Scoping Review Protocol and Template

This template is intended to help plan a scoping review. The information provided in the completed form can be used for the registration of the scoping review. The template incorporates the relevant items listed in Chapter 11 of the JBI Manual for Evidence Synthesis.

History of changes

9/5/2022	Draft
29/06/2022	Completion (data extraction, analysis and evidence)
20/10/2022	Narrow scope to Natural Language Processing - include MCA

Title of the scoping review

Applications of Natural Language Processing for the management of Stroke disorders: a scoping review

Context of the review

First reviewer	Helios De Rosario
Other reviewers	Salvador Pitarch, Marina Vidal (IBV) Ignacio Pedrosa, Beatriz de Otto, Helena García, Lydia Álvarez (CTIC)
Funding source	CERVERA IBERUS

Introduction

Background of the review

Explain the reason for undertaking the scoping review should be clearly stated. If existing scoping reviews or systematic reviews are available on the topic, a justification that specifies how the proposed review will differ from those already conducted should be detailed.

Stroke, called also as a brain attack, happens when blood cannot get to your brain, because of a blocked or burst artery. The impact and effects of the stroke depends on the area of the brain it damages. It can affect to the behaviour, sense perception and/or motor control.

Stroke, along with ischaemic heart disease, is one of the leading causes of mortality and disability. Although stroke mortality is decreasing, the prevalence of people living with the effects of stroke has increased because of the growing and ageing population (Stinear, 2020). So, nowadays stroke still remains a major health-related challenge.

Different economic and social costs, related to hospitalization, treatment and recovery and for stroke patients, are increasing. Consequently, there is an increased need of advanced technologies that can assist in clinical diagnosis, treatment, predictions of clinical events, recommendation of promising therapeutic interventions, rehabilitation programs, etc. (Sirsat, 2020). For example, rapid diagnosis and treatment of stroke is crucial and leads to improved outcomes and prognosis among patients treated within the “golden hour” (Abedi, 2020).

In this context of growing incidence on the worldwide population, novel approaches that complement and go beyond evidence-based medicine are required.

In this regard Artificial Intelligence (AI) has been discussed during last years in the medical literature. AI could be defined as an interdisciplinary science with multiple approaches, but advancements in machine learning (ML) and deep learning (DL) are creating a paradigm shift in virtually every sector of life.

Methods based on machine learning to process electronic health records (EHR) data is resulting in improved understanding of patient clinical trajectories, as well as chronic disease risk prediction. AI can be applied to various types of healthcare data (structured and unstructured). Popular AI techniques include machine learning methods for structured data, such as the classical support vector machine and neural network, and the modern deep learning, as well as natural language processing for unstructured data (Jiang, 2017).

Unlocking the full potential of EHR data is contingent on the development of natural language processing (NLP) methods to automatically transform clinical text into structured clinical data that can guide clinical decisions and potentially delay or prevent disease onset (Sheikhalishahi, 2019).

In recent years, AI techniques have been used in more and more stroke-related studies (Jiang, 2017). The use of this techniques has been applied in research to the principal stages of the Brain stroke: prevention, diagnosis, treatment and prognosis.

Taking the advantage of large amount of data with rich information, AI is expected to help with studying much more complicated yet much closer to real-life clinical questions, which then leads to better decision making in stroke management. Recently, researchers have started endeavours along this direction and obtained promising initial results (Jiang, 2017).

There are a multitude of solutions capable of addressing by AI technologies in health care management challenges; however, there is a paucity of guidance on selecting appropriate methods tailored to the health care industry (Shahid,2019). Some recent reviews show ML approaches employed on various datasets used for solving various stroke problems, better healthcare systems management as well as further research and investigation.

The aim of this review is to **gather the knowledge that might help in that guidance, investigating how NLP is used to deliver a smarter health care in different phases of stroke disorders** (prevention, diagnosis, treatment and prognosis).

Objective and questions

The objective of this scoping review is to investigate how Natural Language Processing (NLP) (AI computer processing system) is used to deliver a “smarter” health care in different phases of stroke disorders (prevention, diagnosis, treatment and prognosis).

The primary questions are:

- In what phases of stroke is NLP used (*prevention, diagnosis, treatment and/or prognosis)*
- In what contexts of stroke management is NLP used (*health systems efficiency, risk alert, screening, medical discharge, radiology analysis, updated medical information support, outcome prediction, personalized care, domiciliary treatment…*)?

- What kind of clinical data are collected and used by NLP algorithms related to stroke process (*clinical notes, laboratory reports, images, physical examinations, questionnaires…)*?
- What are the clinical implications of using this computer processing systems (lower costs, improvement of patients QoL...?)

The review will also include the following secondary questions:

- What are the NLP methods and tools that are used in stroke studies?
- What kind of software and/or architectures are used to collect and process the data?
- Are the algorithms and NLP software specifically tuned for stroke?
- What tools have the best performance, and how do they compare to others?

Inclusion criteria

Types of participants

Important characteristics of participants, including age and other qualifying criteria that make them appropriate for the objectives of the scoping review and for the review question.

adult people that had suffered stroke, and people who were at risk of suffering due to predisposing vascular background or other diseases which increase the risk of developing stroke, like mental illness or ejection fraction.

Concept

The core concept examined by the scoping review to guide the scope and breadth of the inquiry.

