SIIM CMIMI24
Oct 21-22 | Boston, MA
Conference On Machine Intelligence In Medical Imaging

Dissecting the Impact of Data Augmentation on Whole Slide Image Classification

Dagoberto Pulido Arias, PhD, Massachusetts General Hospital; Tiago Gonçalves, MSc; Jaime Cardoso, PhD; Jayashree Kalpathy-Cramer, PhD; Elizabeth Gerstner, MD; Albert Kim, MD; Christopher Bridge, DPhil

Introduction/Background

Machine learning in pathology shows potential for advancing precision oncology across multiple tumor types. However, limited annotated samples and high intra-class variability in histological images constrain these models' clinical potential. Data augmentation enhances model performance and generalizability, especially with limited training data, but whole-slide-image (WSI) classification models require pre-computed tile-level features, limiting the ability to perform on-the-fly augmentations. This study compares the impact of image-level and feature-level data augmentation on WSI classification in pathology at both tile and slide levels, specifically using foundational models for feature extraction. Our goal is to assess the necessity of data augmentation when using foundational models.

Methods/Intervention

We conducted experiments using WSIs from the breast cancer subset of The Cancer Genome Atlas (TCGA), utilizing two pre-trained feature extractor models: ResNet-50, trained on ImageNet, and CONCH (CONtrastive learning from Captions for Histopathology), a vision-language foundational model designed for microscopic pathology and trained on millions of WSIs. We used a Clustering-constrained attention multiple instance learning model for classification. Our study compared different image-level augmentations, such as Hematoxylin-Eosin-DAB (HED) color transformation and tile shifting (see Figure 1), and feature-level augmentation using Pseudo-Bag Mixup (PseMix), applied to features extracted by both models. Feature-level augmentation creates synthetic variations of feature representations to help the model generalize better.

Results/Outcome

The CONCH model, augmented with individual techniques, significantly improved classification accuracy, achieving a test AUC of 0.868 with Pseudo-Bag Mixup. Without augmentation, the baseline model's test AUC was 0.758 with ResNet-50 and 0.846 with CONCH. Image-level augmentations like HED color transformation and tile shifting also improved performance. For example, tile shifting led to a test AUC of 0.835 with ResNet-50, and HED alone with CONCH achieved a test AUC of 0.856 (see Table 1).

Conclusion

Data augmentation techniques are essential for enhancing model performance, addressing the challenges of limited annotated data and high intra-class variability. Both image-level and feature-level augmentations improve predictive performance, providing a robust solution for increasing the accuracy and reliability of computational pathology models.

Statement of Impact

This study underscores the importance and benefits of efficient data augmentation in computational pathology, contributing to the development of robust, high-performing algorithms without needing additional data and resources.

Performance by Augmentation Type

Augmentation	Level	Features	Val AUC	Test AUC
None	-	Resnet 50	0.769	0.758
HED	Slide	Resnet 50	0.784	0.765
Tile shift	Slide	Resnet 50	0.788	0.791
HED	Tile	Resnet 50	0.694	0.804
Tile shift	Tile	Resnet 50	0.824	0.835
Psemix	Features	Resnet 50	0.776	0.807
None	-	CONCH	0.851	0.846
Psemix	Features	CONCH	0.832	0.868
HED	Slide	CONCH	0.805	0.856
HED	Tile	CONCH	0.826	0.868
Tile shift	Slide	CONCH	0.853	0.848
Tile shift	Tile	CONCH	0.822	0.867

Table 1. Validation and Test AUC for different data augmentation techniques applied to ResNet-50 and CONCH models in WSI classification. Augmentations include HED color transformation, tile shifting, and Pseudo-Bag Mixup (PseMix) at both slide and tile levels.

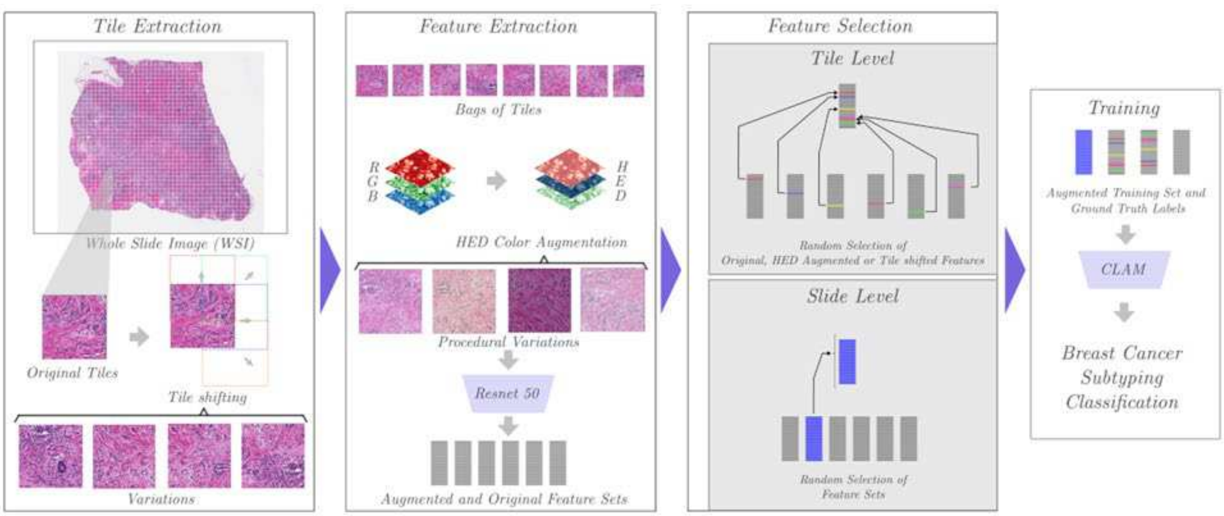

Figure 1. Augmentation pipeline for breast cancer subtyping in WSIs. Tiles are extracted from WSIs, followed by feature extraction with HED color augmentation and tile shifting at both tile and slide levels. These augmented features are used to train a model for breast cancer classification.

Keywords

artificial intelligence; attention mechanisms; breast cancer; computational pathology; data augmentation; whole slide image