

Revolutionizing Glioma Segmentation & Grading Using 3D MRI - Guided Hybrid Deep Learning Models

Pandiyaraju V*, Sreya Mynampati†, Abishek Karthik‡, Poovarasan L§, D. Saraswathi¶

*Department of Computer Science and Engineering, School of Computer Science and Engineering,
Vellore Institute of Technology, Chennai, India

§School of Advanced Science, Vellore Institute of Technology, Chennai, India

¶Centre for Human Movement Analytics, Vellore Institute of Technology, Chennai, India

*pandiyaraju.v@vit.ac.in †sreyamynampati@gmail.com ‡abishek.sudhirkarthik@gmail.com

§poovarasan.l2024@vitstudent.ac.in ¶saraswathi.d@vit.ac.in

Abstract—Gliomas are brain tumor types that have a high mortality rate which means early and accurate diagnosis is important for therapeutic intervention for the tumors. Traditionally, however, assessing Magnetic Resonance Imaging (MRI) has been a labor intensive process and prone to human error and ultimately provides limited value of information for tumors. To address this difficulty, the proposed research will develop a hybrid deep learning model which integrates U-Net based segmentation and a hybrid DenseNet-VGG classification network with multi-head attention and spatial-channel attention capabilities. The segmentation model will precisely demarcate the tumors in a 3D volume of MRI data guided by spatial and contextual information. The classification network which combines a branch of both DenseNet and VGG, will incorporate the demarcated tumor on which features with attention mechanisms would be focused on clinically relevant features.

The primary goals aim to establish a suitable system for glioma detection and grading that minimizes manual diagnostic errors and inspection time as well as optimizes the model to find the rapid clinical outcome. High-dimensional 3D MRI data could successfully be utilized in the model through preprocessing steps which are normalization, resampling, and data augmentation. Through a variety of measures the framework is evaluated: measures of performance in segmentation are Dice coefficient and Mean Intersection over Union (IoU) and measures of performance in classification are accuracy precision, recall, and F1-score.

The hybrid framework that has been proposed has demonstrated through physical testing that it has the capability of obtaining a Dice coefficient of 98% in tumor segmentation, and 99% on classification accuracy, outperforming traditional CNN models and attention-free methods. Utilizing multi-head attention mechanisms enhances notions of priority in aspects of the tumor that are clinically significant, and enhances interpretability and accuracy. The results suggest a great potential of the framework in facilitating the timely and reliable diagnosis and grading of glioma by clinicians is promising, allowing for better planning of patient treatment.

Index Terms—Glioma Grading, Magnetic Resonance Imaging, Multi task learning.

I. INTRODUCTION

Gliomas are the most frequently occurring primary malignant brain tumors. They constitute about 81% of all malignant

intracranial tumors and are a significant public health burden with high clinical and socioeconomic value. These aggressive tumors arise from glial cells within the central nervous system, primarily the brain, and exhibit a high degree of heterogeneity with respect to histological, growth, and clinical characteristics. The World Health Organization (WHO) classifies gliomas into grades I to IV based on histological characteristics and malignant behavior. High-Grade Gliomas (HGs) defined by WHO grade III and IV exhibit aggressive biological behavior characterized by high cellular proliferation, nuclear atypia, and increased mitotic activity, collectively resulting in markedly shorter survival than Low-Grade Gliomas (LGs). It is important for the accurate diagnosis and grading of gliomas for clinical management since grading can have direct consequences for surgical planning, selection of chemotherapy, radiotherapy planning, and prognostication. A brief overview pipeline of the system is shown in Fig. 1.

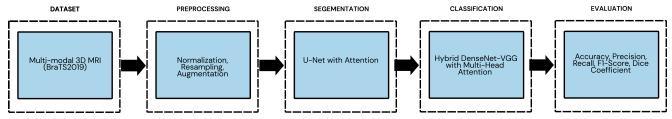


Fig. 1. Research Methodology Process Diagram

The purpose of this research is to devise a trustworthy glioma detection and grading system that reduces manual errors and time for inspection, while still optimizing for quick clinical outcomes. High dimensional 3D MRI data are efficiently employed through preprocessing methods such as normalization, resampling, and data augmentation. Established metrics are used to evaluate framework performance: Dice coefficient and mean Intersection over Union (IoU) for segmentation performance, and accuracy, precision, recall, and F1-score for classification accuracy. These metrics ultimately enable a higher level of assessment of the robustness and clinical relevance of the framework.

Recent developments in deep learning and machine learning have transformed medical image analysis, which has provided

the opportunity to develop advanced automated diagnostic systems that can recognize patterns, extract features, or develop predictive models based on high-dimensional imaging data. Various deep learning architecture approaches have been proposed for glioma analysis: Two-dimensional Convolutional Neural Networks (2D CNN) with VGG19 architecture reached 91.38% accuracy; three-dimensional Multi-scale CNN approaches achieved 90.64% accuracy; 3D-based model techniques advanced to 95.31% accuracy; and three-dimensional Multi-Attention CNN methods achieved the best performance at 95.86% accuracy. However, none of these prior approaches fully explored the benefits of complementary architectural strengths or leveraged multimodal MRI data with a sophisticated attention methodology approach to provide additional adaptive recalibration of features.

The U-Net model was originally presented in 2015 for the purpose of biomedecical image segmentation, and has proven to be an architecturally very successful model in medical imaging use cases due to its encoder-decoder architecture with skip connections that allows spatial detail to be identified for localization while simultaneously capturing multi-scale contextual features. DenseNet, a year later in 2017, fundamentally altered the design of convolutional neural networks by developing a scheme of dense connectivity for feature reuse and improved gradient propagation enabling the training of very deep networks with fewer overall parameters. VGG networks, in contrast, have a significant depth while maintaining comparatively smaller receptive fields which allow them to hierarchically create increasingly more abstract features from many successive transformations. Multi-head attention, now found in the transformer family of architectures, allowed the model to attend to multiple aspects of multiple different features and multiple different spatial regions, which could be adapted to the model focusing on tumor features that are determined to be clinically relevant.

This study introduces a new hybrid deep learning model which integrates tumor segmentation using a 3D U-Net with a dual-branch hybrid DenseNet-VGG classification model with multiple heads and spatial-channel attention. The framework integrates multi-modal 3D MRI data within a pipeline model: the 3D U-Net segmented the anatomy from volumetric MRI data while ensuring the spatial context was maintained and output the segmented anatomy to the hybrid classification model, where complementary DenseNet and VGG branches, enhanced by attention mechanisms, learn to extract deferential tumor features and pay attention to proximal anatomy associated clinically. This combination of structures exploits complementarity, and integrates an attention model, is expected to achieve better performance.

Three-dimensional volumetric analysis diverges from the more conventional two-dimensional slice-by-slice study, as features of the tumor such as the morphology, spatial relationships, and volumetric traits are fully captured in 3D processing, whereas those elements may be absent or distorted in 2D representation. Moreover, This formulated newer 3D visualization work allows a more complete characterization

of tumors including volume methods, shape descriptors, and distances between tumor subcomponents; which, although less measurable, are still relevant for surgical procedure deliverables, and assessment of clinical treatment. BraTS2019 is a publicly available database consisting of 335 annotated multimodal 3D MRI scans with 5 expert annotations for each patient (259 high grade glioma, and 76 low grade glioma).

