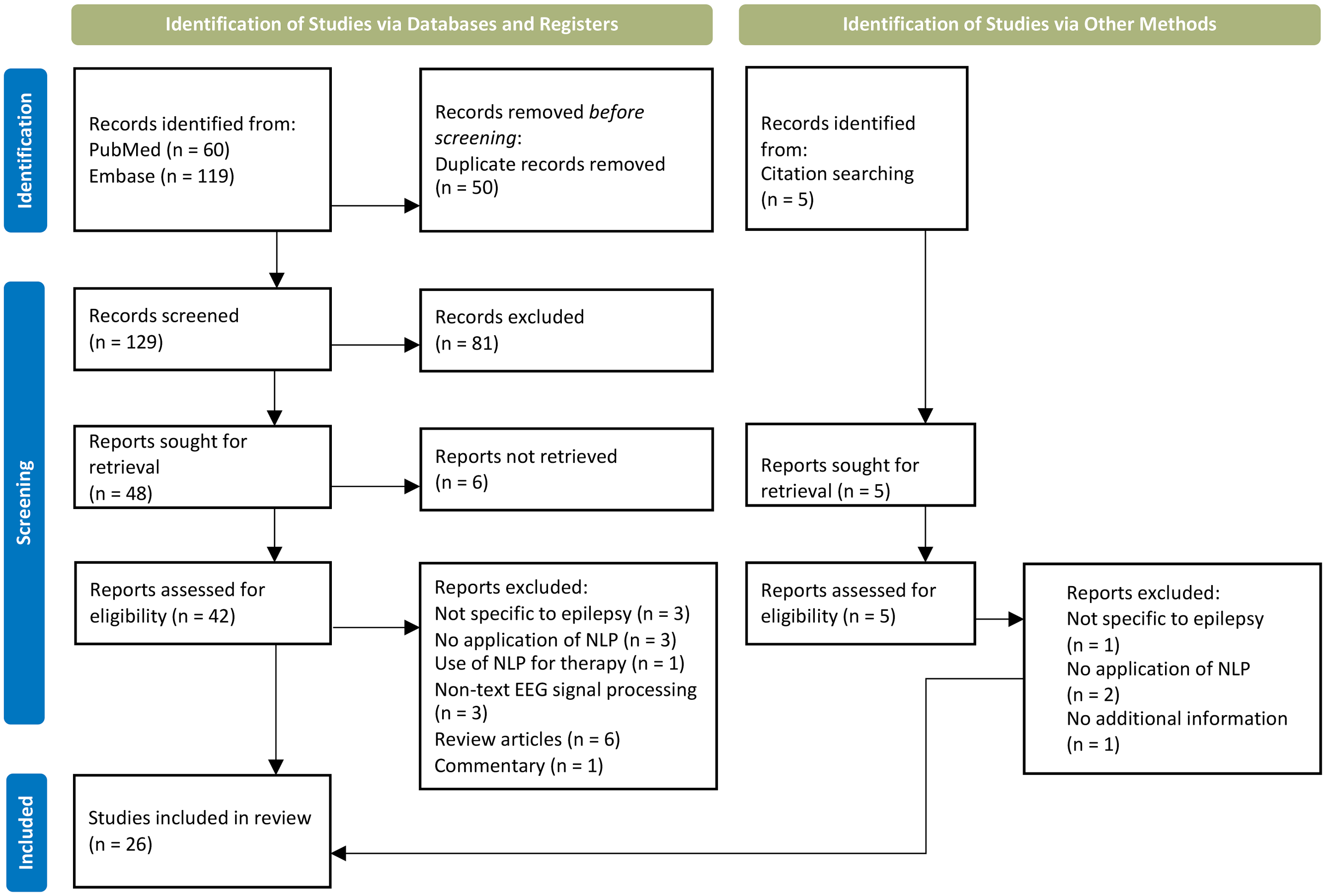

FIGURE 2 PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) flow diagram of the study selection process. EEG, electroencephalographic; NLP, natural language processing

history of epilepsy or seizures based on the presence of relevant features. The authors subsequently reviewed the reports distinguished by the model to manually determine whether an individual's death was attributed to SUDEP.[27]

3.2.2 | Structured information retrieval

NLP was employed in six studies for extracting epilepsy-specific variables from EHRs and scientific articles.[12,32–36] The variables of interest identified from the reports included patients' demographics, epilepsy etiology and diagnoses, brain imaging results, laterality information, seizure semiology and occurrence, and medications. Automated extraction of the information above speeds up the research process by enabling rapid patient cohort identification from EHRs and scientific papers.

Zhang and colleagues developed a platform to capture relevant information on epilepsy by incorporating preexisting NLP tools to extract unstructured retrospective data and integrate this with prospective structured data.[33] The NLP pipeline integrates several components to process the unstructured texts, such as sentence segmentation, tokenization, and part-of-speech tagging.[32,33] A named entity recognition (NER) technique was employed to detect entities of various categories, including epilepsy phenotypes, EEG electrode locations, and medications. The negation detection module further recognizes phrases that indicate denial and uncertainty, such as “is ruled out” and “suspected,” to establish whether the associated concept is negated or undetermined. As the data platform provides a query environment, clinical research can access the patient database to retrieve structured information and for patient cohort search. Finally, the platform enables an expanding integrated database to facilitate large-scale research, thereby circumventing challenges in research involving complex conditions that require the simultaneous identification of several patient characteristics.

