Reviewer’s report

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

Version: 0 Date: 05 May 2020

Reviewer's report:

My recommendation is to reject this manuscript for a publication in MIDM journal, for the three following reasons:

1 - The motivation and the contribution of this work are unclear

I understand that this work is about computer-aided diagnosis of neurological diseases, but the reasons and ways of using distance measurements for this are not really explained.

The sentence "Neurologists lack a mathematical representation of diagnosis stereotypy, diagnosis proximity, and diagnosis difficulty. A distance metric for neurological patients is a step towards representing these ideas in mathematical form." is too vague from my point of view. The authors should further explained why distance metrics are essential/helpful for neurological diagnosis. How can they help the diagnosis?

There are also many other classification methods that could be used with the proposed ontology-based representation of patients (sets or vectors), but the authors do not mention these other options. In my opinion, this proposal would be a lot more convincing if the reasons to use patient distance metrics are carefully explained in the first section and maybe if another approach is included in the experimental part for comparison purpose.

Finally, the novelty of this work compared to existing works is not completely clear to me. The distance measurements have already been published, as well as the representation by ontologies. I think it is necessary to explain more clearly the contributions of this work.

2 - The description of the method is partial

The method is only partially described: Two distance measures are used to compute inter-concept and inter-patient distances but the 'Methods' section does not explain how these measures are exploited afterward. If the goal is to use a k-nearest neighbor method, it seems to me that only the inter-patient distance is necessary. In any case, it is not explained in the manuscript. The 'Methods' section should at least describe a complete procedure for going from patient description to diagnosis.

3 - The experimental protocol is partially described too
There are many aspects of the experimental protocol and analysis that are not given in the manuscript. I give a (non-exhaustive) list below:
- Why to group the different diagnosis into four tranches, and why these tranches? What it brings to the analysis is unclear to me
- How is performed the training and test phase? Are the performance of kNN measured on an independent test set of patient? Several replications?
- Why use the proposed distance measure in particular? How do the alternatives for classification with kNN behave?
- How does the proposed method compare to other classification methods?

**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.

No

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

No

**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.

I declare that I have no competing interests

**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 manuscript 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)