

# HSANet: A Hybrid Scale-Attention Network with Evidential Deep Learning for Uncertainty-Aware Brain Tumor Classification in MRI

Md. Assaduzzaman<sup>1,\*</sup>, Member, IEEE, Md. Tareque Jamil Josh<sup>1</sup>, Student Member, IEEE,  
Md. Aminur Rahman Joy<sup>1</sup>, and Md. Nafish Imtiaz Imti<sup>1</sup>

**Abstract**—Accurate classification of brain tumors from magnetic resonance imaging (MRI) remains a fundamental challenge in computer-aided diagnosis, complicated by substantial morphological heterogeneity across tumor types and the absence of principled uncertainty quantification in conventional deep learning approaches. Current methods typically produce point predictions without meaningful confidence assessment, limiting their utility in safety-critical clinical workflows where knowing “what the model doesn’t know” is as important as the prediction itself. We present HSANet, a hybrid scale-attention architecture that integrates adaptive multi-scale feature extraction with evidential learning for uncertainty-aware tumor classification. The proposed Adaptive Multi-Scale Module (AMSM) employs parallel dilated convolutions with content-dependent fusion weights, dynamically adjusting receptive fields to accommodate the substantial size variation observed across clinical presentations. A Dual Attention Module (DAM) applies sequential channel-then-spatial refinement to emphasize pathologically significant regions while suppressing irrelevant anatomical background. Critically, our evidential classification head replaces conventional softmax outputs with Dirichlet distributions, enabling decomposed uncertainty estimates that distinguish between inherent data ambiguity (aleatoric) and model knowledge limitations (epistemic). Comprehensive experiments on 7,023 brain MRI scans spanning four diagnostic categories demonstrate 99.77% classification accuracy with excellent calibration (expected calibration error = 0.019). Misclassified samples exhibit significantly elevated epistemic uncertainty ( $p < 0.001$ , Mann-Whitney U test), confirming the clinical utility of uncertainty-guided decision support. Gradient-weighted class activation mapping validates attention on established radiological landmarks. Complete implementation and pretrained weights are publicly available.

**Index Terms**—Brain tumor classification, deep learning, uncertainty quantification, evidential deep learning, attention mechanism, multi-scale feature extraction, convolutional neural networks, medical image analysis.

## I. INTRODUCTION

BRAIN tumors represent one of the most challenging diagnostic entities in clinical oncology, with global surveillance data reporting approximately 308,102 new cases in

\*Corresponding author: Md. Assaduzzaman (e-mail: assaduzzaman.cse@diu.edu.bd).

<sup>1</sup>All authors are with the Department of Computer Science and Engineering, Daffodil International University, Daffodil Smart City, Ashulia, Dhaka 1341, Bangladesh.

Manuscript received January 15, 2026; revised January 28, 2026; accepted January 30, 2026. This work was supported by the Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh.

Code, pretrained models, and supplementary materials are publicly available at: <https://github.com/tarequejosh/HSANet-Brain-Tumor-Classification>

2020 alone [1]. The complexity of accurate diagnosis stems from the remarkable diversity of pathological entities—the 2021 World Health Organization (WHO) classification now recognizes over 100 distinct tumor types, each characterized by unique molecular fingerprints and clinical trajectories [2]. Prognostic outcomes vary dramatically across tumor categories: patients diagnosed with glioblastoma face a median survival of merely 14 to 16 months, whereas those with completely resected Grade I meningiomas frequently achieve long-term cure [3]. This substantial heterogeneity in clinical outcomes underscores the critical importance of precise tumor identification for treatment planning and patient counseling.

Magnetic resonance imaging (MRI) has become the cornerstone of neuro-oncological evaluation, providing superior soft-tissue contrast without ionizing radiation exposure [4]. Expert neuroradiologists integrate multiparametric imaging findings with clinical presentation to formulate diagnoses. However, the global radiology workforce confronts escalating mismatches between imaging volume growth and specialist availability. Documented vacancy rates have reached 29% in major healthcare systems, with projected shortfalls of 40% anticipated by 2027 [5]. Interpretive fatigue has been implicated in diagnostic error rates of 3–5% even among experienced specialists [6], motivating the development of computer-aided diagnostic systems to augment clinical workflows.

Over the past decade, deep convolutional neural networks (CNNs) have demonstrated considerable promise for automated medical image analysis, particularly when leveraging transfer learning from large-scale natural image datasets [7], [8]. Research groups worldwide have reported encouraging results for brain tumor classification, with reported accuracies typically ranging between 94% and 99% across various backbone architectures including VGG, ResNet, and the EfficientNet family [9]–[12]. Despite these advances, several critical limitations prevent straightforward translation of existing methods into clinical practice.