As a different approach, Müller and colleagues demonstrated that drug repurposing was feasible in epilepsy through large-scale NLP-aided analysis of scientific papers. NLP benefited the discovery process by enabling rapid extraction of information from >15 million PubMed articles.[36] NER was employed following tokenization and stemming to detect neurological drugs and epilepsy-related

TABLE 1 Common NLP techniques used in the included studies

Category	NLP technique	Description	Textual features
Text boundary detection	Sentence segmentation	Splitting paragraphs into sentences[13]	"Simba was admitted to the hospital" "He was diagnosed with focal epilepsy" "He felt sad"
	Tokenization	Splitting sentences into smaller units, such as words, numbers, and punctuation[6]	"Simba" "was" "admitted" "to" "the" "hospital"
Text normalization	Stemming	Removing word suffixes[66]	"regulated" ➔ "regul" "activated" ➔ "activ"
	Lemmatization	Reducing a word into its dictionary form[66]	"regulated" ➔ "regulate" "activated" ➔ "activate"
Syntactic analysis	Part-of-speech tagging	Labeling words according to their grammatical categories, such as "noun" or "verb"[67]	Noun: "Simba" "epilepsy" Verb: "felt" "was" "admitted" Adjective: "sad"
	Chunking/shallow parsing	Extracting short phrases from texts, such as noun phrases and verb phrases[67]	Noun phrase: "focal epilepsy" Verb phrase: "was admitted"
Semantic analysis	NER	Classifying entities in texts into predefined categories, such as diseases and geographical locations[67]	People's name: "Simba" Disease: "focal epilepsy" Location: "hospital"
	Negation detection	Determining the presence or absence of a concept[68]	Simba was *not* happy He had *no evidence of* focal epilepsy

Note: Text normalization involves transforming a word into its standard form. Syntactic analysis is the task of ascertaining the grammatical structure of sentences. Semantic analysis entails the derivation of meaning from texts.Abbreviations: NER, named entity recognition; NLP, natural language processing.

concepts, such as epilepsy signs and symptoms. The neurological drugs that co-occurred with the epilepsy-related terms could subsequently be identified. Further analyses were conducted to rank the drugs according to their co-occurrence frequency and determine their relations with epilepsy.

3.2.3 | Coping strategy prediction

Five studies implemented NLP techniques to analyze patient and public perceptions of epilepsy portrayed on digital platforms, including social media, health care forums, and an online survey.[37–41] The investigations typically involved analyses of word frequency, sentiment, and themes of public posts to explore the most discussed issues among patients and the underlying connotations of their expressions. In doing so, insights are provided into patients' daily experiences with epilepsy and their concerns regarding epilepsy clinical care. These endeavors to better understand patient perspectives can assist in developing patient-centered care and diminish breaches in communication between health care providers and patients.

A study revealed that it was feasible to process >350 000 patient-generated online posts using NLP to examine patient concerns regarding treatment. The authors employed various modules, including tokenization, part-of-speech tagging, and chunking, followed by NER, to identify medical concepts mentioned by patients. In doing so, researchers gained better insights into patients' perspectives and identified treatment-related issues.[38] Similarly, Lanzone and colleagues performed NLP-based analyses and discovered different word usage patterns between patients and nonepileptic individuals in response to the COVID-19 lockdowns. The authors applied NLP to determine the frequency of terms generated by patients and healthy individuals in an online survey. The underlying connotations of the expressions were further identified using a sentiment analysis algorithm. The analyses revealed significant differences in the frequencies of terms used between the two groups, suggesting that patients with epilepsy had dissimilar coping strategies.[40]

4 | DISCUSSION

4.1 | Study rationale and main results

Medical information is often buried in the ever-increasing number of scientific publications, clinical records, and patient-generated texts, thereby challenging clinicians to utilize these data in clinical decision-making fully. NLP is a promising technique for assisting in identifying information from texts to improve clinical practice and research. This systematic review provides an overview of the current NLP applications in epilepsy research. Most of the included studies demonstrated the

