**Manuscript ID:** BIB-19-0143

**Type:** Review

Title: ACPred-Fuse: fusing multi-view information improves the prediction of anti-cancer peptides

Authors: Bing Rao, Chen Zhou, Guoying Zhang, Ran Su, and Leyi Wei

Date: 14th May 2019

Dear Editor and Reviewers,

**Re: Decision Letter (BIB-19-0143)**

Thanks to the editors and reviewers for reading our manuscripts and for providing some valuable comments that will allow us to understand our deficiencies and modify them.

We have modified our article based on the reviewer's suggestion, and we have also optimized the content of the article. Our detailed responses (highlighted in blue) are given point by point below.

Finally, we would like to thank the editors and reviewers for their comments, which will help to significantly improve our manuscript.

Yours sincerely,

Best regards,

Leyi Wei

(On behalf of all authors)

**Reviewer #1**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Comments***:*

*They did a comprehensive summary of existing feature descriptors. They better further categorize all the features based on the information they used, like compositional features, physical and chemical properties, and so on.*

**Author’s response:**

We appreciate the reviewer’s comment.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**  
**Comments:**

*In datasets section, why they fuse the proteins into a lone sequence and splice the sequence to select non-ACPs. Moreover, in this section, the second paragraph is very difficult to understand. The authors better to rewrite it to make it easier to go through.*

**Author’s response:**

We appreciate the Reviewer’s comment. We fuse proteins into separate sequences and splicing sequences to select non-ACPs. The selected non-ACPs are combined with the original non-ACPs, so that the selected data set can improve the robustness of the prediction model. The second paragraph of this section has been rewritten, and we have carefully checked and corrected the problem and made it consistent.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**  
**Comments:**

*The formula should be numbered.*

**Author’s response:**

We have corrected this mistake. Moreover, we have carefully checked and corrected the problem and made it consistent.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**  
**Comments:***Overall, more insightful discussion should be given in extra section or in the conclusion. For example, why the feature algorithm can capture more distinct characteristics of ACPs?*

**Author’s response:**

We appreciate the Reviewer’s comment. We have discussed more of our experiments in our conclusions and given the corresponding data in supporting information documents to assist the researchers in further research.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Comments:***They presented the predictor namely MLACP in introduction, but don’t compare with this predictor in the performance comparison in results section. They better to add the predictor to make a more comprehensive comparison for existing predictors.*

**Author’s response:**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Reviewer #2**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Comments:**

*In methods and materials, Authors note that they did simple parameter optimization but did not give the details of which parameters they used for optimization. It’s better to investigate deeper to see the influence of the parameters for the overall predictive performance.*

**Author’s response:**

We appreciate the Reviewer’s comment. We have modified the problem, described our parameter optimization in the paper, and gave detailed parameter data in supporting information. In order to show the influence of parameters on the overall prediction performance, we have analyzed and explained in the paper.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**  
**Comments:***In results and discussions, I suggest to compare the proposed feature representation learning algorithm with the original feature representation learning algorithm, which I notice is also proposed by the same group. Besides, the results should be presented in more details. Especially, the first two sections of the results are not clear enough.*

**Author’s response:**

We appreciate the Reviewer’s comment. We have revised the results and discussion sections, introduced the results in more detail, and revised the parts of the previous two sections that are unclear. We scrutinized and corrected the problem and made it easy to understand.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**  
**Comments:***At last, the English should be polished by a native English speaker. It would definitely help in a better understanding of this study.*

**Author’s response:**

We appreciate the Reviewer’s comment. We have also asked a native English speaker to help us improve the presentation of this manuscript.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Reviewer #3**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Comments:**

*"We used 29 different sequence-based feature...":   
Those 29 features needs some justification, did author try other thresholds? The influence of the quality of “29 different sequence-based feature" on the predictive performance should be evaluated.*

**Author’s response:**

Thank the reviewer for such a good suggestion. We have added a comparison of the prediction performance of 29 different sequence-based features in the paper. Because our feature extraction method will filter the features, so we add more features to it, and finally It is the addition of 29 features.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**  
**Comments:**

*In feature selection, the authors should describe how and why the RF algorithm determines the number of trees. This is a very important precondition of RF algorithms.*

**Author’s response:**

We appreciate the Reviewer’s comment. We have modified the problem, described our parameter optimization in the paper, and gave detailed parameter data in supporting information. In order to show the influence of parameters on the overall prediction performance, we have analyzed and explained in the paper.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**  
**Comments:***"...containing 332 ACPs and 1023 non-ACPs...":   
According to the paper, "the 250 ACPs" and "82 experimentally validated ACPs" are derived from one dataset. The results of "Independent test" need statistical interpretation. Moreover, the authors should describe the information how remove the batch effect of data.*

**Author’s response:**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**  
**Comments:***The classification experiments showed that the average AUC and accuracy obtained by ACPred-Fuse were higher than other methods. However, whether the difference is statistically significant is not reported (e.g., Wilcoxon signed-rank test).*

**Author’s response:**

We appreciate the Reviewer’s comment. To show whether the difference is statistically significant, we calculated and analyzed the p-value.

**Comments:***"Anti-cancer peptides (ACPs) recently have emerged as promising therapeutic agents for cancer treatment."   
There is no biology explanation or application for the classification results.*

**Author’s response:**

We appreciate the Reviewer’s comment. We have fixed the problem of improper description in the paper.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Comments:***"They collected experimentally validated ACPs from three main resources including Chen’s work [12], Tyagi’s work [13] and the largest ACP database CancerPPD [20] as positive samples."   
Authors should mention the normalization technique*

**Author’s response:**

We appreciate the Reviewer’s comment. We have detailed descriptions of the standardization techniques of data sets and added them to the paper.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Comments:***The authors should carefully check the manuscript for typos including supplementary material. Numerous examples of English grammar that could be better.*  
 

**Author’s response:**

We appreciate the Reviewer’s comment. We have also asked a native English speaker to help us improve the presentation of this manuscript.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Comments:***It's confusing that all the lines are dots, such as Fig. 4B and Fig. 4D.*

**Author’s response:**

We appreciate the Reviewer’s comment. We have modified Figures to solve the problem that all lines in the diagram are points.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**End of Response**