Using MONAI Pre-Trained Models for Colorectal Tissue Type Phenotyping: A Feasibility Study to Integrate Deep Learning Model Results using the Medical Extension OMOP CDM

Shijia Zhang, M.B.I[1], Woo Yeon Park, M.S.[1], Blake Dewey, Ph.D.[2], Paul Nagy, Ph.D.[1]

[1]Biomedical Informatics and Data Science, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

[2]Department of Neurology, Johns Hopkins School of Medicine, Baltimore, MD, USA.

Background:
The development of deep learning models has led to significant achievements in medical imaging analysis[1]. Medical Open Network for AI (MONAI) is an open-source initiative built on the Python PyTorch library to enable deep learning research within medical imaging research. Our study illustrates the utility of pre-trained machine learning models to effectively extract structured and coded imaging features stored in the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) medical imaging extension[2]. The OMOP CDM medical imaging extension allows the systematic storage of imaging findings and their provenances.

Methods:
This study adopts a transfer learning paradigm by fine-tuning MONAI's pre-existing pathology tumor detection model, which is based on ResNet-18[5]. The model was fine-tuned using the PathMNIST Colon Pathology dataset[3,4]. This dataset categorizes colon pathology images into nine distinct classes, each representing a unique type of pathological tissues. The OMOP CDM with medical imaging extension (MI-CDM) includes two additional tables—Image_occurrence and Image_feature. These tables are designed to record provenance of the imaging features. The outputs generated by the model were subsequently stored in the Conditions_occurrence table (Figure1).

To exam the effectiveness of the fine-tuned model, we compared its performance against the benchmark ResNet-18 model on the overall classification task. We also recorded the model's performance on phenotyping different tissue types, including adipose, background, debris, lymphocytes, mucus, smooth muscle, normal colon mucosa, cancer-associated stroma, and colorectal adenocarcinoma epithelium.

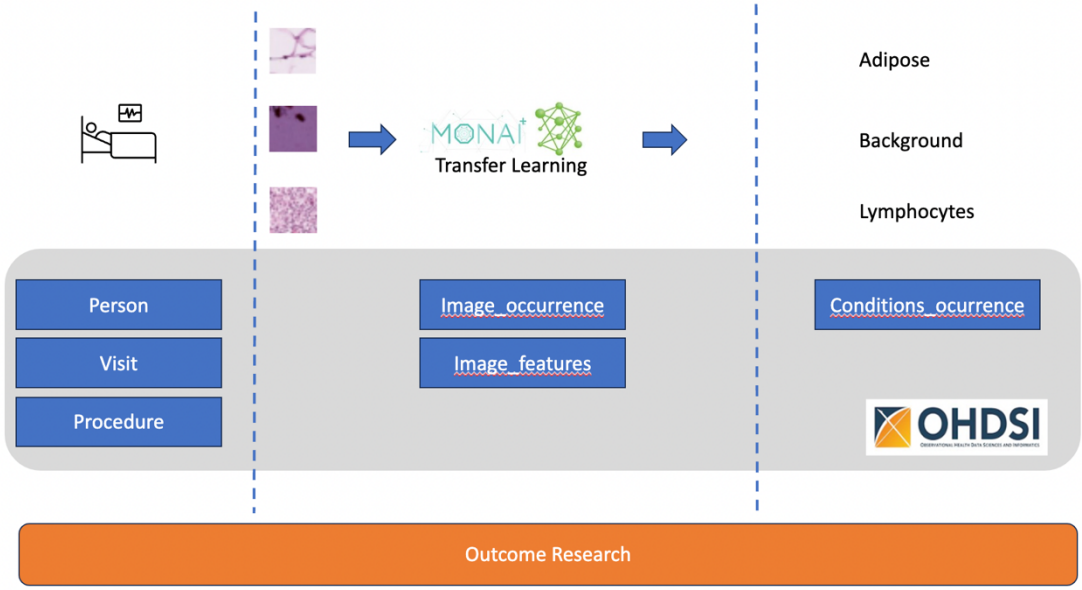

Figure1 - Visual Representations OMOP data encoding

Results:
Preliminary findings corroborate our hypothesis that the MONAI pre-trained models, once fine-tuned, surpass the performance of the benchmark ResNet-18 model on PathMNIST testing dataset on overall AUC (0.995 vs. 0.989) and Accuracy (0.928 vs. 0.909) and the model has performed well on most of the tissue types (Table1). These results underpin the efficacy and reliability of using the adjusted MONAI's pre-trained model. Furthermore, the model results, colon tissue types, are recorded in the Conditions_occurrence table along with its provenance recorded in the Image_occurrence and Image_feature tables. This expands phenotype definitions by adding organ level details to the OMOP CDM.

	Precision	Recall	F1-Score	Num Cases
Adipose	0.96	0.95	0.95	1338
Background	0.95	1	0.98	847
Debris	0.79	0.9	0.84	339
Lymphocytes	0.97	1	0.98	634
Mucus	0.99	0.89	0.94	1035
Smooth Muscle	0.74	0.92	0.82	592
Normal Colon Mucosa	0.96	0.96	0.96	741
Cancer-Associated Stroma	0.95	0.53	0.68	421
Colorectal Adenocarcinoma Epithelium	0.94	0.97	0.96	1233

Table1 - Classification Performance by Tissue Types by MONAI's model

Conclusion:
The results of this study illustrate three critical findings. First, this elaborates on the potential of using machine learning models to populate evidence from pathology images to the OMOP CDM. This process significantly streamlined the recording and storing of medical imaging data. Second, this allows incorporating imaging findings without parsing through the unstructured reports and leads to a more organized and efficient system for interpreting image data. Lastly, the structured medical imaging findings will enhance phenotype definitions to include cell-level or organ-level details and enhance the specificity and accuracy of medical diagnoses and prognoses, contributing to personalized medicine.

Reference:

1. Jiang Y, Yang M, Wang S, Li X, Sun Y. Emerging role of deep learning-based artificial intelligence in tumor pathology. *Cancer Commun Lond Engl*. 2020;40(4):154-166. doi:10.1002/cac2.12012

2. Data Standardization – OHDSI. Accessed June 8, 2023. https://www.ohdsi.org/data-standardization/

3. MONAI Pathology Tumor Detection | NVIDIA NGC. NVIDIA NGC Catalog. Accessed June 8, 2023. https://catalog.ngc.nvidia.com/orgs/nvidia/teams/monaitoolkit/models/monai_pathology_tumor_detection

4. Yang J, Shi R, Wei D, et al. MedMNIST v2 - A large-scale lightweight benchmark for 2D and 3D biomedical image classification. *Sci Data*. 2023;10(1):41. doi:10.1038/s41597-022-01721-8

5. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. Published online December 10, 2015. doi:10.48550/arXiv.1512.03385