Weakly Supervised Segmentation of Malignant Epithelium in Digital Breast Pathology

Background

Tumor segmentation in digital pathology plays a crucial role in breast cancer diagnosis and prognosis [1], [2]. Precise delineation of malignant epithelial regions in hematoxylin and eosin (H&E)-stained or immunohistochemistry (IHC)-stained slides enables downstream analyses, such as cellularity estimation and biomarker quantification for diagnostic pathological examination, therapeutic response assessment, treatment selection, and survival prediction [3]–[8]. Deep learning-based segmentation approaches overcome the inefficiency of manual assessment, enabling high-throughput analysis of histopathological datasets. However, current approaches predominantly rely on supervised learning, which requires labor-intensive pixel-level manual annotations that are impractical at scale [9]–[11]. Weakly supervised learning has emerged as a promising alternative, leveraging coarse-grained labels to reduce annotation burdens. Yet, existing solutions are constrained by the restriction to whole-slide image (WSI)-level classification [12], reliance on partial cell-level annotations [13], and unproven generalizability across diverse breast cancer cohorts and staining protocols [14], [15]. These challenges underscore the need for a weakly supervised segmentation method that is trained using only image-level annotations while achieving pixel-level precision in malignant epithelium delineation and generalizing to heterogeneous breast cancer datasets.

Specific tasks

- Literature study to get familiar with the different topics.
- Perform data preprocessing, including extracting patches from whole slide images, applying color deconvolution to separate the Hematoxylin stain from H&E and IHC images using ImageJ, and applying data augmentation techniques such as flipping, rotation, and adjusting brightness and contrast to address class imbalance.
- Implement prevalent convolutional neural network (CNN) and Transformer models, as described in Table 4 and Table 5 of Ref. [16], and conduct training and inference of these models using Python, preferably with PyTorch.
- Validate the segmentation results predicted by these models across various breast cancer datasets, including H&E and IHC images, by comparing them to

the ground truth segmentation mask (e.g., on the MHCI and BCSS datasets) or the ground truth cellularity (e.g., on the BreastPathQ and Post-NAT-BRCA datasets).

- [Optional] Develop multiple instance learning (MIL) techniques to improve segmentation performance across diverse breast cancer datasets, aiming to achieve accuracy comparable to that of supervised semantic segmentation methods.

Preliminary results

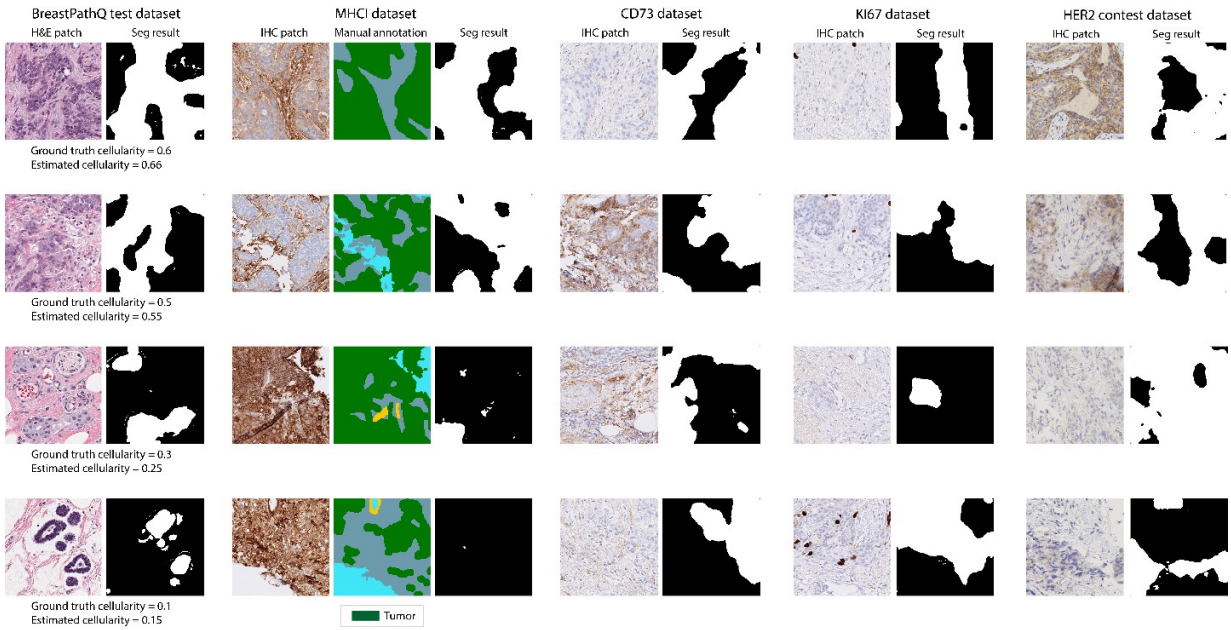

Figure: Tumor segmentation results (white: tumor; black: non-tumor) on the BreastPathQ test dataset and multiple external IHC datasets, obtained using a standard U-Net trained on the BreastPathQ training dataset with image-level cellularity labels.

