Mathematical Methods and Models in Biosciences
June 18-23, 2023, Pomorie, Bulgaria
https://biomath.math.bas.bg/biomath/index.php/bmcs

Predicting hereditary *BRCA1/2* mutations using publicly available data

Ekaterina Auer[1], Lorenz Gillner[1], Wolfram Luther[2]

[1]Department of Electrical Engineering and Computer Science,
University of Applied Sciences Wismar, Germany
ekaterina.auer@hs-wismar.de
lorenz.gillner@hs-wismar.de

[2]Department of Computer Science and Applied Cognitive Science,
University of Duisburg-Essen, Germany
luther@inf.uni-due.de

Over the past decade, biomedical sciences have become more open towards scrutiny by the general public. Modern journals publish articles open access while increasingly many research-related institutions demand disclosure of the related data sets (e.g., via websites) for reproducibility, information exchange, and validation. Still, a certain amount of sensitive data has to be kept under restricted access because of data protection guidelines, for example, data on genetic samples containing germline variants (i.e., hereditary mutations). A further difficulty is that data on mutations can be also quite opaque wrt. its provenance. However, access not only to such data but also to the corresponding metadata is often crucial in risk assessment for inherited medical conditions.

In this contribution, we focus on the hereditary breast and ovarian cancer syndrome (HBOC) and present a novel approach to assess the combined personal/familial risk of carrying a pathogenic *BRCA1/2* variant. Using a combination of the Dempster-Shafer theory [1] and interval analysis [2], we improve a model from [3] towards taking into account uncertainty about persons' ages. Because public germline samples are currently unavailable, we use combined findings from various open access publications on HBOC-related mutation probabilities as our factual basis.

While being computationally simple, our model yields results comparable to those of established, more complex models (relying on undisclosed data). Additionally, we give an outlook on the way to automate the predominantly manual process of information extraction from relevant publications using context awareness and pattern recognition. The so obtained data set could be appropriately released without violating privacy regulations.

Keywords: open access, HBOC, Dempster-Shafer theory, risk assessment

References

[1] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, 1976.

[2] W. Tucker, *Validiated Numerics*, Princeton University Press, 2011.

[3] E. Auer, W. Luther, Uncertainty Handling in Genetic Risk Assessment and Counseling, *Journal of Universal Computer Science*, 27(12): 1347-1370, 2021.