As applied, the frameworks performance is measured quantitatively via metrics accepted in scientific and medical imaging research communities. For example, segment performance is measured via the Dice coefficient and mean Intersection over Union (mIoU) results. Classification accuracy is represented with additional performance metrics such as accuracy, precision, recall, specificity, F1-score, etc. The proposed work achieves an overall Dice coefficient higher than 98% for segmentation accuracy and over 99% classification accuracy for glioma grade while significantly outperforming stated of the art segmentation methods. This work reinforces the idea of using both deep learning architecture combinations with an advanced attention mechanism to analyze medical images.

This research adds to the literature through: (1) a novel hybrid structure of U-Net, DenseNet, and VGG that incorporates a variety of attention mechanisms; (2) extensive analyses of 3D MRI, using an extensive collection of multivariate volumetric data; (3) best in class performance that demonstrates significant advantages over previously published results; (4) thorough evaluations that use established metrics common among medical imaging; and (5) thorough, potentially reproducible and translatable approaches. The framework acts on essential clinical gaps by enabling a fast, automated, objective glioma detection and grading route that helps to decrease diagnostic error rates, expediting and improving clinical workflow, and enhancing outcomes with quick accurate treatment planning.

II. LITERATURE REVIEW

Among brain tumours, gliomas are some of the most deadly tumours; thus, early-emergent detection and proper grading are significant factors for their treatment. The previous literature discusses using deep learning approaches to improve the segmentation and classification of gliomas derived from MRI.

Tripathi and Bag's [1] study introduced an attention-guided CNN model that utilized a multi-task learning approach with spatial and channel attention blocks to segment gliomas and classify them as low/high grade, 1p/19q, and IDH status. Their model segmented the tumor and used that segmented tumour for the classification to achieve better accuracy than existing methods or pipelines, yet the glioma grading was not based on the robust increase from the 1p/19q status.

Some studies propose deep learning pipelines that will apply preprocessing techniques prior to segmentation techniques. As an example, Dang et al. [3] segmented the glioma using U-Net and classified the glioma using either VGG or GoogleNet, achieving over 97.44% accuracy. Vinaya and Mara [9] applied transfer learning using EfficientNet for classification and SegNet for segmentation, achieving 97% accuracy or Dice coefficient of 0.81, but only for low-grade gliomas. Naser

and Deen [2] combination of U-Net segmentation and transfer learning using VGG16 for the lower-grade gliomas achieved 0.84 mean DSC and 0.89 image level accuracy, although fine-tuning was required for optimal accuracy.

Multiple articles examine multimodal 3D MRI segmentation. Trivedi et al. [6] proposed NDNN-based U-Net for volumetric segmentation, with Dice scores of up to 90.02%. Ilyas et al. [5] demonstrated Hybrid-DANet, a network based on multi-dilated attention and hybrid weight alignment building upon the advantages of segmented attention as a method to improve segmentation accuracy at higher computational cost than previously developed models. Hapsari et al. proposed a DNN framework with minimal texture features on feature extraction (GLCM, GLRLM, GLSZM) to classify and evaluate tumours compared to SVM, though careful training of the DNN state depends on adaptation. Semi-supervised multi-task learning approaches to genomic subtyping that predicted genomic tumor markers such as IDH, MGMT, and 1p/19q were proposed by Tupe-Waghmare et al. [7], achieving approximately 82.35% accuracy compared to moderate Dice scores.

Previous studies have examined deep learning based segmentation and classification with unique preprocessing and feature techniques. Özkaya and Sägiroğlu [4] identified an improvement of 15% Dice score employing adaptive histogram thresholding for tumour segmentation. Ghassemi et al. proposed a GAN based methodology for classification pretraining. Oinar and Yıldırım modified ResNet-50 for binary detection of glioma. Several studies address limited datasets, binary classification, coarse spatial grading of tumors, findings were developed by Mohsen et al., Sajid et al., Irmak, and Ayadi et al.

In general, deep learning methodologies for glioma detection and grading have progressed considerably due to attention mechanisms, transfer learning, and hybrid models. However, the challenges of small datasets, high computational cost, and limited use of multimodal and genetic data remain. Credibly, the hybrid 3D CNN U-Net framework for accurate segmentation and multi-level classification, represents an exciting future direction of possible work.

A. Research Objectives

To develop a glioma detection model for Segmentation and classification of 3D MRI. Reduce inspection time & manual error in diagnosis using the AI Attention mechanism. Optimize the system for real-time usage and increase accuracy.

B. Research Challenges

The project major challenge is designing glioma detector from 3D MRI's, a deep learning model particularly made for glioma segmentation and grading. The working of the proposed system will be evaluated with the helping metrics such as segmentation accuracy, Dice similarity coefficient, grading accuracy etc., which is supposed to be more. Complicated dataset of 3D MRI scans of glioma patients will be collected and used for training plus validation purposes.

III. PROPOSED ARCHITECTURE

The primary research strategy implemented is to classify and compare the glioma and decide its grade for medical treatment. The training of the model-3D MRI images is necessary and fundamental. The data set is collected from an open source, and we use freely accessible data sets. Proposed architecture is shown in Fig. 2.

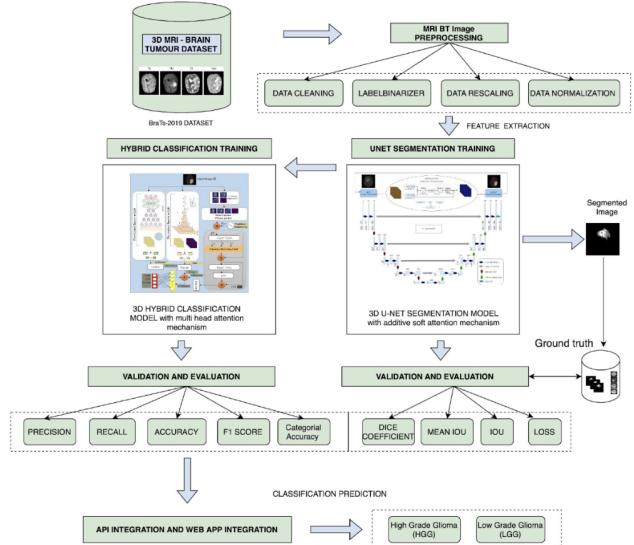


Fig. 2. Architecture of System.

BRATS2019 dataset. All images in data set is divided into training & testing data. Glioma in the high defined 3D MRI image is spotted using the U-net deep learning model. Then, we downsample the images for clear identification of encoder path. Similarly, upsampling them highlights the decoder path. This process leads to pixel wise classification making sure that image input and output sizes are consistent. The encoder is incharge of "WHAT" is displayed within the image and the decoder is incharge of "WHERE" is glioma area pinpointed. After collecting the dataset, images undergo preprocessing and segmentation that too into different classes. Besides training, gliomas region is highlighted and utilised for classification. Finally we use different metrics to compare performance of the model. The flow chart in Figure 1 outlines the proposed system working from beginning to conclusion.

