SIIM CAIMI25
Conference on Artificial Intelligence in Medical Imaging
OCT 20-21 | UNIVERSITY OF CALIFORNIA, IRVINE

# Applying a Pediatric Liver Fibrosis Classification Tool to Adult Trichrome-Stained Liver Biopsies: A Validation Study

Zachary Taylor, Cincinnati Children's Hospital; Anas Bernieh; Sarangarajan Ranganathan, MD, MBBS; Christopher Woods; Madhieh Shabanian, MS; Jennifer Picarsic, MD; Elanchezhian Somasundaram, PhD; Jonathan Dillman, MD, MSc

## Introduction/Background

AI tools and pipelines have been developed to provide preliminary assessments of liver fibrosis severity in trichrome-stained liver tissue, distinguishing between normal/mild and moderate/severe fibrosis. This classification can assist pathologists in diagnosis, as fibrosis grading is a tedious process with significant inter-observer variability. In this study, we validate a model originally developed using a pediatric dataset on an external adult validation cohort.

## Methods/Intervention

In this study, we applied the CLAM (Clustering-constrained Attention Multiple Instance Learning) pipeline (Mahmood et al.) to classify liver fibrosis severity in trichrome-stained pathology slides. A previously published five-fold model set, trained on 217 pediatric slides rated by two board-certified pathologists using a 70–10–20 train-tune-test split and CONCH attention weights, was evaluated on an external adult dataset to assess generalizability. The adult cohort included trichrome-stained slides annotated by a board-certified pathologist using METAVIR (0–4; n=164) and ISHAK (0–6; n=169) scores. Slides were binarized at a threshold of 2: scores 0–2 as low-stage (label A) and ≥2 as high-stage fibrosis (label B). This evaluation tests the model's robustness and clinical applicability across age groups and data sources.

## Results/Outcome

The CONCH models across their five folds reached an average accuracy of 0.83±0.03 and AUC of 0.92±0.02 for the 164 external METAVIR rated slides, and an average accuracy of 0.79±0.02 and AUC of 0.91±0.004 for the 169 external ISHAK rated slides. This compares to the earlier performance of the models on its internal cross-validation dataset reaching an accuracy of 0.80±0.07 and AUC of 0.88±0.07 for METAVIR and an accuracy of 0.85±0.07 and AUC of 0.92±0.04 for ISHAK across five folds of cross-validation.

## Conclusion

This study highlights the CLAM pipeline's effectiveness in binary classification of liver fibrosis in both pediatric and adult cases, outperforming internal benchmarks on the METAVIR scale. Limitations include a small dataset and a lack of multiple pathologist ratings. Future work should incorporate larger, more diverse datasets with multi-pathologist annotations to address inter-observer variability and improve model performance.

## Statement of Impact

This study shows the viability of applying a preliminary internal pediatric liver fibrosis classification model set to adult cohorts from outside sources, saving time during determination of liver disease progression.

