

Review

Applications of Natural Language Processing for the Management of Stroke Disorders: Scoping Review

Helios De Rosario¹, PhD; Salvador Pitarch-Corresa¹, PT, AD; Ignacio Pedrosa², PhD; Marina Vidal-Pedros¹, PT, MSc; Beatriz de Otto-López², PhD; Helena García-Mieres², PhD; Lydia Álvarez-Rodríguez², MSc

¹Instituto de Biomecánica de Valencia, Universitat Politècnica de València, Valencia, Spain

²CTIC Centro Tecnológico de la Información y la Comunicación, Gijón, Spain

Corresponding Author:

Helios De Rosario, PhD

Instituto de Biomecánica de Valencia

Universitat Politècnica de València

Camino de Vera s/n, Ed. 9C

Valencia, 46022

Spain

Phone: 34 961111170

Email: helios.derosario@ibv.org

Abstract

Background: Recent advances in natural language processing (NLP) have heightened the interest of the medical community in its application to health care in general, in particular to stroke, a medical emergency of great impact. In this rapidly evolving context, it is necessary to learn and understand the experience already accumulated by the medical and scientific community.

Objective: The aim of this scoping review was to explore the studies conducted in the last 10 years using NLP to assist the management of stroke emergencies so as to gain insight on the state of the art, its main contexts of application, and the software tools that are used.

Methods: Data were extracted from Scopus and Medline through PubMed, using the keywords “natural language processing” and “stroke.” Primary research questions were related to the phases, contexts, and types of textual data used in the studies. Secondary research questions were related to the numerical and statistical methods and the software used to process the data. The extracted data were structured in tables and their relative frequencies were calculated. The relationships between categories were analyzed through multiple correspondence analysis.

Results: Twenty-nine papers were included in the review, with the majority being cohort studies of ischemic stroke published in the last 2 years. The majority of papers focused on the use of NLP to assist in the diagnostic phase, followed by the outcome prognosis, using text data from diagnostic reports and in many cases annotations on medical images. The most frequent approach was based on general machine learning techniques applied to the results of relatively simple NLP methods with the support of ontologies and standard vocabularies. Although smaller in number, there has been an increasing body of studies using deep learning techniques on numerical and vectorized representations of the texts obtained with more sophisticated NLP tools.

Conclusions: Studies focused on NLP applied to stroke show specific trends that can be compared to the more general application of artificial intelligence to stroke. The purpose of using NLP is often to improve processes in a clinical context rather than to assist in the rehabilitation process. The state of the art in NLP is represented by deep learning architectures, among which Bidirectional Encoder Representations from Transformers has been found to be especially widely used in the medical field in general, and for stroke in particular, with an increasing focus on the processing of annotations on medical images.

(*JMIR Med Inform* 2023;11:e48693) doi: [10.2196/48693](https://doi.org/10.2196/48693)

KEYWORDS

stroke; natural language processing; artificial intelligence; scoping review; scoping; review methods; review methodology; NLP; cardiovascular; machine learning; deep learning

Introduction

Stroke, also called “brain attack,” is a medical emergency that occurs when blood flow to a part of the brain is disrupted caused by a clot blocking an artery or by a cerebral hemorrhage due to a ruptured artery. Stroke can result in a range of symptoms and complications depending on the area of the brain that is affected, having impacts on perception, motor control (typically weakness or paralysis on one side of the body, dizziness or difficulty with balance), or behavior (difficulty in speaking or understanding speech), which is a life-threatening emergency that requires immediate medical attention. Although mortality from stroke is decreasing in developed, high-income countries, it remains one of the leading causes of mortality and disability along with ischemic heart disease, and the prevalence of people living with the effects of stroke is increasing due to the growing and aging population [1].

Therefore, the economic and social costs related to the hospitalization, treatment, and recovery of stroke patients are increasing, and there is a growing demand for advanced technologies that can assist in clinical diagnosis, treatment, predictions of clinical events, intervention recommendations, rehabilitation programs, and related factors [2]. For instance, a quick diagnosis and treatment of stroke is crucial as it leads to improved outcomes and prognosis among patients treated within the so-called “golden hour” [3].

In this context, novel approaches that complement and go beyond evidence-based medicine are required. Tools based on artificial intelligence (AI), with their ability to process large amounts of data, have been widely discussed in recent years as one of the proposed approaches to improve the care of stroke, assisting in diagnosis, prognosis, treatment, and prevention [3,4].

AI is an interdisciplinary science with multiple approaches, which in recent years has experienced a significant growth in the fields of machine learning (ML) and deep learning (DL). ML and DL algorithms can learn from data and improve their performance over time without being explicitly programmed, and these methods can deal with very large and complex data sets. DL is considered a recent specialization of ML, which uses artificial neural networks to extract complex representations and features from data. Throughout the manuscript, a distinction is made between DL, used for algorithms based on multilayered neural networks, and traditional ML based on other techniques.

The application of AI to the management of stroke is a topic that has gained a lot of traction in the general field of health informatics [5], partly owing to the remarkable impact of stroke in public health and the subsequent high demand for effective and efficient tools to diagnose and treat stroke. Moreover, the complexity and variety of stroke casuistry make it a good target for AI solutions, which are especially suited to process large amounts of data from a wide range of sources, identify patterns and trends in large data sets, and learn and adapt to new data.

A domain where those advances have produced particularly good results is natural language processing (NLP), which is a promising tool for medicine to unlock the full potential of

electronic health records (EHRs), since it might be used to automatically transform clinical text into structured clinical data that can guide clinical decisions [6,7]. The potential of NLP in the analysis of EHR data is particularly appealing given the great quantity of data contained in these records. Notwithstanding their importance, such data are intractable with conventional mathematical methods, since they are recorded in clinical reports, prescriptions, annotations on medical images, and generally unstructured texts [8].

NLP can assist in the identification of patterns and trends in large data sets, which can improve the understanding of factors that contribute to the development of diseases and can in turn help to define more effective prevention and treatment strategies. NLP can also be used in the analysis of particular cases to guide decisions and potentially delay or prevent the onset of the disease. NLP can also be used to develop intelligent systems to find relevant information in the medical literature [9].

Nevertheless, NLP poses particular challenges, including the protection of privacy in the extraction of data, since personal information is often mixed with other data; the variety of the quality and format of EHR data, which depend on the source and software used to collect them; and the difficulty of annotating data samples for training [10]. Therefore, to unlock the potential of NLP in the exploitation of EHRs, researchers and developers need to combine different advanced ML techniques, apply careful data management, and gain a deep understanding of the clinical domain. There is, however, a paucity of guidance on selecting appropriate methods tailored to the health care industry [11].

This scoping review aimed to gather the knowledge that might help in that guidance by investigating how NLP is used to deliver a smarter health care in different phases of stroke disorders (prevention, diagnosis, treatment, and prognosis). The primary questions that served as a guide for the review are: (1) In which phases or contexts of stroke management is NLP used (prevention, diagnosis, treatment, and/or prognosis)? (2) Which are the main benefits of applying NLP to stroke management, related to clinical, social, and economic factors? and (3) What types of clinical data are collected and used by NLP in stroke management (ie, demographic data, medical notes, physical and functional examination, reports of laboratory or medical devices)?

This review also focused on the following secondary questions: (1) What NLP methods, AI algorithms, and tools are used in stroke studies? (2) Which AI techniques or frameworks are used to process and analyze the data? (3) Are there algorithms and NLP software specifically tuned for stroke? and (4) Which tools have the best performance and how do they compare to others?

Methods

Design

The unregistered protocol for this review was created following the PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) guidelines [12] and the JBI Manual for Scoping Reviews [13].

Inclusion Criteria

The target patient population of this scoping review included adults that had suffered stroke and people at risk of stroke due to a history of predisposing vascular background or other conditions that increase the risk of developing stroke, including mental illness or heart diseases such as a reduced ejection fraction.

The main concept of interest was the use of NLP in stroke management in public or private health care systems, including use cases and the data and technologies involved in those applications. We considered both the application of NLP for monitoring and decision-making of individual patients as well as for the planification of care resources in the management of stroke cases.

We were interested in any context where prevention, treatment, or rehabilitation of stroke might take place, ranging from early detection outside or inside clinical settings, diagnosis and evaluation of cases, clinical decision-making, administration and monitoring of rehabilitation, and postrehabilitation management.

The types of evidence sources taken into account included articles from peer-reviewed journals, books, and conference papers, considering both primary research studies and systematic or scoping reviews, as well as reports from scientific, medical, or government institutions.

Search Strategy

The search was performed in the electronic databases of Scopus and Medline through PubMed, using the keywords “natural language processing” and “stroke,” restricted to articles published in the last 10 years, between 2013 and 2022.

Selection Process

The results of the search were imported into the Zotero Reference Manager software (Corporation for Digital Scholarship, Virginia), which was used to filter out duplicate records. Titles and abstracts of the filtered list were screened independently by two reviewers to ascertain their eligibility according to the inclusion criteria. Disagreements were resolved in a discussion session between the reviewers to obtain a consensus.

The full text of the papers was read by two independent reviewers to extract the relevant data as described below. An internal cross-validation by three other experts on the topic was also considered. Works whose content did not meet the eligibility criteria or did not contain sufficient information to

answer the primary questions were excluded and those that reported the same results from the same study were treated as duplicates. The record of rejected works was shared by the reviewers to confirm the decisions of either part.