TABLE 2 Baseline characteristics

Authors	NLP model	NLP resource	Algorithm	Type of dataset
Patient identification				
Hamid et al., 2013	NA	Yale cTAKES extension (YTEX)	Machine learning	Progress notes
Connolly et al., 2014	NA	NV	Machine learning	Progress notes
Rijo et al., 2014	NA	GATE & FreeLing	Machine learning	Portuguese electronic medical records
Sullivan et al., 2014	NA	NA	Machine learning	Discharge summaries & EEG reports
Biswal et al., 2015	NA	NA	Machine learning	EEG reports
Cohen et al., 2016	NA	NA	Machine learning	Clinical notes
Wissel et al., 2019	NA	NA	Machine learning	Progress notes
Barbour et al., 2019	NA	NA	Rule-based	Electronic medical records
Glauser et al., 2020	NA	LWIC software	Rule-based & machine learning	Neuropsychiatric interview
Wissel et al., 2020	NA	NA	Machine learning	Progress notes
Keller et al., 2020	NA	Python Software Foundation	Rule-based	Forensic autopsy reports
Lo Barco et al., 2021	NA	DrWarehouse at Necker Pediatric Hospital	NV	EHRs
Pevy et al., 2021	NA	NLTK	Machine learning	Verbatim transcripts of doctor–patient conversations
Alim-Marvasti et al., 2021	NA	NA	Machine learning	EHRs
Wissel et al., 2021	NA	NA	Machine learning	Neurology notes, EEG and MRI reports
Structured information retrieval				
Cui et al., 2012	EpiDEA	cTAKES	Rule-based & machine learning	Discharge summaries
Zhang et al., 2014	MEDCIS	EpiDEA	Rule-based & machine learning	Clinical free text
Cui et al., 2014	PEEP	MetaMap	Rule-based	Discharge summaries
Fonferko-Shadrach et al., 2019	ExECT	GATE	Rule-based & machine learning	Clinic letters
Xie et al., 2022	NA	BERT, Bio_ClinicalBERT, RoBERTa	Deep learning	Progress notes
Müller et al., 2022	NA	UIMA	NV	PubMed articles
Coping strategy prediction				
Meng et al., 2017	NA	NA	NV	Social media posts

Output	NLP application	Performance (*F*-score, unless otherwise specified)
People with probable or definite PNES versus people with epilepsy	Text classification	96%
Presence or absence of partial epilepsy, generalized epilepsy, unclassified epilepsy	Text classification	70.80%
Presence or absence of epilepsy, complex focal seizure, generalized convulsive epilepsy	Text classification	85.53%[a]
Presence or absence of infantile spasms	Text classification	71.40%
Presence or absence of seizures and epileptiform discharges	Text classification	AUC = 98.29%[a]
Surgical candidacy score	Text classification	82%
Surgical candidacy score	Text classification	AUC = 94%
Presence or absence of SUDEP risk variables (i.e., generalized tonic–clonic seizure, refractory epilepsy, potential or previous surgical candidacy)	Text classification	85.67%[a]
Psychiatric comorbidities	Text classification	AUC = 57%–78%
Surgical candidacy score	Text classification	AUC = 79%
Presence or absence of past epilepsy or seizure experience	Text classification	Sensitivity = 98.5% Specificity = 99.7%
Phenotypes related to Dravet syndrome and febrile seizures	Information extraction	NA
Epileptic versus psychogenic nonepileptic seizures	Text classification	67%
Seizure semiology features, hippocampal sclerosis features, and epileptogenic zone features	Information extraction	91%
Surgical candidacy score	Text classification	AUC = 93.65%
Sex, age, epileptogenic zone, etiology, EEG pattern, past and current antiepileptic medications	Information extraction	88.53%
Query results: gender, age, epileptogenic zone, EEG pattern, and anatomical locations that are clinically relevant	Information extraction & query answering	NA
Epilepsy phenotypes (i.e., epileptogenic zone, seizure semiology, lateralizing sign, interictal EEG pattern, and ictal EEG pattern) and pairs of candidate phenotype and anatomical location	Information extraction	89.15%[a]
Clinic date, date of birth, epilepsy diagnosis, epilepsy type, focal seizures, generalized seizures, seizure frequency, medication, CT, MRI, and EEG results	Information extraction	86.10%
Presence or absence of recent seizures, seizure frequency, and date of last seizure	Text classification & information extraction	Bio_ClinicalBERT: 86.30% RoBERTa: 85.65%[a]
Neurological drugs and various epilepsy-related concepts (e.g., signs, symptoms, and seizure types)	Information extraction	NA
Themes related to seeking support, requesting information, providing support, appreciation, and advertisement	Word frequency analysis & thematic analysis	NA

TABLE 2 (Continued)

Authors	NLP model	NLP resource	Algorithm	Type of dataset
He et al., 2019	NA	Natural language toolkit, standard part-of-speech tagger, & MetaMap	NV	Patient online posts
Falcone et al., 2020	NA	NA	Machine learning	Patients' digital conversation
Lanzone et al., 2020	NA	Tidytext & affin lexicon	NV	Patients' single word responses collected from a survey
Fazekas et al., 2021	NA	NA	NV	Patients' digital conversations

Abbreviations: AUC, area under the receiver operating characteristic (i.e., sensitivity versus 1 − specificity) curve; BERT, bidirectional encoder representations from transformers; CT, computed tomography; cTAKES, Clinical Text Analysis and Knowledge Extraction System; EEG, electroencephalographic; EHR, electronic health record; EpiDEA, Epilepsy Data Extraction and Annotation; ExECT, Extraction of Epilepsy Clinical Text; GATE, General Architecture for Text Engineering; LWIC, Linguistic Inquiry and Word Count; MEDCIS, Multimodality Epilepsy Data Capture and Integration System; MRI, magnetic resonance imaging; NA, not applicable; NLP, natural language processing; NLTK, Python Natural Language Toolkit; NV, not available; PEEP, Phenotype Extraction in Epilepsy; PNES, psychogenic nonepileptic seizures; RoBERTa, robustly optimized BERT approach; SUDEP, sudden unexplained death in epilepsy; UIMA, Unstructured Information Management Architecture; YTEX, Yale cTAKES extension.

