L-Università ta' Malta

Centre for Molecular Medicine & Biobanking

University of Malta
Msida MSD 2080, Malta

Tel: +356 2340 2288
cmmb@um.edu.mt

www.um.edu.mt

Dr Jean-Paul Ebejer
Office 320,
Centre for Molecular Medicine and Biobanking,
Biomedical Sciences Building,
University of Malta,
Msida, MSD2080,
Malta.

+356 79555556
jean.p.ebejer@um.edu.mt

20th September 2022.

Prof. Feng Zhu
Associate Editor
Journal of Chemical Information and Modeling

RE: Response to reviewers for manuscript identifier ci-2022-00779z

Dear Prof. F. Zhu,

Reference is made to following manuscript:

Manuscript ID:	**ci-2022-00779z**
Title:	**Few-Shot Learning for Low-Data Drug Discovery**
Author(s):	Vella, Daniel; Ebejer, Jean-Paul

Many thanks for your and the reviewer's helpful comments on our submission. We believe this process has helped improve the overall quality of our manuscript. We addressed the requested changes as commented next (reviewers' comments in shaded areas).

Reviewer #1 Comments

1a. The article is generally well written and clear. However, there is significant duplication, particularly in the method descriptions which are repeated in the Related Work and Methods sections. So much so that some sentences/paragraphs are identical. Perhaps much of the detail could be moved to methods to make the Related Work section more precise?

We have made an extensive effort to read the article thoroughly and remove the duplications referred to by the reviewer. We believe that this streamlining has greatly improved the readability of the manuscript. However, we disagree that we should move some detail from the *Related Work* to *Methodology* as the goal of these sections are different. *Related Work* explains the work which has been done by others in the field of few-shot learning whilst *Methodology* describes our contribution to this area. Evenso, after removing of the duplicate text and the rewording of parts of the manuscript the reviewer should be satisfied we addressed this point successfully.

1b. As another example, Figure 11 is redundant given Figure 10.

We have now removed Figure 11. We have also added a new figure (now Figure 10) to better explain the learning methodology.

2a. The language is sometimes difficult to follow. E.g. p. 25 line 44/45. This talks about training on a query prediction. P. 23 line 17 talks about training on samples from test with remaining data used for testing. It would be better to used prediction in the latter case.

This suggestion is taken on board and we reviewed the whole manuscript to make it more clear. In the specific examples highlighted by the reviewer we have reworded the sentences accordingly.

2b. P. 21 talks about “learning” the embeddings. What is being learned?

A SMILES molecule is transformed to a graph representation (including node attributes) which is used as input to our network architecture. The embedding is essentially a vector of size 128 for neural network consumption. The model learns how to embed a molecule into latent space (i.e. this vector). This is now explained in Graph Convolutional Networks section in Methodology.

2c. Are all “test” molecules included in the initial GCN embedding or just the support set for “training” the few-shot model? i.e. is this step imparting “test” information into the model?

For training and testing we are using distinct, non-overlapping targets (related but not identical). Therefore we are training on a set of tasks (e.g. nuclear receptors from Tox21) and then testing on previously unseen tasks (e.g. stress responses from Tox21). This implies that the same molecule could be encountered in both training and testing, but the class label for the molecule is specifically determined by each individual target (which are not the same for training and testing).

3a. On Tox21, the size of the support set has very little impact on the results. The ROC_AUC varies in the third decimal place and PR-AUC by about 0.01. This doesn't seem logical. Even if one accepts the premise that a network can learn from a single example in each class, increasing the data ten-fold must have a more significant impact. Do the authors have an explanation for this?

Note that we have evaluated our methods on Tox21 to be able to compare to the state-of-the-art work of Altae-Tran et al. (Reference [3]).

Additionally, we have added the following explanation for this in the Tox21 section in Results:

"One observation which can be made is that there is not a significant improvement in performance from one-shot learning to 10-shot learning. This consistency may be attributed to the methodology used for training. Few-shot learning conditions during training must match the ones during testing, but during training, we use a sequence of episodes, in which we match the few-shot conditions during testing in each episode. Having a sequence of episodes means that training sees many molecules, albeit in a few-shot scenario. Hence, the model itself may already be maximally informative due to the number of episodes it is exposed to. Thus, when we get to the testing stage, presenting one-shot or 10-shot support sets to fine-tune the model does not seem to make an impactful difference."

3b. What is the correlation between the training and test endpoints in Tox21, particularly the SR assays used in Test and Train? i. Compound overlap/similarity ii. Activity overlap

This is related to the previous point. There could be an unreported bias in Tox21 which gives enough signal to ML models to discriminate between the positive and negative classes across the different classes. Whilst a review of the Tox21 dataset is outside the scope of this study we have added the following comment:

"Another possible explanation for the insignificant impact of the support sets size on the performance of the few-shot learning models is that there is an inherent bias across targets in the dataset which the methods are able to pick upon. If this is the case, it would warrant future research on the composition of the Tox21 dataset when used to evaluate Machine Learning models."

