Final Project Proposal Form (CN6000)

"Illuminating the Black Box: An Explainable AI Framework for Brain Tumor Segmentation Using Volumetric Data Interpretation in 3D CNNs"

Yogi Amitkumar Patel   u2536809

Proposed Aim

The primary goal is to create and assess a new Explainable AI (XAI) framework for volumetric brain tumour segmentation by interpreting 3D Convolutional Neural Network models. By creating 3D voxel-based density maps, this framework will turn opaque 3D models into transparent systems. A significant innovation is the visualisation of these explanations in an immersive Virtual Reality (VR) environment, giving doctors a user-friendly and interactive decision-support tool.

Proposed objectives

1. To identify and evaluate the various processes for 3D, **voxel-based XAI** in order to quantify their contribution to 3D CNN segmentation predictions and localise pertinent features from multimodal 3D MRI volumes.
2. To implement and train a 3D CNN segmentation model using **PyTorch** and **MONAI** on the available dataset subset.
3. To create a **VR visualisation pipeline** that uses 3D Slicer and the SlicerVR extension to export the original 3D MRI, the 3D segmentation mask, and the 3D XAI saliency volume into NIfTI format (.nii.gz) for interactive, immersive analysis.
4. To integrate the trained 3D CNN with the established 3D XAI techniques to produce a single explainable hybrid model that can produce 3D prominent volumes.
5. to **quantitatively** examine the basic trade-off between the 3D model's spatial transparency and segmentation accuracy. This includes confirming the 3D XAI feature attributions' quantitative robustness.
6. to compare the visual saliency maps with accepted clinical information in order to qualitatively assess the accuracy of the 3D XAI output.

7. Using XAI's 3D interpretability insights, **prune** the 3D CNN model with the goal of reducing parameters while maintaining good segmentation accuracy on unknown 3D medical data.

Rationale

The **"Black Box"** nature of 3D CNNs, such as the **3D U-Net**, restricts clinical trust and safe adoption, despite its great accuracy in segmenting volumetric brain tumours. Standard 2D XAI heatmaps are not adequate for volumetric analysis, but clinicians need explanations that align with their 3D workflow.

By creating a 3D XAI framework that produces 3D saliency volumes (such as **3D Grad-CAM**) to explain the 3D CNN's voxel-level decisions, this study fills this gap. The main innovation is the development of a pipeline that uses 3D Slicer with SlicerVR to visualise these 3D explanations in an immersive virtual reality (VR) environment.

By enabling physicians to intuitively understand the model's logic, our 3D XAI-VR approach boosts confidence, facilitates decision-making, and offers the insights required for XAI-driven model pruning.

Facilities Required

Dataset

- Primary Dataset: BraTS 2021, a large dataset of multimodal 3D MRI scans (T1, T1ce, T2, FLAIR) in NIfTI and DICOM format.

Computational Requirement

- GPU: A local or cloud-based GPU with 16GB+ VRAM is required for training 3D CNN models on the full dataset.

Software Requirements

- Deep Learning Framework: PyTorch and MONAI
- Explainability Libraries: SHAP, LIME, and TorchCAM (or other 3D Grad-CAM implementations).
- VR Visualization: 3D Slicer with the SlicerVR extension
- 3D Rendering Utilities: ITK-SNAP

Hardware Requirements

- VR Headset: A Meta Quest 3

Supervisor

Maimoona Sharif