Reviewer’s report

Title: Machine Learning Based Refined Differential Gene Expression Analysis of Pediatric Sepsis

Version: 1 Date: 21 Apr 2020

Reviewer's report:

1. How is the survival to non-survival ratio in the dataset of 199 pediatric patients? If it is an imbalanced dataset, the threshold should be adjusted and not 0.5. And the ACC, Sn, Sp and MCC metrics are affected by the threshold.

2. Why do authors select the LR model with top seven DEGs not 2 DEG? Page 9, Line 14, should it be 7 features not 2 features based model is dominated in the leftmost region (in Figure S3)? The top seven DEGs in feature selection is not the top DEGs ranked by the absolute fold change, so the formulation of "the top seven DEGs" is irrelevancy and the authors should show the list of the seven DEGs in feature selection.

3. There is no comparison of the model based on seven DEGs with the models using 10 DEGs. In this work, what is the purpose or significance of the LR model using seven DEGs by feature selection?

4. In the Section of Discussion, one of the two limitations is the computation time, how long is the running time? If the time is not very long, this limitation may be ignored.

5. Could the analysis of enriched pathway for the top 10 genes ranked by machine learning and the significance of these 10 genes for pediatric sepsis be added in the study?

6. Page 6, Line 39, 10 runs -> 100 runs, or Line 48, 100 runs -> 10 runs? They (10 and 100 runs) are confused in the paper.

7. AUC scores of the models should be added in the ROC curves after the name of models, for example, if Figure 5, RF100 (AUC = 0.86). The names of model in Table 3 are inconsistent with the names in Figure 5.

8. Could authors provide the codes or web-serve of the prediction model for the researches?

9. The numbers of significant digits retained should be consistent, three or four?

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