Resources

1. BreastPathQ dataset: a public dataset consisting of 69 H&E stained WSI collected from the resection specimens of 37 post-neoadjuvant therapy patients with

invasive residual breast cancer. 2579 image patches with ROI of 512 × 512 pixels are manually annotated with estimated cellularity ranging between [0, 1].

2. Other public datasets: https://github.com/maduc7/Histopathology-Datasets

3. IHC datasets in NEOCHECKRAY. There are 109 IHC patches stained with an MHC-I antibody with pixel-level manual annotations.

Contact

Dr. Ir. Jennifer Dhont (jennifer.dhont@hubruxelles.be), Head of Data Science & AI Research Unit at Hopital Universitaire de Bruxelles (Erasme campus)

References

[1] D. Yan, X. Ju, et al., "Tumour stroma ratio is a potential predictor for 5-year disease-free survival in breast cancer," *BMC Cancer*, vol. 22, no. 1, p. 1082, Oct. 2022.

[2] L. Priya C V, B. V G, V. B R, and S. Ramachandran, "Deep learning approaches for breast cancer detection in histopathology images: A review," *Cancer Biomarkers*, vol. 40, no. 1, pp. 1–25, May 2024.

[3] M. Peikari, S. Salama, et al., "Automatic Cellularity Assessment from Post-Treated Breast Surgical Specimens," *Cytometry A*, vol. 91, no. 11, pp. 1078–1087, Nov. 2017.

[4] S. Akbar, M. Peikari, et al., "Automated and Manual Quantification of Tumour Cellularity in Digital Slides for Tumour Burden Assessment," *Sci Rep*, vol. 9, no. 1, p. 14099, Oct. 2019.

[5] X. Catteau, E. Zindy, et al., "Comparison Between Manual and Automated Assessment of Ki-67 in Breast Carcinoma: Test of a Simple Method in Daily Practice," *Technol Cancer Res Treat*, vol. 22, p. 15330338231169603, Jan. 2023.

[6] E. H. Allott, S. M. Cohen, et al., "Performance of Three-Biomarker Immunohistochemistry for Intrinsic Breast Cancer Subtyping in the AMBER Consortium," *Cancer Epidemiology, Biomarkers & Prevention*, vol. 25, no. 3, pp. 470–478, Mar. 2016.

[7] T. Vougiouklakis, B. J. Belovarac, et al., "The diagnostic utility of EZH2 H-score and Ki-67 index in non-invasive breast apocrine lesions," *Pathology - Research and Practice*, vol. 216, no. 9, p. 153041, Sep. 2020.

[8] J. R. Brown, M. P. DiGiovanna, et al., “Quantitative assessment Ki-67 score for prediction of response to neoadjuvant chemotherapy in breast cancer,” *Laboratory Investigation*, vol. 94, no. 1, pp. 98–106, Jan. 2014.
[9] W. Bulten, P. Bándi, et al., “Epithelium segmentation using deep learning in H&E-stained prostate specimens with immunohistochemistry as reference standard,” *Sci Rep*, vol. 9, no. 1, p. 864, Jan. 2019.
[10] Y.-R. Van Eycke, C. Balsat, et al., “Segmentation of glandular epithelium in colorectal tumours to automatically compartmentalise IHC biomarker quantification: A deep learning approach,” *Medical Image Analysis*, vol. 49, pp. 35–45, Oct. 2018.
[11] D. Firmbach, M. Benz, et al., “Tumor–Stroma Ratio in Colorectal Cancer—Comparison between Human Estimation and Automated Assessment,” *Cancers*, vol. 15, no. 10, p. 2675, Jan. 2023.
[12] G. Campanella, M. G. Hanna, et al., “Clinical-grade computational pathology using weakly supervised deep learning on whole slide images,” *Nat Med*, vol. 25, no. 8, pp. 1301–1309, Aug. 2019.
[13] A. Rakhlin, A. Tiulpin, et al., “Breast Tumor Cellularity Assessment Using Deep Neural Networks.”
[14] K. Li, Z. Qian, Y. Han, et al., “Weakly supervised histopathology image segmentation with self-attention,” *Medical Image Analysis*, vol. 86, p. 102791, May 2023.
[15] Z. Jia, X. Huang, et al., “Constrained Deep Weak Supervision for Histopathology Image Segmentation,” *IEEE Trans. Med. Imaging*, vol. 36, no. 11, pp. 2376–2388, Nov. 2017.
[16] R. Wang, Y. Qiu, et al., “MIHIC: a multiplex IHC histopathological image classification dataset for lung cancer immune microenvironment quantification,” *Front Immunol*, vol. 15, p. 1334348, Feb. 2024.