Image analysis in oncology research: an institutional experience in how to overcome obstacles and obtain reliable data

A.C. Ortiz[1], B.V. Jardim-Perassi[2], M.R. Tomaszewski[2], J.O. Johnson[1], R.J. Gillies[2] and M.M. Bui[3]
[1]Analytic Microscopy Core, [2]Department of Cancer Physiology, [3]Department of Anatomic Pathology
H. Lee Moffitt Cancer Center & Research Institute, Tampa, Florida

MOFFITT CANCER CENTER

Abstract

Image analysis is motivated by the desire to extract relevant biological data from digital images in order to answer underlying questions about the pathogenesis of disease, and other biological phenomena. With the advent of automated whole slide scanning came the demand for image analysis software equipped with machine learning and artificial intelligence. Thus, ensued new issues and obstacles in image analysis and for those who perform it. Here, we present common difficulties in image acquisition and analysis on a cohort of tissue samples processed through the Analytic Microscopy Core at Moffitt Cancer Center. We compare pros and cons between two types of image analysis software on images acquired on an Aperio AT2 whole slide scanner. We provide a realistic workflow which addresses and circumvents common bottlenecks in data exploitation from pre- to post-acquisition. In addition, we offer strategies to overcome complications like stitching artifacts from image acquisition, tissue heterogeneity and damage, variability in stain penetrance, and border definition during segmentation which all result in variation between samples. Being mindful of these obstacles and verifying data post-processing can lead to robust, reliable, reproducible results, and biologically relevant data.

Objective

Image analysis from whole slide imaging data is susceptible to bottlenecks that occur throughout in the entirety of the image analysis workflow that limits the amount of useable data. Our objective is to foster a robust image analysis plan that limits the magnitude of this bottleneck.

Image Analysis Bottleneck

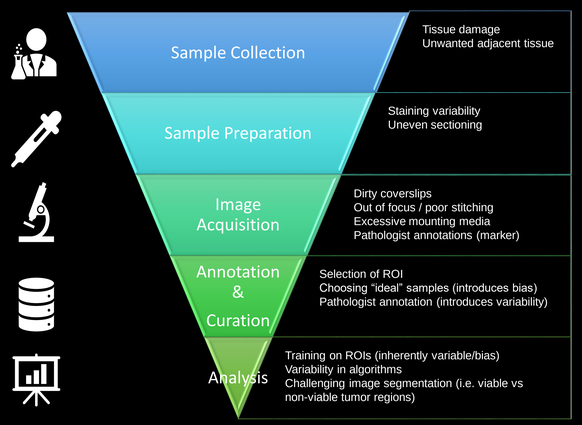

Figure 1. A schematic representation of the bottleneck generated in the image analysis workflow. From sample preparation to quantitative data generation there are various factors arise that impact the amount of exploitable data.

An "Ideal" Specimen

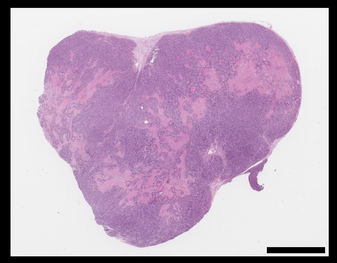

Figure 2. An "ideal" H&E sample. An example of a H&E stained breast cancer xenograft tumor specimen with even staining, no adjacent tissue, clean coverslip, no tissue damage, and discernible regions of interest. Scale bar = 3 mm.

The Harsh Reality

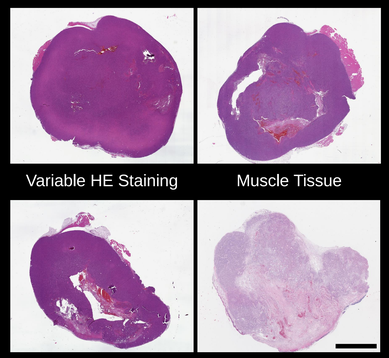

Figure 3. The harsh reality of H&E samples. An example of H&E stained pancreatic tumor (Panc02) specimens with variable staining, adjacent tissue, tissue damage, and indiscernible regions of interest. All with visible image stitching artifacts. Scale bar = 4 mm.