[a]We computed the average performance of the NLP algorithm in conducting different tasks.

feasibility of using NLP to classify clinical reports into categories of interest based on the differences in textual features between records. This approach enabled better discrimination between patients based on their clinical history and conditions for enhancing diagnosis and treatment. Two other categories of studies focused on either the extraction of epilepsy-related variables from texts for rapid literature analysis and patient cohort identification or the exploration of public and patient perceptions of epilepsy and clinical care through thematic, sentiment, and word frequency analyses of public posts. Overarching opportunities and challenges in the field are discussed below.

4.2 | How does NLP contribute to epilepsy research and clinical care?

Cohort selection in conventional epidemiological research often relies on manual chart reviews that are time-consuming, labor-intensive, and error-prone.[10,42] NLP circumvents these challenges by enabling rapid and systematic automated extraction of patient characteristics from datasets. NLP-aided large-scale exploration of patient databases increases the statistical power of research and boosts investigations of rare conditions by improving the detection of infrequent events.[43] The quality of data collected is further enhanced as NLP minimizes the rate of false-negative patient selections and nonrandom missing data during information retrieval.[44] Depending on research needs, NLP algorithms can be tailored to recognize variables specific to the condition of interest. Examples presented in this review include NLP-aided accurate differentiation of epileptic seizures from PNES and simultaneous identification of multiple SUDEP-relevant characteristics from EHRs.[12,16]

Traditional qualitative research is often liable to researchers' subjectivity, as the process entails manual content analysis and the formulation of survey questions based on researchers' perspectives.[38,45] NLP mitigates the risk of investigator bias by minimizing the need for human involvement in defining thematic categories and processing patient-reported outcomes from interviews and questionnaires.[45] It also offers an alternative to examining patient perspectives through online posts, thereby circumventing the need for survey questions. Nevertheless, subsequent manual analyses are often required to supplement the NLP-generated outcomes with important details and context, as NLP cannot detect all nuances in text.[46] For instance, NLP-aided clustering of online posts into thematic categories revealed prominent treatment-related issues discussed among patients, and subsequent in-depth analyses specified patient concerns regarding managing treatment side effects.[41]

In clinical settings, NLP has the potential for earlier and improved detection of patient conditions to reduce

Output	NLP application	Performance (*F*-score, unless otherwise specified)
Medical concepts, frequency of words, and co-occurrence relation between words	Information extraction, word frequency analysis, & thematic analysis	NA
Data on speakers' age group (i.e., teenagers or adults), types of sites used, topics related to epilepsy and suicide, psychographic mindset of speakers, polarity of topics discussed (i.e., positive, negative, or neutral)	Information extraction, thematic analysis, & sentiment analysis	NA
Frequency and polarity of words	Word frequency analysis & sentiment analysis	NA
Themes related to disease area, time, treatment, support, and sleep	Thematic analysis & sentiment analysis	NA

the time to diagnosis and treatment. NLP algorithms can detect implicit textual patterns predictive of a medical condition that often go unnoticed by physicians. Furthermore, it may contribute to the simultaneous analysis of different clinical records, such as physician notes and brain imaging reports, to substantiate a diagnosis or treatment decision. For example, Wissel and colleagues validated the ability of an NLP model to identify pediatric surgical patients up to 2 years before the patients were referred to presurgical evaluations.[24] Ideally, one would have direct access to the information provided by these NLP algorithms. To achieve this, a digital infrastructure that supports automatic data capture methods from EHRs embedded in NLP algorithms is essential.[47] One can think of many potential uses when such digital infrastructure is available and easy to use without time-consuming additional steps for the physician. For example, a clinician receives real-time feedback with regard to additional investigations or history taking while documenting clinical information from the patient. This feedback could lead to more efficient ancillary investigations, epilepsy monitoring, and treatment strategies. Differently, when confronted with patients after a paroxysmal event suspected of epilepsy, a physician could directly receive a probability of epilepsy based on the clinical notes. This way, referral policies and patient counseling can be improved (for more examples of integrating NLP in clinical care, see the following Section 4.3).

4.3 | What are the current opportunities and challenges?

NLP has garnered increasing popularity in various medical subfields.[10,45,48,49] Achievements in other domains could serve as exemplars for advancements in epilepsy-specific NLP applications, such as quality assessment of clinical practice,[50] clinical outcome prediction,[51] patient record summarization,[52] and secondary use of scientific papers.[53] The identification and connection of information across scientific papers for novel discoveries, such as the extraction and ranking of neurological drugs for drug repurposing,[36] is relatively unexplored in the field of epilepsy but has provided valuable insights in other research fields.[6,53,54] To this end, our research group is currently working on developing an NLP model that extracts standardized information from epilepsy-related publications. The eventual aim of this research effort is to identify current trends and possible future directions for epilepsy research. Another opportunity for new research initiatives is to take raw and unprocessed text produced by the patient (i.e., surface content) as a starting point rather than using the information content. Surface realization typically includes properties of texts such as punctuation, word ordering, and formatting information that are typically overlooked when focusing on disease-related content only.[55] The work of Pevy and colleagues is an example of how surface realization can have an added value in clinical decision-making.[29]