Data Extraction and Presentation of Results

The reviewers filled out a table with the following data from each work included in the final selection: type of study, primary diagnosis, related diseases that were used either as inclusion criteria or as predictors in the data analysis, sample size (if suitable), and qualitative responses to the primary and secondary questions.

Works were classified depending on whether or not they reported experimental studies, and those that did were further subclassified as clinical trials or different types of observational studies: cross-sectional, retrospective or prospective, and cohort or case-control studies.

A dictionary of terms was defined for the tabulated records of the primary and secondary questions and their relative frequencies were calculated. In addition, the relationships between answers were analyzed in two different multiple correspondence analyses (MCAs), which can be employed to detect and represent underlying structures in categorical data sets (ie, frequent co-occurrence of specific categories in two or more variables) [14]. One of the MCAs focused on the primary questions, seeking relationships between the context of application (eg, classification of diagnostics, prognosis of outcomes) and the types of data that were processed. The other MCA focused on the secondary questions, seeking relationships between NLP methods and software tools. In both analyses, the type of AI models (general ML, DL, or rule-based algorithms) was also included as a variable. The analysis was performed in R [15], using the packages *factoMineR* [16] and *factoextra* [17] for MCA and its graphical representation.

Results

General Description of the Studies

A total of 115 unique papers were identified out of 223 records obtained in the search; 29 studies were eventually included for data extraction and analysis after screening by title and abstract and reading of the full text (see the flow diagram in [Figure 1](#)).

The general characteristics of the 29 reviewed studies (year, type of study, target diseases, and sample size), together with the items extracted from the primary and secondary questions are respectively presented in [Tables 1](#), [2](#), and [3](#).

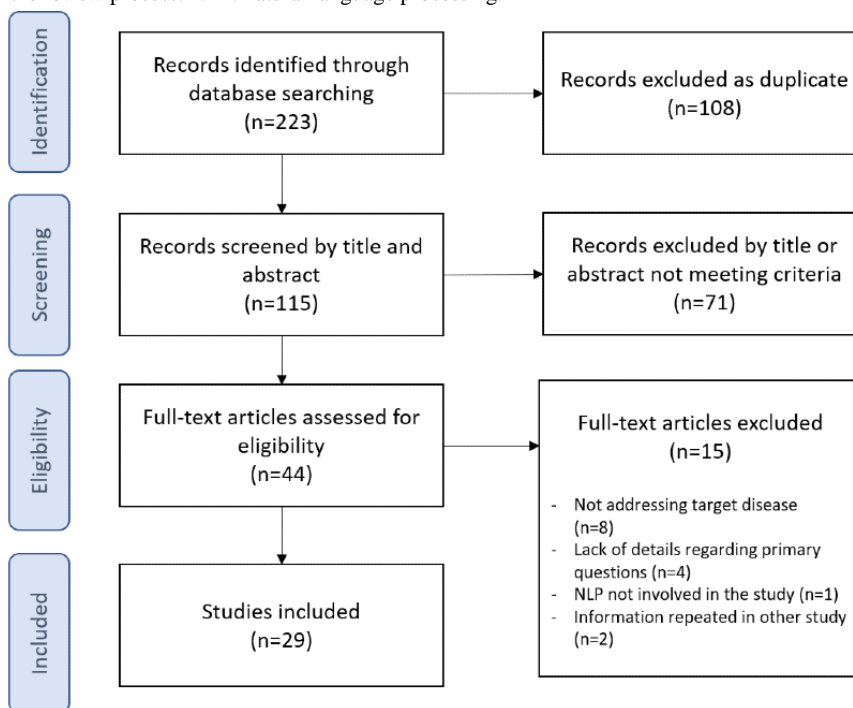
Figure 1. Flow diagram of the review process. NLP: natural language processing.

Table 1. Summary of the included studies: study type, sample size, type of stroke, and other diseases or conditions taken into account.

Reference	Year	Type of study	Sample size ^a	Type of stroke	Other conditions
Zhao et al [18]	2021	Cohort study	4914	Transient ischemic attack, hemorrhagic stroke	AF ^b
Zanotto et al [19]	2021	Retrospective cross-sectional cohort study	188	Ischemic stroke	AF, CAD ^c , DM ^d , dyslipidemia, hypertension, smoking, other ^e
Sung et al [20]	2022	Retrospective cohort study	3847	Acute ischemic stroke	AF, CHF ^f , DM, cancer, hyperlipidemia, hypertension
Sung et al [21]	2021	Retrospective cohort study	3847	Acute ischemic stroke	AF, CHF, DM, cancer, hyperlipidemia, hypertension
Miller et al [22]	2022	Retrospective cohort study	918	Ischemic stroke	Other
Mayampurath et al [23]	2021	Cohort study	965	Acute ischemic stroke, hemorrhagic stroke	Other
Lineback et al [24]	2021	Retrospective cohort study	2855	Ischemic stroke, hemorrhagic stroke	AF, CAD, CHF, DM, cancer, hyperlipidemia, hypertension, other
Kogan et al [25]	2020	Retrospective cohort study	7149	Ischemic stroke, hemorrhagic stroke, transient ischemic attack	None
Heo et al [26]	2020	Retrospective cohort study	1810	Acute ischemic stroke	DM, dyslipidemia, hyperglycemia, hypertension, smoking, other
Deng et al [27]	2022	Feasibility study	1000 (simulated)	Hemorrhagic stroke	DM, hypertension
Bacchi et al [28]	2019	Cohort study	2201	Transient ischemic attack	None
Yu et al [29]	2021	Cohort study	1320	Ischemic stroke, hemorrhagic stroke	None
Wheater et al [30]	2019	Cohort study	2160	Ischemic stroke, hemorrhagic stroke	None
Sung et al [31]	2020	Cohort study	4640	Acute ischemic stroke	None
Sung et al [32]	2018	Feasibility study	90	Acute ischemic stroke	Hyperglycemia, other
Shek et al [33]	2021	Cohort study	2327	Stroke comorbidities	AF, CHF, DM, hypertension
Rannikmäe et al [34]	2021	Cohort study	207	Intracerebral hemorrhage, subarachnoid hemorrhage, and ischemic stroke	None
Ong et al [35]	2020	Cohort study	721	Acute ischemic stroke	None
Mowery et al [36]	2016	Cohort study	498	Ischemic stroke	CAD, CHF, DM, hypertension
Li et al [37]	2021	Cohort study	3971	Acute or subacute ischemic stroke	None
Leung et al [38]	2021	Cohort study	182	Not applicable	Other
Kim et al [39]	2019	Cohort study	3204	Acute ischemic stroke	None
Kent et al [40]	2021	Retrospective cohort study	261,960	Ischemic stroke	AF, CAD, CHF, DM, hyperlipidemia, hypertension, other
Lin et al [41]	2021	Retrospective cohort study	1700	Acute ischemic stroke	Other
Guan et al [42]	2021	Cohort study	1598	Ischemic stroke	CHF, other
Garg et al [43]	2019	Cohort study	1091	Ischemic stroke	AF, CAD, DM, hyperlipidemia, hypertension
Farran et al [44]	2022	Retrospective cohort study	16,916	Not applicable	AF
Elkin et al [45]	2021	Cohort study	96,681	Not applicable	AF
Bacchi et al [46]	2022	Cohort study	438	Ischemic stroke, hemorrhagic stroke	None

^aNumber of patients involved.

^bAF: atrial fibrillation.

^cCAD: coronary artery disease.

^dDM: diabetes mellitus.

^eOther refers to conditions that are not already listed in the table.

^fCHF: coronary heart failure.

The vast majority were cohort studies that analyzed clinical aspects, along with societal or economic aspects of the disease in some cases, at the moment of data gathering. Approximately one third of the papers (n=10) also included a retrospective analysis and 2 of them were limited to feasibility studies. Although the search included a time span of 10 years, only one of the studies included in the review was older than 5 years [36] and most studies (n=19) had been published in the last 2 years (2021 or 2022).

Most studies (n=24) focused on ischemic stroke (either acute, subacute, or transient); the second most frequent type of stroke was hemorrhagic stroke (n=9), which in the majority of cases was in addition to and not excluding ischemic stroke (only 2 papers dealt exclusively with hemorrhagic stroke). Many studies considered other clinical conditions that were used to select the patients or were included as information taken into account by

the models. The most common conditions were atrial fibrillation, diabetes mellitus, and hypertension; each of them was considered in one third of the reviewed papers (n=10). Other diseases that were considered with smaller frequency were hyper- or dyslipidemia, hyperglycemia, hypercholesterolemia, coronary heart failure, smoking, or cancer.

The sample size of the cohort studies was highly varied, ranging between 182 patients [38] and more than 260,000 patients [40], with a median sample size of 2160 patients. The two feasibility studies were conducted either with simulated cases [27] or with a smaller sample of 90 patients [32].

Table 4 shows the frequency of each category used to classify the answers to the primary and secondary questions, except for the question about the specificity of algorithms and NLP tools for stroke, since there was little variability in those answers.

Table 2. Summary of the answers to the primary questions.

