Author’s response to reviews

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

Authors:

Yi-Shian Chen (ilovebabowbow@gmail.com)
Che-Yi Chou (cychou.chou@gmail.com)
Arbee L.P. Chen (arbee@asia.edu.tw)

Version: 1 Date: 19 Jan 2020

Author's response to reviews:

Response to Review Comments

We are thankful to the reviewers for their thorough and helpful reviews. We have revised the paper according to their suggestions and comments as follows.

Response to Reviewer #1 Comments

Comment 1:

Sentence problem: Starting from line 42 on page 8 and ending on line 7 on page 9, there is a repeated explanation of the same problem. This problem also exists between line 42 on page 14 and line 30 on page 15. Please check the errors carefully and correct them.

Response:

Thank you for pointing out these problems. We have corrected them accordingly.

Comment 2:

In line 52 on page 10, the author mentions " We consider two ratios for these two parts, i.e. 8:2 and 5:5.". After reading the full text, I still cannot understand why such a division is made, and what is the significance of such a division?
Response:

The dataset is often divided into the training set and the test set with a ratio of 8:2. In order to understand whether different ratios affect the performance, we consider not only the 8:2 ratio, but also the 5:5 ratio and compare the performance. According to the results, the performance of the 8:2 ratio and the 5:5 ratio is similar.

Comment 3:

There are many ways to express similarity. Why do the authors choose the difference between the two to express the similarity between the target inpatient and other inpatients?

Response:

Yes, there are many ways to express similarity. We use two different similarity measures for the last value and the trend of the sequence. For the last value, we simply choose the difference of the symbolic values between the target inpatient and another inpatient; the smaller the difference is, the more similar the two inpatients are.

Comment 4:

At the end of page 15, there is such a sentence "Finally, we classify the inpatient to be positive if there exist one (laboratory test, type) pair which classify the inpatient to be positive.". From the context, I still cannot understand the conclusion. Can the author explain in detail what this conclusion is based on?

Response:

Each inpatient may have different laboratory test items. We use a feature selection method to select most effective (laboratory test, type) pairs and use these pairs to classify the inpatients. Since each most effective (laboratory test, type) pair is carefully selected, we define the inpatient to be positive if there exists one of the most effective (laboratory test, type) pairs, which classifies the inpatient to be positive.
Comment 5:

In the "Comparison of existing works" section on page 16, did the author use the same data set for comparison experiments? If not, this part of the experiment is meaningless, and you need to add corresponding experiments to illustrate the superiority of this model.

Response:

Yes, we implemented the most relevant method of Cheng et al. [25] and used our dataset for the comparison, which shows our method has a better performance.

Comment 6:

There is such a sentence at the beginning of page 8: we call an encounter and an inpatient hereafter interchangeably. Please explain the sentence in detail

Response:

An inpatient may have multiple admissions (encounters). In some admissions the AKI might occur, and in some others it might not. We define the exclusion criteria to identify the inpatients who develop AKI in a specific admission, i.e. it ensures the inpatients did not acquire AKI when they were admitted to the hospital, and developed AKI during the hospitalization. Therefore, we treat each admission individually and use the corresponding laboratory tests to do the prediction. We have added these descriptions in Description of the dataset (page 8) to explain this assumption.

Comment 7:

In this study, authors generated 8 data sets, However, Table 3 only shows the experimental results of a set of balanced data. So, which set of data is this?

Response:

We generated 8 balanced datasets. Table 3 shows the average result of the 8 balanced datasets. We have added an explanation of the experimental results in Analysis of different parameters (page 19) to make it clear.
Response to Reviewer #2 Comments

Comment 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.

Response:

An inpatient may have multiple admissions (encounters). In some admissions the AKI might occur, and in some others it might not. We define the exclusion criteria to identify the inpatients who develop AKI in a specific admission, i.e. it ensures the inpatients did not acquire AKI when they were admitted to the hospital, and developed AKI during the hospitalization. Therefore, we treat each admission individually and use the corresponding laboratory tests to do the prediction. We have added these descriptions in Description of the dataset (page 8) to explain this assumption.

Comment 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".

Response:

Thank you for pointing out this mistake. We have followed your suggestion to correct this mistake in Similarity measures (page 13) and K-nearest neighbor classification (page 15).

Comment 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.
Response:

Thank you for the suggestion. We have added detailed descriptions and an example for DTW in Similarity measures (pages 13-15).

Comment 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.

Response:

A discussion has been provided in Discussions (page 23). Our model outperforms [25] for the following reasons: (1) We consider the last value and the trend of the sequence for each laboratory test while [25] only uses the last recorded value before the AKI event for the prediction. (2) We define the exclusion criteria to identify the inpatients who developed AKI during hospitalization and flexibly set the length of the data collection window up to 5 days. (3) We individually select the most effective (laboratory test, type) pairs to do the prediction for different days of early prediction while [25] just selects a fixed set of laboratory tests.

Comment 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.

Response:

In order to avoid minor differences, we transform the last value and the trend of the sequence into symbolic values. We have added more descriptions in Segmentation and data representation (pages 11-12) to clarify our motivation. The $s_j=(v_{j+1}-v_j)/(t_{j+1}-t_j)$ is the slope between two adjacent laboratory values. If the two adjacent laboratory values have a longer time interval, then the slope will be smaller. Yes, the s_j is sensitive w.r.t. time.
Comment 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?

Response:
The Receiver Operating Characteristic (ROC) curve is generated by setting different probability thresholds. Since the result of our method is not a probability value, we use the precision, recall and F1-score as the performance measures. The recall metric is especially important because it measures the fraction of the correctly predicted inpatients. Although some precision results of our method are lower than the logistic regression, the recall results of our method are constantly better. We also use the comprehensive measure F1-score to evaluate the performance, which shows our method outperforms others. Regarding the superiority of our method, a discussion has been provided in Discussions (page 23).

Comment 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.

Response:
Thank you for pointing it out and we have checked the whole paper and corrected the errors.

Comment 8:
Some related work are missing. For example,


Response:

Thank you for providing these references. We have added their descriptions in Background (page 5) and listed them in References (pages 33-34).