Reviewer’s report

Title: A validated natural language processing algorithm for brain imaging phenotypes from radiology reports in UK electronic health records

Version: 0 Date: 30 Jun 2019

Reviewer: Sunyang Fu

Reviewer's report:

The paper demonstrated using NLP to identify stroke-related phenotypes from radiology reports. Overall the work is meaningful, but requires revision and editorial assistance.

General comments: The author should at least provide an overall summary of the rule-based features.

Specific comments:

Abstract:
- Please rephrase line 21 - 26 "We developed and tested..."

Introduction:
- Lines 45 - 50: the author should directly summarize the related work. For example, the NLP technique was applied to which problem and achieved % score.
- Please consider citing the following paper as the related work to introduction and background:


Method:
- Lines 46 - 49: please add more details about the experiment design. I.e. how did you divide the datasets for training and test?

- In the introduction, you mentioned about the inter-observer variation as one of the challenges of reading brain image, however, the study achieved extremely high Cohen's kappa. Please discuss and explain the possible reasons, such as good practice, use of any abstraction protocols? How did you measure the kappa score of the total 7186 entities, and what's the definition of each entity?

- Understand the tool may not be open source but suggest to provide important features, lexicon and rules used for the study to address the limitation of generalizability and portability.

Result:
- Suggest to add micro average f-score to the evaluation
- Suggest to add the comparison between NLP and ICD (or combined) to show the advantage of using unstructured text

Discussion:
- Please discuss the potential limitation due to the discrepancy between the imaging report and the actual images.
- Please discuss the potential limitation of NLP portability and generalizability

Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

No

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Yes

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

Yes
Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics

Quality of written English
Please indicate the quality of language in the manuscript:

Needs some language corrections before being published

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?
2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?
3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?
4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?
5. Do you have any other financial competing interests?
6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal