Reviewer’s report

Title: Importance of Medical Data Preparation in Predictive Modeling and Risk Factor Discovery for the Frailty Syndrome

Version: 2 Date: 11 Oct 2018

Reviewer: Klaus-Hendrik Wolf

Reviewer's report:

There is one concern that already appears in the abstract, but is present throughout the paper.

Abstract-Conclusions: while term 'predictive-model' is technically the correct choice, the models try to distinguish non-frail from pre-frail and frail patients using current state of a patient, thus 'diagnostic-model' and especially 'diagnostic-factor' might be a better term. This recurs in the objectives of the paper. IMHO 'predictors' should be 'symptoms', 'indicators' or 'signs'. To my understanding the model only relies on data taken at the same time as the diagnosis is made. No temporal data mining approaches are applied. Neither is data used to predict the outcome or future state of patients.

With regard to the models you might want do discuss the interpretability of the resulting models. E.g. with decision trees it is a lot easier to interpret and understand how a classification is reached.

The last sentence of results is actually future work and it would be good to express the need to validate the model found in retrospective data via prospective trials.

It is quite a broad statement, that the work 'contributes towards the possibility of obtaining predictive models that can anticipate the onset of age-related deterioration'. I would strongly recommend to formulate more decently. It is for sure a step in the right direction, but there is paramount work ahead.

Overall, with the necessary focussing on the main objectives it became more apparent to me, that the overall contribution of the paper IMHO is quite limited.

Minor comments:

- Please introduce abbreviations (e.g. RF).
- There are still parts where the language could (and should) be improved
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.

Yes

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:

Not suitable for publication unless extensively edited

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