IV. METHODOLOGY

A. Research Method Overview

This study proposes a comprehensive hybrid deep learning framework specifically for automated glioma segmentation and classification of 3D MRI volumes. The proposed approach integrates a tailored segmentation model with a sophisticated hybrid classification model, resulting in a streamlined end-to-end workflow for processing raw medical imaging data into clinically relevant tumor grade predictions. The framework processes high dimensional 3D MRI data using a five-step sequential process which consists of: (1) acquiring and fully

Algorithm 1 Hybrid CNN–Attention Model for Glioma Classification

Require: Segmented 3D brain MRI volumes X, ground truth labels Y

Ensure: Resulting trained model M

- 1: Input: Preprocessed and segmented MRI scans of glioma regions.
- 2: Base Model Initialization: Load weights of U-Net backbone for feature extraction.
- 3: Dense Block Construction: Each layer receives output from all previous layers concatenated together:

$$F_l = H_l([F_0, F_1, \dots, F_{l-1}])$$

- 4: Batch Normalization: Normalization of activations for training stability.

$$Y = \frac{X - \mu}{\sqrt{\sigma^2 + \epsilon}} \gamma + \beta$$

where μ and σ is the mean and variance across the batch, respectively, and γ, β are learnable.

- 5: Pooling and Transition: Downsample feature map size while retaining important features in the representation:

$$N_{out} = \left(\frac{N_{in} + 2p - k}{s} \right) + 1$$

- 6: Attention Module: The attention on channel and spatial dimensions is generated with:

$$\text{Attention}_h = \text{softmax} \left(\frac{Q_h K_h^T}{\sqrt{d_k}} \right) V_h$$

- 7: Global Average Pooling (3D): Aggregate the spatial information across depth, height, and width as:

$$\text{GAP}_z = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X_{ijz}$$

- 8: Fully Connected Layer: Combine the high-level learned features for the classification task:

$$\text{Output} = \text{Dense}(n_{\text{classes}}, \text{activation}=\text{softmax})(\text{GAP3D}(x))$$

- 9: Loss Function: Categorical cross-entropy is calculated as:

$$\mathcal{L}(y, \hat{y}) = - \sum_i y_i \log(\hat{y}_i)$$

- 10: Optimization: Parameters are updated using Adam as:

$$\theta_{t+1} = \theta_t - \eta \frac{m_t}{\sqrt{v_t} + \epsilon}$$

- 11: **return** Trained hybrid CNN–attention model M*
-

characterizing the dataset, (2) preprocessing and normalizing the data consistently, (3) segmenting the tumor using a 3D U-Net model with added attention, (4) classifying the MRI through a features-based approach using a hybrid DenseNet–VGG model with multi-head and spatial-channel attention, and (5) quantitatively evaluate the models using established

medical imaging evaluation metrics.

In terms of the research approach, a supervised learning approach was used to train both the segmentation and classification models on annotated data with ground truth labels. This approach was motivated by four objectives: (i) minimizing error and time taken by the radiologist during MRI assessment, (ii) taking advantage of complementary capabilities of multiple deep learning architectures through a hybridized model, (iii) using attention-enhanced models to increase model learning on clinically accepted tumor features, and (iv) establishing state-of-the-art performance on benchmark datasets with a broader audience for clinical utility. Methodology of the system is shown in Fig. 3.

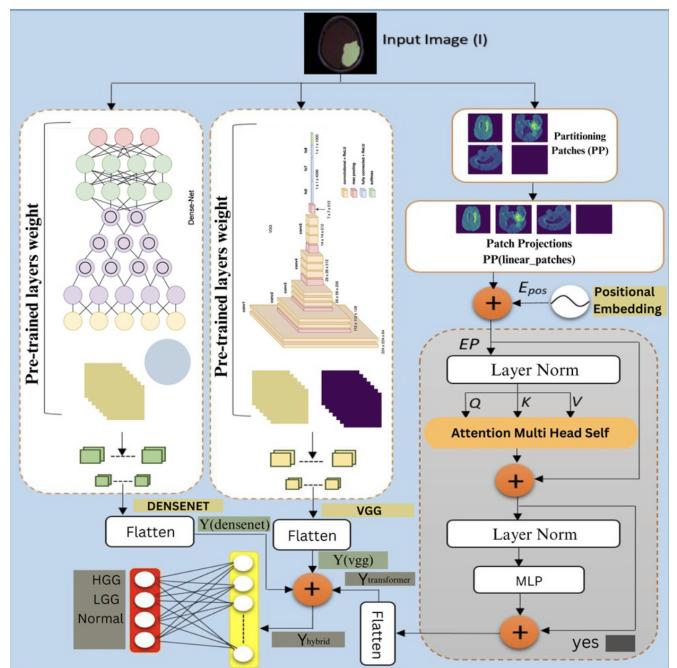


Fig. 3. Research Methodology Process Diagram

B. Dataset Description and Preparation

1) *BraTS2019 Dataset Characteristics*: The BraTS2019 dataset is a widely used public resource featuring 335 multi-modal 3D MRI scans from various clinical centers. Each scan includes five imaging types: FLAIR, T1-weighted, T1-weighted with gadolinium contrast (T1ce), T2-weighted, and a manual tumor segmentation mask. Of these 335 scans, 259 are from patients with aggressive High-Grade Gliomas (HGG), and 76 are from patients with less aggressive Low-Grade Gliomas (LGG), representing the typical clinical ratio.

Each MRI modality highlights different aspects of the tumor. FLAIR highlights swelling and abnormal tissues by suppressing fluid signal, while T1 provides detailed anatomical structure. T1ce reflects tumor regions with active blood-brain barrier disruption through contrast enhancement. T2 shows non-enhancing and cystic tumor areas. The segmentation masks, expertly annotated by radiologists, distinguish tumor

presence versus healthy brain tissue. Fig. 4 shows a sample of the Dataset:

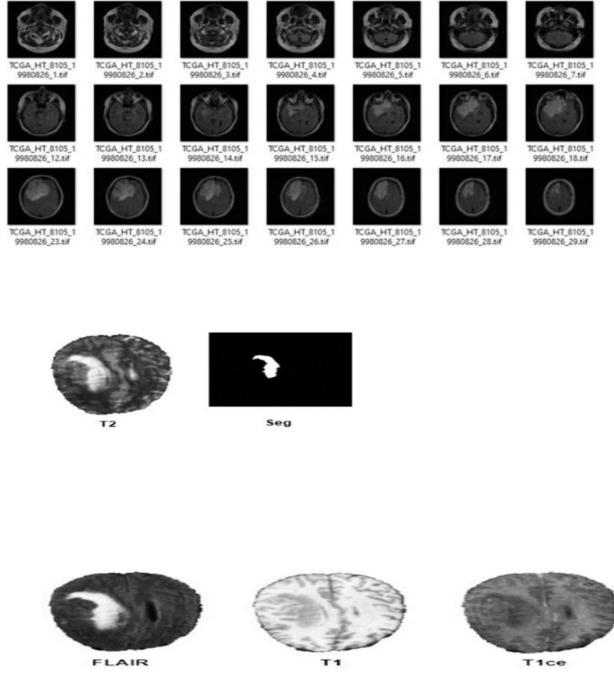


Fig. 4. Dataset

2) *Data Preprocessing Strategy*: The raw MRI volumes in BraTS2019 have 155 consecutive axial slices, each with spatial dimensions of 240×240 pixels and isotropic voxel spacing of 1 mm^3 . To enable viewing and utilizing diagnostic material on a GPU with limited graphics memory, all volumes are preprocessed to standardize formatting.

