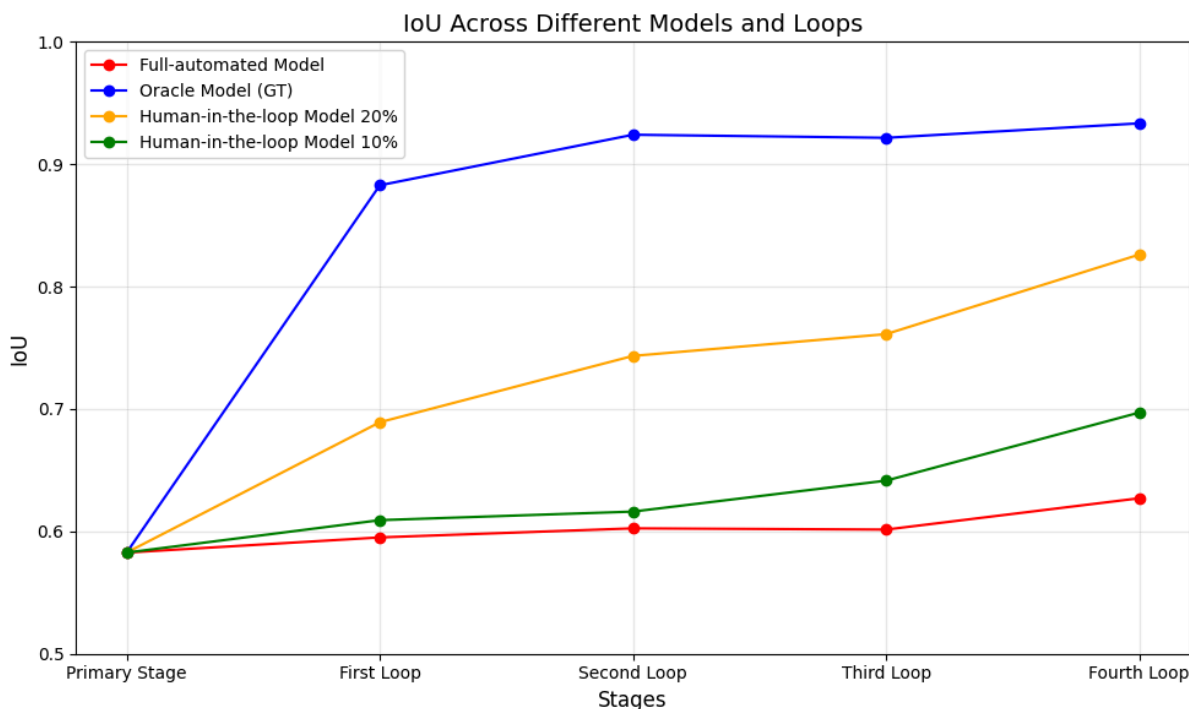# HUMAN-IN-THE-LOOP FOR PATHOLOGY DATASET EXPERIMENT REPORT

Team PathOs

[Figure 1]

This experiment was conducted to evaluate the efficacy of a human-in-the-loop framework. In Figure 1, the horizontal axis denotes the number of completed loops, while the vertical axis indicates the Intersection over Union (IoU) score. Four experimental conditions were compared: (1) the Oracle Model, in which all samples are paired with ground-truth labels at every loop; (2) the Full-automated Model, in which, apart from the Primary Stage, all samples at each loop are labeled by the model's segmentation output; and (3–4) the Human-in-the-loop Models with 10% and 20% refinement, respectively, in which, again excluding the Primary Stage, 10% (or 20%) of the training samples at each loop are randomly replaced with true labels before retraining. We used the Warwick-Q dataset [1] for training and evaluation. DeepLabV3+-resnet101 utilized for the experiment.

The results demonstrate that the human-in-the-loop approach yields steadily increasing performance at each loop, ultimately converging to a level comparable to that of the Oracle Model. This finding substantiates that, with only a small amount of refinement provided directly by a pathologist, one can train a model that achieves superior performance even when limited data are available.

References

[1] Sirinukunwattana, Korsuk, Josien P. W. Pluim, Hao Chen, Xiaojuan Qi, Pheng-Ann Heng, Yun Bo Guo, Li Yang Wang, Bogdan J. Matuszewski, Elia Bruni, Urko Sanchez, Anton Böhm, Olaf Ronneberger, Bassem Ben Cheikh, Daniel Racoceanu, Philipp Kainz, Michael Pfeiffer, Martin Urschler, David R. J. Snead, and Nasir M. Rajpoot. "Gland Segmentation in Colon Histology Images: The GlaS Challenge Contest." *Medical Image Analysis* 35 (January 2017): 489–502.