A Novel CNN-ViT Approach to Endoscopic Ultrasound Image Analysis to Advance Diagnosis of Early-Stage Pancreatic Cancer

Shaye, Nina (School: Kalamazoo Area Mathematics and Science Center)

Pancreatic cancer is one of the deadliest cancers in the world, largely due to late-stage diagnosis. Current imaging diagnostic tools like CT and MRI scans lack the accuracy and sensitivity to detect early-stage tumors less than 1cm. Endoscopic ultrasound (EUS) is one of the most sensitive imaging techniques for detecting small pancreatic tumors, but interpretation of EUS images remains difficult and highly subjective. These ultrasound images are noisy and unclear, making it difficult to interpret and accurately diagnose. This study investigates whether combining two models can improve the detection of early-stage pancreatic cancer in EUS images. Three models were developed and evaluated: an EfficientNet-B0 convolutional neural network (CNN), a Vision Transformer (ViT), and a hybrid CNN–Vision Transformer architecture designed to combine local feature extraction with global contextual analysis. The models were trained and evaluated using a public dataset of 3,500 EUS images. Model performance was assessed using accuracy, recall, and area under the receiver operating characteristic curve (AUC). The hybrid CNN–ViT model achieved the strongest overall performance, with an accuracy of 94.2%, a recall of 93.4%, and an AUC of 0.96, demonstrating strong ability to distinguish pancreatic cancer from non-cancer images. Attention visualization further showed that the model focused on clinically relevant regions of the ultrasound images, suggesting that predictions were based on meaningful anatomical features. By improving the accuracy and consistency of EUS interpretation, this hybrid CNN–ViT architecture enables more reliable detection of early-stage pancreatic cancer, potentially reducing late-stage diagnoses.

Awards Won: