Reviewer’s report

Title: Predictive Models for Diabetes Mellitus Using Machine Learning Techniques

Version: 0 Date: 19 Mar 2019

Reviewer: Martin Macaš

Reviewer's report:

Four different machine learning methods were compared on Diabetes Mellitus detection data consisting of records from Canadian patients. Regarding AROC criterion computed using hold-out error estimate, they found that Gradient Boosting and Logistic Regression outperformed random forest and Rpart model. Authors found that their results were comparable to Wilson et. al., Mashayekhi et al. and some scoring systems without using a parental history as predictor. Such a results does not sound convincingly. I do not want to force the authors to obtain 90% AROC, but would like to ask them to point out, what is their added value and why such a result is success and should be published.

It is not clear if the random forest and decision tree were also pre-tuned using cross-validation or if some default parameters were used. For GBM, at least four parameters were tuned (n.trees, interaction.depth, n.minobsinnode, shrinkage). It is nor described in the article, if the parameters of the other methods were also tuned. I expect that no. This can lead to invalid conclusion that GBM is better method. Authors must clearly describe the tuning process for all models or they have to perform tuning for the other methods as well. Both the decision trees and random forests are highly sensitive on the parameters settings.

Quality of figures is very poor, authors should use EPS format for images.

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.

Yes
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