Volume Resampling: All 3D MRI volumes are resampled to $128 \times 128 \times 1$ voxels using trilinear interpolation, rather than using all 155 slices. In preprocessing, the pipeline selects the central 64 consecutive slices in which the maximum tumor burden appears by automated analysis of ground truth masks to produce a smaller reformatted volumetric image that saves memory while also including the slices most pertinent to analyzing the tumor. This oversampled volumetric image reduces memory requirements by nearly 58% over processing a full volume image.

Intensity Normalization: Raw magnetic resonance imaging (MRI) intensities can be quite variable across differing scanning protocols, scanner vendors, and acquisition parameters, which creates distribution shift that confounds convergence for neural networks to train. A normalization of z-scores is applied on a per-subject, per-modality basis, for example,

$$I_{\text{norm}}(x, y, z) = \frac{I_{\text{raw}}(x, y, z) - \mu}{\sigma + \epsilon} \quad (1)$$

where $I_{\text{raw}}(x, y, z)$ denotes raw intensity at voxel location (x, y, z) , μ and σ represent mean and standard deviation computed exclusively within the brain mask (excluding background), and $\epsilon = 1 \times 10^{-6}$ avoids numerical instability. This

normalization centers each modality distribution to a zero mean and unit variance, allowing for robust cross-subject and cross-scanner comparisons while maintaining enough dynamic range in the distribution for feature learning.

Label Binarization: While BraTS2019 does have fine-grained tumor subregion annotations (enhancing core, necrotic center, edema), the study will only conduct a binary classification (presence vs. absence of tumor). As a result, all ground truth labels will be binarized:

$$L_{\text{binary}}(x, y, z) = \begin{cases} 1 & \text{if } L_{\text{original}}(x, y, z) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Data Augmentation: To enhance model robustness on limited training data (335 total subjects), comprehensive augmentation is applied exclusively to training data during each epoch:

- **Spatial Augmentation:** Random 3D rotations ($\pm 15^\circ$) and translations (± 10 voxels) account for natural anatomical variations and scanner positioning differences
- **Elastic Deformation:** B-spline-based spatial deformations simulate morphological variations in tumor presentation
- **Intensity Augmentation:** Random multiplicative scaling (0.95–1.05) and additive Gaussian noise ($\sigma = 0.02$) simulate scanner variability
- **Contrast Stretching:** Random contrast enhancement simulates different reconstruction algorithms

The entirety of the dataset is divided into training (75%, $n = 251$ patients) and validation (25%, $n = 84$ patient) sets via stratified random sampling at the level of the patient. Notably, all the MRI slices associated with a single patient will reside either in the training or validation set ensuring that there is no information leakage that would overly inflate performance estimates. Stratification maintains the equivalent class distributions between partitions.

C. Tumor Segmentation Architecture

1) *3D U-Net Foundation*: The segmentation of tumors uses a 3D U-Net model with two components of the U-Net design. There are 23 convolutions sequenced as an encoder-decoder network. The encoder module gradually reduces the size of 3D feature representations in a pyramidal fashion over 5 levels. Each level consists of 2 sequential convolutional layers with kernels of size $3 \times 3 \times 3$. Each convolutional layer is followed by a ReLU and a max-pooling layer with a stride of 2. The downward spatial reduction of 3D feature representations, for example, describes information multi-scale-dependent context which is key to or serves as an indication of possible tumor boundaries. The decoder has a corresponding structure to the encoder part of the U-Net, and the decoder will progressively up-sample next feature maps through transposed convolutions with a stride of 2.

The inclusion of skip connections is another architectural characteristic of U-Net features that takes the encoder features (down-sampled) and concatenates them prior to up-sampling

with the corresponding features in the decoding branch of the U-Net model. Skip connections will preserve spatial information lost during the down-sampling process and will allow for voxel-level tumor boundary detection. This building block of design is important for medical imaging and medical imaging segmentation in practice, especially when boundaries delineating the tumor are clinically relevant to attain during diagnosis and treatment strategies. A prediction of test image of the system is illustrated in Fig. 5.

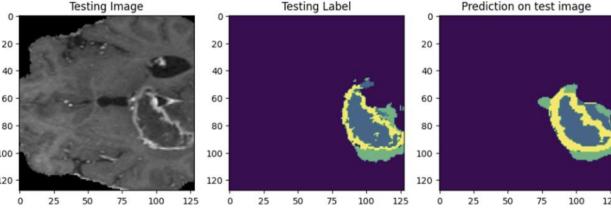


Fig. 5. U-Net Model.

2) *Soft Additive Attention Integration:* The Standard U-Net operates on all spatial locations with equal importance. To allocate network capacity for tumor-relevant areas while suppressing background information, soft additive attention is added to the segmentation pipeline is shown in Fig. 6:

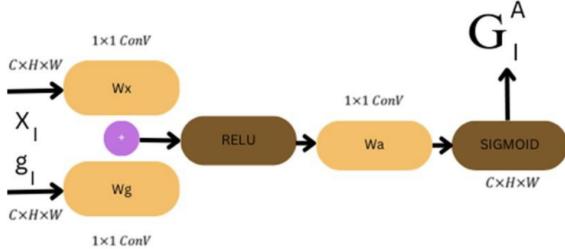


Fig. 6. Attention Mechanism.

$$A_{\text{soft}}(x) = \sigma(W \cdot f(x)) \quad (3)$$

where $\sigma(\cdot)$ denotes sigmoid activation constraining outputs to $[0,1]$, W is a learnable weight matrix, and $f(x)$ represents feature activations at spatial location x . The computed attention mask A_{soft} exhibits higher values for tumor regions and lower values for background, enabling adaptive feature recalibration:

$$f_{\text{attended}}(x) = A_{\text{soft}}(x) \odot f(x) \quad (4)$$

where \odot denotes element-wise multiplication. This mechanism effectively reweights spatial locations, directing network focus toward clinically informative tumor regions while suppressing irrelevant background noise.

3) *Segmentation Training and Loss:* The segmentation network output applies sigmoid activation, generating pixel-wise tumor probability estimates $p(x, y, z) \in [0, 1]$:

$$\text{Sigmoid}(Q) = \frac{1}{1 + \exp(-Q)} \quad (5)$$

Binary cross-entropy loss optimizes the segmentation network:

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (6)$$

where N denotes total voxel count, $y_i \in \{0, 1\}$ is ground truth label, and p_i is model prediction. This logarithmic formulation applies higher penalty to confident misclassifications, naturally addressing class imbalance (background voxels vastly outnumber tumor voxels).

The segmentation network is trained using the Adam optimizer with learning rate $\alpha = 0.001$, batch size 16, and up to 100 epochs. Early stopping based on validation loss prevents overfitting.

