**Reviewer's report**

**Title:** Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone

**Version:** 0  **Date:** 04 Oct 2019

**Reviewer:** Carrie Daymont

**Reviewer's report:**

This manuscript describes a secondary analysis of a published dataset of 299 heart failure patients from Pakistan. Analyses of this data have previously been published using traditional survival analysis models. The authors sought to use machine learning to evaluate the binary outcome of whether patients survived the follow-up period. A strength of the study is including and comparing multiple machine learning methods. However, unfortunately, analyzing the dataset in this manner does not lead to meaningful results. In the original publication on the dataset, subjects die at a relatively steady rate throughout the study period. Therefore, whether a subject dies during or after the study period is in large part a function of how long the study period is, rather than the subject's risk of dying. It is possible to use machine learning methods for survival analysis, which the authors should pursue if they believe that may lead to improvements in understanding these patients compared to what is currently known.

Even disregarding concerns about survival analysis vs binary outcomes, the conclusions of the paper are overstated in multiple places. Specifically, the conclusion statement that doctors can discard clinical information beyond creatinine and ejection fraction is absolutely not supported by the study.

Additional suggestions on the manuscript are below:

1) I would recommend that the authors work with at least one clinician who can provide more guidance regarding the clinical information in the paper, and also provide guidance regarding how such data could be used in practice.

2) The introduction should clearly lay out what is known about prediction of death in patients with heart failure and what the gaps in evidence are.

3) Although more details of the dataset are available elsewhere, important information is left out of this report. For example, creatinine is evaluated as a predictor - are patients with primary kidney disease included in the cohort? It is also crucial to know what type of follow-up was done and, if known, why patients were lost to follow-up?

4) On page 7 Line 14 it says "Regarding machine learning, we focused only on Random Forests" but the paper goes on to describe multiple other machine learning methods.
5) Several continuous variables (hemoglobin, BP) were collapsed into binary variables in the dataset without a clear rationale. Was this what was available in the dataset, or did the authors choose to analyze the data in this way?

**Are the methods appropriate and well described?**
If not, please specify what is required in your comments to the authors.

No

**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.

No

**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.