**Reviewer's report**

**Title:** Rare Disease Knowledge Enrichment through a Data-Driven Approach

**Version:** 1  **Date:** 11 Dec 2018

**Reviewer:** Rafael Berlanga Llavori

**Reviewer's report:**

This revised version of the manuscript covers most of the reviewers' concerns as well as it introduces some improvements in the evaluation of the approach by including a new gold standard. Results are now more convincing, but some aspects can be further improved. The first one is the inclusion of the Jaccard scores in Table 5 so that the reader can get an idea of how far each proposed disease is from the reference disease. The second one is the validation by either experts or the literature of the identified common diseases in the differential diagnoses of this table.

Another issue comes from the sentence in lines 206-208. How is it possible that the degree of certain diseases gets lower after enriching the graph? Are some phenotypes removed when enriching the graph?

Other minor comments are as follows:

Reference [17], what is this? PhD, Technical Report?

The term performance is uncountable in the context it is used

Line 204: combining two datasets -> combining the two datasets

Line 211: diseases with high degrees -> diseases with the highest degrees

Line 213: to compare performances -> to compare the performance

Line 247: it is not necessary to include references for bigrams and trigrams

Line 271: naive Bayes -> Naïve Bayes

Reference [47], what is this? Journal, conference?

Many references are badly formatted, for example in many journals they do not include volume, number etc.
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:

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