Self-supervised learning and human-in-the-loop methods for detection and differentiation of circulating tumor cells in liquid biopsy data

Abstract Submitter: Hümeyra Husseini-Wüsthoff, Germany*

Co-Authors: Sabine Riethdorf, Andreas Schneeweiss, Andreas Trumpp, Klaus Pantel, Harriet Wikman, Maximilian Nielsen, René Werner

*Institute for Applied Medical Informatics, University Medical Center Hamburg-Eppendorf

Abstract

Background: Despite the utility of the CellSearch® system (CS), challenges persist, such as the labor-intensive manual examination of extensive image galleries, especially from patients with metastatic disease. To tackle these issues, recent efforts, including machine-learning techniques, have focused on achieving greater, if not complete, automation in circulating tumor cell (CTC) detection.

Objective: Since a model cannot always be fully confident in its predictions, we focus on efficiently selecting training samples, particularly from areas of uncertainty, to enable automation even with a limited number of samples. Through the targeted integration of human expertise to label selected samples, this approach aims to improve classifier accuracy while minimizing time consumption.

Methods: Datasets from 90 metastatic breast cancer patients were collected via CS. StarDist segmentation yielded 1,321,951 (three-channel: DAPI, CK, CD45) single-cell images. Self-supervised training on 60 patients with 999,285 unlabeled images provided a learned representation to a machine-learning classifier for binary classification on 30 remaining patients. Training the classifier on 10 patients with a small GT, consisting of 1,509 CTCs and 1,129 non-CTCs, classifier uncertainties were identified with clustering for 10 test patients. With a human-in-the-loop strategy, targeted sampling and expert labeling of images from latent space clusters with higher uncertainties was carried out. After four loops, the final model was applied to a hold-out set of 10 patients, and a comparison was made between expert-labeled CTCs proposed by the model and the CS events.

Results: Cluster analysis revealed distinct clusters with varying classification performances, each representing images with similar cell characteristics. The iterative, cluster-specific human-in-the-loop strategy improved classification accuracy. The final model identified a similar number of actual CTCs as CS but with higher positive predictive value, reducing false positives.

Conclusion: By combining automation with human expertise, this approach addresses limitations of the CS system, allows for an expert-guided iterative optimization of CTC detection and requires fewer annotation time of the expert – which could be also beneficial in clinical application perspective when classifier uncertainty is present. Future work will expand this strategy's applicability to data with lower CTC counts and to other tumor entities.

Do you have any conflicts of interest?

No, I do not have a conflict of interest.