Natural Language Processing algorithms used for stroke applications in health systems or individual health. Use cases, data and technologies.

Context

This may include, but is not limited to, consideration of cultural factors, such as geographic location and/or specific social, cultural, or gender-based interests, and specific settings.

Any context where prevention, treatment or rehabilitation of stroke might take place, ranging from early detection and out or inside the hospital, diagnosis and evaluation of cases, clinical decision-making, administration and monitoring of rehabilitation, and post-rehabilitation management.

Types of evidence sources

For the purposes of a scoping review, the "source" of information can include any existing literature, e.g. primary research studies, systematic reviews, meta-analyses, letters, guidelines, websites, blogs, etc.

Any type of scientific literature, and reports from scientific, medical or government institutions.

Search strategy

Electronic databases and websites

The search will be done primarily in PubMed, using "AND" combinations of keyword categories, each category being defined by an "OR" combination of synonyms or related vocabulary. The keyword categories with their respective synonyms are:

- Natural Language Processing
- Stroke

Other methods

Google Scholar will also be used to retrieve grey literature, taking the 200 first results, as recommended by Haddaway et al.[1]

Methods of review

Management and documentation of search results

How will search results be managed & documented? ie which reference management software, how duplicates dealt with.

The results will be managed in a Zotero library, where duplicates will be identified and removed. Titles and abstracts will be exported into a spreadsheet, which evaluators will use to mark the works according to their suitability to the review.

Selection process

Number of reviewers, how agreements to be reached and disagreements dealt with, etc.

All titles and abstracts after removal of duplicates will be screened by two reviewers independently, where those that do not meet the inclusion criteria will be excluded. Any disagreement will be discussed between the two reviewers in order to arrive to a common judgement.

The selected works will be distributed between the two reviewers and three further revieweres, in order to determine their eligibility based on the full text, recording reasons for rejecting them. The record of rejected works will be shared between the reviewers, in order to confirm the decisions of either part. The final selection of papers will be redistributed between the reviewers for data extraction

Data extraction

Explain what information is to be collected on each included study. If databases or forms on Word or Excel are used, were these piloted and how is this recorded.

A table shall be filled out with the following data from each work included in the final selection:

- Year of publication

[1] Haddaway, N. R., Collins, A. M., Coughlin, D., & Kirk, S. (2015). The Role of Google Scholar in Evidence Reviews and Its Applicability to Grey Literature Searching. PLOS ONE, 10(9), e0138237. https://doi.org/10.1371/journal.pone.0138237

- Type of study (cohort study, RCT, etc.)
- Sample size
- Type of stroke considered
- Other medical conditions besides stroke considered
- Context (diagnosis, prognosis, treatment)
- Clinical benefits expected by the data processing and analysis
- Societal benefits expected by the data processing and analysis
- Economic benefits expected by the data processing and analysis
- Data analysis architecture (Machine Learning, Deep Learning, etc.)
- NLP methods used to analyse text
- Other statistical methods used in the study
- Software packages used (if named)
- Metrics used to compare methods (if there are comparisons)
- Best method (if there are comparisons)

Analysis of the evidence

Explain what kind of analysis will be done to the extracted data: codification, how data will be aggregated, etc.

A quantitative analysis will be done of frequency of occurrence of the different contexts, data sources, statistical and Machine Learning methods, with particular focus on NLP tools, in order to assess the most frequent types of studies.

Relationships between answers to the different questions will be done through Multiple Correspondence Analysis (MCA).

Presentation of the results

How results and data will be presented (e.g. draft chart, figures or tables).

The results will contain:

- Tables with the extracted data for the selected papers.
- Frequency tables for the different categories of answers.
- MCA diagrams to discern qualitatively if there are questions whose particular answers concur frequently.

Bibliography

Reference bibliography cited in the text

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. Epub 2020 Jan 28. PMID: 32004440.

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. Epub 2020 Jul 28. PMID: 32912543.

Sheikhalishahi S, Miotto R, Dudley J, Lavelli A, Rinaldi F, Osmani V. **Natural Language Processing of Clinical Notes on Chronic Diseases: Systematic Review**. JMIR Med Inform 2019;7(2):e12239. URL: https://medinform.jmir.org/2019/2/e12239. DOI: 10.2196/12239

Abedi V, Khan A, Chaudhary D, Misra D, Avula V, Mathrawala D, Kraus C, Marshall KA, Chaudhary N, Li X, Schirmer CM, Scalzo F, Li J, Zand R. **Using artificial intelligence for improving stroke diagnosis in emergency departments: a practical framework**. Ther Adv Neurol Disord. 2020 Aug 25; 13:1756286420938962. doi: 10.1177/1756286420938962. PMID: 32922515; PMCID: PMC7453441.

Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, Wang Y, Dong Q, Shen H, Wang Y. **Artificial intelligence in healthcare: past, present and future**. Stroke Vasc Neurol. 2017 Jun 21;2(4):230-243. doi: 10.1136/svn-2017-000101. PMID: 29507784; PMCID: PMC5829945.

Shahid N, Rappon T, Berta W. **Applications of artificial neural networks in health care organizational decision-making: A scoping review**. PLoS One. 2019 Feb 19;14(2): e0212356. doi: 10.1371/journal.pone.0212356. PMID: 30779785; PMCID: PMC6380578.