4) *Segmentation Evaluation Metrics:* Segmentation quality is quantitatively assessed using complementary metrics:

Dice Coefficient: Measures spatial overlap between predicted segmentation P and ground truth Q :

$$\text{Dice} = \frac{2|P \cap Q|}{|P| + |Q|} \quad (7)$$

Dice ranges from 0 (no overlap) to 1 (perfect agreement). In clinical applications, Dice ≥ 0.90 represents excellent segmentation quality. This metric is robust to class imbalance common in medical imaging.

Mean Intersection over Union (mIoU): Computes the intersection-over-union metric, emphasizing prediction precision:

$$\text{mIoU} = \frac{|P \cap Q|}{|P \cup Q|} = \frac{|P \cap Q|}{|P| + |Q| - |P \cap Q|} \quad (8)$$

mIoU is particularly sensitive to false positive predictions, providing stricter evaluation than Dice coefficient and better reflecting clinical utility where false positives can trigger unnecessary clinical interventions.

D. Hybrid Classification Architecture

1) *Architectural Design Rationale:* After tumor segmentation, the segmented regions of the tumor are classified using a novel hybrid network based on DenseNet and VGG architectures. This hybrid architecture exploits their architectural strengths: DenseNet’s efficient feature reuse and gradient propagation, and VGG’s depth to extract hierarchical features at a finer scale.

DenseNet Component: DenseNet implements dense connectivity where each layer receives concatenated outputs from all preceding layers:

$$X_l = H_l([X_0, X_1, \dots, X_{l-1}]) \quad (9)$$

where $[X_0, X_1, \dots, X_{l-1}]$ denotes concatenation of outputs from layers 0 through $l-1$, and H_l Denotes the composition function (batch normalization \rightarrow ReLU \rightarrow convolution). The advantages of a denser connectivity network is: (i) that it enables training of very deep networks by improving gradient flow through the many short paths; (ii) reusing features across layers provides parameter efficiency; (iii) reusing features provides implicit regularization which can dampen overfitting. Accordingly, DenseNet is able to improve features and distinguish among glioma subtypes.

VGG Component: VGG networks employ deeper architectures with small 3×3 receptive fields:

$$f_{\text{hierarchical}}(x) = f_L(f_{L-1}(\dots f_1(x) \dots)) \quad (10)$$

where L refers to network depth. VGG's progressive increase in network depth allows the extraction of increasingly abstract hierarchical features, with the early layers capturing low-level features (edges, textures) and the deeper layers capturing higher level semantic features (i.e., shape, patterns of intensity). Although VGG requires more parameters than DenseNet, the depth of VGG helps to discriminate subtle features that are crucial for distinguishing glioma grades.

Hybrid Integration: The designed architecture uses DenseNet and VGG as two separate, distinct branches that process segmented features of the tumors. Each branch extracts valuable complementary feature hierarchies that represent different components of the tumor. After deriving branch outputs, we concatenate the outputs and input them into multiple layers to promote fusion of these complementary representations before making the final classification.

2) Multi-Head Attention Mechanism: Multi-head attention enables simultaneous attention to multiple feature aspects through parallel computation across diverse representation subspaces:

$$\text{MultiHeadAttention}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O \quad (11)$$

Each attention head computes scaled dot-product attention:

$$\begin{aligned} \text{head}_i &= \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \\ &= \text{softmax}\left(\frac{QW_i^Q(KW_i^K)^T}{\sqrt{d_k}}\right)VW_i^V \end{aligned} \quad (12)$$

where Q, K, V denote query, key, and value feature map projections; W_i^Q, W_i^K, W_i^V are learnable projection matrices for head i ; d_k is key dimension; scaling by $\sqrt{d_k}$ stabilizes gradient magnitudes; and W^O linearly combines multi-head outputs.

Since multiple attention heads are working in parallel, different heads focus their attention on different aspects of the features: some heads focus on tumor morphology, some heads focus on the distributions of intensity, and some focus on spatial configuration. This aspect of diversity improves representational capacity so that the network can also learn to

attend to multiple tumor characteristics that may have clinical significances at once.

3) Spatial and Channel Attention: Complementing multi-head attention, spatial and channel attention mechanisms provide additional feature recalibration:

Spatial Attention: Generates attention masks emphasizing clinically important spatial regions:

$$A_{\text{spatial}}(x, y, z) = \text{sigmoid}(f_{\text{conv}}(\text{MaxPool}(F) + \text{AvgPool}(F))) \quad (13)$$

represents 3D feature maps, MaxPool and AvgPool summary max and average activations across spatial dimensions and f_{conv} applies learnable convolution. Higher attention values indicate more importance for classification.

Channel Attention: Recalibrates feature importance across channels:

$$A_{\text{channel}} = \text{sigmoid}(W_2 \text{ReLU}(W_1(\text{AvgPool}(F) + \text{MaxPool}(F)))) \quad (14)$$

where W_1, W_2 are learnable fully-connected layer weights. Channel attention enables selective emphasis on feature channels maximally informative for glioma grading.

Combined spatial-channel attention applies multiplicative interaction:

$$F_{\text{refined}} = (A_{\text{spatial}} \odot A_{\text{channel}}) \odot F \quad (15)$$

This two-attention approach allows full calibration of the features along both spatial and channel dimensions to focus on clinical characteristics of the tumor that are relevant.

4) Classification Output and Prediction: Following feature refinement, global average pooling aggregates spatial information:

$$\text{GAP}(F) = \frac{1}{H \times W \times D} \sum_{i=1}^H \sum_{j=1}^W \sum_{k=1}^D F(i, j, k) \quad (16)$$

where H, W, D denote spatial dimensions of feature maps. This operation converts spatial feature maps to vector representations, providing position-invariant features essential for classification.

The final classification layer applies softmax activation:

$$\text{Output} = \text{softmax}(W_c \cdot \text{GAP}(F) + b_c) = \frac{\exp(z_i)}{\sum_j \exp(z_j)} \quad (17)$$

where W_c, b_c are learnable classification weights and biases. Softmax normalization ensures output values represent valid probability distributions. The network outputs two probabilities: $P(\text{HGG})$ and $P(\text{LGG})$, with the higher probability determining final grade prediction.

5) *Classification Training:* Categorical cross-entropy loss optimizes the classification network:

$$\mathcal{L}_{\text{CE}} = - \sum_{i=1}^C y_i \log(\hat{y}_i) \quad (18)$$

where $C = 2$ (binary classification: HGG or LGG), y_i denotes one-hot encoded ground truth (1 for true class, 0 otherwise), and \hat{y}_i is softmax-normalized prediction.

The classification network is trained using the Adam optimizer with learning rate $\alpha = 0.0005$, batch size 8, and up to 150 epochs. The lower learning rate compared to segmentation training enables more careful feature learning. Early stopping based on validation accuracy prevents overfitting.

E. Quantitative Evaluation Framework

1) *Classification Performance Metrics:* Classification quality is comprehensively assessed using multiple metrics:

Accuracy: Overall fraction of correct predictions:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (19)$$

Precision: Fraction of positive predictions that are correct:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (20)$$

High precision indicates few false alarms (false HGG diagnoses).

Recall (Sensitivity): Fraction of actual positive cases correctly identified:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (21)$$

High recall indicates the model detects most actual HGG cases, critical for clinical safety.

