Reviewer’s report

Title: nRC: non-coding RNA Classifier based on structural features

Version: 0 Date: 25 May 2017

Reviewer: Davide Chicco

Reviewer's report:

The paper presents a classifier for non-coding RNA, which takes advantage of feature extraction and convolutional neural networks.

The paper is interesting and well written, but some improvements are required.

Authors correctly split the dataset into training set, validation set, and test set, using an untouched data subset for the latter. This way they prevent a very common mistake in the machine learning community: using the same data in the training or optimization, and in the testing phase.

== Scientific comment ==

Authors comment their results on standard statistical scores (accuracy, sensitivity, specificity, precision, F score (F1?)), but they do not consider the most important and relevant statistical score: the Matthews correlation coefficient (MCC)

https://en.wikipedia.org/wiki/Matthews_correlation_coefficient

MCC is the only score able to represent the quality of the prediction in all the four confusion matrix classes (TP, TN, FP, FN), and therefore must be analyzed in every classification projects.

Authors should add the values of the MCC to Table 2, Table 3, Table 4, Figure 8.

Authors should comment the results of the MCC scores in the paper text, too.

== Style comments ==

Two main points:

(1) Authors should avoid the usage of references (citations, tables, figures) in the syntactic construct of the sentences.
For example:

"To do that, as described in Figure 7, we first predicted the ..." WRONG

"To do that, we first predicted the ... (Figure 7)" CORRECT

"In [27], the authors demonstrated that feature selection could be obtained from neural network training." WRONG

"LeCun and colleagues demonstrated that feature selection could be obtained from neural network training [27]." CORRECT

The text of a paper should be correctly readable also in case the reader preferred to omit all the references to citations, tables, figures.

(2) There are too many passive sentences. All the passive sentences must be rephrased into active sentences. For example:

"The first comparison was done to consider..." WRONG

"We did the first comparison to consider..." CORRECT

== Specific comments ==

- The Abstract should explain why non-coding RNAs are important for cancer

- In the Background, authors should add references to other papers which use deep learning in a bioinformatics context ("Deep autoencoder neural networks for Gene Ontology annotation predictions" by Pierre Baldi, "Predicting the sequence specificities of DNA-and RNA-binding proteins by deep learning" by Brendan Frey, and others)

- In the Background, the justification of the usage of CNN is missing. Why did the author decide to use this deep learning algorithm? Please explain

- In the Results, authors should add a table with the formulas of MCC, accuracy, sensitivity, specificity, precision, F score (F1?)

- In the Results, authors should specify if they use the F1 score or another F score

- In the Methods, authors mention they chose 13 ncRNA classes. How were these classes chosen? Why?
- In the Methods, authors define deep learning as having "three or four layers". Authors should clearly specify if the layers must be three or four.

- In the Conclusion, authors mention nothing about the future developments. What is the future of this project? Please explain.

- References should not contain the doi (I suggest to use BibTex package for LaTeX and to download all the BibTex citations from Google Scholar)

== Minor points ==

- Authors should include the URL of Weka when they first introduce it

- In Conclusion, points (1) and (2) should be renamed (i) (ii), to avoid confusion with reference numbers

Level of interest
Please indicate how interesting you found the manuscript:

An article whose findings are important to those with closely related research interests

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

Do you want to get recognition for reviewing this manuscript? Add a record of this review to Publons to track and showcase your reviewing expertise across the world’s journals. Signing up is quick, easy and free!

Yes