Reviewer’s report

Title: Application of three statistical models for predicting the risk of diabetes

Version: 1 Date: 10 Sep 2019

Reviewer: Nicole Hobbs

Reviewer's report:

Application of three statistical models in the prediction of diabetes mellitus
With high rates of undiagnosed type 2 diabetes and the intricacies observed in the Asia-Pacific population, an accurate method specific to the residents of China is valuable. This manuscript describes three models used to predict the occurrence of type 2 diabetes and risk factors identified to be significant are discussed. The developed neural network model outperformed the decision-tree and logistic regression model in the reported performance metrics.

Abstract


Background

1. In your sentence, "At present, the proportion of the potential population with diabetes in Chinese adults is as high as 15.5%, and 60.7% of these individuals are unable to obtain effective treatment and prevention education as early as possible without a prior diagnosis [4]" the final phrase is unclear. When you say 60.7% do not receive the education as early as possible without a prior diagnosis, do you mean that they had not previously received a diagnosis of pre-diabetes? Please revise this portion for clarity.

2. Line 13 - "Diabetes complications can be avoided…" - This is not completely correct. There are genetic factors at play in attention to how well their diabetes is controlled. Maybe state that "in some cases, diabetes complications can be avoided when a person with diabetes is diagnosed early, treated and maintains tight control of their blood sugar levels."

3. Line 32 - Remove etc

Methods

1. Line 30: Why were individuals with diagnosed type 2 diabetes excluded?

2. Line 31: State "of which 4177 met the inclusion criteria."
3. Line 48: What does it mean when waist circumference was measured with 1 cm in width?

4. Line 18: This sentence is confusing. "The participants were not diagnosed with dyslipidemia and were not taking lipid-lowering drugs." What was the criteria for dyslipidemia?

5. Did you consider using a stratified training and testing set? What percentage of people with type 2 diabetes fell in each set?

6. Line 29: Verify was spelled incorrectly.

Results

1. In this sentence, "Substituting the prediction model into the original data, with a critical value of 0.5, the results show that the prediction accuracy of the model was 90.8% and the area under the ROC curve was 0.711 (95% Cl: 0.697-0.725)." Should it say substituting into the testing data set?

2. You have a large portion on training the hyperparameters of the NN model. Did you consider separating into a training, testing and validation set? Please describe your reasoning to me and in the manuscript for why you did not do so.

3. In your model comparison section, did you complete 3 t-tests? Why not use ANOVA and Tukey’s HSD to reduce the risk of type 1 errors. If you did so, please specify.

Previous Comments & Responses:

1. "The height measurement tool used in this study has a range of 2 m, and the weight measurement tool has a range of 150 kg. The results show that there was no case outside the range."

a. If there were no measurements outside the range, it may be clearer to remove these maximum values. The accuracies of the measurement tools are likely sufficient alone.
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.

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.