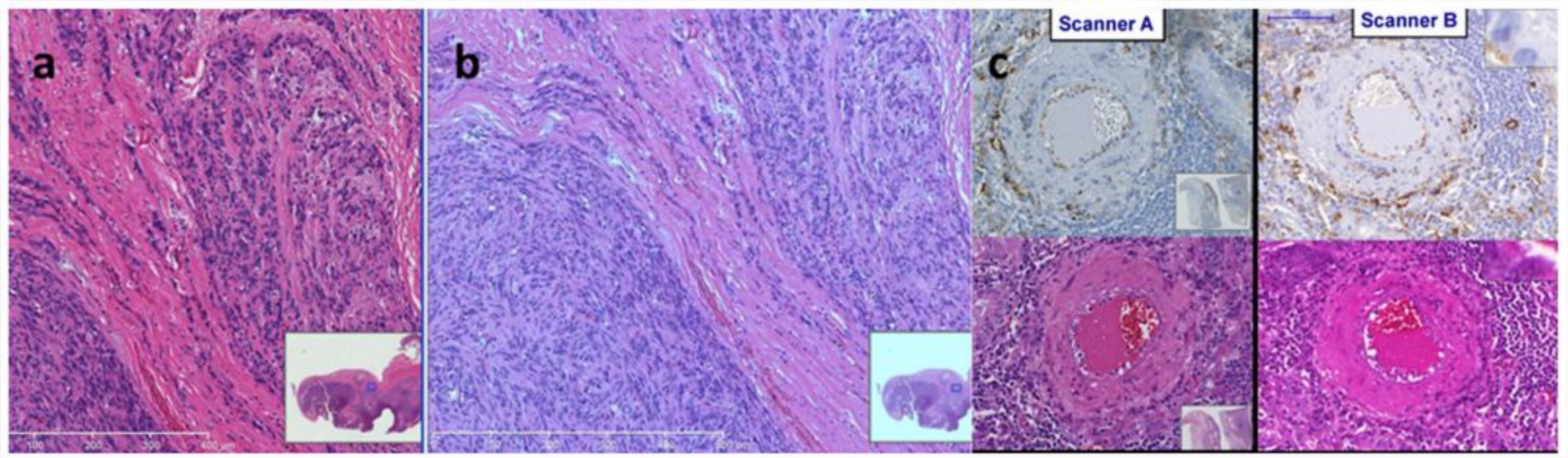
Scanner Variability

🔬 SeeGene Project Report 🔬

Willmer R. Quinones R.
2021.09.17

Scanner Variability | **What's the issue?**

Each time the scanner is changed to digitize tissue images, **the properties of the whole slide image change**. Although for the pathologists these changes are meaningless, **they might have a great effect on the machine learning model**.



Badano, Aldo, et al. "Consistency and Standardization of Color in Medical Imaging: a Consensus Report." Journal of digital imaging. 28. (2014).

Scanner Variability | **Literature Review**

- **Conversion to grayscale** ignores the important information regarding the color representation.
- **Color normalization** tries to adjust the color distribution of the source image to that of a reference image. However, this method might be expensive because we need to adjust for every reference data.
- **Color augmentation** is performed by applying random hue, saturation, brightness, and contrast. The advantage of color augmentation lies in the easy implementation regardless of the object being analyzed.

* Komura, Daisuke, et al. "Machine Learning Methods for Histopathological Image Analysis." Computational and Structural Biotechnology Journal. (2018).

Scanner Variability | **Literature Review**

Lafarge et al. deals with dataset variabilities caused by differences in pathology labs (e.g., scanners).

1. Color Augmentation
2. Staining Normalization (Macenko)
3. Domain-Adversarial Neural Network

CA		■			■	■		■
SN			■		■		■	■
DANN				■		■	■	■
ITS	$.61 \pm .02$	$.61 \pm .01$	$.57 \pm .06$	$.61 \pm .02$	$.55 \pm .01$	$\mathbf{.62 \pm .02}$	$.61 \pm .01$	$.57 \pm .01$
ETS	$.33 \pm .08$	$.58 \pm .03$	$.46 \pm .02$	$.55 \pm .05$	$.48 \pm .08$	$\mathbf{.62 \pm .00}$	$.51 \pm .02$	$.53 \pm .03$

Maxime W. Lafarge et al. "Domain-Adversarial Neural Networks to Address the Appearance Variability of Histopathology Images." Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. DLMIA (2017)

