

30-Jan-2025

Dear Editor-in-Chief:

We are submitting our manuscript entitled “Toward reliable false discovery rate control in classification problems under distribution shift” for consideration for publication in *Scientific Reports*.

**Affiliations:**

1. The corresponding author is: Prof. Attila Kertesz-Farkas, [akerteszfarkas@hse.ru](mailto:akerteszfarkas@hse.ru),
  2. Co-author, first author: Andrey Borevsky.
- We are both part of the same laboratory: Laboratory on AI for Computational Biology, Faculty of Computer Science, HSE University, Moscow Russian Federation

We note that our institution is not under sanctions and we are not involved in any research related to the Russian military, etc.

**Statement of why this manuscript is appropriate for Scientific Reports:**

We present our study on controlling false discovery rates (FDR) under data distribution shift in classification problems in the biomedical domain. In this study, we draw attention to a very important but often neglected problem; namely, the performance of deep-learning based classifiers is assumed to be the same in application. However, it often can be much worse than thought due to data distribution shift (a.k.a. covariate shift), data label distribution shift, batch effects caused by some confounding factors, and overfitting. This performance drop of a machine learning system often remains undetected, resulting in risk for people in health care and clinical applications.

In this manuscript, we present a simple, fast and robust method to adjust the test prediction scores so that the latent test null distribution will be more similar to the train null distribution. This will result in more accurate FDR control.

Our method operates in the 1-dimensional prediction score space. Therefore, it is not hindered by high-dimensional space, data sparsity (i.e. curse-of-dimensionality), and correlated features. In addition, our method does not rely on data clustering or graphical causal models.

We tested our method with four public benchmark datasets commonly used in biomedical image analysis: PatchCamelyon (PCam), Chest x-ray (CheXpert), TissueNet, and Breast Cancer Semantic Segmentation (BCSS). With our method, we obtained much more accurate FDR control under data distribution shifts in the test data compared to a recent algorithm, called Learn-then-test, developed by Michael Jordan’s lab from Stanford University.

This manuscript has been approved by both authors. This manuscript has never been submitted to any other journals, and we have not had any prior discussion with members of the *Scientific Reports* Editorial Board about this work. The manuscript will be properly formatted for the journal if it is needed for the reviewing stage or once it is accepted. We did not use AI chat bots in manuscript preparation or in this study in general.

**Suggested reviewers:**

1. Nikita Moshkov, PhD, Institute of Biochemistry, Biological Research Center. Szeged, Hungary, Deep Learning specialist in biological image analysis, cell segmentation, etc. [nikita.moshkov@brc.hu](mailto:nikita.moshkov@brc.hu)
2. Peter Horvath, PI at AI for Health, Helmholtz Munich, Germany; PI at Biological Research Center, Szeged, Hungary [horvath.peter@brc.hu](mailto:horvath.peter@brc.hu)
3. Pavel Sulimov, PhD, ZHAW School of Engineering, Switzerland, [pavel.sulimov@zhaw.ch](mailto:pavel.sulimov@zhaw.ch)
4. Yang Lu, PhD, Assistant Professor, University of Waterloo, Canada, [yanglu@uwaterloo.ca](mailto:yanglu@uwaterloo.ca)

We hope that this manuscript provides a useful and interesting study for the readers of *Scientific Reports*.

Best regards,

Attila Kertesz-Farkas,  
Professor

Head of the Laboratory on AI for Computational Biology, Faculty of Computer Science  
HSE University, Moscow, Russia