Online Resource 4: Calibration reliability diagrams

Article: When Does Supervision Beat Reconstruction? Magnification Shift and the Robustness of AI Quality Control in Digital Pathology

Journal: Journal of Medical Systems

Authors: Konstantinos Moutselos, Ilias Maglogiannis

Corresponding author: Konstantinos Moutselos, Department of Digital Systems, University of Piraeus, Piraeus, Greece; kmouts@unipi.gr

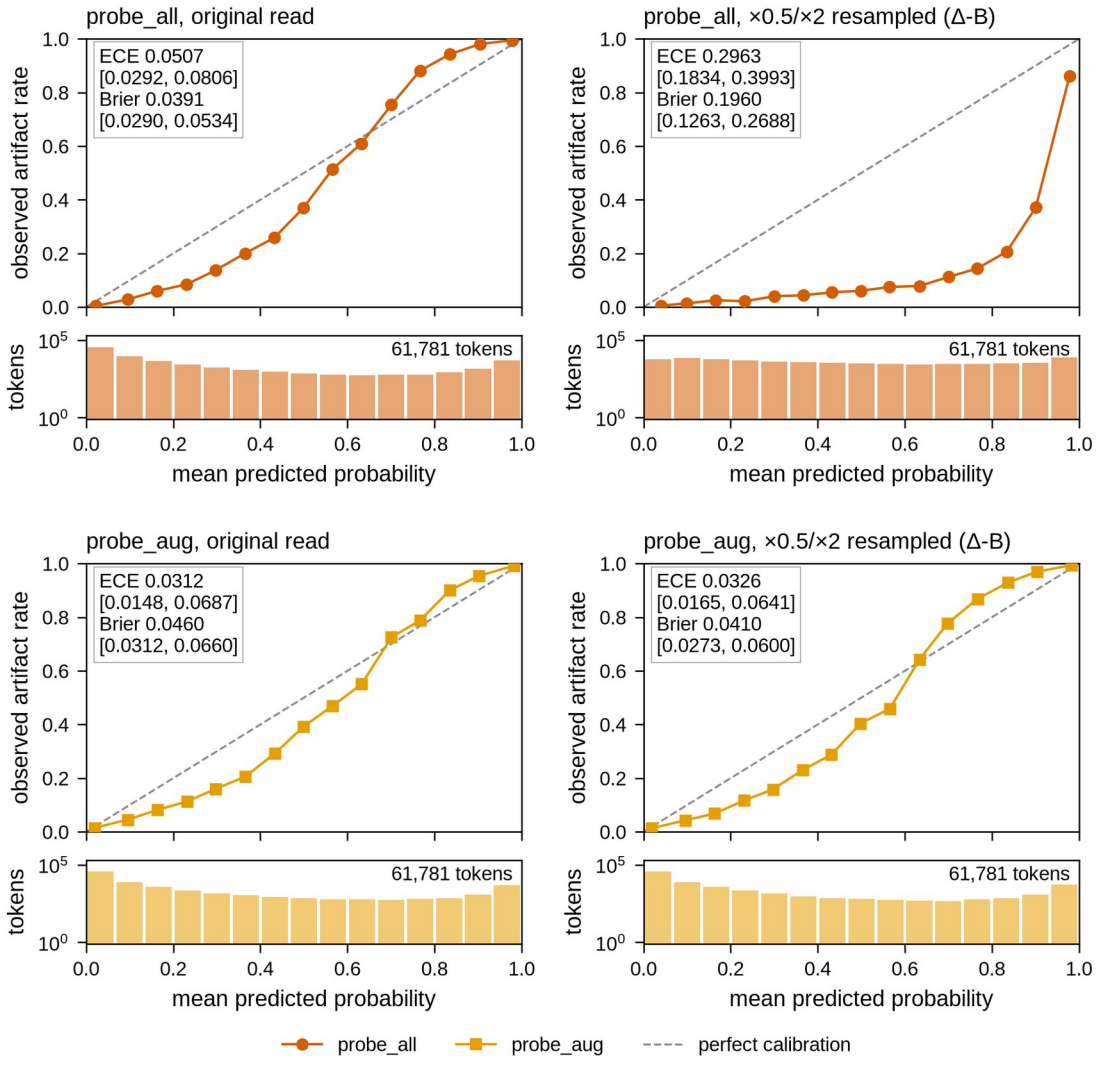


Fig. S1 Reliability diagrams of the UNI2-h probe (circles) and the scale-augmented probe (squares) on the development benchmark, at the original read (left) and after ×0.5/×2 resampling (right). Each point is one of 15 equal-width probability bins: observed artifact rate against mean predicted probability; the dashed line is perfect calibration. The strip below each panel shows the number of tokens per bin (logarithmic scale), and ECE and Brier score with 95% intervals are given in each panel. After

resampling, the probe's curve falls far below the diagonal and its tokens move out of the lowest bins; the scale-augmented probe is unchanged