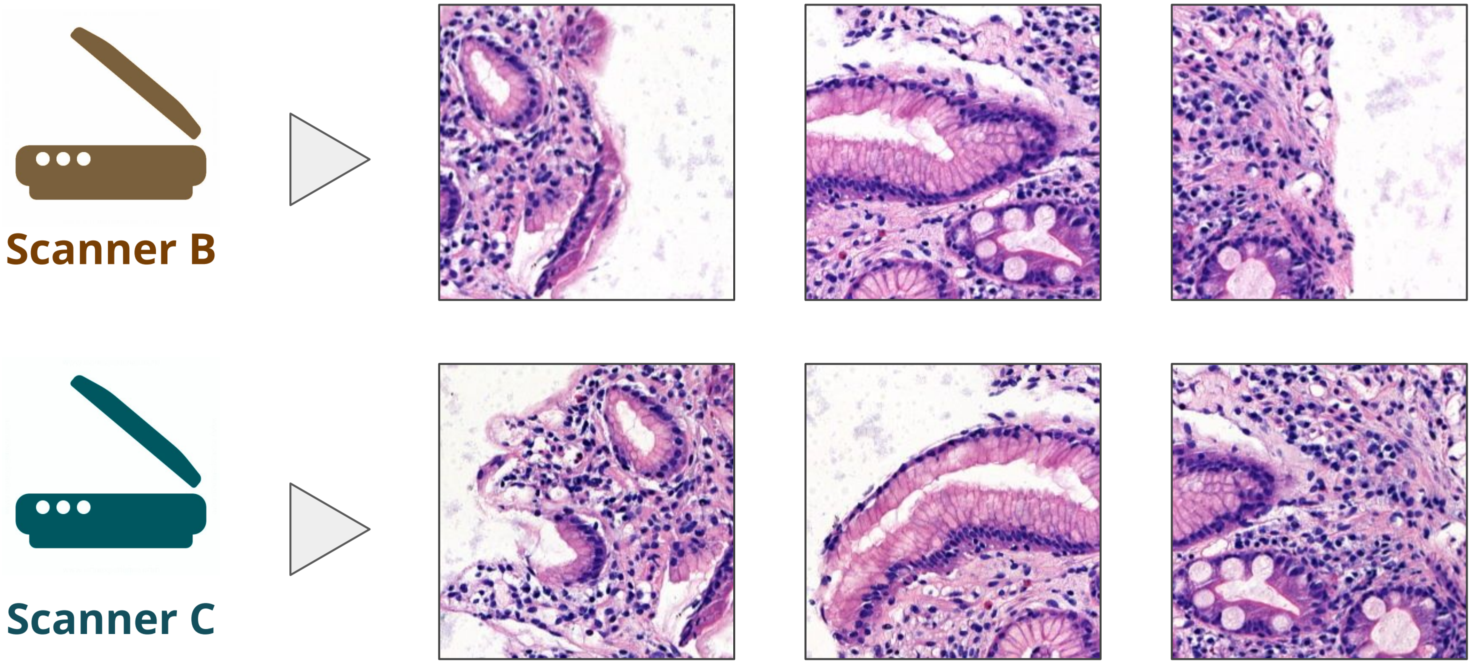
Scanner Variability | **First Approach**

There are some approaches to deal with this kind of variability, which include **adversarial training** and **color augmentation**. Color Augmentation has been shown useful and effective in some domain (according to the literature review), hence:

- Split the dataset (WSI) into:
  - Train → Scanner A
  - Test → Scanner B and Scanner C
- **Implement color augmentation during the training process**
- If the model performs well on **scanner B** and **scanner C**, having been trained with **Scanner A**, then we can have more confidence that the model can be robust to scanner variability.

