Explainable Vision Transformer for Automated Liver Fibrosis Staging: A Deep Learning Approach with Attention Visualization

Research Paper Abstract

ABSTRACT

Liver fibrosis is a progressive condition characterized by excessive accumulation of extracellular matrix proteins, primarily resulting from chronic liver diseases such as hepatitis B and C, alcoholic liver disease, and non-alcoholic steatohepatitis (NASH). Accurate staging of liver fibrosis (F0-F4) is crucial for clinical decision-making, treatment planning, and prognosis assessment. Traditional methods rely on invasive liver biopsies preceded by histopathological analysis, which are subject to sampling variability and inter-observer disagreement. This study presents an automated, non-invasive deep learning framework for liver fibrosis staging using histopathological images with explainable artificial intelligence (XAI) capabilities.

We propose a Vision Transformer (ViT-B-16) based architecture with pre-trained ImageNet weights, fine-tuned for five-class liver fibrosis classification (F0: No fibrosis, F1: Portal fibrosis, F2: Periportal fibrosis, F3: Bridging fibrosis, F4: Cirrhosis). The input images undergo Contrast Limited Adaptive Histogram Equalization (CLAHE) preprocessing to enhance local contrast and highlight fibrotic tissue patterns. The model employs dropout regularization (p=0.3) to prevent overfitting and utilizes AdamW optimizer with label smoothing cross-entropy loss for robust training. Early stopping with patience monitoring ensures optimal model convergence without overtraining.

To address the critical need for interpretability in medical AI systems, we integrate Gradient-weighted Class Activation Mapping (Grad-CAM) visualization to generate attention heatmaps highlighting regions of the input image most influential for the model's predictions. This explainability component enables clinicians to validate model decisions and builds trust in automated diagnostic systems. The attention visualizations demonstrate that the model correctly focuses on fibrotic tissue patterns, portal areas, and architectural distortions characteristic of different fibrosis stages.

The model was validated using stratified train-validation splits to ensure balanced class representation. Performance evaluation includes multi-class accuracy, weighted F1-score, confusion matrix analysis, and Cohen's Kappa score with quadratic weighting to account for the ordinal nature of fibrosis staging. The quadratic-weighted Kappa is particularly appropriate as misclassifications between adjacent stages (e.g., F1 vs F2) are penalized less severely than distant misclassifications (e.g., F0 vs F4), reflecting clinical significance. K-fold cross-validation with 95% confidence intervals ensures statistical robustness of reported metrics.

The proposed framework demonstrates the potential of transformer-based architectures for automated liver fibrosis staging while maintaining clinical interpretability through explainable AI techniques. This approach offers a scalable, reproducible, and objective alternative to traditional histopathological assessment, potentially reducing diagnostic variability and improving patient outcomes through early and accurate fibrosis detection.

Keywords: Liver Fibrosis Staging, Vision Transformer (ViT), Deep Learning, Explainable AI, Grad-CAM, Medical Image Classification, CLAHE Preprocessing, Cohen's Kappa, Histopathology Analysis, Computer-Aided Diagnosis