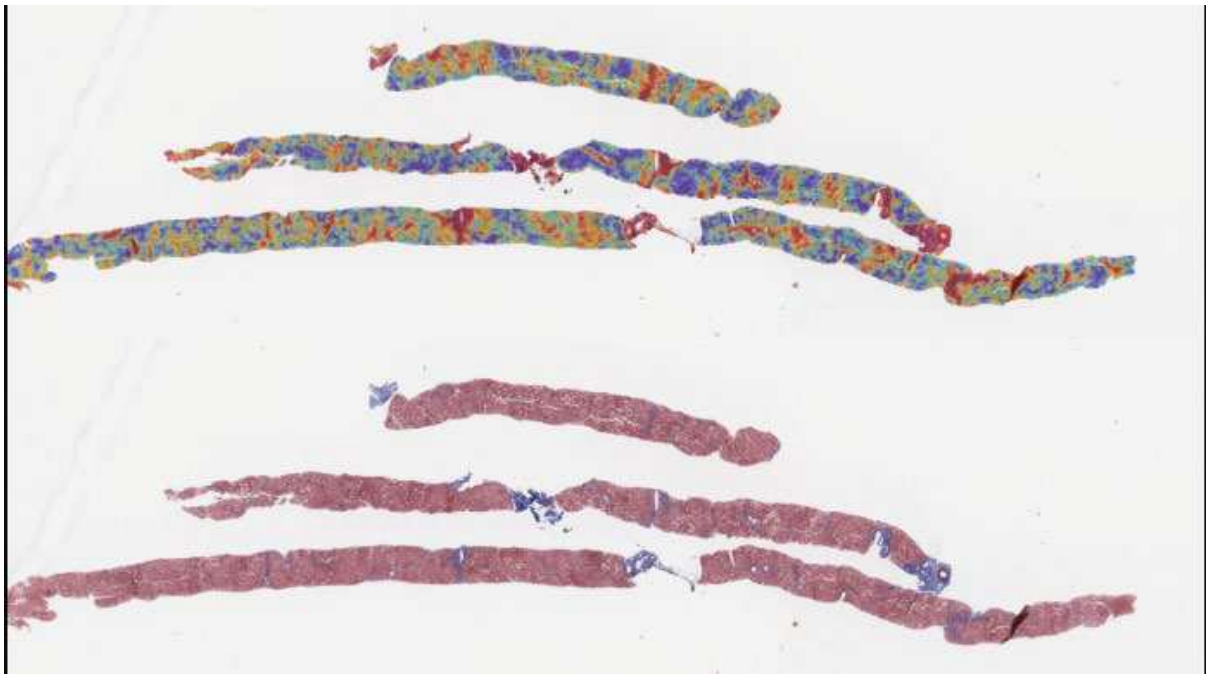

Figure 1. CLAM-generated model attention heatmap, (top) created from correctly categorized high-fibrosis regions on trichrome-stained liver tissue slide (bottom).

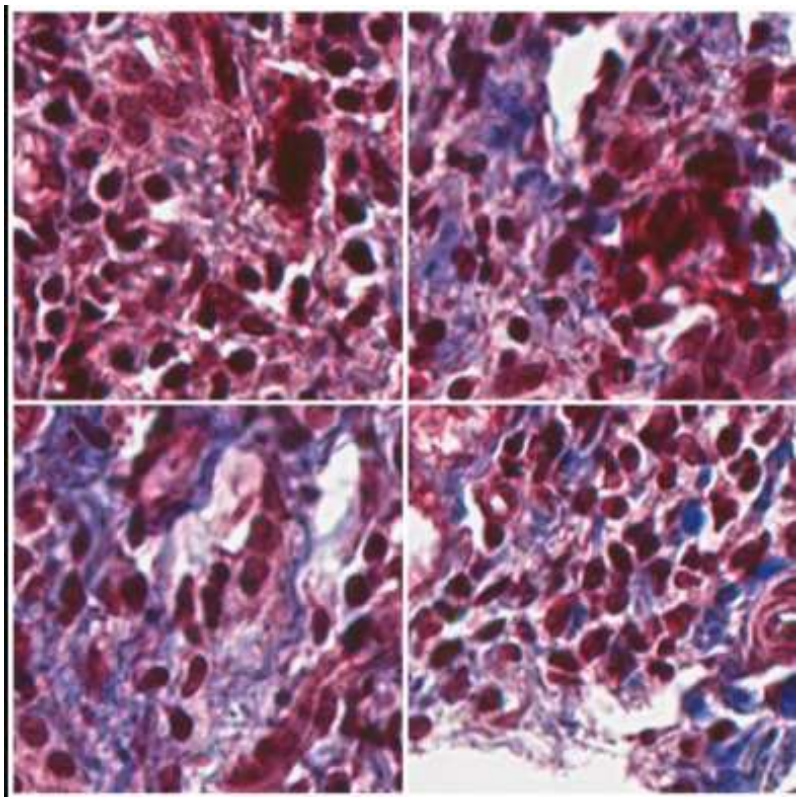

Figure 2. A selection of patches from liver tissue slide in Figure 1 with highest attention towards high-fibrosis.


## Keywords

Pathology; Classification; Liver Fibrosis; Histopathology; Pediatrics; Model Validation