Challenges facing the domain of NLP concern the transparency and reproducibility of the algorithms. NLP approaches may appear less appealing for implementation in clinical and research settings, as users are often blinded to the underlying information used to generate the output.[56] Consequently, clinicians cannot trace and verify the recommendations provided by NLP, as researchers may not specify how each patient's data was processed before analysis for research replication. The complexity of the algorithmic processes could also complicate the reproducibility of NLP-aided studies, as the same model operating on a specific task may generate different outcomes when applied to new datasets.[56–58] In this respect, rule-based algorithms may be more favorable than machine learning models, as the former offer more transparency and are not subjected to random statistical fluctuations.[58] Machine learning models that produce probabilistic scores may also allow for better verification than binary classifiers,[56] such as the score-generating algorithm developed by Wissel and colleagues.[24]

Another challenge in NLP applications pertains to the generalizability of the models, as most algorithms were trained and validated on textual data from a single dataset or institution. Variations in clinical practice settings, EHR templates, and terminology used across different institutions may impede the generalizability of the models. Thus, external validation is warranted before model implementation.[25] Possible approaches that could increase model generalizability include using training datasets that cover a diverse population and wide variability in textual features,[25] and avoiding overadaptation of the algorithms to textual details of the training set.[59] Data-sharing initiatives similar to OpenNeuro,[60] a platform where neuroimaging and neurophysiological datasets are stored according to an international guideline and freely available, could prove helpful here. Equally important to consider here are the efforts to harmonize standardization in clinical research, which includes the development of Common Data Elements (CDEs) for clinical information, a project initiated by the National Institute of Neurological Disorders and Stroke in 2006. Over the past years, several epilepsy-specific CDEs were developed on the following topics: antiepileptic drugs and other antiepileptic therapies, comorbidities, electrophysiology, imaging, neurological examination, neuropsychology, quality of life, seizure and syndromes, surgery, and pathology.[61] Apart from data harmonization, these CDEs can catalyze data-gathering strategies that will facilitate NLP-based research.

Finally, most NLP algorithms currently used—and included in our systematic review—are constricted to a maximum length of 1000 tokens (or characters). When longer texts are used as "data entry" for a study—for example, the length of an abstract or longer—texts are split into shorter text formats. As a result, NLP algorithms will need disproportionally more time and lead to inferior results when dealing with longer texts. To this aim, new NLP methods are being developed to circumvent this limitation. An example is the "multidocument approach," a system that generates query-oriented multidocument summaries trained on a selection of PubMed search queries. For each query, this system generates an overview summary consisting of a number of paragraphs, aggregating over a large number of publications relevant to that query.[62] A second illustration of how NLP can be used on longer strands of text is the "Digital Scribe," a tool that automatically populates health records based on doctor–patient conversations by speech recognition technology with medical knowledge in real time.[63] By classifying these texts into medical entities, Digital Scribe can potentially link unprocessed data with subsequent steps of diagnostic modeling or provide suggestions for ancillary investigations, creating a so-called "end-to-end output" (i.e., from extracted concepts to valid relations). Research dedicated to extracting medication-related information and adverse drug events from EHRs to assist the clinician in choosing the most appropriate drug treatment is an illustrative example of such an end-to-end research pipeline.[64,65]

5 | CONCLUSIONS AND FUTURE DIRECTIONS

NLP enables rapid, large-scale, and reliable derivation of information from epilepsy-related texts to identify patients for diagnosis and treatment, build epidemiological cohorts, examine patient perspectives, and analyze scientific articles. The field could further benefit from NLP by adopting successes in other domains, such as NLP-aided quality evaluation of clinical practice, seizure and surgical outcome prediction, and clinical record summarization. Future research on NLP-aided literature management and analysis could lead to novel discoveries, such as new uses for medications.[36] Additional investigations are warranted to transcend the current proof-of-concept findings to actual NLP implementations in clinical settings, such as repeated training and prospective or external validation of the models, increased model transparency for clinical verification, and use of score-based rather than binary outcomes. To facilitate NLP implementations in a clinical setting, we believe it is invaluable to include a data scientist with an NLP background on the research team. To this end, increasingly research-oriented medical centers are opening data science departments or starting collaborative initiatives with data science institutions. Furthermore, comparative studies tackling the relative strengths and weaknesses of rule-based and machine learning methods could provide insights into the suitability of each technique for use in different contexts. With the unprecedented escalation in NLP applications, standardized guidelines are warranted to ensure clinical safety and research reproducibility of the models, such as guidelines for reporting,[58] clinical training of the models, and the minimal model performance required for clinical use.

AUTHOR CONTRIBUTIONS

Willem M. Otte and Eric van Diessen designed and initiated the study. All authors performed data collection and interpretation, wrote the manuscript, and read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

This research was supported by the EWUU grant "AI for Health" (2021). We further thank the anonymous reviewers for constructive feedback.