Reference	Context for NLP ^a use	Expected benefits	Types of clinical data ^b
Zhao et al [18]	Prevention and diagnosis (classification)	CLINICAL: improved triage	Demographic data, laboratory test results, medical history, medication
Zanotto et al [19]	Prognosis (outcomes)	CLINICAL: care information management, characterize patients, prediction of outcomes, risk assessment; SOCIETAL: supporting research studies; ECONOMIC: public health management	Diagnostic reports
Sung et al [20]	Prognosis (outcomes)	CLINICAL: prediction of outcomes	Annotated medical images, clinical scales, demographic data, diagnostic reports, medical history, patient treatments
Sung et al [21]	Prognosis (outcomes)	CLINICAL: prediction of outcomes, risk assessment	Annotated medical images, clinical scales, demographic data, diagnostic reports, functional outcomes data
Miller et al [22]	Prognosis (outcomes)	CLINICAL: prediction of outcomes, risk assessment	Annotated medical images, diagnostic reports
Mayampurath et al [23]	Diagnosis (classification)	CLINICAL: improved triage	Diagnostic reports
Lineback et al [24]	Prognosis (recurrence)	CLINICAL: care information management	Demographic data, diagnostic reports, medical history, medication, patient treatments
Kogan et al [25]	Prognosis (outcomes)	CLINICAL: administration of treatments, care information management, improved triage, prediction of outcomes	Demographic data, clinical scales, medical history, patient treatments, medication
Heo et al [26]	Prognosis (outcomes)	CLINICAL: prediction of outcomes	Annotated medical images, diagnostic reports
Deng et al [27]	Diagnosis (details); treatment	CLINICAL: administration of treatments	Annotated medical images, clinical scales, diagnostic reports, medical history
Bacchi et al [28]	Diagnosis (classification)	CLINICAL: stroke cause prediction	Annotated medical images, diagnostic reports, medical history, medication
Yu et al [29]	Diagnosis (details)	CLINICAL: improved triage; ECONOMIC: public health management	Annotated medical images, diagnostic reports
Wheater et al [30]	Diagnosis (classification)	CLINICAL: disease surveillance, improved triage; ECONOMIC: public health management	Annotated medical images, diagnostic reports
Sung et al [31]	Prevention and diagnosis (classification)	CLINICAL: administration of treatments, care information management, disease surveillance; ECONOMIC: public health management	Diagnostic reports
Sung et al [32]	Diagnosis (details); treatment	CLINICAL: administration of treatments	Diagnostic reports, laboratory test results, medical history
Shek et al [33]	Diagnosis (comorbidities)	CLINICAL: care information management	Demographic data, medical history
Rannikmäe et al [34]	Diagnosis (classification)	CLINICAL: improved triage	Annotated medical images, diagnostic reports
Ong et al [35]	Diagnosis (details)	CLINICAL: administration of treatments, prediction of outcomes; SOCIETAL: supporting research studies	Annotated medical images, diagnostic reports
Mowery et al [36]	Prevention	CLINICAL: risk assessment	Diagnostic reports
Li et al [37]	Diagnosis (classification)	CLINICAL: improved triage	Annotated medical images, diagnostic reports
Leung et al [38]	Diagnosis (details)	CLINICAL: care information management, characterize patients	Annotated medical images, diagnostic reports
Kim et al [39]	Diagnosis (classification)	CLINICAL: care information management, characterize patients	Annotated medical images, laboratory results, demographic data, diagnostic reports, functional outcomes data

Reference	Context for NLP ^a use	Expected benefits	Types of clinical data ^b
Kent et al [40]	Prognosis (outcomes)	CLINICAL: care information management, characterize patients, stroke cause prediction	Annotated medical images, diagnostic reports
Lin et al [41]	Diagnosis (details); prognosis (recurrence)	SOCIETAL: supporting research studies	Diagnostic reports
Guan et al [42]	Diagnosis (classification)	CLINICAL: improved triage	Clinical scales, diagnostic reports
Garg et al [43]	Diagnosis (classification)	CLINICAL: improved triage, risk assessment	Annotated medical images, diagnostic reports, medical history
Farran et al [44]	Diagnosis (classification); prognosis (outcomes)	CLINICAL: stroke cause prediction, disease surveillance; ECONOMIC: public health management	Clinical scales, demographic data, medical history, patient treatments
Elkin et al [45]	Diagnosis (classification)	Not applicable	Clinical scales, demographic data
Bacchi et al [46]	Diagnosis (classification)	Not applicable	diagnostic reports, patient treatment

^aNLP: natural language processing.

^bSee [Multimedia Appendix 1](#) for the definitions of clinical data types, following Jiang et al [6].

Table 3. Summary of the answers to the secondary questions.

Reference	AI ^a technique	NLP ^b methods ^c	Other statistical methods ^c	Software packages ^{c,d}	Performance metrics ^c	Best performing methods
Zhao et al [18]	ML ^e	Regular expressions	LR ^f , RF ^g	MedTagger, Weka	PPV ^h , NPV ⁱ , F1, sensitivity	RF
Zanotto et al [19]	ML	Ontologies (OWL ^j), BERT ^k , BOW ^l , TF-IDF ^m	CNN ⁿ , K-NN ^o , RF, SVM ^p , naïve Bayes	spaCy	PPV, F1, sensitivity	SVM ontological rules
Sung et al [20]	ML	Negation extraction ontologies (UMLS ^q)	Gradient boosting	Jazzy spell checker, MetaMap, XGBoost ^r	AUC ^s , IDI ^t , NRI ^u	Not applicable
Sung et al [21]	DL	BOW, BERT (Clinical-BERT)	Not applicable	Jazzy spell checker	AUC, IDI, NRI	Not applicable
Miller et al [22]	DL rule-based	BOW, negation extraction, TF-IDF, BERT (BioClinical-BERT)	LASSO ^v , K-NN, RF, MLP ^w	scikit-learn	AUC, PPV, sensitivity, specificity	BioClinical-BERT (except for rare and continuous outcomes)
Mayampurath et al [23]	ML	N-grams (1- or 2-)	SVM	Not applicable	AUC, PPV, NPV, sensitivity, specificity	Not applicable
Lineback et al [24]	ML	N-grams (1- or 2-), TF-IDF, Word-embedding (Word2Vec)	LASSO, LR, PCA ^x , RF, SVM, gradient boosting, naïve Bayes	XGBoost	AUC	ML methods in general
Kogan et al [25]	ML rule-based	Not applicable	RF, gradient boosting, MLP	Not applicable	Correlations, RMSE ^y	Not applicable
Heo et al [26]	DL	BOW, Word-embedding (sent2vec, BioWordVec)	Decision trees, CNN, LASSO, LSTM ^z , MLP, RF, SVM	Quanteda, NLTK ^{aa} , Tensorflow, Keras	AUC	Document-level methods, CNN
Deng et al [27]	DL rule-based	BERT	Not applicable	Not applicable	AUC, PPV, NPV, sensitivity, specificity	Not applicable
Bacchi et al [28]	DL	BOW, negation extraction	Decision trees, CNN, LSTM, RF	Not applicable	AUC, PPV, NPV, sensitivity, specificity	CNN
Yu et al [29]	Rule-based	Regular expressions	Not applicable	CHARTextract	PPV, NPV, accuracy, sensitivity, specificity	Not applicable
Wheater et al [30]	Rule-based	Regular expressions, grammatical analysis, ontologies (custom), negation extraction	Not applicable	BRAT rapid annotation tool	PPV, sensitivity, specificity	Not applicable
Sung et al [31]	ML rule-based	Grammatical analysis (part-of-speech), negation extraction, ontologies (UMLS)	Decision trees (CART ^{bb}), K-NN, LR, RF, SVM	Google spell checker, MetaMap, Weka	Accuracy, κ	Mixed results
Sung et al [32]	Not applicable	Grammatical analysis (part-of-speech), negation extraction, ontologies (UMLS)	Not applicable	Google spell checker, MetaMap, Stata	NPV, F1, sensitivity, specificity	Document-level methods
Shek et al [33]	DL	Grammatical analysis, Negation extraction, Ontologies (SNOMED ^{cc})	Not applicable	MedCAT	NPV, F1, sensitivity, specificity	Not applicable
Rannikmäe et al [34]	ML rule-based	Ontologies (UMLS)	Not applicable	SemEHR	PPV, sensitivity	Mixed results
Ong et al [35]	DL	BOW, TF-IDF, Word-embedding (GloVe ^{dd})	Decision trees (CART), K-NN, LR, LSTM, RF	scikit-learn, Tensorflow	AUC, F1, accuracy, sensitivity, specificity	GloVe + LSTM
Mowery et al [36]	Rule-based	Regular expressions	Not applicable	pyConText	PPV, NPV, sensitivity, specificity	Not applicable
Li et al [37]	ML	BOW, N-gram (2- and 3-), negation extraction	RF	scikit-learn, NLTK	F1, accuracy	Not applicable

Reference	AI ^a technique	NLP ^b methods ^c	Other statistical methods ^c	Software packages ^{c,d}	Performance metrics ^c	Best performing methods
Leung et al [38]	DL rule-based	Not applicable	Not applicable	MedTagger	PPV, NPV, accuracy, sensitivity, specificity	Not applicable
Kim et al [39]	ML	N-gram (1- and 2-), TF-IDF	Decision trees, LR, naïve Bayes, RF, SVM	Quanteda	AUC, F1	Single decision trees
Kent et al [40]	DL rule-based	Ontologies (named entity recognition)	Not applicable	MedTagger	PPV, NPV, accuracy, sensitivity, specificity	Not applicable
Lin et al [41]	DL	BERT (ClinicalBERT, StrokeBERT)	Not applicable	spaCy	AUC, F1	StrokeBERT
Guan et al [42]	ML	Regular expressions, negation extraction	Decision trees (CART), K-NN, LR, RF, SVM	Quanteda	AUC, PPV, NPV, F1, accuracy, specificity	RF
Garg et al [43]	ML	BOW, N-grams (1- to 3-)	Decision trees, K-NN, stacking LR, PCA, RF, SVM, gradient boosting	cTAKES, spaCy, XG-Boost	AUC, sensitivity, κ	Stacking, LR, gradient boost
Farran et al [44]	ML	Ontologies (SNOMED), negation extraction	Not applicable	MedCAT	Accuracy	Not applicable
Elkin et al [45]	ML	Ontologies (SNOMED)	Not applicable	HD-NLP ^{ee}	PPV, NPV, sensitivity, specificity	Not applicable
Bacchi et al [46]	ML	BOW, N-grams (1- to 3-), negation extraction	Decision trees, LR, RF	scikit-learn, NLTK	AUC, PPN, NPP, sensitivity, specificity	RF

^aAI: artificial intelligence.

