Automatic Detection and Severity Assessment of Brain Tumors Using Deep Learning Approaches

Summary

The manuscript proposes applying a ResU-Net architecture to 2D MRI slices from the Kaggle LGG Segmentation dataset for brain tumor segmentation and a simple "severity assessment" based on tumor area. The approach is compared to U-Net, FC-DenseNet, and DeepLabv3+ under a uniform training regime, reporting ResU-Net as best on Dice (0.92 test) and tumor area estimation error. The work emphasizes clinical utility via area estimation, supplemented by per-image (tumor vs. no tumor) confusion matrices.

Remark 1 (Introduction)

The introduction provides sufficient background on brain tumors, the role of MRI-based diagnosis, and the development of deep learning models in medical image segmentation. However, it is still not clearly articulated why this study is necessary and which research gaps in previous works the paper aims to address. In addition, the manuscript describes the architectures of several models (U-Net, ResU-Net, FC-DenseNet, DeepLabV3+), but these descriptions are not well connected to the problem the paper focuses on, making the introduction lean more toward listing information rather than establishing a coherent argument.

Remark 2 (Methodology and Experimental Design)

The research methodology is presented in considerable detail, including the segmentation pipeline, training, and evaluation; however, several important technical and experimental issues remain. The use of DPI to estimate tumor area on MRI is not appropriate, as MRI provides voxel spacing information, and 3D volumetric calculations should be based on slices with correct spacing and thickness/gap. Data splitting at the slice level instead of the patient level needs to be adjusted to avoid data leakage and inflated performance metrics. The use of a 2D architecture with 3-channel input also needs clarification, particularly regarding the use of pretrained backbones and potential domain shift. Additionally, the manuscript does not compare with modern 3D models or state-of-the-art baselines such as nnU-Net, Swin UNETR, TransBTS, or new fusion/modal models, and lacks synthetic data augmentation and boundary-aware training methods. Therefore, although the study has potential clinical significance, the current claims regarding quantification and performance are not fully reliable, and methodological corrections (patient-wise splitting, voxel-spacing-based volumetry) are necessary for it to become a valuable applied study.

In summary, this manuscript could be considered for publication provided that the major revisions outlined above are completed.