CONFLICT OF INTEREST

The authors declare no conflict of interest. We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

ORCID

Willem M. Otte https://orcid.org/0000-0003-1511-6834
Eric van Diessen https://orcid.org/0000-0002-7773-1990

REFERENCES

1. Keezer MR, Sisodiya SM, Sander JW. Comorbidities of epilepsy: current concepts and future perspectives. Lancet Neurol. 2016;15(1):106–15.
2. Pitkänen A, Löscher W, Vezzani A, Becker AJ, Simonato M, Lukasiuk K, et al. Advances in the development of biomarkers for epilepsy. Lancet Neurol. 2016;15(8):843–56.
3. Thijs RD, Surges R, O'Brien TJ, Sander JW. Epilepsy in adults. Lancet. 2019;393(10172):689–701.
4. van Donselaar CA, Stroink H, Arts WF. How confident are we of the diagnosis of epilepsy? Epilepsia. 2006;47(Suppl 1):9–13.
5. Cowie MR, Blomster JI, Curtis LH, Duclaux S, Ford I, Fritz F, et al. Electronic health records to facilitate clinical research. Clin Res Cardiol. 2017;106(1):1–9.
6. Friedman C, Rindflesch TC, Corn M. Natural language processing: state of the art and prospects for significant progress, a workshop sponsored by the National Library of medicine. J Biomed Inform. 2013;46(5):765–73.
7. Zhou M, Duan N, Liu S, Shum HY. Progress in neural NLP: modeling, learning, and reasoning. Engineering. 2020;6(3):275–90.
8. Savova G, Pestian J, Connolly B, Miller T, Ni Y, Dexheimer JW, et al. Natural language processing: applications in pediatric research. Translational Bioinformatics. Volume 10. Singapore: Springer; 2016. p. 231–50.
9. Buchlak QD, Esmaili N, Bennett C, Farrokhi F. Natural language processing applications in the clinical neurosciences: a machine learning augmented systematic review. Acta Neurochir Suppl. 2022;134:277–89.
10. Pons E, Braun LMM, Hunink MGM, Kors JA. Natural language processing in radiology: a systematic review. Radiology. 2016;279(2):329–43.
11. Childs LC, Enelow R, Simonsen L, Heintzelman NH, Kowalski KM, Taylor RJ. Description of a rule-based system for the i2b2 challenge in natural language processing for clinical data. J Am Med Inform Assoc. 2009;16(4):571–5.
12. Cui L, Sahoo SS, Lhatoo SD, Garg G, Rai P, Bozorgi A, et al. Complex epilepsy phenotype extraction from narrative clinical discharge summaries. J Biomed Inform. 2014;51:272–9.
13. Fu S, Chen D, He H, Liu S, Moon S, Peterson KJ, et al. Clinical concept extraction: a methodology review. J Biomed Inform. 2020;109:103526.
14. Pouyanfar S, Sadiq S, Yan Y, Tian H, Tao Y, Reyes MP, et al. A survey on deep learning. ACM Comput Surv. 2019;51(5):1–36. https://doi.org/10.1145/3234150
15. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. Statement: an updated guideline for reporting systematic reviews. Int J Surg. 2020;2021:88.
16. Hamid H, Fodeh SJ, Lizama AG, Czlapinski R, Pugh MJ, LaFrance WC, et al. Validating a natural language processing tool to exclude psychogenic nonepileptic seizures in electronic medical record-based epilepsy research. Epilepsy Behav. 2013;29(3):578–80.
17. Connolly B, Matykiewicz P, Cohen KB, Standridge SM, Glauser TA, Dlugos DJ, et al. Assessing the similarity of surface linguistic features related to epilepsy across pediatric hospitals. J Am Med Inform Assoc. 2014;21(5):866–70.
18. Rijo R, Silva C, Pereira L, Gonçalves D, Agostinho M. Decision support system to diagnosis and classification of epilepsy in children. J Univ Comput Sci. 2014;20(6):907–23.
19. Sullivan R, Yao R, Jarrar R, Buchhalter J, Gonzalez G. Text classification towards detecting misdiagnosis of an epilepsy syndrome in a pediatric population. AMIA Annu Symp Proc. 2014;2014:1082–7.
20. Biswal S, Nip Z, Moura V, Bianchi MT, Rosenthal ES, Westover MB. Automated information extraction from free-text EEG reports. Annu Int Conf IEEE Eng Med Biol Soc. 2015;2015:6804–7.
21. Cohen KB, Glass B, Greiner HM, Holland-Bouley K, Standridge S, Arya R, et al. Methodological issues in predicting pediatric epilepsy surgery candidates through natural language processing and machine learning. Biomed Inform Insights. 2016;8:BII.S38308.
22. Wissel BD, Greiner HM, Glauser TA, Mangano FT, Santel D, Pestian JP, et al. Investigation of bias in an epilepsy machine learning algorithm trained on physician notes. Epilepsia. 2019;60(9):e93–8.
23. 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.
24. Wissel BD, Greiner HM, Glauser TA, Pestian JP, Kemme AJ, Santel D, et al. Early identification of epilepsy surgery candidates: a multicenter, machine learning study. Acta Neurol Scand. 2021;144(1):41–50.

