Reviewer’s report

Title: Evaluation of standard and semantically-augmented distance metrics for neurology patients

Version: 0 Date: 01 May 2020

Reviewer's report:

It's a decent paper, not sure it's completely novel, but okay. I have some technical issues with the paper though, such as:

1) There are ton of patient-distance metrics already in existence using similar calculations. The authors seem to gloss over this. The literature review here is completely inadequate. The authors are not the first ones to do this. This needs to be fixed.

2) I would like to see more details about how the "case abstraction" was done exactly. SNOMED and other medical ontologies often have several similar and competing terms for any given "symptom", so one of the challenges is figuring out exactly how to map clinician notes to various codes. I would like to know how the authors dealt with that challenge, in situations where mapping was not so cut & dry?

3) Table 3 is fine, but if you are going to use some machine learning technique to test your distance metric model, then you need to compare its performance with other distance metrics somehow. Otherwise, there is no context for your accuracy and F1 scores. They are just meaningless metrics in a vaccum. That's now how ML is done. You have to provide enough information so the readers can interpret your results. Needs to be fixed.

4) What's the purpose of Figures 10 and 11 and the "problem of proximity" section? That some diagnoses are more similar to each other than others? I think people already know that. Whatever it is you are trying to communicate in that section is not clear, or if it is it is not a novel finding. The authors need to clarify.

5) In the discussion section, the authors state: "These accuracies are comparable to the bedside diagnostic accuracy of human experts"? So you are saying that after all this work, your model basically doesn't perform any better than humans already do ... so again what is the takeaway for the reader? This approach doesn't seem to be an improvement then, or am I misreading that section?

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

Unable to assess

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

Reviewer Publication Consent. 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) if this manu-
script is accepted for publication. Any comments that I do not wish to be included in the published report have been included as confidential comments to the editor, which will not be published. If you are not happy for us to publish this report, please contact the editorial office before completing the review. If you wish, you can include your name in your published report. Please note you must decide whether to include your name at the start of the process and confirm this decision whenever prompted. If you change your decision later, we will not be able to publish your name.

I agree to the terms of the CC-BY 4.0 license; please do not publish my name with my report.

(default)