^bNLP: natural language processing.

^cSee brief descriptions of the NLP tools, statistical methods, software packages, and performance metrics in [Multimedia Appendix 2 \[47-51\]](#).

^dExcluding general programming frameworks like Python or R.

^eML: machine learning.

^fLR: logistic regression.

^gRF: random forest.

^hPPV: positive predictive value.

ⁱNPV: negative predictive value.

^jOWL: Web Ontology Language.

^kBERT: Bidirectional Encoder Representations from Transformers.

^lBOW: bag-of-words.

^mTF-IDF: term frequency-inverse document frequency.

ⁿCNN: convolutional neural network.

^oK-NN: K-nearest neighbor.

^pSVM: support vector machine.

^qUMLS: Unified Medical Language System.

^rXGBoost: extreme gradient boosting.

^sAUC: area under the curve.

^tIDI: integrated discrimination index.

^uNRI: Net Reclassification Index.

^vLASSO: least absolute shrinkage and selection operator.

^wMPL: multilayer perceptron.

^xPCA: principal component analysis.

^yRMSE: root mean squared error.

^zLSTM: long short-term memory.

^{aa}NLTK: Natural Language Processing toolkit for Python.

^{bb}CART: classification and regression tree.

^{cc}SNOMED: Systematized Nomenclature of Medicine.

^{dd}GLoVe: Global Vectors for Word Representation.

^{ee}HD-NLP: high-definition natural language processing.

Table 4. Frequencies of distinctive items found in primary and secondary questions among the included studies (N=29).^a

Variable and category ^b	Studies, n (%)
Context	
Diagnostic (classification)	13 (45)
Diagnostic (details)	6 (21)
Prognostic (outcomes)	8 (28)
Prognostic (recurrence)	2 (7)
Prevention	3 (10)
Treatment	2 (7)
Clinical benefits	
Improved triage	9 (31)
Care information management	8 (28)
Prediction of outcomes	7 (24)
Administration of treatments	5 (17)
Risk assessment	5 (17)
Patient characterization	4 (14)
Disease surveillance	3 (10)
Stroke causes	3 (10)
Data sources	
Diagnostic reports	24 (83)
Annotated images	15 (52)
Medical history	10 (34)
Demographic data	9 (31)
Clinical scales	7 (24)
Treatments	5 (17)
Medication	4 (14)
Laboratory results	3 (10)
Functional outcomes data	2 (7)
Artificial intelligence technique	
ML ^c	15 (52)
DL ^d	10 (34)
Rule-based	10 (34)
Natural language processing tools	
Negation extraction (NEGEX)	11 (38)
Ontologies	10 (34)
Bag-of-words (BOW)	
<i>n</i> -grams	6 (21)
Bidirectional Encoder Representations from Transformers (BERT)	5 (17)
Regular expressions (REG-EXPR)	5 (17)
TF-IDF ^e	5 (17)
Grammatical analysis	4 (14)
Word-embedding	3 (10)
Other statistical tools	

Variable and category ^b	Studies, n (%)
Random forest (RF)	14 (48)
Decision trees	8 (28)
Support vector machine (SVM)	7 (24)
Logistic regression (LR)	7 (24)
K-nearest neighbor (K-NN)	6 (21)
Gradient boosting	4 (14)
Naïve Bayes	3 (10)
Multilayer perceptron (MLP)	3 (10)
Long short-term memory (LSTM)	3 (10)
Principal component analysis (PCA)	2 (7)
Software packages	
scikit-learn	4 (14)
NLTK ^f	3 (10)
spaCy	3 (10)
Quanteda	3 (10)
MedTagger	3 (10)
MetaMap	3 (10)
XGBoost ^g	3 (10)
MedCAT	2 (7)
Weka	2 (7)
Tensorflow	2 (7)
Performance metrics	
Based on ratios (PPV ^h , NPV ⁱ , F1, accuracy, sensitivity, or specificity)	23 (79)
Based on ROC ^j curves (AUC ^k , C-statistic)	14 (48)
Differential measures (NRI ^l , IDI ^m)	2 (7)

^aOnly the items that occurred more than once are reported in this table; however, since different items often overlapped in each study, the frequencies of each variable normally sum to more than 100%.

^bSee brief descriptions of the NLP tools, statistical methods, software packages, and performance metrics in [Multimedia Appendix 2](#) [47-51].

^cML: machine learning.

^dDL: deep learning.

^eTF-IDF: term frequency-inverse document frequency.

^fNLTK: Natural Language Processing toolkit for Python.

^gXGBoost: extreme gradient boosting.

^hPPV: positive predictive value.

ⁱNPV: negative predictive value.

^jROC: receiver operating characteristic.

^kAUC: area under the curve.

^lNRI: Net Reclassification Index.

^mIDI: integrated discrimination index.

The most frequent context of stroke in which the studies were applied was the diagnostic phase, followed by the prognosis of outcomes. The potential benefit of the results on clinical processes (eg, improving the triage of patients depending on the type or severity of stroke, more efficient management of care information) was the main focus of all studies but one [41], which chiefly focused on the societal aspect of supporting

research studies, similar to two other studies that also evaluated that aspect along with clinical applications. Five of the 29 studies (17%) also considered the potential economic benefit of NLP, in terms of reducing the costs of stroke for the public health sector.

The most frequent source of data for NLP models was diagnostic reports (n=24), followed in many cases by annotations on medical images such as radiographs and scans (n=15). General ML models were used more frequently than DL or rule-based algorithms to process the data (n=15 for ML vs n=10 papers for either DL or rule-based techniques). NLP tools, other statistical methods, and the software packages that were used to implement them highly varied across papers, although there were some associations with the AI technique and other variables (see the next subsection).

In nearly all studies, the AI architectures and algorithms had been adapted to deal with stroke-related data, except for one study that used an ML model for patients with severe mental illness at risk of stroke [44]. One of the studies actually used a software tool that was specifically designed for stroke [41], StrokeBERT, which is a language representation model based on Google's Bidirectional Encoder Representations from Transformers (BERT) [47]. Other studies used models that were adapted to broader medical terminology, including ClinicalBERT [52], BioClinicalBERT [53], and BioWordVec [54], or models tuned with standard medical vocabularies such as Systematized Nomenclature of Medicine (SNOMED) [55] or Unified Medical Language System (UMLS) [56].

The methods used to compare the performance of the models were also highly varied, although in the greatest majority of cases (n=23) they were metrics based on the ratios of true/false-positive or -negative values (positive predictive value, negative predictive value, sensitivity, specificity, F1 score, or accuracy), and many were based on the receiver operating characteristic curve (n=14); a few studies (n=2) also used measures of classification improvements such as the net reclassification index and the integrated discrimination index

[48], and only one study used other statistics such as correlation coefficients or the root mean squared error [25].

Owing to the variety of methods and tools used in the studies, there were few coincidences in the selection of the best ones. The only methods that were chosen as the best performing in more than one study were random forest (n=3), convolutional neural network (n=2), and BERT (n=2).

Multiple Correspondence Analysis

Figures 2 and 3 show the proximity of the categories that exhibited the closest relationships in the two first dimensions obtained in the MCA.

The common variable used in the analysis (AI technique) was clearly distinguished in the first two dimensions of the MCA plot, which on the one hand separated rule-based techniques from ML and DL and on the other hand separated general ML from DL.

In the first MCA (Figure 2), it could be observed that the studies focusing on the classification of diagnostics (often used for the triage of patients) and prospects of recurrent stroke were often those that also used ML techniques with demographic data and information on treatments. Although the other categories were less tightly related, the text associated with clinical tests and the annotations on images were related more closely to prognostics of outcomes than to other contexts of application, with annotated images also being used to ascertain details of the stroke episode. Both types of studies were frequently approached by DL and sometimes by rule-based techniques.

In the other MCA (Figure 3), AI techniques were separated between ML, DL, and rule-based methods in the two main dimensions of the projected space, although only general ML and DL were closely related to other items.

