



Oct 04, 2020

NLP of radiology reports: systematic review protocol

Beatrice Alex¹, Arlene Casey¹, Emma Davidson¹, Hang Dong¹, Daniel Duma¹, Andreas Grivas¹, Claire Grover¹, Victor Suarez Paniagua¹, [Michael Tin Chung Poon](#)¹, Richard Tobin¹, William Whiteley¹, Honghan Wu¹

¹University of Edinburgh

1 *Works for me* dx.doi.org/10.17504/protocols.io.bmwhk7b6

Clinical NLP group, University of Edinburgh

Michael Tin Chung Poon

DOI

dx.doi.org/10.17504/protocols.io.bmwhk7b6

DOCUMENT CITATION

Beatrice Alex, Arlene Casey, Emma Davidson, Hang Dong, Daniel Duma, Andreas Grivas, Claire Grover, Victor Suarez Paniagua, Michael Tin Chung Poon, Richard Tobin, William Whiteley, Honghan Wu 2020. NLP of radiology reports: systematic review protocol. **protocols.io**
<https://dx.doi.org/10.17504/protocols.io.bmwhk7b6>

LICENSE

This is an open access document distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

CREATED

Sep 30, 2020

LAST MODIFIED

Oct 04, 2020

DOCUMENT INTEGER ID

42665

DISCLAIMER:

DISCLAIMER – FOR INFORMATIONAL PURPOSES ONLY; USE AT YOUR OWN RISK

The protocol content here is for informational purposes only and does not constitute legal, medical, clinical, or safety advice, or otherwise; content added to [protocols.io](#) is not peer reviewed and may not have undergone a formal approval of any kind. Information presented in this protocol should not substitute for independent professional judgment, advice, diagnosis, or treatment. Any action you take or refrain from taking using or relying upon the information presented here is strictly at your own risk. You agree that neither the Company nor any of the authors, contributors, administrators, or anyone else associated with [protocols.io](#), can be held responsible for your use of the information contained in or linked to this protocol or any of our Sites/Apps and Services.

Review title

Natural language processing (NLP) of radiology reports from 2015-2019: a systematic review

Anticipated or actual start date

February 2020

Anticipated completion date

September 2020

Named contact

Michael Poon

Organisational affiliation of the review

University of Edinburgh

Review team members and their organisational affiliations

(In alphabetical order)

Beatrice Alex

Arlene Casey

Emma Davidson

Hang Dong

Daniel Duma

Andreas Grivas

Claire Grover

Victor Suarez Paniagua

Michael Poon

Richard Tobin

William Whiteley

Honghan Wu

Funding sources/sponsors

The Alan Turing Institute

Review question

- (1) What methods are used in natural language processing of radiology reports?
- (2) Has the use of these methods changed over between 2015 and 2019?
- (3) What are the clinical applications of these natural language processing methods?
- (4) Are datasets and codes of natural language processing available in published studies 2015-2019?

Searches

We developed an automated search strategy in Google Scholar with additional metadata collected from crossref, pubmed, semantic scholar, arxiv and unpaywall. Search terms include ("radiology" OR "radiologist") AND ("natural language" OR "text mining" OR "information extraction" OR "document classification" OR "word2vec") NOT patient. Citations retrieved underwent a snowballing process whereby a reference list of the previous systematic review (Pons et al) is also included. We then applied an automatic exclusion of citations if the citation is: (1) non-English; (2) a patent document; (3) published before 2015; (4) a review article; (5) about radiology images only rather than reports; (6) not relating to radiology; (7) not natural language processing; (8) not available in full text; (9) duplicates; (10) reviews, conference abstracts, comments, or editorials; and (11) case reports.

Condition or domain being studied

Natural language processing of radiology reports

Participants/population

Patients who underwent any radiological investigation from which a report was generated. We exclude studies of patients in whom radiological images were analysed without radiology reports.

Interventions, exposures

No specified exposure

Comparators/control

There are two main groups of comparators. Studies may compare their NLP systems with expert annotated radiology reports or may compare it with another NLP systems. We anticipate some studies comparing several NLP systems or in combination with expert-annotated reports. Expert-annotated reports would be considered the control.

Types of study to be included

We include cross-sectional studies where a corpus of radiology reports was annotated and/or analysed. Studies that adopted a pseudo-case-control where they included radiology reports from patients with or without a disease of interest are included. We also include cohort studies where outputs from text analytic systems were used as exposures/outcomes.

We exclude a study if it is: (1) a case report; (2) published before 2015; (3) in language other than English; (4) relating to radiology images only; (5) a review/conference abstract/comments/editorial; (6) not reporting descriptive/outcomes of interest; (7) not relating to radiology report; (8) not using natural language processing methods; (9) not available in full text; (10) duplications.

Main outcomes

The main outcome is the performance of the NLP being used to perform the designated task. However, this may not be applicable to studies that used NLP for cohort selection in epidemiological studies.

Measures of effect

The main outcomes are the precision (positive predictive value), recall (sensitivity), and F1 score (the harmonic mean of precision and recall) associated with the NLP being used to perform the designated task (where applicable).

Additional outcomes

Not applicable

Measures of effect

Not applicable

Data extraction

Three reviewers screened all titles and abstracts of potentially eligible studies from the search strategy using Rayyan online platform. We discussed and resolved citations where two reviewers excluded it and one reviewer included it. All other citations were included for full eligibility assessment. A team of six reviewers assessed eligibility of the resulting citations with each study. We plan to double-review the included studies. We use a pre-specified data collection tool for recording eligibility assessment outcomes.

Data items extracted from studies include: study primary objective, data source(s), study period, language of radiology reports, anatomical region, imaging modality, disease area, size of dataset, annotated set size, training set size, validation set size, test set size, external validation performed, domain expert used, number of annotators, inter-annotator agreement, natural language processing technique(s) used, best reported recall, best reported precision, best reported F1 score, availability of dataset, and availability of code. Data extraction is performed by two reviewers independently. Recording of data extraction is done through a shared data collection tool. Any disagreement will be resolved by discussion during 2-weekly review team meeting.

Risk of bias (quality) assessment

There is currently no risk of bias tools applicable to the anticipated heterogeneous types of studies in this review. However, we

adopted aspects of the ROBINS-E tool to assess epidemiological aspects of the studies, where appropriate. Risk of bias measures relating to technical aspects were developed by the review team who has considerable knowledge on text analytic methods.

Strategy for data synthesis

Our objective is to provide descriptive data. We do not plan to summarise data using meta-analysis because of the anticipated heterogeneous study design, objectives, natural language processing techniques, and reported outcomes.

Analysis of subgroups or subsets

We will report descriptive statistics stratified by disease areas, NLP techniques, and year of publication. No meta-analysis will be performed.