3c. The authors explain away the MUV results by the fact that the compounds are deliberately chosen to be dissimilar. But isn't this the general use case where Hit ID has discovered new lead molecules that are distinct from anything else. What are the limits of applicability of these methods in a real-world setting?

The reviewer is right in the sense that these models are useful if they find novel chemical moieties rather than a marginal, incremental change in the molecule's structure. However, like all multidimensional optimization problems, this is a fine balance to strike. Depart too far from an active molecule and you loose activity (or decrease potency). Change an active molecule in a small manner (e.g. add a methyl) and you gain little new information and knowledge on what gives a

molecule its activity. The usefulness of these methods can only be shown in a prospective study. We firmly believe and agree what Altae-Tran et al. claimed in their last sentence of their article:

"The use of one-shot learning in chemistry can only be validated experimentally, but we hope that our results will provide the impetus for such work."

3d. Would an alternative explanation be that the methods do not work and that there is some inherent bias in the Tox21 data as for DUD-E?

We think we have addressed this in points 3a and b.

Reviewer #2 Comments

The related work section was suggested to move to the supporting information.

As discussed earlier, we think that the material in the *Related Work* section is useful to the reader who may not be familiar with these less-popular Machine Learning models. Without them the whole paper is harder to follow. Reviewer #1 suggested we move *Related Work* (presumably the same parts) to *Methodology*. We also are against this suggestion as this is work which has been described in *Related work* is existing work by other researchers, whereas *Methodology* highlights work we have undertaken. In the first iteration of the article there was significant repetition between related work and methodology. This has now been addressed and we've improved the manuscript's readability considerably.

It is better to redrawing the Figures 8 and 9 using Figure 7 as a reference.

Figure 9 was a sub-part of Figure 8, and we removed this due to this redundancy. Figure 8 is now explained in the caption using Figure 7 as a reference, as suggested by the reviewer.

Some discussion sentences in the conclusion section should be move to result and discussion section.

After having reread and edited the manuscript thoroughly, the article now reads better and we present **no** new ideas in the *Conclusion* section.

Editor's Comments

X) NEED Table of Contents (TOC) Graphic. This graphic should capture the reader's attention and provide a quick visual impression of your topic. Guidelines for the preparation of effective TOC graphics are available at: http://pubs.acs.org/paragonplus/submission/toc_abstract_graphics_guidelines.pdf. The TOC graphic should appear on the last page of your manuscript and should not have a caption.

The TOC graphic has now been moved to the last page and the caption removed.

Funding Sources: Authors are required to report ALL funding sources and grant/award numbers relevant to this manuscript. Enter all sources of funding for ALL authors relevant to this manuscript in BOTH the Open Funder Registry tool in ACS Paragon Plus and in the manuscript to meet this requirement.

The funding source may be found in the Acknowledgement Section (pg. 37), as well as in the ACS Paragon Plus Open Funder Registry tool during manuscript submission.

ORCID: Authors submitting manuscript revisions are required to provide their own validated ORCID iDs before completing the submission, if an ORCID iD is not already associated with their ACS Paragon Plus user profiles.

The ORCID iDs for the two authors Daniel Vella (0000-0001-9404-6576) and Jean-Paul Ebejer (0000-0003-0888-2637) were provided during manuscript submission as well as associated with their ACS Paragon Plus profiles.

The reviewers indicate that the quality of the written English could be improved.

We have thoroughly read the manuscript and fixed remaining typos and grammar, as well as removed unnecessary duplication.

JCIM is now accepting images to appear on the front cover of the journal. If you have a visually arresting and scientifically interesting image, please upload it and a brief (80 word) description of the image as part of your revised submission.

We are submitting a front cover image (ci-2022-00779z-coverpage-art.png) together with our revised submission. The 80-word blurb is as follows:

Supervised Machine learning is typically data-hungry, which is in stark contrast to the limited data for new disease targets. We explore few-shot machine learning which aims to "learn how to learn" from just a few examples. Our proposed architecture makes use of embeddings created through graph convolutional networks, achieving better results than the state-of-the-art on Tox21 data. We classified molecules as active or otherwise, within a previously unseen experimental assay using only 1-10 molecules and a model trained on related assays.

JCIM has a Twitter feed (@JCIM_ACS; https://twitter.com/jcim_acs) covering the latest news and highlights in the field. If you would like us to feature your manuscript on this platform, please share a Tweet of no more than 140 characters.

We suggest the following tweet:

Few-shot learning using GCNs and Prototypical Networks tested on Tox21, MUV and DUD-E datasets. @uniofmalta @dr_jpe @danvlla

Once again, we thank you and the reviewers for their helpful and constructive feedback: it has strengthened the manuscript, and we look forward to its publication in JCIM.

If you require further details or clarifications, please do not hesitate to get in touch.

Sincerely yours,

Dr Jean-Paul Ebejer
Principal Investigator
Centre for Molecular Medicine and Biobanking, University of Malta