Unsupervised vs Supervised Segmentation

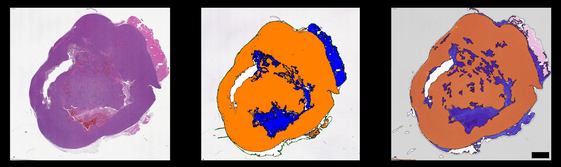

Figure 4. Segmentation variability between image analysis algorithms. An example of an H&E stained pancreatic tumor (Panc02) specimen. The original specimen (left) is segmented to viable (orange) and non-viable (blue) tumor regions by unsupervised/untrained K-means clustering (middle), and decision forest (right), which requires manual training. K-means clustering saves time but is less accurate than machine learning segmentation. Scale bar = 2 mm.

Addressing the Bottleneck

Figure 5. A diagram showing steps in the image analysis workflow with recommendations to limit bottlenecking.

Methods

Histology slides were scanned using the Aperio™ ScanScope AT2 (Leica Biosystems, Vista, CA) with a 20x/0.8NA objective lens. In order to demonstrate image analysis data is variable depending on the amount of training provided to the machine learning software (Definiens Tissue Studio 2.7), 8 H&E slides of pancreatic tumors (Panc02) with varying levels of necrosis were analyzed with minimal (1 training area) and maximal training (12 training areas). Images were annotated by a pathologist prior to analysis. From each slide, 1000x1000 pixel regions of 5 viable and 5 non-viable tissue were excised for scoring for a total of 80 scored images.

To evaluate performance of each algorithm, we used the following definitions for performance metrics:

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+TN+FP}$$

Whereby, TP = True Positive, TN = True Negative, FP = False positive, and FN = False Negative. Formulas are from Xu et al., 2020 (DOI: 10.4103/jpi.jpi_68_19).

Results

Amount of training impacts image segmentation. Segmentation of viable and non-viable tumor areas are impacted by the level of training and representative areas provided to the machine learning software (Figure 6). Maximal training provides additional detection and segmentation of non-viable tumor areas.

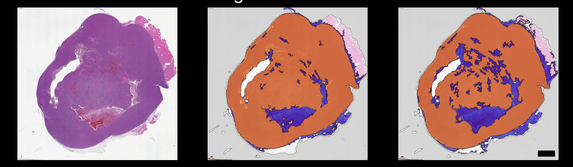

Figure 6. Segmentation of viable and non-viable histological tumor sections are impacted by amount of training. A tumor section stained with H&E (left), and respective masks of viable (orange) and non-viable (blue) tumor regions with minimal training (middle), and maximal training (right). Scale bar = 2 mm.

Algorithm performance is determined by amount of training. Performance of each algorithm was evaluated by comparing the analysis segmentation versus the annotation provided by experts. We observed that maximal training areas increased the sensitivity and accuracy of non-viable tumor areas while decreasing in viable tumor areas (Table 1).

	Viable		Non-Viable	
	Minimal	Maximal	Minimal	Maximal
Sensitivity	63%	38%	73%	82%
Accuracy	63%	38%	74%	83%

Table 1. Summary of algorithm performance across all histological slides for viable and non-viable tumor areas.

Conclusions & Future Directions

Image analysis software equipped with machine learning algorithms and artificial intelligence has been a buzzword in the scientific community. However, these software are susceptible to inherent bias based on user training and pathologist annotation. We observed that while increased training enhanced sensitivity and accuracy of non-viable tumor areas, it can impact the ability to detect some viable tumor areas. As shown here, automated image analysis is a balancing act with the final goal of attaining the best possible segmentation. Meaning, at times you need to compromise the accuracy and sensitivity of one parameter to improve the results of the other. In addition, deep learning algorithms are currently in development and could potentially overcome these limitations and improve accuracy in histopathologic diagnosis. In order to improve on current analysis methods, imaging scientists should reduce factors that negatively impact the digital pathology workflow, which will help create robust reproducible and reliable results.

Acknowledgements

This work is supported by the Analytic Microscopy Core Facility at the H. Lee Moffitt Cancer Center & Research Institute, an NCI designated Comprehensive Cancer Center (A. Ortiz, J. Johnson, M. Bui; P30-CA076292). In addition, by a NIH grant awarded through the NCI (R.J. Gillies; 5R01CA187532).