

Uncertainty Quantification with Conformal Prediction as a Tool for Fairness in Genomic Medicine

Leandra Bräuninger^{1,2,3}, Tapabrata Chakraborti², Christopher R. S. Banerji², Ariane Marandon-Carlhian², and Brieuc Lehmann^{1,2}

¹*Department of Statistical Science, University College London, London, UK*

²*The Alan Turing Institute, The British Library, London, UK*

³*Please reach out to: ucaklbr@ucl.ac.uk*

June 27, 2024

Abstract.

We use conformal prediction to provide individualised uncertainties for personalised medicine tools. Using a publicly available TCGA breast cancer dataset, we compute conformal prediction sets for the Oncotype DX recurrence score and compare both performance and uncertainty across ancestral subpopulations. The analysis shows small differences in uncertainty between subpopulations with large differences in accuracy. Uncertainty is a relevant metric, but better social population stratification categories need to be established. This model-agnostic method allows for post-hoc evaluation and is thus readily implementable as a fairness metric on pre-established methods.

Keywords: conformal prediction; genomic medicine; algorithmic fairness.

1 Introduction

This document is a brief appendix to the poster on this project with the main aim of pointing at some of the methodology. To read the report and find out more, please reach out via email to Leandra Bräuninger via the address specified above.

2 Method

1. calculate approx. ODX score from data
2. create 100 samples: split data into train, calib and test
3. create bins: true survival from data, ODX-predicted from survival model cutpoints
4. compute the conformal scores ($\alpha = 0.05$) and prediction intervals based on them
5. check accuracy of bin allocation
6. compare accuracy and interval length (uncertainty) for each subgroup and globally

3 Limitations & Future Work

1. ODX is proprietary, our approximation may be (too) inaccurate
2. Bins are rather arbitrary, a sensitivity analysis on the choice of cutpoints should inform where exactly to place them.
3. Currently, conclusions are more informative on accuracy, class conformal can provide further comparative equity insights.
4. Similarly, explore different conformal set-ups.

Acknowledgments

This work was supported by the Engineering and Physical Sciences Research Council [grants EP/T517793/1 and EP/W524335/1]. Furthermore, this work was conducted in close collaboration, but without financial support of the Turing-Roche partnership.