25. Barbour K, Hesdorffer DC, Tian N, Yozawitz EG, McGoldrick PE, Wolf S, et al. Automated detection of sudden unexpected death in epilepsy risk factors in electronic medical records using natural language processing. Epilepsia. 2019;60(6):1209–20.
26. Glauser T, Santel D, DelBello M, Faist R, Toon T, Clark P, et al. Identifying epilepsy psychiatric comorbidities with machine learning. Acta Neurol Scand. 2020;141(5):388–96.
27. Keller AE, Ho J, Whitney R, Li SA, Williams AS, Pollanen MS, et al. Autopsy-reported cause of death in a population-based cohort of sudden unexpected death in epilepsy. Epilepsia. 2021;62(2):472–80.
28. lo Barco T, Kuchenbuch M, Garcelon N, Neuraz A, Nabbout R. Improving early diagnosis of rare diseases using natural language processing in unstructured medical records: an illustration from Dravet syndrome. Orphanet J Rare Dis. 2021;16(1):309.
29. Pevy N, Christensen H, Walker T, Reuber M. Feasibility of using an automated analysis of formulation effort in patients' spoken seizure descriptions in the differential diagnosis of epileptic and nonepileptic seizures. Seizure. 2021;91:141–5.
30. Alim-Marvasti A, Pérez-García F, Dahele K, Romagnoli G, Diehl B, Sparks R, et al. Machine learning for localizing epileptogenic-zone in the temporal lobe: quantifying the value of multimodal clinical-semiology and imaging concordance. Front Digit Health. 2021;3:559103.
31. Fisher RS, Cross JH, French JA, Higurashi N, Hirsch E, Jansen FE, et al. Operational classification of seizure types by the international league against epilepsy: position paper of the ILAE commission for classification and terminology. Epilepsia. 2017;58(4):522–30.
32. Cui L, Bozorgi A, Lhatoo SD, Zhang GQ, Sahoo SS. EpiDEA: extracting structured epilepsy and seizure information from patient discharge summaries for cohort identification. AMIA Annu Symp Proc. 2012;2012:1191–200.
33. Zhang GQ, Cui L, Lhatoo S, Schuele SU, Sahoo SS. MEDCIS: multi-modality epilepsy data capture and integration system. AMIA Annu Symp Proc. 2014;2014:1248–57.
34. Fonferko-Shadrach B, Lacey AS, Roberts A, Akbari A, Thompson S, Ford DV, et al. Using natural language processing to extract structured epilepsy data from unstructured clinic letters: development and validation of the ExECT (extraction of epilepsy clinical text) system. BMJ Open. 2019;9(4):e023232.
35. Xie K, Gallagher RS, Conrad EC, Garrick CO, Baldassano SN, Bernabei JM, et al. Extracting seizure frequency from epilepsy clinic notes: a machine reading approach to natural language processing. J Am Med Inform Assoc. 2022;29(5):873–81.
36. Müller B, Castro LJ, Rebholz-Schuhmann D. Ontology-based identification and prioritization of candidate drugs for epilepsy from literature. J Biomed Semantics. 2022;13(1):3.
37. Meng Y, Elkaim L, Wang J, Liu J, Alotaibi NM, Ibrahim GM, et al. Social media in epilepsy: a quantitative and qualitative analysis. Epilepsy Behav. 2017;71:79–84.
38. He K, Hong N, Lapalme-Remis S, Lan Y, Huang M, Li C, et al. Understanding the patient perspective of epilepsy treatment through text mining of online patient support groups. Epilepsy Behav. 2019;94:65–71.
39. Falcone T, Dagar A, Castilla-Puentes RC, Anand A, Brethenoux C, Valleta LG, et al. Digital conversations about suicide among teenagers and adults with epilepsy: a big-data, machine learning analysis. Epilepsia. 2020;61(5):951–8.
40. Lanzone J, Cenci C, Tombini M, Ricci L, Tufo T, Piccioli M, et al. Glimpsing the impact of COVID19 lock-down on people with epilepsy: a text mining approach. Front Neurol. 2020;11:870.
41. Fazekas B, Megaw B, Eade D, Kronfeld N. Insights into the real-life experiences of people living with epilepsy: a qualitative netnographic study. Epilepsy Behav. 2021;116:107729.
42. Kaur H, Sohn S, Wi CI, Ryu E, Park MA, Bachman K, et al. Automated chart review utilizing natural language processing algorithm for asthma predictive index. BMC Pulm Med. 2018;18(1):34.
43. Banerji A, Lai KH, Li Y, Saff RR, Camargo CA, Blumenthal KG, et al. Natural language processing combined with ICD-9-CM codes as a novel method to study the epidemiology of allergic drug reactions. J Allergy Clin Immunol. 2020;8(3):1032–1038.e1.
44. Stephens KA, Au MA, Yetisgen M, Lutz B, Suchsland MZ, Ebell MH, et al. Leveraging UMLS-driven NLP to enhance identification of influenza predictors derived from electronic medical record data. bioRxiv. 2020; 2020.04.24.058982.
45. Abbe A, Grouin C, Zweigenbaum P, Falissard B. Text mining applications in psychiatry: a systematic literature review. Int J Methods Psychiatr Res. 2016;25(2):86–100.
46. Guetterman TC, Chang T, DeJonckheere M, Basu T, Scruggs E, Vinod Vydiswaran VG. Augmenting qualitative text analysis with natural language processing: methodological study. J Med Internet Res. 2018;20(6):e231.
47. Kaggal VC, Elayavilli RK, Mehrabi S, Pankratz JJ, Sohn S, Wang Y, et al. Toward a learning health-care system – knowledge delivery at the point of care empowered by big data and NLP. Biomed Inform Insights. 2016;8s1:BII.S37977.
48. Yang LWY, Ng WY, Foo LL, Liu Y, Yan M, Lei X, et al. Deep learning-based natural language processing in ophthalmology: applications, challenges and future directions. Curr Opin Ophthalmol. 2021;32(5):397–405.
49. Juhn Y, Liu H. Artificial intelligence approaches using natural language processing to advance EHR-based clinical research. J Allergy Clin Immunol. 2020;145(2):463–9.
50. Ip IK, Mortele KJ, Prevedello LM, Khorasani R. Focal cystic pancreatic lesions: assessing variation in radiologists' management recommendations. Radiology. 2011;259(1):136–41.
51. Barber EL, Garg R, Persenaire C, Simon M. Natural language processing with machine learning to predict outcomes after ovarian cancer surgery. Gynecol Oncol. 2021;160(1):182–6.
52. Hirsch JS, Tanenbaum JS, Gorman SL, Liu C, Schmitz E, Hashorva D, et al. HARVEST, a longitudinal patient record summarizer. J Am Med Inform Assoc. 2015;22(2):263–74.
53. Perera N, Dehmer M, Emmert-Streib F. Named entity recognition and relation detection for biomedical information extraction. Front Cell Dev Biol. 2020;8:673.
54. Wang LL, Lo K. Text mining approaches for dealing with the rapidly expanding literature on COVID-19. Brief Bioinform. 2021;22(2):781–99.
55. Reiter E, Dale R. Building applied natural language generation systems. Nat Lang Eng. 1997;3(1):57–87.
56. Velupillai S, Suominen H, Liakata M, Roberts A, Shah AD, Morley K, et al. Using clinical natural language processing for health outcomes research: overview and actionable suggestions for future advances. J Biomed Inform. 2018;88:11–9.
57. Digan W, Névéol A, Neuraz A, Wack M, Baudoin D, Burgun A, et al. Can reproducibility be improved in clinical natural language processing? A study of 7 clinical NLP suites. J Am Med Inform Assoc. 2021;28(3):504–15.

