Reviewer’s report

Title: Predictive Models for Diabetes Mellitus Using Machine Learning Techniques

Version: 0 Date: 22 Mar 2019

Reviewer: Seth Berkowitz

Reviewer's report:

This is a well written paper describing development of a diabetes mellitus prediction model. Mostly the methods are appropriate but there seem to be some significant oversights that render their conclusions unsupported.

The authors claim that the GBM model is superior to the logistic regression model and the random forest model based on a nominally larger AUC, but they do not account for uncertainty in the estimates and do not provide any confidence intervals. I am not convinced that the GBM model is truly superior and not acknowledging the uncertainty of their estimates is a major oversight. They also claim machine learning methods outperform current prediction rules, but did not compare their models to the current prediction rules they cite (because they do not have sufficient data).

Additional issues:

Excluding patients who are on insulin seems unjustified as many patients with type 2 diabetes receive insulin treatment. The authors should explore other methods for identifying individuals with type 1 diabetes if they need to remove them.

The authors do not describe how the diagnosis of diabetes was made

The authors do not describe the time period that proceeded the predictor data used and the diagnosis of diabetes

The authors mention hba1c in the formulation of their dataset but then do not use it as a predictor, which seems unusual given that it is a strong predictor of diabetes. If the reason is that many individuals did not have this measured, why include it in the dataset in the first place?

Why did the authors treat age as a categorical variable only for the logistic regression model but not for the GBM model?
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 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