



Open Access This file is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

Editorial Note: This manuscript has been previously reviewed at another journal that is not operating a transparent peer review scheme. This document only contains reviewer comments and rebuttal letters for versions considered at Nature Communications.

REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author):

I would like to thank Calderaro and colleagues for revising their manuscript and addressing mine and other reviewers' comments.

While I still see the exposure of some test data during the training phase as a potential issue, which the scientific community should be carefully considering in the future, given the success of self-supervised models, I do agree with the authors in that this might affect an intermediary result and not the results on the external validation set.

Response to Reviewer #1

Reviewer #1 (Remarks to the Author):

I would like to thank Calderaro and colleagues for revising their manuscript and addressing mine and other reviewers' comments.

While I still see the exposure of some test data during the training phase as a potential issue, which the scientific community should be carefully considering in the future, given the success of self-supervised models, I do agree with the authors in that this might affect an intermediary result and not the results on the external validation set.

Response: Thank you very much. No action required.