Figure 2. Projection of the scores of the categories in the first two dimensions of the multiple correspondence analysis plot involving context of application, data sources, and artificial intelligence technique. DL: deep learning; ML: machine learning.

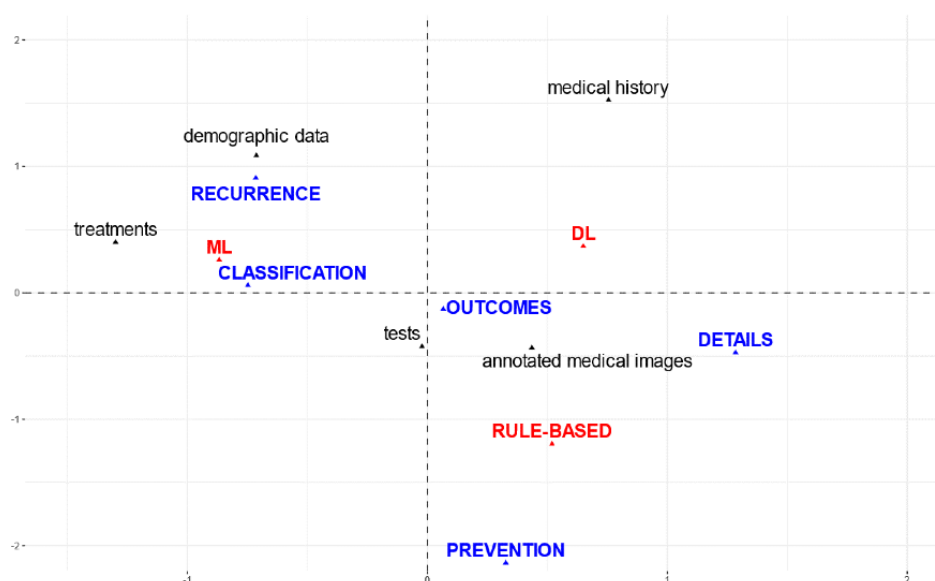
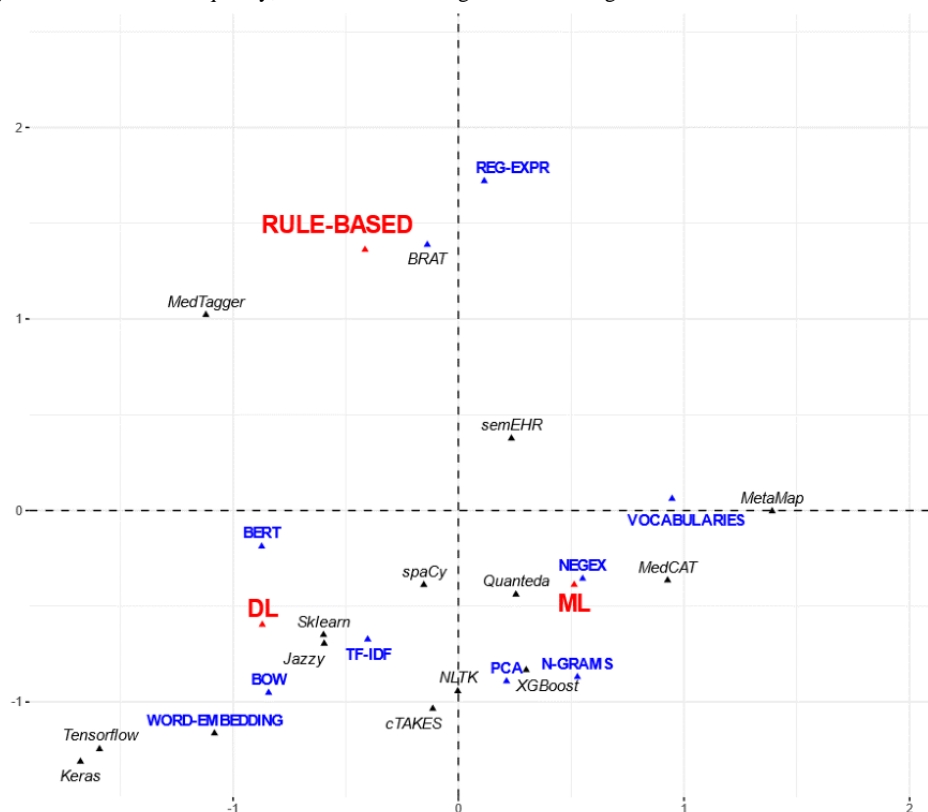


Figure 3. Projection of the scores of the categories in the first two dimensions of the multiple correspondence analysis plot involving natural language processing methods, software, and artificial intelligence techniques. See brief descriptions of the methods and software in [Multimedia Appendix 2](#). BERT: Bidirectional Encoder Representations from Transformers; BOW: Bag-of-words; BRAT: Browser-based Rapid Annotation Tool; DL: deep learning; ML: machine learning; NEGEX: Negation extraction; NLTK: Natural Language Processing toolkit for Python; REG-EXPR: regular expressions; TF-IDF: term frequency-inverse document frequency; XGBoost: extreme gradient boosting.



ML was related to NLP methods that are used in the first steps of the processing pipeline, such as the extraction of text tokens in the form of n -grams, detection of negated terms, and use of standard vocabularies. This was mostly performed with software tools such as MetaMap, MedCAT, Quanteda, and extreme gradient boosting.

Conversely, DL was more associated with the usage of BERT, a language representation model based on transformers [47], and NLP methods applied to numerical and vectorized representations of the language tokens, such as the “bag-of-words,” term frequency-inverse document frequency word embeddings, and other word embeddings. This was chiefly performed with software packages such as Tensorflow through Keras and scikit-learn. Other software packages that are often used for NLP, such as Natural Language Processing toolkit for Python, were observed in the middle of the primary axis of the MCA plot, halfway between the general ML and DL architectures.

Discussion

The research on AI for stroke management has gained greater interest and impact in the last few years [5], and the growing rate of publications found in this scoping review reveals that the same trend is occurring in research on NLP, which is a particular field of AI, applied to the same clinical condition. However, in other aspects, the studies focused on NLP show their own specific trends.

Although the search for this scoping review was very broad, and did not limit the type and phase of stroke to be studied, the vast majority of studies were focused on ischemic stroke in its acute, subacute, or transient stage, and the purpose of using NLP was to improve processes in a clinical context. This focus on clinical contexts is related to the relevance that is attributed to the unstructured information contained in EHRs, (ie, in notes, reports, and annotated images) as predictors of outcomes and complications, which are crucial for proper decision-making, together with the difficulty of processing that information automatically with traditional tools. The deployment of NLP models integrated in the pipelines of an EHR, programmed to automatically ingest and process incoming records [57], or even the patients’ commentaries in emergency through voice-to-text [58], may be used to identify patients at high risk and requiring prompt access to specific treatments; find signs to anticipate impending stroke; or evaluate its severity, type, and risks of complications.

Efficient triage of patients in emergency and early consultations, more accurate diagnostics, or prognostics of outcomes and recurrence were the main intended applications of NLP models in the reviewed studies. Accordingly, the main sources of information exploited by NLP algorithms were clinical data of the patients obtained from their history, especially the diagnostic reports of the current stroke episode. Administration and monitoring of rehabilitation, or postrehabilitation management, were not dealt with in the final selection of studies that were the object of the review.

NLP is itself a broad concept, which involves many types of computational techniques. In its more general sense, NLP comprises all methods and tools that can be used to analyze texts in order to represent human languages, based either on theory of language constructs, semantic mappings, or emulation of linguistic processes occurring in the human brain [59]. The relationships between these tools, types of statistical and ML models, data sources, and applications found by the MCA help to understand how each subset of techniques can be used to solve different problems, and can also help to interpret some trends in the evolution of this technology applied to the clinical management of stroke.

Some of these methods rely on text-processing algorithms that use predefined rules and vocabularies, such as the tokenization of long texts into smaller items, categorization of those items in parts of speech, and construction of syntactic structures, and they have been widely used since long before the recent revolution of big data and DL fields. What this revolution has provided to the field of NLP is the maturity of more complex representations of language data, such as the word embeddings into large-dimensional numeric vectors and their effective processing through deep neural networks, as well as the exploitation of huge databases of texts, such as the Common Crawl data set that includes petabytes of text data, crawled monthly from dozens of billions of web pages [60].

In this context, the state of the art in NLP is represented by DL architectures such as GPT, XLNet, or BERT [61]. Among these, BERT has been found to be particularly widely used in the medical field in general, and for stroke in particular, along with specialized versions fitted to these applications that improve their performance [22,41]. More basic ML algorithms and hybrid approaches with rule-based techniques are still more present than advanced DL networks in the recent research on NLP for stroke, and in some cases, tailored rule-based systems outperformed BERT and its derivatives [19,22]. Support vector machine methods were also found to perform better than BERT in one study [19], although random forest was reported to have

the best performance more frequently than any other ML method in the set of reviewed studies [18,42,46]. Some of these results may seem unexpected, given the remarkable performance of DL in general, and particularly large language models (LLMs), in other areas. However, the computational complexity and large data sets needed to train LLMs can limit their current scalability, not outperforming other ML methods that work better on limited training data such as the data sets of the mentioned studies.

