Dear Editors,

In a recent paper in *NPJ Breast Cancer*, Hart et al.[1] built a highly accurate model for predicting damaging missense mutations in *BRCA1* and *BRCA2* genes. As there are thousands of missense variants of unknown significance (VUSs), the clinical meanings of which are not yet known, the computational predictions of missense VUSs are highly demanded to interpret VUSs identified by genetic testing and to support clinical decision-making for their carriers. The authors trained the BRCA-ML machine learning algorithm based on results from functional assays and 25 *in silico* prediction models. For *BRCA1*, the sensitivity and specificity of functional annotation were as high as 89.5% and 91.5%, respectively. The precision-recall curve produced by BRCA-ML was remarkably improved compared with other prediction algorithms. This great prediction performance of BRCA-ML might provide reliable evidence for clinical usage. However, we would like to point out 2 critical concerns.

The first concern is that the authors evaluated the probability of pathogenicity of substitutions from amino acid 1 to 1500 in BRCA1, whereas BRCA1 encodes a protein of 1863 amino acids. The authors failed to assess the substitutions from amino acid 1501 to 1863 including the BRCT domain (amino acid 1642–1855). The BRCT domain is essential for BRCA1 function and more than one-third of missense variants in the BRCT domain shows significant impact on function[2]. We found that mutated amino acid positions of BRCA1 variants were incorrectly assigned in their raw data (Supplementary Data Set 1). Accordingly, their missense prediction scores for BRCA1 are incorrect (see the attached figure).

The second concern is double dipping of the data. Double dipping is a methodological problem that both building and evaluating a prediction model on the same dataset[3]. This error leads to overfitting of the model and inappropriately high predictive power. The authors trained the BRCA-ML model with 80% of variants, and then predicted all the variants including training dataset. That is double dipping. To avoid double dipping, we re-analyzed the same model with the holdout method. As a result, the sensitivity and specificity of functional annotation for BRCA1 variants were 70.0% and 91.5%, respectively.

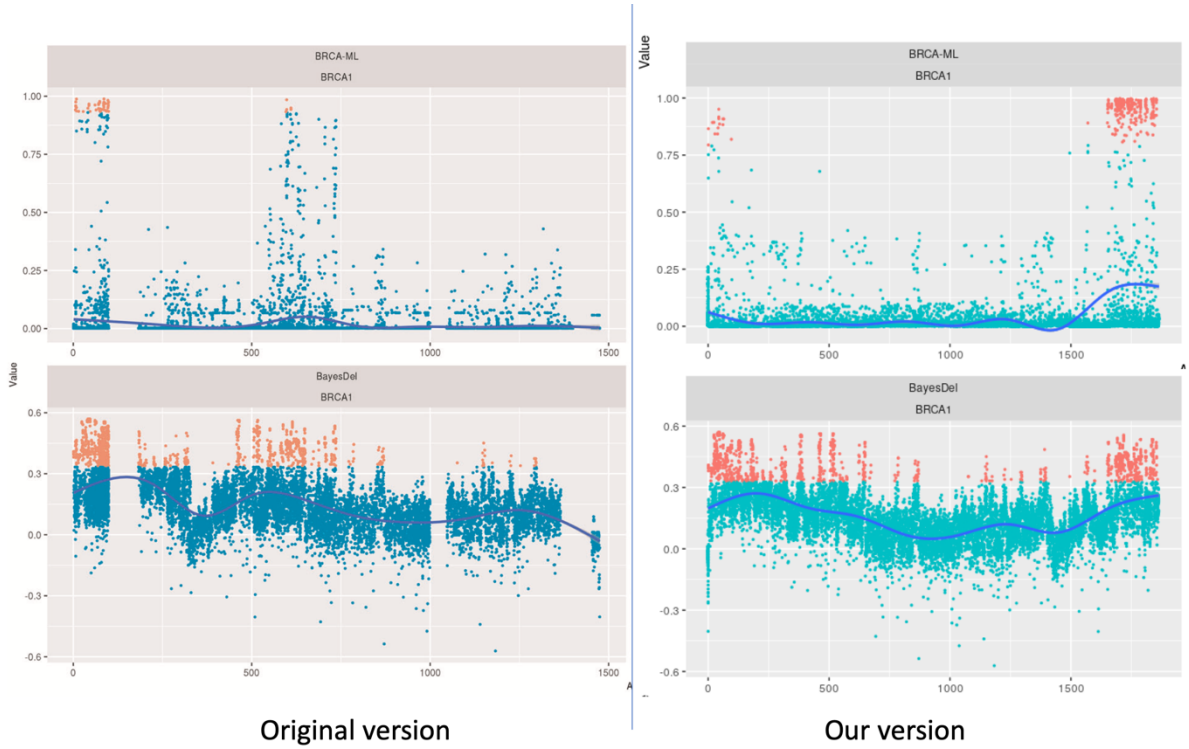
We think that the BRCA-ML model is a powerful prediction tool, but there is still room for refinement.

The data and code are openly available via our GitHub website, at https://github.com/MANO-B/BRCA-ML.

References

1. Hart, S.N. *et al.* Prediction of the functional impact of missense variants in BRCA1 and BRCA2 with BRCA-ML. *NPJ Breast Cancer* **6,** 13 (2020).
2. Fernandes, V. C. *et al.* Impact of amino acid substitutions at secondary structures in the BRCT domains of the tumor suppressor BRCA1: implications for clinical annotation. *J. Biol. Chem*. (2019).
3. Ball, T.M. *et al.* Double dipping in machine learning: problems and solutions. *Biol Psychiatry Cogn Neurosci Neuroimaging* **5,** 261-263 (2020).

Attached figure: Fig. 2 in the original paper.



Amino acid positions are corrected in the figure of our version.