

Towards Explainable and Fair Uncertainty Modeling in Healthcare via Set-Based Methods

Ekaterina Auer¹, Lorenz Gillner¹ and Wolfram Luther²

¹ University of Applied Sciences Wismar
Philipp-Müller-Straße 14, 23966 Wismar, Germany
{ekaterina.auer,lorenz.gillner}@hs-wismar.de

² University of Duisburg-Essen
Forsthausweg 2, 47057 Duisburg, Germany
wolfram.luther@uni-due.de

Keywords: Interval analysis, Dempster-Shafer theory, *BRCA1/2* mutation probability, Explainability, Fairness

Introduction

Explainability and fairness are widely recognized as key requirements for trustworthy healthcare prediction systems and AI applications more broadly. However, neither concept has a universally accepted definition [4, 8], and their practical implementation remains challenging. In [6], we propose a meta-level perspective on fairness (meta-fairness), in which explainability can be regarded as one of the quality criteria supporting fair decision-making. Alternatively, some conceptual frameworks, such as structured explanation taxonomies [10], include fairness as one of the components of explainability in AI systems.

In medical risk prediction, explainability and fairness are deeply intertwined with uncertainty. In particular, the way uncertainty is modeled and communicated not only affects the reliability of predictions but also their interpretability and the perceived fairness of decisions. While predictive models typically capture aleatory uncertainty associated with inherent variability, epistemic uncertainty arising from

incomplete knowledge of the data, model parameters, or model structure is often overlooked.

In this talk, we investigate how richer representations of epistemic uncertainty can improve the explainability of medical risk predictions and whether this increased transparency can support fairer decision-making. As a case study, we focus on estimating *BRCA1/2* mutation probabilities in hereditary breast and ovarian cancer syndrome. We compare a framework based on Dempster–Shafer theory [9] (DST) with logistic regression and a DST-based interpretation of logistic regression using a synthetic dataset. We extend this DST model using interval arithmetic [7], Yager functions [2], and thick intervals [1], emphasizing explicit epistemic uncertainty representation to enhance algorithmic transparency and support fairness at a meta-level. A light user study further explores how different uncertainty representations are perceived by non-expert users, providing insights into explainability.

Uncertainty versus Explainability and Fairness

Meaningful explanations should communicate not only the rationale behind a prediction but also its uncertainty. Without uncertainty information, predictions may create a false sense of precision, especially in settings with scarce or noisy data. Explicit epistemic uncertainty representations improve transparency by revealing model limitations and indicating when predictions should be interpreted with caution. Recent work has also argued that uncertainty itself can serve as an explainability mechanism [11].

The relationship between uncertainty and fairness is particularly important in healthcare, where predictive models support diagnosis, risk assessment, and treatment decisions. Due to differences in data availability, data quality, healthcare access, and historical exclusion, epistemic uncertainty is often unevenly distributed across patient populations. As a result, underrepresented groups may receive predictions based on weaker evidence than better-represented groups. Explicit uncertainty representation can help identify such limitations, revealing potential inequities in the underlying evidence base and supporting

fairer decision-making [3].

Although there is a growing body of work on the interplay between uncertainty and explainability or fairness [5], interval- and set-based approaches for epistemic uncertainty remain relatively underexplored in this context. Likewise, frameworks such as DST, which can represent both aleatory and epistemic uncertainty, have received limited attention from the perspective of their usefulness for explainability and fairness.

Summary

In this talk, we provide an overview of our work on representing epistemic uncertainty in hereditary breast and ovarian cancer syndrome risk prediction using interval analysis and DST. A particular challenge in this domain is that germline mutation status constitutes sensitive genetic information and is therefore typically unavailable in publicly accessible datasets. Consequently, we employ synthetic data, as our primary objective is to investigate uncertainty representation, algorithmic explainability, and fairness-related aspects rather than predictive performance itself. In particular, we examine how uncertainty modeling relates to explainability from an algorithmic perspective and to fairness through the lens of meta-fairness. Implementations of the proposed models, together with the synthetic datasets used in our experiments, are publicly available at

<https://github.com/lorenzgillner/BRCA-DST>.

References

- [1] Reda Boukezzoula, Luc Jaulin, and Laurent Foulloy. Thick gradual intervals: An alternative interpretation of type-2 fuzzy intervals and its potential use in type-2 fuzzy computations. *Engineering Applications of Artificial Intelligence*, 85:691–712, 2019.

- [2] E. S. Lee and Q. Zhu. An interval Dempster–Shafer approach. *Computers & Mathematics with Applications*, 24(7):89–95, 1992.
- [3] Michelle Seng Ah Lee, Kirtan Padh, David Watson, Niki Kilbertus, and Jatinder Singh. Fairness under uncertainty in sequential decisions. *arXiv preprint arXiv:2604.21711*, 2026.
- [4] Zachary C. Lipton. The mythos of model interpretability. *Queue*, 16(3):31–57, 2018.
- [5] Helena Löfström, Tuwe Löfström, Anders Hjort, and Fatima Rabbia Yapicioglu. Concerning uncertainty – a systematic survey of uncertainty-aware xai. *arXiv preprint arXiv:2603.26838*, 2026.
- [6] W. Luther and E. Auer. Towards fair medical risk prediction software. *Future Internet*, 17(11):491, 2025.
- [7] R. E. Moore, R. B. Kearfott, and M. J. Cloud. *Introduction to Interval Analysis*. Society for Industrial and Applied Mathematics, Philadelphia, 2009.
- [8] Otto Sahlgren. What’s impossible about algorithmic fairness? *AI and Ethics*, 2024.
- [9] Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
- [10] The Alan Turing Institute. Transparency and explainability. <https://alan-turing-institute.github.io/turing-commons/skills-tracks/aeg/chapter5/transparency/>, 2025-03-08. AI Ethics and Governance Skills Track.
- [11] Arthur Thuy and Dries F. Benoit. Explainability through uncertainty: Trustworthy decision-making with neural networks. *European Journal of Operational Research*, 317(2):330–340, 2024.