The prevalence of studies based on traditional ML methods over those that use DL neural networks may be partly due to the recency of the more complex DL architectures, as well as to the need of larger sets of data to train those models, which raises the bar to conduct studies with that approach. However, it is also interesting to observe that the choice of the AI technique also relates to the type of data that are processed and the context of application of NLP, such that DL is more closely related to studies that involve medical imaging with annotations to prognosticate the outcomes of stroke.

Taking into account these pieces of evidence, and considering the future of NLP in stroke, further development of LLMs in the biomedical field may be expected. LLMs emerged in 2018 as a class of language models that use neural networks with billions of parameters trained on huge amounts of unlabeled text data through self-supervised learning. LLMs are often based on transformers, a self-attention mechanism to compute contextual relationships between the input tokens [62]. However, innovation in the NLP field will come from the development of these models for medical specialties such as stroke. These biomedical LLMs can be trained not only with data sources from EHRs but also from scientific and clinical publications and social network posts from specialized fields. The particularity is that these models need to be trained on much larger databases than those used by classical ML algorithms to achieve adequate performance metrics. This involves combining computational resources and very large data sources, an option that is not always available for the existing resources in research.

Acknowledgments

This review was conducted within the framework of the IBERUS project Technological Network of Biomedical Engineering Applied to Degenerative Pathologies of the Neuromusculoskeletal System in Clinical and Outpatient Settings (CER-20211003) and the CERVERA Network financed by the Ministry of Science and Innovation through the Center for Industrial Technological Development (CDTI), charged to the General State Budgets 2021 and the Recovery, Transformation, and Resilience Plan.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Categories of clinical data.

[\[DOCX File, 15 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Description of artificial intelligence (AI), natural language processing (NLP), and statistical tools.

[\[DOCX File, 20 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

PRISMA-ScR checklist.

[\[PDF File \(Adobe PDF File\), 103 KB-Multimedia Appendix 3\]](#)

References

1. Stinear CM, Lang CE, Zeiler S, Byblow WD. Advances and challenges in stroke rehabilitation. *Lancet Neurol* 2020 Apr;19(4):348-360 [doi: [10.1016/S1474-4422\(19\)30415-6](https://doi.org/10.1016/S1474-4422(19)30415-6)] [Medline: [32004440](https://pubmed.ncbi.nlm.nih.gov/32004440/)]
2. Sirsat MS, Fermé E, Câmara J. Machine learning for brain stroke: a review. *J Stroke Cerebrovasc Dis* 2020 Oct;29(10):105162 [doi: [10.1016/j.jstrokecerebrovasdis.2020.105162](https://doi.org/10.1016/j.jstrokecerebrovasdis.2020.105162)] [Medline: [32912543](https://pubmed.ncbi.nlm.nih.gov/32912543/)]
3. Abedi V, Khan A, Chaudhary D, Misra D, Avula V, Mathrawala D, et al. Using artificial intelligence for improving stroke diagnosis in emergency departments: a practical framework. *Ther Adv Neurol Disord* 2020 Aug 25;13:1756286420938962 [FREE Full text] [doi: [10.1177/1756286420938962](https://doi.org/10.1177/1756286420938962)] [Medline: [32922515](https://pubmed.ncbi.nlm.nih.gov/32922515/)]
4. Thompson MP, Fanaroff AC, Parker JD, Vallabhajosyula S, Sterling MR. Focusing on the future of cardiovascular outcomes research: highlights From the American Heart Association/American Stroke Association Quality of Care and Outcomes Research 2018 Scientific Sessions. *Circ Cardiovasc Qual Outcomes* 2018 Jun;11(6):e004871 [doi: [10.1161/CIRCOUTCOMES.118.004871](https://doi.org/10.1161/CIRCOUTCOMES.118.004871)] [Medline: [29903937](https://pubmed.ncbi.nlm.nih.gov/29903937/)]
5. Luvizutto GJ, Silva GF, Nascimento MR, Sousa Santos KC, Appelt PA, de Moura Neto E, et al. Use of artificial intelligence as an instrument of evaluation after stroke: a scoping review based on international classification of functioning, disability and health concept. *Top Stroke Rehabil* 2022 Jul 11;29(5):331-346 [doi: [10.1080/10749357.2021.1926149](https://doi.org/10.1080/10749357.2021.1926149)] [Medline: [34115576](https://pubmed.ncbi.nlm.nih.gov/34115576/)]
6. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol* 2017 Dec;2(4):230-243 [FREE Full text] [doi: [10.1136/svn-2017-000101](https://doi.org/10.1136/svn-2017-000101)] [Medline: [29507784](https://pubmed.ncbi.nlm.nih.gov/29507784/)]
7. Sheikhalishahi S, Miotto R, Dudley JT, Lavelli A, Rinaldi F, Osmani V. Natural language processing of clinical notes on chronic diseases: systematic review. *JMIR Med Inform* 2019 Apr 27;7(2):e12239 [FREE Full text] [doi: [10.2196/12239](https://doi.org/10.2196/12239)] [Medline: [31066697](https://pubmed.ncbi.nlm.nih.gov/31066697/)]
8. Adnan K, Akbar R, Khor S, Ali ABA. Role and challenges of unstructured big data in healthcare. In: Sharma N, Chakrabarti A, Balas VE, editors. *Data management, analytics and innovation. Advances in intelligent systems and computing*. Singapore: Springer; 2020:301-323
9. Sneiderman CA, Rindfleisch TC, Aronson AR. Finding the findings: identification of findings in medical literature using restricted natural language processing. *Proc AMIA Annu Fall Symp* 1996:239-243 [FREE Full text] [Medline: [8947664](https://pubmed.ncbi.nlm.nih.gov/8947664/)]
10. Li I, Pan J, Goldwasser J, Verma N, Wong WP, Nuzumlali MY, et al. Neural natural language processing for unstructured data in electronic health records: a review. *Comput Sci Rev* 2022 Nov;46:100511 [doi: [10.1016/j.cosrev.2022.100511](https://doi.org/10.1016/j.cosrev.2022.100511)]
11. Shahid N, Rappon T, Berta W. Applications of artificial neural networks in health care organizational decision-making: a scoping review. *PLoS One* 2019;14(2):e0212356 [FREE Full text] [doi: [10.1371/journal.pone.0212356](https://doi.org/10.1371/journal.pone.0212356)] [Medline: [30779785](https://pubmed.ncbi.nlm.nih.gov/30779785/)]
12. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med* 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: [10.7326/M18-0850](https://doi.org/10.7326/M18-0850)] [Medline: [30178033](https://pubmed.ncbi.nlm.nih.gov/30178033/)]
13. Peters M, Godfrey C, McInerney P, Munn Z, Tricco A, Khalil H. Chapter 11: Scoping reviews. In: Aromataris E, Munn Z, editors. *JBIManual for Evidence Synthesis*. Adelaide, Australia: JBI Collaboration; 2020.
14. Husson F, Josse J. Multiple correspondence analysis. In: Blasius J, Greenacre M, editors. *Visualization and verbalization of data*. Boca Raton, FL: Chapman and Hall/CRC; 2014.
15. R Core Team. R: A Language and Environment for Statistical Computing. 2020. URL: <http://www.R-project.org/> [accessed 2022-12-12]
16. Lê S, Josse J, Husson F. FactoMineR: an R package for multivariate analysis. *J Stat Soft* 2008;25(1):1-18 [doi: [10.18637/jss.v025.i01](https://doi.org/10.18637/jss.v025.i01)]
17. Kassambara A, Mundt F. Factoextra: extract and visualize the results of multivariate data analyses. CRAN R Project. 2020. URL: <https://CRAN.R-project.org/package=factoextra> [accessed 2022-12-12]
18. Zhao Y, Fu S, Bielinski SJ, Decker PA, Chamberlain AM, Roger VL, et al. Natural language processing and machine learning for identifying incident stroke from electronic health records: algorithm development and validation. *J Med Internet Res* 2021 Mar 08;23(3):e22951 [FREE Full text] [doi: [10.2196/22951](https://doi.org/10.2196/22951)] [Medline: [33683212](https://pubmed.ncbi.nlm.nih.gov/33683212/)]
19. Zanutto BS, Beck da Silva Etges AP, Dal Bosco A, Cortes EG, Ruschel R, De Souza AC, et al. Stroke outcome measurements from electronic medical records: cross-sectional study on the effectiveness of neural and nonneural classifiers. *JMIR Med Inform* 2021 Nov 01;9(11):e29120 [FREE Full text] [doi: [10.2196/29120](https://doi.org/10.2196/29120)] [Medline: [34723829](https://pubmed.ncbi.nlm.nih.gov/34723829/)]
20. Sung S, Hsieh C, Hu Y. Early prediction of functional outcomes after acute ischemic stroke using unstructured clinical text: retrospective cohort study. *JMIR Med Inform* 2022 Feb 17;10(2):e29806 [FREE Full text] [doi: [10.2196/29806](https://doi.org/10.2196/29806)] [Medline: [35175201](https://pubmed.ncbi.nlm.nih.gov/35175201/)]