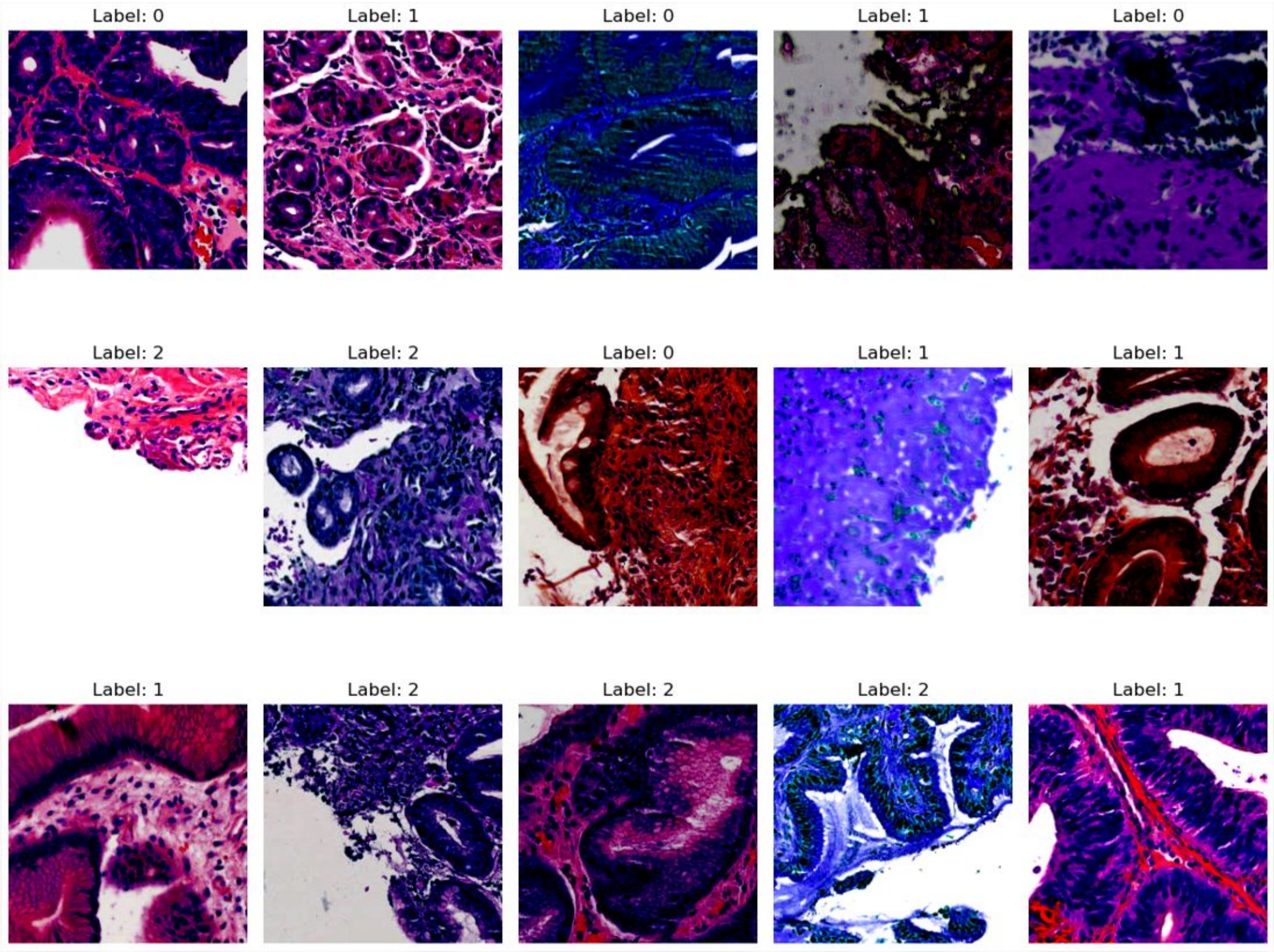
Scanner Variability | **First Approach**

Depending on the scanner, the tiling code outputs different patches for the same sample:



First, second, and third patches extracted from scanner 1 and scanner 2

Scanner Variability | **First Approach**



Scanner Variability | **First Approach**

Backbone: ResNet50

Adam Optimizer - lr: 0.001 | Exponential Decay - lrd: 0.85

Epochs: 20 | Early Stopping - Patience: 3 epochs

	Same Scanner	**Scanner B**	**Scanner C**
Baseline	**0.768**	?	?
Color Augmented	0.750	?	?

Scanner Variability | **First Approach**

```
Backbone: ResNet50
Adam Optimizer - lr: 0.001  |  Exponential Decay - lrd: 0.85
Epochs: 20  |  Early Stopping - Patience: 3 epochs
```

	Same Scanner
Baseline	**0.768**
Color Augmented	0.750

New slides are not "fine-annotated", but weakly-annotated

Cannot perform tile-level evaluation
🙅‍♂️

Scanner Variability | **First Approach**

Backbone: ResNet50

Adam Optimizer - lr: 0.001 | Exponential Decay - lrd: 0.85

Epochs: 20 | Early Stopping - Patience: 3 epochs

Slide Level Prediction:

D if *#_patches_predicted_as_D* > *#_patches_predicted_as_M* else *M*

	Same Scanner	Scanner B	Scanner C
Baseline	**0.885**	0.795	0.663
Color Augmented	0.855	**0.819**	**0.837**

Scanner Variability | **First Approach**

Backbone: ResNet50

Adam Optimizer - lr: 0.001 | Exponential Decay - lrd: 0.85

Epochs: 20 | Early Stopping - Patience: 3 epochs

Slide Level Prediction:

D if *#_patches_predicted_as_D* > *#_patches_predicted_as_M* else *M*

	Same Scanner	Scanner B	Scanner C
Baseline	**0.885**	0.795	0.663
Color Augmented	0.855	**0.819**	**0.837**

Although this slide-level classification is somehow heuristic, there is some evidence of <u>robustness</u> 🤔

~ THANK YOU ~