Toward Clinical Transparency: A Comparative Analysis of CNN and Vision Transformer Architectures for Explainable Brain Tumor MRI Classification

Areenee Hayeepador[1], **Patipat Sitpasert**[1], **Atchara Namburi**[1,*]
[1]Department of Computer and Information Science, Science and Engineering Faculty, Kasetsart University, Sakon Nakhon, Thailand

**Corresponding Author Email: csnarn@ku.ac.th*

Abstract

Brain tumors remain a significant global health challenge, necessitating early and precise diagnosis to improve patient survival and treatment outcomes. While deep learning has demonstrated remarkable success in automating MRI-based classification, clinical adoption is often hindered by the "black-box" nature of these models and the inherent class imbalance in medical datasets. This research proposes an explainable deep learning framework to classify brain tumors into four categories: glioma, meningioma, pituitary tumor, and healthy cases. We conduct a rigorous comparative study between state-of-the-art Convolutional Neural Networks (CNNs), such as ResNet-50 and EfficientNet, and the emerging Vision Transformer (ViT) architecture to evaluate their diagnostic efficacy.

To address data scarcity and class imbalance, synthetic data augmentation and SMOTE-based techniques are employed. To foster clinical trust, the framework integrates Gradient-weighted Class Activation Mapping (Grad-CAM) to generate visual saliency maps, providing radiologists with interpretable evidence of the model's decision-making process. Preliminary results indicate that while CNNs exhibit high precision in local feature extraction, Vision Transformers excel in modeling global contextual relationships within complex MRI slices. This study bridges the gap between high-performance computational modeling and clinical accountability, offering a scalable solution for AI-assisted neuro-oncology.

Keyword: Brain Tumor, MRI Classification, Convolutional Neural Networks, Vision Transformers, Explainable AI, Grad-CAM.