21. Sung S, Chen C, Pan R, Hu Y, Jeng J. Natural language processing enhances prediction of functional outcome after acute ischemic stroke. *J Am Heart Assoc* 2021 Dec 21;10(24):e023486 [FREE Full text] [doi: [10.1161/JAHA.121.023486](https://doi.org/10.1161/JAHA.121.023486)] [Medline: [34796719](https://pubmed.ncbi.nlm.nih.gov/34796719/)]
22. Miller MI, Orfanoudaki A, Cronin M, Saglam H, So Yeon Kim I, Balogun O, et al. Natural language processing of radiology reports to detect complications of ischemic stroke. *Neurocrit Care* 2022 Aug 09;37(Suppl 2):291-302 [FREE Full text] [doi: [10.1007/s12028-022-01513-3](https://doi.org/10.1007/s12028-022-01513-3)] [Medline: [35534660](https://pubmed.ncbi.nlm.nih.gov/35534660/)]
23. Mayampurath A, Parnianpour Z, Richards CT, Meurer WJ, Lee J, Ankenman B, et al. Improving prehospital stroke diagnosis using natural language processing of paramedic reports. *Stroke* 2021 Aug;52(8):2676-2679 [FREE Full text] [doi: [10.1161/STROKEAHA.120.033580](https://doi.org/10.1161/STROKEAHA.120.033580)] [Medline: [34162217](https://pubmed.ncbi.nlm.nih.gov/34162217/)]
24. Lineback CM, Garg R, Oh E, Naidech AM, Holl JL, Prabhakaran S. Prediction of 30-day readmission after stroke using machine learning and natural language processing. *Front Neurol* 2021 Jul 13;12:649521 [FREE Full text] [doi: [10.3389/fneur.2021.649521](https://doi.org/10.3389/fneur.2021.649521)] [Medline: [34326805](https://pubmed.ncbi.nlm.nih.gov/34326805/)]
25. Kogan E, Twyman K, Heap J, Milentijevic D, Lin JH, Alberts M. Assessing stroke severity using electronic health record data: a machine learning approach. *BMC Med Inform Decis Mak* 2020 Jan 08;20(1):8 [FREE Full text] [doi: [10.1186/s12911-019-1010-x](https://doi.org/10.1186/s12911-019-1010-x)] [Medline: [31914991](https://pubmed.ncbi.nlm.nih.gov/31914991/)]
26. Heo TS, Kim YS, Choi JM, Jeong YS, Seo SY, Lee JH, et al. Prediction of stroke outcome using natural language processing-based machine learning of radiology report of brain MRI. *J Pers Med* 2020 Dec 16;10(4):286 [FREE Full text] [doi: [10.3390/jpm10040286](https://doi.org/10.3390/jpm10040286)] [Medline: [33339385](https://pubmed.ncbi.nlm.nih.gov/33339385/)]
27. Deng B, Zhu W, Sun X, Xie Y, Dan W, Zhan Y, et al. Development and validation of an automatic system for intracerebral hemorrhage medical text recognition and treatment plan output. *Front Aging Neurosci* 2022 Apr 8;14:798132 [FREE Full text] [doi: [10.3389/fnagi.2022.798132](https://doi.org/10.3389/fnagi.2022.798132)] [Medline: [35462698](https://pubmed.ncbi.nlm.nih.gov/35462698/)]
28. Bacchi S, Zerner T, Oakden-Rayner L, Kleinig T, Patel S, Jannes J. Deep learning in the prediction of ischaemic stroke thrombolysis functional outcomes: a pilot study. *Acad Radiol* 2020 Feb;27(2):e19-e23 [doi: [10.1016/j.acra.2019.03.015](https://doi.org/10.1016/j.acra.2019.03.015)] [Medline: [31053480](https://pubmed.ncbi.nlm.nih.gov/31053480/)]
29. Yu AXY, Liu ZA, Pou-Prom C, Lopes K, Kapral MK, Aviv RI, et al. Automating stroke data extraction from free-text radiology reports using natural language processing: instrument validation study. *JMIR Med Inform* 2021 May 04;9(5):e24381 [FREE Full text] [doi: [10.2196/24381](https://doi.org/10.2196/24381)] [Medline: [33944791](https://pubmed.ncbi.nlm.nih.gov/33944791/)]
30. Wheeler E, Mair G, Sudlow C, Alex B, Grover C, Whiteley W. A validated natural language processing algorithm for brain imaging phenotypes from radiology reports in UK electronic health records. *BMC Med Inform Decis Mak* 2019 Sep 09;19(1):184 [FREE Full text] [doi: [10.1186/s12911-019-0908-7](https://doi.org/10.1186/s12911-019-0908-7)] [Medline: [31500613](https://pubmed.ncbi.nlm.nih.gov/31500613/)]
31. Sung S, Lin C, Hu Y. EMR-based phenotyping of ischemic stroke using supervised machine learning and text mining techniques. *IEEE J Biomed Health Inform* 2020 Oct;24(10):2922-2931 [doi: [10.1109/jbhi.2020.2976931](https://doi.org/10.1109/jbhi.2020.2976931)]
32. Sung S, Chen K, Wu DP, Hung L, Su Y, Hu Y. Applying natural language processing techniques to develop a task-specific EMR interface for timely stroke thrombolysis: A feasibility study. *Int J Med Inform* 2018 Apr;112:149-157 [doi: [10.1016/j.ijmedinf.2018.02.005](https://doi.org/10.1016/j.ijmedinf.2018.02.005)] [Medline: [29500013](https://pubmed.ncbi.nlm.nih.gov/29500013/)]
33. Shek A, Jiang Z, Teo J, Au Yeung J, Bhalla A, Richardson MP, et al. Machine learning-enabled multitrust audit of stroke comorbidities using natural language processing. *Eur J Neurol* 2021 Dec 29;28(12):4090-4097 [doi: [10.1111/ene.15071](https://doi.org/10.1111/ene.15071)] [Medline: [34407269](https://pubmed.ncbi.nlm.nih.gov/34407269/)]
34. Rannikmäe K, Wu H, Tominey S, Whiteley W, Allen N, Sudlow C, et al. Developing automated methods for disease subtyping in UK Biobank: an exemplar study on stroke. *BMC Med Inform Decis Mak* 2021 Jun 15;21(1):191 [FREE Full text] [doi: [10.1186/s12911-021-01556-0](https://doi.org/10.1186/s12911-021-01556-0)] [Medline: [34130677](https://pubmed.ncbi.nlm.nih.gov/34130677/)]
35. Ong CJ, Orfanoudaki A, Zhang R, Caprasso FPM, Hutch M, Ma L, et al. Machine learning and natural language processing methods to identify ischemic stroke, acuity and location from radiology reports. *PLoS One* 2020 Jun 19;15(6):e0234908 [FREE Full text] [doi: [10.1371/journal.pone.0234908](https://doi.org/10.1371/journal.pone.0234908)] [Medline: [32559211](https://pubmed.ncbi.nlm.nih.gov/32559211/)]
36. Mowery DL, Chapman BE, Conway M, South BR, Madden E, Keyhani S, et al. Extracting a stroke phenotype risk factor from Veteran Health Administration clinical reports: an information content analysis. *J Biomed Semantics* 2016 May 10;7(1):26 [FREE Full text] [doi: [10.1186/s13326-016-0065-1](https://doi.org/10.1186/s13326-016-0065-1)] [Medline: [27175226](https://pubmed.ncbi.nlm.nih.gov/27175226/)]
37. Li M, Lang M, Deng F, Chang K, Buch K, Rincon S, et al. Analysis of stroke detection during the COVID-19 pandemic using natural language processing of radiology reports. *AJNR Am J Neuroradiol* 2021 Mar;42(3):429-434 [FREE Full text] [doi: [10.3174/ajnr.A6961](https://doi.org/10.3174/ajnr.A6961)] [Medline: [33334851](https://pubmed.ncbi.nlm.nih.gov/33334851/)]
38. Leung LY, Fu S, Luetmer PH, Kallmes DF, Madan N, Weinstein G, et al. Agreement between neuroimages and reports for natural language processing-based detection of silent brain infarcts and white matter disease. *BMC Neurol* 2021 May 11;21(1):189 [FREE Full text] [doi: [10.1186/s12883-021-02221-9](https://doi.org/10.1186/s12883-021-02221-9)] [Medline: [33975556](https://pubmed.ncbi.nlm.nih.gov/33975556/)]
39. Kim C, Zhu V, Obeid J, Lenert L. Natural language processing and machine learning algorithm to identify brain MRI reports with acute ischemic stroke. *PLoS One* 2019;14(2):e0212778 [FREE Full text] [doi: [10.1371/journal.pone.0212778](https://doi.org/10.1371/journal.pone.0212778)] [Medline: [30818342](https://pubmed.ncbi.nlm.nih.gov/30818342/)]
40. Kent DM, Leung LY, Zhou Y, Luetmer PH, Kallmes DF, Nelson J, et al. Association of silent cerebrovascular disease identified using natural language processing and future ischemic stroke. *Neurology* 2021 Sep 28;97(13):e1313-e1321 [FREE Full text] [doi: [10.1212/WNL.00000000000012602](https://doi.org/10.1212/WNL.00000000000012602)] [Medline: [34376505](https://pubmed.ncbi.nlm.nih.gov/34376505/)]

