**Reviewer’s report**

**Title:** A classification framework for exploiting sparse multi-variate temporal features with application to adverse drug event detection in medical records

**Version:** 0  **Date:** 26 Jun 2018

**Reviewer:** Mei Liu

**Reviewer's report:**

The authors presented a sophisticated classification framework to exploit the temporality and sparsity of electronic medical records for adverse drug events classification. The proposed framework transforms sparse and multi-variate time series features into a single-valued feature representation that can be used with any classifier. The study also evaluated three different strategies for incorporating sparsity information in the transformation. Following are some major and minor concerns.

**Major Concerns:**

1. Based on the title, the study focus seems to be adverse drug events (ADE) classification using EMR data; however, the background description is geared towards general EMR data analysis with temporal data mining. If ADE is the central research contribution, then overview of the post-marketing ADE surveillance methods and how the proposed framework differs should be presented. If latter is the case, title should be modified accordingly.

2. Overall, how does the proposed framework perform in comparison to other simple methods that consider temporal information? This is essentially a question regarding potential performance and knowledge gain from more complex model over simpler models.

3. Authors had a very interesting way of defining ADE where positive samples are patients who have been diagnosed with a specific ADE code defined by ICD10-SE code, while negative samples are patients who have been given diagnosis code that belongs to the same disease taxonomy that shares the same first three levels of the ICD-10 hierarchy, but is not an ADE. How accurate is the ADE coding in the studied hospital? ADEs are often missed by physicians, which is a part of the reason for under-reporting ADEs. Thus, there may be many unrecognized ADEs, how will that affect classification performance and how to overcome that?
4. The authors used laboratory tests to predict ADEs. Why not use medication information? Is it because laboratory tests are time-dependent variables and medication is not? One would think that medication information is more useful here as it may generate new knowledge on drug-ADE associations, which may have a significant impact on patient safety. By only using laboratory test, you lose the opportunity to discover new drug-ADE associations.

Minor Concerns:
1. How many laboratory tests were included in the classification? What are they and how were they selected?

2. Is the classification made for each encounter with the hospital? Inpatient or outpatient encounters or both?

3. How far back do you collect the data for classification? In other words, do you only use data between the start of an encounter until the time of ADE (positive samples) or other diagnoses (negative samples)?

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