Reviewer’s report

Title: Use of Natural language processing to improve predictive models for imaging utilization in children presenting to the Emergency Department

Version: 1 Date: 08 Jul 2019

Reviewer: Jiaping Zheng

Reviewer's report:

The authors presented a study that evaluates the effectiveness of including unstructured data in predicting imaging utilization in pediatric ED. The study compared the performance of using only structured variables, only unstructured data, and combining the two sources. The overall design of the study is reasonable. However, there are some concerns I have regarding the details.

1. There is no baseline system to compare against the various systems. Although the main focus of the paper is the effectiveness of the unstructured data, having a competitive baseline ensures that NLP methods provide value beyond what is available in the structured variables, and not merely compensating for the low performance of an algorithm.

2. Line 117, how are the word pairs extracted? Are all bigrams included as features or only two-word phrases included? The cutoff the sparse term seems rather high (99.9% on line 118). Please provide more details.

3. It seems (from table 3) that the structured data barely adds any predictive power. The authors many want to discuss why this is the case.

4. The authors spent a large amount of the paper discussing the characteristics of the patients in the dataset. I wonder how this is relevant to the methodology or the results. Unless the authors discuss how it may impact the system design or performance, I don't think this is necessary.

5. The limitation of NLP reads like a generic description, rather than limitations in the methods employed by the authors.
Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

Yes

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

Unable to assess

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:

Acceptable

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.

None

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.