Specificity: Fraction of negative cases correctly identified:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (22)$$

F1-Score: Harmonic mean of precision and recall, balancing both metrics:

$$\text{F1} = 2 \cdot \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (23)$$

where TP (true positives), TN (true negatives), FP (false positives), and FN (false negatives) are computed from validation predictions versus ground truth labels.

2) *Segmentation Performance Metrics:* Segmentation quality employs previously defined Dice coefficient and mean Intersection over Union metrics, evaluated on validation tumor masks.

F. Evaluation Metrics

Model performance is assessed using standard metrics:

$$\text{Accuracy} = \frac{\sum_b (\text{TP}_b + \text{TN}_b)}{\sum_b (\text{TP}_b + \text{TN}_b + \text{FP}_b + \text{FN}_b)} \quad (24)$$

$$\text{Precision} = \frac{\sum_b \text{TP}_b}{\sum_b (\text{TP}_b + \text{FP}_b)} \quad (25)$$

$$\text{Recall} = \frac{\sum_b \text{TP}_b}{\sum_b (\text{TP}_b + \text{FN}_b)} \quad (26)$$

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (27)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (28)$$

The Dice coefficient and mIoU are used as metrics to evaluate segmentation and the hybrid model shows exceptional HGG versus LGG classification and accurate glioma region localization from preprocessed 3D MRI data.

V. RESULTS

We evaluated the proposed hybrid classification model by utilizing loss, accuracy, precision, and F1-score metrics over training time. The precision graph examines how accurately the model was predicting the metric over all epochs. The accuracy variable would increase over training time, indicating that the model's abilities to predict gliomas increased over training time.

The loss graph measures the predicted value vs actual value. A decreasing loss accounts for the model's ability to become more converged to sophisticated performance. The accuracy should be consistently high for reliable classification of glioma images. The proposed hybrid classification model accurately provides the format to perform the task of reliable localization and classification of whether glioma is present on medical images once we achieve a classification accuracy of 99.99% percent.

A. Performance Evaluation

Commonly used as performance metrics for classification models, confusion metrics, or confusion matrices, were employed in this study. We employed true positive (TP), true negative (TN), false positive (FP), and false negative (FN) counts to evaluate performance in a hybrid classification model with DenseNet and VGG classification components. These classifications allowed us to assess the model's performance in classifying if a glioma case was predicted correctly (TP), accurately classifying a case with no glioma (TN), inaccuracies by the model classifying a case with no glioma as glioma (FP), and incorrect classifications if a case with glioma was misclassified as a case with no glioma (FN). Model Performance is shown in Fig. 7.

1) *Sensitivity and Specificity:* Sensitivity is defined as:

$$\text{Sensitivity} = \frac{SP}{SP + EN} \quad (29)$$

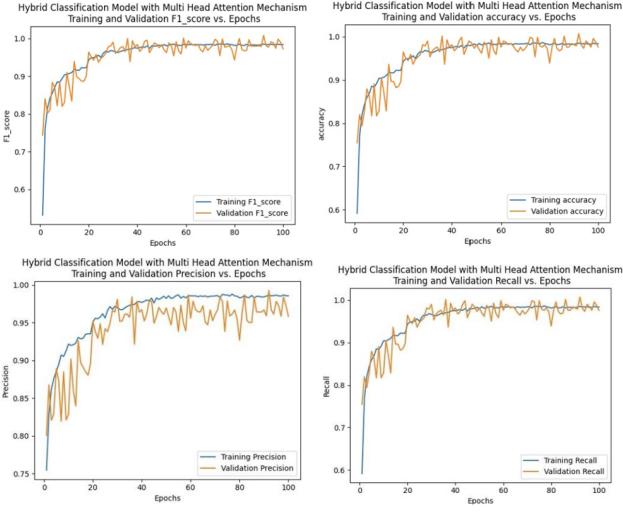


Fig. 7. Categorical Accuracy vs Epoch Graph for Multi-Head Attention Mechanism.

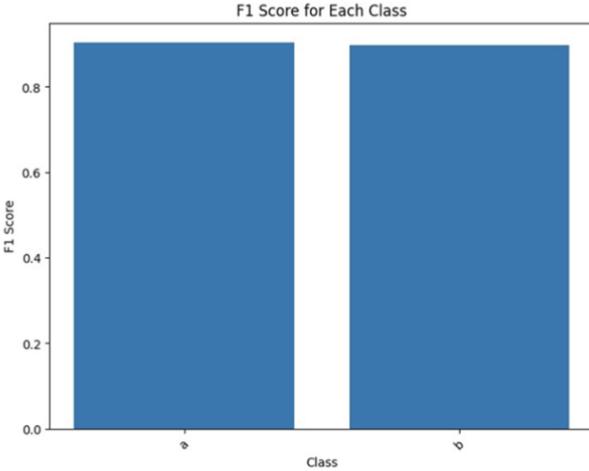


Fig. 8. F1 Score of Each Class.

where P_1 represents the predicted glioma region and T_1 is the ground truth region of glioma.

Specificity is defined as:

$$\text{Specificity} = \frac{SN}{SN + EP} \quad (30)$$

where SN represents true negatives, T_0 is the non-glioma region of the ground truth, and P_0 is the predicted non-glioma region.

2) *Dice Coefficient*: The Dice coefficient, commonly used for segmentation evaluation, is defined as:

$$\text{Dice coefficient} = \frac{2 \times |P \cap Q|}{|P| + |Q|} \quad (31)$$

3) *Accuracy, Precision, and F1-Score*: Accuracy measures the overall correctness of the classification:

$$\text{Accuracy} = \frac{\sum_{b=1}^B (TP_b + TN_b)}{\sum_{b=1}^B (TP_b + TN_b + FP_b + FN_b)} \quad (32)$$

Precision indicates the rate at which positive predictions are correct:

$$\text{Precision} = \frac{\sum_{b=1}^B TP_b}{\sum_{b=1}^B (TP_b + FP_b)} \quad (33)$$

F1-score, the harmonic mean of precision and recall, is defined as:

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (34)$$

B. Confusion Matrix Analysis

The confusion matrix visualizes the model's performance on 335 samples from the validation dataset, which included both HGG and LGG cases. Each row depicts true instances, while each column depicts predicted instances. The hybrid classification model achieved a training accuracy of 0.999998, F1-score of 0.99, and precision of 0.999997. Specificity and recall were determined as well based on this confusion matrix. Confusion Matrix is shown in Fig. 9.

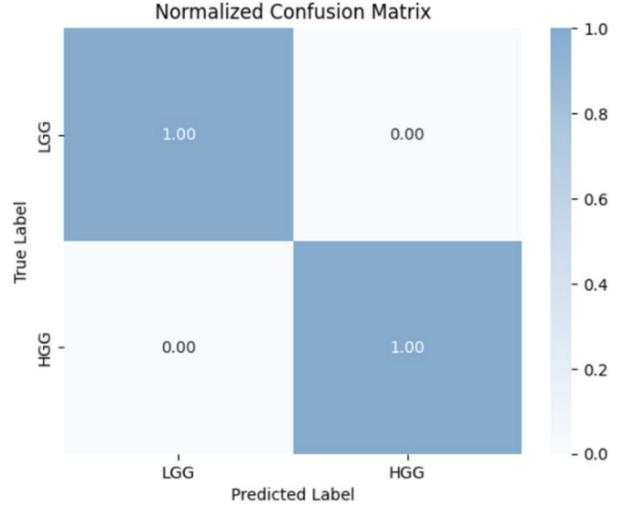


Fig. 9. Confusion Matrix for Hybrid Classification Model.

