Dear Editor,

Thank you for the opportunity to submit my paper “A comprehensive review of classifier probability calibration metrics” to the Data Science journal.

There has been an intense interest in the concept of probability calibration and associated metrics over the last seven years. Understanding probability calibration is important for a number of reasons, including the assurance of machine learning models in safety or business critical contexts, for combining the outputs of multiple systems, and for users to gain trust in models.

This paper gives a review of 82 probability calibration metrics for classifier and object detection models, organising them according to a number of different categorisations to help understand the relationships between them. This paper is believed to be the most comprehensive survey of probability calibration metrics to-date, providing descriptions of more metrics than other surveys combined. The most recent survey of similar depth on this topic was published in the Machine Learning journal (Silva Filho et al., 2023) and was first submitted to the journal more than three years ago. The present paper includes several metrics not included in that survey, as well as many new metrics that have been published since then. Where comparisons between metrics have been given in the literature, these are summarised in the present paper to help researchers decide which metrics are best for a given situation. The paper also provides practical equations for computing each metric where available, enabling ease of understanding, implementation, and comparison by future researchers.

In line with the Sage Preprint Sharing policy, I would like to let you know that I have submitted the paper to the arXiv preprint service under the standard arXiv non-exclusive license to distribute, at https://doi.org/10.48550/arXiv.2504.18278.

Permission from Tobias Kuhn (joint Editor-in-Chief) has kindly been given for the normal word limit for a survey paper to be exceeded. Comments from his quick editor check have been addressed in the version submitted for formal review.

I believe this paper would be a good addition to the journal and I look forward to hearing your decision.

Yours sincerely,

Dr Richard Lane CEng MIET FIMA

Principal Data Scientist, QinetiQ