41. Lin C, Hsu K, Liang C, Lee T, Liou C, Lee J, et al. A disease-specific language representation model for cerebrovascular disease research. *Comput Methods Programs Biomed* 2021 Nov;211:106446 [FREE Full text] [doi: [10.1016/j.cmpb.2021.106446](https://doi.org/10.1016/j.cmpb.2021.106446)] [Medline: [34627022](https://pubmed.ncbi.nlm.nih.gov/34627022/)]
42. Guan W, Ko D, Khurshid S, Trisini Lipsanopoulos AT, Ashburner JM, Harrington LX, et al. Automated electronic phenotyping of cardioembolic stroke. *Stroke* 2021 Jan;52(1):181-189 [FREE Full text] [doi: [10.1161/STROKEAHA.120.030663](https://doi.org/10.1161/STROKEAHA.120.030663)] [Medline: [33297865](https://pubmed.ncbi.nlm.nih.gov/33297865/)]
43. Garg R, Oh E, Naidech A, Kording K, Prabhakaran S. Automating ischemic stroke subtype classification using machine learning and natural language processing. *J Stroke Cerebrovasc Dis* 2019 Jul;28(7):2045-2051 [doi: [10.1016/j.jstrokecerebrovasdis.2019.02.004](https://doi.org/10.1016/j.jstrokecerebrovasdis.2019.02.004)] [Medline: [31103549](https://pubmed.ncbi.nlm.nih.gov/31103549/)]
44. Farran D, Bean D, Wang T, Msosa Y, Casetta C, Dobson R, et al. Anticoagulation for atrial fibrillation in people with serious mental illness in the general hospital setting. *J Psychiatr Res* 2022 Sep;153:167-173 [FREE Full text] [doi: [10.1016/j.jpsychires.2022.06.044](https://doi.org/10.1016/j.jpsychires.2022.06.044)] [Medline: [35816976](https://pubmed.ncbi.nlm.nih.gov/35816976/)]
45. Elkin PL, Mullin S, Mardekian J, Crouner C, Sakilay S, Sinha S, et al. Using artificial intelligence with natural language processing to combine electronic health record's structured and free text data to identify nonvalvular atrial fibrillation to decrease strokes and death: evaluation and case-control study. *J Med Internet Res* 2021 Nov 09;23(11):e28946 [FREE Full text] [doi: [10.2196/28946](https://doi.org/10.2196/28946)] [Medline: [34751659](https://pubmed.ncbi.nlm.nih.gov/34751659/)]
46. Bacchi S, Gluck S, Koblar S, Jannes J, Kleinig T. Automated information extraction from free-text medical documents for stroke key performance indicators: a pilot study. *Intern Med J* 2022 Feb 20;52(2):315-317 [doi: [10.1111/imj.15678](https://doi.org/10.1111/imj.15678)] [Medline: [35187820](https://pubmed.ncbi.nlm.nih.gov/35187820/)]
47. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv. 2019 May 24. URL: <https://arxiv.org/abs/1810.04805> [accessed 2022-12-12]
48. Pencina MJ, D'Agostino RB, D'Agostino RB, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008 Jan 30;27(2):157-72; discussion 207 [doi: [10.1002/sim.2929](https://doi.org/10.1002/sim.2929)] [Medline: [17569110](https://pubmed.ncbi.nlm.nih.gov/17569110/)]
49. Harkema H, Dowling JN, Thornblade T, Chapman WW. ConText: an algorithm for determining negation, experienter, and temporal status from clinical reports. *J Biomed Inform* 2009 Oct;42(5):839-851 [FREE Full text] [doi: [10.1016/j.jbi.2009.05.002](https://doi.org/10.1016/j.jbi.2009.05.002)] [Medline: [19435614](https://pubmed.ncbi.nlm.nih.gov/19435614/)]
50. Resnick MP, LeHouillier F, Brown SH, Campbell KE, Montella D, Elkin PL. Automated modeling of clinical narrative with high definition natural language processing using Solor and Analysis Normal Form. *Stud Health Technol Inform* 2021 Nov 18;287:89-93 [FREE Full text] [doi: [10.3233/SHTI210822](https://doi.org/10.3233/SHTI210822)] [Medline: [34795088](https://pubmed.ncbi.nlm.nih.gov/34795088/)]
51. Wu H, Toti G, Morley KI, Ibrahim ZM, Folarin A, Jackson R, et al. SemEHR: A general-purpose semantic search system to surface semantic data from clinical notes for tailored care, trial recruitment, and clinical research. *J Am Med Inform Assoc* 2018 May 01;25(5):530-537 [FREE Full text] [doi: [10.1093/jamia/ocx160](https://doi.org/10.1093/jamia/ocx160)] [Medline: [29361077](https://pubmed.ncbi.nlm.nih.gov/29361077/)]
52. Huang K, Altosaar J, Ranganath R. ClinicalBERT: modeling clinical notes and predicting hospital readmission. arXiv. 2020 Nov 29. URL: <https://arxiv.org/abs/1904.05342> [accessed 2022-12-12]
53. Alsentzer E, Murphy J, Boag W, Weng WH, Jindi D, Naumann T, et al. Publicly available clinical BERT embeddings. 2019 Presented at: 2nd Clinical Natural Language Processing Workshop; June 2019; Minneapolis, MN [doi: [10.18653/v1/w19-1909](https://doi.org/10.18653/v1/w19-1909)]
54. Zhang Y, Chen Q, Yang Z, Lin H, Lu Z. BioWordVec, improving biomedical word embeddings with subword information and MeSH. *Sci Data* 2019 May 10;6(1):52 [doi: [10.1038/s41597-019-0055-0](https://doi.org/10.1038/s41597-019-0055-0)] [Medline: [31076572](https://pubmed.ncbi.nlm.nih.gov/31076572/)]
55. Stearns MQ, Price C, Spackman KA, Wang AY. SNOMED clinical terms: overview of the development process and project status. *Proc AMIA Symp* 2001:662-666 [FREE Full text] [Medline: [11825268](https://pubmed.ncbi.nlm.nih.gov/11825268/)]
56. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res* 2004 Jan 01;32(Database issue):D267-D270 [FREE Full text] [doi: [10.1093/nar/gkh061](https://doi.org/10.1093/nar/gkh061)] [Medline: [14681409](https://pubmed.ncbi.nlm.nih.gov/14681409/)]
57. Afshar M, Sharma B, Dligach D, Oguss M, Brown R, Chhabra N, et al. Development and multimodal validation of a substance misuse algorithm for referral to treatment using artificial intelligence (SMART-AI): a retrospective deep learning study. *Lancet Digit Health* 2022 Jun;4(6):e426-e435 [FREE Full text] [doi: [10.1016/S2589-7500\(22\)00041-3](https://doi.org/10.1016/S2589-7500(22)00041-3)] [Medline: [35623797](https://pubmed.ncbi.nlm.nih.gov/35623797/)]
58. Cho A, Min IK, Hong S, Chung HS, Lee HS, Kim JH. Effect of applying a real-time medical record input assistance system with voice artificial intelligence on triage task performance in the emergency department: prospective interventional study. *JMIR Med Inform* 2022 Aug 31;10(8):e39892 [FREE Full text] [doi: [10.2196/39892](https://doi.org/10.2196/39892)] [Medline: [36044254](https://pubmed.ncbi.nlm.nih.gov/36044254/)]
59. Chowdhary KR. Natural language processing. In: *Fundamentals of artificial intelligence*. India: Springer; 2020:603-649
60. Patel JM. Introduction to common crawl datasets. In: *Getting structured data from the internet: running web crawlers/scrapers on a big data production scale*. New York: Apress; 2020:277-324
61. Topal MO, Bas A, van Heerden I. Exploring transformers in natural language generation: GPT, BERT, and XLNet. arXiv. 2021 Feb 16. URL: <https://arxiv.org/abs/2102.08036> [accessed 2022-12-12]
62. Fan L, Li L, Ma Z, Lee S, Yu H, Hemphill L. A bibliometric review of large language models research from 2017 to 2023. arXiv. 2023 Apr 03. URL: <https://arxiv.org/abs/2304.02020> [accessed 2023-08-03]

Abbreviations

AI: artificial intelligence

BERT: Bidirectional Encoder Representations from Transformers

DL: deep learning

EHR: electronic health record

LLM: large language model

MCA: multiple correspondence analysis

ML: machine learning

NLP: natural language processing

PRISMA-ScR: Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews

SNOMED: Systematized Nomenclature of Medicine

UMLS: Unified Medical Language System

Edited by C Lovis; submitted 03.05.23; peer-reviewed by J Heo, SF Sung; comments to author 05.06.23; revised version received 26.07.23; accepted 28.07.23; published 06.09.23

Please cite as:

De Rosario H, Pitarch-Corresa S, Pedrosa I, Vidal-Pedrós M, de Otto-López B, García-Mieres H, Álvarez-Rodríguez L

Applications of Natural Language Processing for the Management of Stroke Disorders: Scoping Review

JMIR Med Inform 2023;11:e48693

URL: <https://medinform.jmir.org/2023/1/e48693>

doi: [10.2196/48693](https://doi.org/10.2196/48693)

PMID: [37672328](https://pubmed.ncbi.nlm.nih.gov/37672328/)

©Helios De Rosario, Salvador Pitarch-Corresa, Ignacio Pedrosa, Marina Vidal-Pedrós, Beatriz de Otto-López, Helena García-Mieres, Lydia Álvarez-Rodríguez. Originally published in JMIR Medical Informatics (<https://medinform.jmir.org>), 06.09.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <https://medinform.jmir.org/>, as well as this copyright and license information must be included.