Reviewer’s report

Title: Predictive Models for Diabetes Mellitus Using Machine Learning Techniques

Version: 1 Date: 17 May 2019

Reviewer: Seth Berkowitz

Reviewer's report:

The authors have significantly revised the manuscript, however they have not fully addressed my concerns. My main concern was that, despite the title of the paper stating machine learning methods outperformed current models, this conclusion was not supported by their results. Though the authors have added a procedure for quantifying uncertainty through repeated cross-validations, I am not sure that this is an appropriate method for confidence interval construction (and the authors given no citation or rationale or proof for why wald-style CI's derived from their procedure would be valid). I am also not certain that their t-test results (esp. the 1-sided results) on a sample of 30 means all resampled from the same data is the best way to compare the AUC's (they also give no citation for this). However, even if we accept these, the authors note that the differences, even if statistically significant, are not practically meaningful. Further, the AUC in the Pima Indian cohort is larger for logistic regression than GBM (at least for the first test set that they report in the text), and even though the authors did not compare their models to any current DM prediction rules (as they do not have the data to do so), the authors still claim that machine learning methods outperform current prediction rules. This remains unsupported in my opinion.

Other issue:

I had previously asked the authors to provide the definition of diabetes used for their analysis which is critical for understanding and being able to reproduce their work. They do not provide this in the manuscript or in the response letter. They add a reference at the end of the reference section that does not appear (at least as far as I can tell) to correspond to any statement in the manuscript text, and thus it is not even clear, from the manuscript, that this reference contains the definition used. I do not think it is acceptable to not provide the definition of the outcome used in the study, which should appear in the methods 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.

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:

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 organisation 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.