Manuscript 2021 Journal of Pathology Informatics

2021 Journal of Pathology Informatics

Title: “A pathologist-annotated dataset for validating artificial intelligence: A project description and pilot study”

- **Author list:** Sarah N Dudgeon, Si Wen, Matthew G Hanna, Rajarsi Gupta, Mohamed Amgad, Manasi Sheth, Hetal Marble, Richard Huang, Markus D Herrmann, Clifford H Szu, Darick Tong, Bruce Werness, Evan Szu, Denis Larsimont, Anant Madabhushi, Evangelos Hytopoulos, Weijie Chen, Rajendra Singh, Steven N Hart, Ashish Sharma, Joel Saltz, Roberto Salgado, Brandon D Gallas

Abstract:

- **Purpose:** Validating artificial intelligence algorithms for clinical use in medical images is a challenging endeavor due to a lack of standard reference data (ground truth). This topic typically occupies a small portion of the discussion in research papers since most of the efforts are focused on developing novel algorithms. In this work, we present a collaboration to create a validation dataset of pathologist annotations for algorithms that process whole slide images. We focus on data collection and evaluation of algorithm performance in the context of estimating the density of stromal tumor-infiltrating lymphocytes (sTILs) in breast cancer.
- **Methods:** We digitized 64 glass slides of hematoxylin- and eosin-stained invasive ductal carcinoma core biopsies prepared at a single clinical site. A collaborating pathologist selected 10 regions of interest (ROIs) per slide for evaluation. We created training materials and workflows to crowdsource pathologist image annotations on two modes: an optical microscope and two digital platforms. The microscope platform allows the same ROIs to be evaluated in both modes. The workflows collect the ROI type, a decision on whether the ROI is appropriate for estimating the density of sTILs, and if appropriate, the sTIL density value for that ROI.
- **Results:** In total, 19 pathologists made 1645 ROI evaluations during a data collection event and the following 2 weeks. The pilot study yielded an abundant number of cases with nominal sTIL infiltration. Furthermore, we found that the sTIL densities are correlated within a case, and there is notable pathologist variability. Consequently, we outline plans to improve our ROI and case sampling methods. We also outline statistical methods to account for ROI correlations within a case and pathologist variability when validating an algorithm.
- **Conclusion:** We have built workflows for efficient data collection and tested them in a pilot study. As we prepare for pivotal studies, we will investigate methods to use the dataset as an external validation tool for algorithms. We will also consider what it will take for the dataset to be fit for a regulatory purpose: study size, patient population, and pathologist training and qualifications. To this end, we will elicit feedback from the Food and Drug Administration via the Medical Device Development Tool program and from

the broader digital pathology and AI community. Ultimately, we intend to share the dataset, statistical methods, and lessons learned.

- Link to journal article page