Progress Report 4: Cross-Architectural Knowledge Distillation in Medical Imaging: Multi-Scale Geometric Feature Fusion for MRI Scan Classification

Mahiyat Nawar Mantaqa
2122455042

MD. Shafat Islam Khan
2121517642

I. TRAINED HYBRID ViT TEACHER MODEL

Two Vision Transformer (ViT) models: ViT-B-16 (Base) and ViT-L-16 (Large) is applied. Both models are initialized with pretrained weights. Each ViT model processes the input image separately, producing classification outputs of size: num_classes. These outputs are then concatenated to form a combined feature representation, which is passed through a final linear classifier to produce the final prediction. This approach leverages the strengths of both ViT architectures, with ViT-B-16 providing efficiency and ViT-L-16 capturing richer features due to its larger parameter count. The 'initialize_model' function ensures that the model is set up correctly for training with an input size of 224x224 pixels.
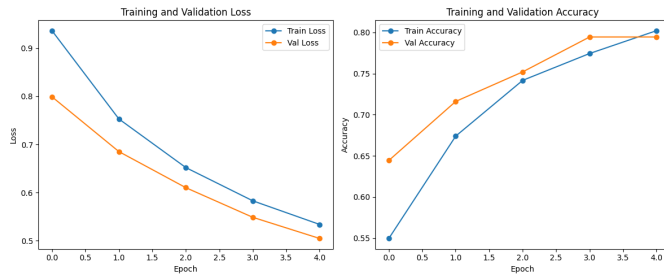

Fig. 1. Hybrid ViT Loss and Accuracy Curve

Fig. 2. Knowledge Distillation Loss and Accuracy Curve

II. KNOWLEDGE DISTILLATION AFTER HYBRID ViT

The hybrid teacher models ViT-B and ViT-L were then used to perform the knowledge distillation process as applied before to the EfficientNetB0 student model.
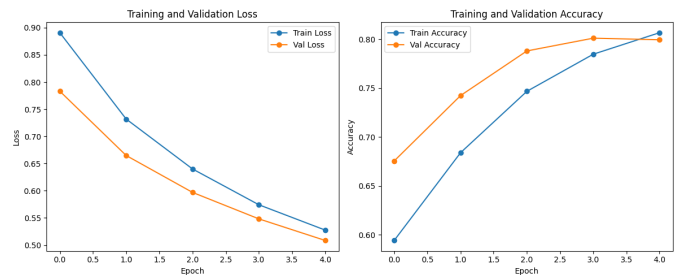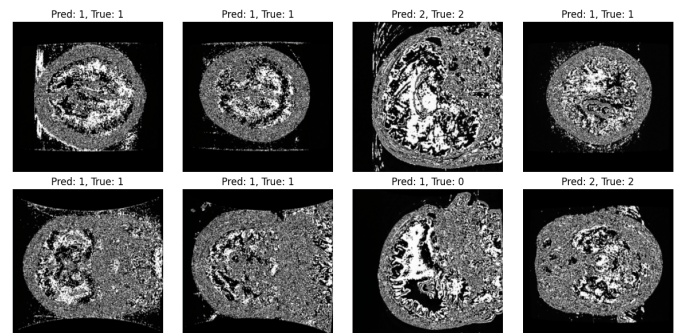

Fig. 3. Knowledge Distillation Prediction Results