Automated Quality Control and Stain Classification of Whole-Slide Images in Pediatric Brain Tumors: Developing Scalable Harmonization Methods for the Children’s Brain Tumor Network Dataset

High-resolution whole-slide images (WSIs) are increasingly central to pediatric brain tumor research, yet large-scale quality control (QC) and metadata curation remain persistent bottlenecks. As part of the Kids First program, the Children's Brain Tumor Network (CBTN) repository provides a large dataset of WSIs across pediatric brain tumor histopathologies (2,277 patients, 2,620 tumor samples, 19,176 WSIs). Given this collection consists of WSIs acquired through clinical protocols, available stain types across samples can differ due to their diagnostic cohort and thus the clinical relevance of specific stain markers. WSIs can also exhibit considerable variability in tissue quality and digitization artifacts, often without reliable annotations. These inconsistencies limit downstream applications in computational pathology and multi-modal integration. We propose testing unsupervised and supervised machine learning methods to address two critical challenges: (1) automated detection of poor-quality or outlier WSIs, and (2) classification of stain type (e.g., H&E, Ki-67, GFAP, etc.) in the absence of metadata. Leveraging dimensionality reduction and patch-level feature extraction via pretrained convolutional neural networks (e.g., ResNet, CLIP, or foundational digital pathology models), we will cluster WSIs or tile patches into quality- and stain-coherent groups. Clustering results will be validated against available metadata, expert annotation, and slide-level inspection. Our approach provides a scalable, annotation-light method to improve data hygiene in CBTN’s extensive pathology archive. By identifying low-quality or mislabeled images and surfacing underrepresented staining types, this project supports more reliable use of CBTN pathology data in downstream machine learning pipelines, biomarker discovery, and computational pathology studies.