Reviewer’s report

Title: Early Prediction of Acquiring Acute Kidney Injury for Older Inpatients Using Most Effective Laboratory Test Results

Version: 0 Date: 20 Dec 2019

Reviewer: Huiyuan Chen

Reviewer's report:

In this paper, the author proposed an machine learning approach based on K-nearest neighbor (KNN) to predict the Acute Kidney Injury (AKI) by using the electronic health records (EHR). More importantly, they found that the earlier the AKI is predictive (up to 5 days prior to the AKI event) if more features (laboratory test, type) are included, and the Blood Urea Nitrogen is a very important factor in the prediction. In the experiments, the authors also compare with current classification method, including Logistic Regression, Random Forest, and AdaBoostM1. The results on a real-life EHR dataset, show the effectiveness of the proposed algorithm.

Positive Points:

1) The research topic is interesting and valuable. Using machine learning to investigate AKI is very important in clinical settings. The predictive results physicians adequate time for earlier intervention and improved patient outcomes.

2) Prediction of Acute Kidney Injury using Electronic Health Record is promising due to the emerging of EHR data.

Negative Points:

1) The technical contribution is limited since KNN classification is a well-built technique.

2) The paper assesses model performance only on patients older than the age of 60. Therefore, it may not be generalizable to other populations (age <60).

3) Some details are not clear. See below.

In the Revision, I would like to hear the feedback from the authors for the follow questions.
1) In page 7, "Although an inpatient may have multiple admissions (encounters), in this study, different encounters of the same inpatient are treated independently for the prediction". This is a very strong assumption. Any statistical test or related reference shows similar assumption.

2) In page 13, The study regards LR = |r_target - r_other| as similarity score. However, mathematically, LR is one-dimensional Manhattan distance e.g., L1 norm. In page 14, the claim "In this study, the distance is the similarity score between each pair of inpatients," is not true. "Distance" and "Similarity" are totally two different things. Not sure how this distance measurement can be used in KNN. One suggestion: transfer "Distance" to "Similarity".

3) In page 13, some details about the trend similarity of sequence (e., DTW) will be more clear. And some data visualization of EHR data will be helpful to readers about what is the "similarity of sequence". Also, a Table about statistics of dataset is suggested.

4) [25] built models to predict AKI 0 to 5 days prior to the possible AKI event for inpatients aged 18-64. A discussion between the proposed model and [25] should be provided.

5) The segmentation and data representation in page 11 are not clear. The motivation for r_last, r_j and s_j are not well described. The s_j seems to be sensitive w.r.t. time.

6) Instead of Table 1/Table 3, The area under ROC cure plot may be more useful. Also, any discussion why the KNN can outperform the ensemble learner? In Table 1, some results of Logistic regression are better than the proposed method, any explanation?

7) Some English written errors in the paper. For example, the sentence "In addition, we strictly extract the inpatients who develop AKI during hospitalization to make the results more convincing" appears twice in line 39 and line 52 in Page 8.

8) Some related work are missing. For example,


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.

No

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:

Needs some language corrections before being published

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