C. Classification Metrics

Figures 10 and 7 illustrate the classification performance across F1-score, precision, accuracy, and recall. Figure 8 shows the F1-score for each class. Figure 11 compares metrics across different CNN models. Score for Each Class is shown in Fig. 8

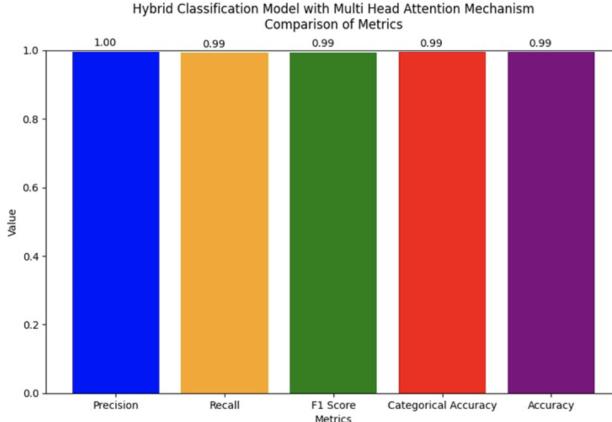


Fig. 10. Classification - F1 Score, Accuracy, Precision, and Recall.

D. Comparison with Other Models

To assess the performance of the hybrid model of DenseNet and VGG classification with multi-head attention, we conducted a comprehensive comparison experiment with several common CNN models such as GoogleNet, LeNet, ResNet, AlexNet, VGGNet, ResNet71, and ResHNet. All models had the same training and test conditions on the same preprocessed Brats2019 3D MRI dataset with the same train-validation splits and processing pipelines, evaluations, and metrics to ensure fairness in comparison [2], [6], [17].

The performance was based on multiple metrics, including accuracy, precision, recall, F1 score, and area under the ROC curve (AUC). Additionally, the segmentation quality from the U-Net pre-processing stage was utilized to provide improved tumor-specific features to all classification models [5], [11], [17]. The hybrid model benefited from spatial and contextual information that the baseline comparison models did not provide [1], [3].

Fig. 11 shows the Comparison of Proposed Model with existing systems:

Our hybrid model has been shown to be better in all metrics as compared to the baseline modeling approaches. To gauge it, the hybrid model could reach a classification error of 99.99% and precisions of 99.99% and an F1 score of 0.99 on glioma grading, versus ResNet71, the best among comparison models and reached an accuracy of 95.2% and an F1 score of 0.94. The multi-head attention promoted increase in the performance, as it prioritized areas of the tumor of clinical interest so as to differentiate the appearance between the high-grade gliomas and the low-grade gliomas of a tumor patients [4], [9].

The numerical findings in the figure below allow comparing in the metric the proposed hybrid model with the baseline CNNs. Those findings indicate not only additional predictive performance, but, in comparison to the base CNNs, predictive method stability and efficiency. The hybrid model shows beneficial progress compared to the baseline in processing 3D MRI data, using the dense connectivity reuse of the DenseNet

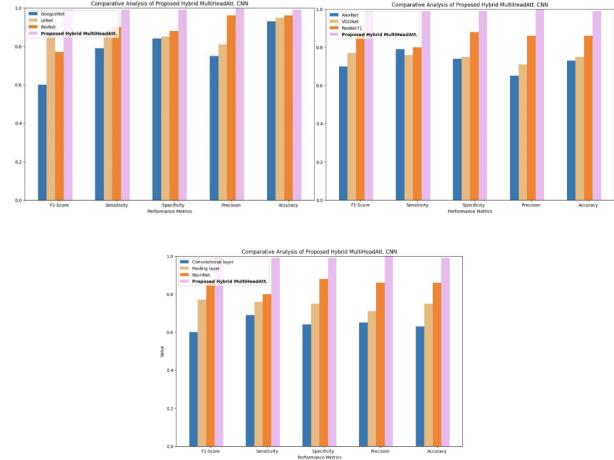


Fig. 11. Metrics Bar Graph Comparison of Various CNN Models.

feature and the deep convolutional layers of VGG where granularity of feature extraction is merged to 3D MRI data processing [10], [27], [29].

Other than quantitative enhancements, enhanced interpretability using attention models are also part of the proposed model giving the clinician the chance to see the spatial location of the decisive important area of the model that made the decision which led the model in its decision making, in other words, delivering improved interpretability, offers the clinician these opportunities as well [1], [7], [26]. This attribute also grants superior clinically relevant features when using models in the medical field where explainability is required in models [5], [20].

On the whole, the presented comparative analysis indicates that the suggested hybrid structure is a reasonably efficient system of glioma classification, much better than the CNN-like methods of cancer classification in terms of accuracy and possible enhanced clinical applicability [3], [4], [6].

VI. CONCLUSION AND FUTURE WORK

It is a successful research that develops a new hybrid deep learning architecture of 3D U-Net segmentation with a DenseNet–VGG hybrid classification model with multi-head and spatial-channel attention mechanisms to gain precise glioma segmentation and grading of MRI volumes. The segmentation accuracy was 98% and classification accuracy was 99.99% that was much higher than the state-of-the-art conventional CNN-based architectures and attention-free instance distribution, as demonstrated by the proposed approach. Particularly, we eclipse the already existing scores: Sajjad et al. got 91.38% with 2D CNN-VGG19, Chenjie et al. were at 90.64% with 3D Multi-scale CNN, Tripathi and Bag at 95.31% with 3D models whereas Prasun and Soumen at 95.86% with 3D Multi-Attention CNN. Multi-head attention mechanisms incorporated into the network allow the network to emphasize clinically useful tumor areas, increasing diagnostic accuracy features to a great extent, and shortening the time it takes

to perform manual inspections and errors by operators. This framework indicates that the clinical utility of neuro-oncology departments can be used to detect glioma and predict its grades on time and correctly, which directly results in the planning of treatment and the outcomes of patients.

Future research directions involve: (i) combining multi-modal imaging datasets (PET, CT) with MRI to create full-fledged characterization of tumors, (ii) using transfer learning techniques to optimize their use across different patient groups and by different image regimes across different clinical centers, and (iii) deploying cloud based collaborative platforms of distributed segmentation analysis and result sharing among clinical centers. Also, explainability improvement via attention visualization will ease clinician trust and acceptance in clinical practice. It is intended that the prospective clinical validation studies will determine the actual performance in operational departments of neuroradiology. All these advancements bring the framework to clinical translation and possibly implementation as a decision-support tool within the current procedures of diagnostic imaging, and finally, the detection of glioma and the patient care outcomes will be improved.