58. Beam AL, Manrai AK, Ghassemi M. Challenges to the reproducibility of machine learning models in health care. J Am Med Assoc. 2020;323(4):305–6.
59. Azad TD, Ehresman J, Ahmed AK, Staartjes VE, Lubelski D, Stienen MN, et al. Fostering reproducibility and generalizability in machine learning for clinical prediction modeling in spine surgery. Spine J. 2021;21(10):1610–6.
60. Markiewicz CJ, Gorgolewski KJ, Feingold F, Blair R, Halchenko YO, Miller E, et al. The OpenNeuro resource for sharing of neuroscience data. Elife. 2021;10:10.
61. Loring DW, Lowenstein DH, Barbaro NM, Fureman BE, Odenkirchen J, Jacobs MP, et al. Common data elements in epilepsy research: development and implementation of the NINDS epilepsy CDE project. Epilepsia. 2011;52(6):1186–91.
62. Sanchez-Gomez JM, Vega-Rodríguez MA, Pérez CJ. A multi-objective memetic algorithm for query-oriented text summarization: medicine texts as a case study. Expert Syst Appl. 2022;198:116769.
63. Coiera E, Kocaballi B, Halamaka J, Laranjo L. The digital scribe. NPJ Digit Med. 2018;1:1.
64. Henry S, Buchan K, Filannino M, Stubbs A, Uzuner O. 2018 n2c2 shared task on adverse drug events and medication extraction in electronic health records. J Am Med Inform Assoc. 2020;27(1):3–12.
65. Chen L, Gu Y, Ji X, Sun Z, Li H, Gao Y, et al. Extracting medications and associated adverse drug events using a natural language processing system combining knowledge base and deep learning. J Am Med Inform Assoc. 2020;27(1):56–64.
66. Liu H, Christiansen T, Baumgartner WA, Verspoor K. BioLemmatizer: a lemmatization tool for morphological processing of biomedical text. J Biomed Semantics. 2012;3(1):3.
67. Connolly B, Miller T, Ni Y, Cohen KB, Savova G, Dexheimer JW, et al. Natural language processing – overview and history. Pediatr Biomed Inform. 2016;203–30.
68. Slater LT, Bradlow W, Motti DFA, Hoehndorf R, Ball S, Gkoutos GV. A fast, accurate, and generalisable heuristic-based negation detection algorithm for clinical text. Comput Biol Med. 2021;130:104216.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Yew ANJ, Schraagen M, Otte WM, van Diessen E. Transforming epilepsy research: A systematic review on natural language processing applications. Epilepsia. 2023;64:292–305. https://doi.org/10.1111/epi.17474