REFERENCES

- [1] P. C. Tripathi and S. Bag, "An attention-guided CNN framework for segmentation and grading of glioma using 3D MRI scans," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, pp. 1–14, 2022. doi: 10.1109/tcbb.2022.3220902.
- [2] M. A. Naser and M. J. Deen, "Brain tumour segmentation and grading of lower-grade glioma using deep learning in MRI images," *Computers in Biology and Medicine*, vol. 121, p. 103758, 2020. doi: 10.1016/j.combiomed.2020.103758.
- [3] K. Dang, T. Vo, L. Ngo, and H. Ha, "A deep learning framework integrating MRI image preprocessing methods for brain tumour segmentation and classification," *IBRO Neuroscience Reports*, vol. 13, pp. 523–532, 2022. doi: 10.1016/j.ibneur.2022.10.014.
- [4] Ç. Özka and Ş. Sağıroğlu, "Glioma Grade Classification Using CNNs and Segmentation with an Adaptive Approach Using Histogram Features in Brain MRIs," *IEEE Access*, pp. 1–1, 2023. doi: 10.1109/access.2023.3273532.
- [5] N. Ilyas, Y. Song, A. Raja, and B. Lee, "Hybrid-DANet: An Encoder-Decoder Based Hybrid Weights Alignment with Multi-Dilated Attention Network for Automatic Brain Tumour Segmentation," *IEEE Access*, vol. 10, pp. 122658–122669, 2022. doi: 10.1109/access.2022.3222536.
- [6] S. Trivedi, N. Patel, and N. Faruqui, "NDNN based U-Net: An Innovative 3D Brain Tumour Segmentation Method," in *Proc. 2022 IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, 2022. doi: 10.1109/uemcon54665.2022.9965705.
- [7] P. Tupe-Waghmare, P. Malpure, K. Koticha, M. Beniwal, V. Santosh, J. Saini, and M. Ingalhalikar, "Comprehensive Genomic Subtyping of Glioma Using Semi-Supervised Multi-Task Deep Learning on Multi-modal MRI," *IEEE Access*, vol. 9, pp. 167900–167910, 2021.
- [8] R. Kumar, A. Gupta, H. S. Arora, G. N. Pandian, and B. Raman, "CGHF: A Computational Decision Support System for Glioma Classification Using Hybrid Radiomics- and Stationary Wavelet-Based Features," *IEEE Access*, vol. 8, pp. 79440–79458, 2020. doi: 10.1109/access.2020.2989193.
- [9] V. R. M. Vinaya and G. C. Mara, "Transfer Learning Approach to Detect and Classify Glioma Tumours," in *Proc. 2022 8th Int. Conf. Advanced Computing and Communication Systems (ICACCS)*, 2022.
- [10] P. Afshar, A. Mohammadi, and K. N. Plataniotis, "BayesCap: A Bayesian Approach to Brain Tumour Classification Using Capsule Networks," *IEEE Signal Processing Letters*, vol. 27, pp. 2024–2028, 2020.
- [11] N. Noreen, S. Palaniappam, A. Qayyum, I. Ahmad, and M. Imran, "Attention-Guided Version of 2D UNet for Automatic Brain Tumour Segmentation," *IEEE Access*, vol. 8, 2020.
- [12] Navoneel, "Brain MRI Images for Brain Tumour Detection," Kaggle, 2020. [Online]. Available: <https://www.kaggle.com/navoneel/brain-mri-images-for-brain-tumour-detection/metadata>
- [13] F. Özyurt, E. Sert, and D. Avci, "An expert system for brain tumour detection: Fuzzy C-means with super resolution and convolutional neural network with extreme learning machine," *Med. Hypotheses*, vol. 134, p. 109433, 2020.
- [14] X. Lei, X. Yu, J. Chi, Y. Wang, J. Zhang, and C. Wu, "Brain tumour segmentation in MR images using a sparse constrained level set algorithm," *Expert Systems With Applications*, 2020.
- [15] B. Leena and A. Jayanthi, "Brain tumour segmentation and classification via adaptive CLFAHE with hybrid classification," *Int J Imaging Syst Technol*, 2020.
- [16] N. Bansal, P. Dawande, S. Shukla, and S. Acharya, "Effect of lifestyle and dietary factors in the development of brain tumours," *J Family Med Prim Care*, vol. 9, no. 10, pp. 5200–5204, 2020.
- [17] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.
- [18] A. F. Abien, "Deep Learning using Rectified Linear Units (ReLU)," arXiv:1803.08375, 2018.
- [19] S. K. Ghosh, B. Biswas, and A. Ghosh, "Restoration of Mammograms by Using Deep Convolutional Denoising Auto-Encoders," in *Computational Intelligence in Data Mining*, Springer, 2020, vol. 990.
- [20] W. Weng and X. Zhu, "INet: convolutional networks for biomedical image segmentation," *IEEE Access*, vol. 9, pp. 16591–16603, 2021. doi: 10.1109/ACCESS.2021.3053408.
- [21] S. Banerjee, S. Mitra, F. Masulli, and S. Rovetta, "Deep Radionomics for Brain Tumour Detection and Classification from Multi-Sequence MRI," 2019. [Online]. Available: <http://arxiv.org/abs/1903.09240>
- [22] M. U. Rehman, S. Cho, J. H. Kim, and K. T. Chong, "Bu-net: brain tumour segmentation using modified u-net architecture," *Electronics*, vol. 9, no. 12, 2020. doi: 10.3390/electronics9122203.
- [23] M. Asim, A. Rashid, and T. Ahmad, "Scour modeling using deep neural networks based on hyperparameter optimization," *ICT Express*, vol. 8, pp. 357–362, 2021. doi: 10.1016/j.icte.2021.09.012.
- [24] P. S. Shammuga, R. S. Saran, B. Surendiran, and N. Arulmurgaselvi, "Brain Tumour Detection in MRI Using Deep Learning," in *Evolution in Computational Intelligence*, Springer, 2020, pp. 395–403. doi: 10.1007/978-981-15-5788-0_38.
- [25] M. Siar and M. Teshnehab, "Brain tumour detection using deep neural network and machine learning algorithm," in *Proc. 9th Int. Conf. Computer and Knowledge Engineering (ICCKE)*, 2019, pp. 363–368. doi: 10.1109/ICCKE48569.2019.8964846.
- [26] P. M. Siva Raja and A. V. Rani, "Brain tumour classification using a hybrid deep autoencoder with Bayesian fuzzy clustering-based segmentation approach," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 1, pp. 440–453, 2020. doi: 10.1016/j.bbce.2020.01.006.
- [27] M. F. I. Soumik and M. Ali Hossain, "Brain Tumour Classification with Inception Network Based Deep Learning Model Using Transfer Learning," in *2020 IEEE Region 10 Symposium (TENSYMP)*, June 2020, pp. 1018–1021. doi: 10.1109/TENSYMP50017.2020.9230618.
- [28] B. Srikanth and S. Venkata Suryanarayana, "Multi-Class classification of brain tumour images using data augmentation with deep neural network," *Materials Today: Proceedings*, 2021. doi: 10.1016/j.matpr.2021.01.601.
- [29] G. S. Tandel, A. Tiwari, and O. G. Kakde, "Performance optimisation of deep learning models using majority voting algorithm for brain tumour classification," *Computers in Biology and Medicine*, vol. 135, 2022.
- [30] P. Ding, J. Li, L. Wang, M. Wen, and Y. Guan, "HYBRID-CNN: an efficient scheme for abnormal flow detection in the SDN-based smart grid," *Secure Common Network*, 2020, pp. 1–20. doi: 10.1155/2020/8850550.