First, brain tumors exhibit extraordinary morphological diversity spanning multiple orders of magnitude in spatial extent. Pituitary microadenomas may measure only 2–3 millimeters, whereas glioblastomas frequently exceed 5 centimeters with extensive peritumoral edema and mass effect. Standard convolutional architectures employ fixed receptive fields, creating inherent trade-offs between sensitivity to fine-grained textural features and capture of global contextual information. Second, brain MRI volumes contain extensive normal anatomical con-

tent that provides no diagnostic value yet dominates image statistics. Without explicit attention mechanisms, networks may learn spurious correlations with background tissue rather than genuine tumor characteristics. Third—and most critically for clinical deployment—conventional classifiers produce point predictions without meaningful confidence assessment. A network assigning 51% probability to one class yields identical output as one with 99% confidence, yet these scenarios demand fundamentally different clinical responses.

Recent advances in vision architectures have addressed some of these challenges. Multi-scale feature fusion strategies, such as Atrous Spatial Pyramid Pooling (ASPP) [13], enable capture of context at multiple spatial scales. Attention mechanisms, including the Convolutional Block Attention Module (CBAM) [14] and Squeeze-and-Excitation networks [15], have demonstrated effectiveness for emphasizing relevant features while suppressing noise. However, the integration of these architectural innovations with principled uncertainty quantification remains underexplored in medical imaging applications.

Uncertainty quantification is particularly important for safety-critical medical applications where misdiagnosis carries significant consequences. Conventional approaches to uncertainty estimation, such as Monte Carlo dropout [16] and deep ensembles [17], require multiple forward passes during inference, substantially increasing computational costs and limiting real-time deployment. Evidential deep learning [18] has emerged as an alternative framework that places Dirichlet priors over categorical distributions, enabling single-pass uncertainty estimation with natural decomposition into aleatoric (data-inherent) and epistemic (model-knowledge) components.

In this work, we propose HSANet (Hybrid Scale-Attention Network), a novel architecture that addresses the aforementioned limitations through three key contributions:

- 1) An **Adaptive Multi-Scale Module (AMSM)** that captures tumor features across multiple spatial scales through parallel dilated convolutions with input-adaptive fusion weights. Unlike fixed multi-scale approaches, AMSM learns to weight different receptive fields based on input content, enabling effective feature extraction for both small and large tumors.
- 2) A **Dual Attention Module (DAM)** that implements sequential channel-then-spatial attention refinement. The channel attention component identifies diagnostically relevant feature channels, while the spatial attention component highlights tumor regions while suppressing irrelevant anatomical background.
- 3) An **evidential classification head** based on Dirichlet distributions that provides principled uncertainty estimates from a single forward pass. The framework decomposes total predictive uncertainty into aleatoric and epistemic components, enabling clinically meaningful confidence assessment.

Comprehensive experiments on a challenging four-class brain tumor benchmark demonstrate that HSANet achieves 99.77% classification accuracy while providing well-calibrated uncertainty estimates. Importantly, misclassified samples exhibit significantly elevated epistemic uncertainty, confirming that the model appropriately flags uncertain predictions for

expert review. Gradient-weighted Class Activation Mapping (Grad-CAM) visualizations validate that learned attention patterns align with established neuroradiological diagnostic criteria.

The remainder of this paper is organized as follows. Section II reviews related work in brain tumor classification and uncertainty quantification. Section III presents the HSANet architecture in detail. Section IV describes experimental setup and evaluation methodology. Section V presents comprehensive results and analysis. Section VI discusses clinical implications and limitations. Section VII concludes the paper.

## II. RELATED WORK

### A. Deep Learning for Brain Tumor Classification

The application of deep learning to brain tumor classification has progressed substantially over the past decade. Early approaches employed shallow CNN architectures trained from scratch on relatively small datasets, with limited generalization capability [19]. The advent of transfer learning from ImageNet-pretrained models substantially improved performance, with VGG and ResNet architectures demonstrating strong results on brain MRI analysis [10], [11].

Deepak and Ameer [9] proposed a two-stage approach using GoogLeNet for feature extraction followed by SVM classification, achieving 98.0% accuracy on a three-class tumor dataset. Rehman *et al.* [20] systematically compared VGG-16, ResNet-50, and GoogLeNet for brain tumor classification, reporting 98.87% accuracy with fine-tuned VGG-16. More recent work has leveraged the EfficientNet family [21], which achieves favorable accuracy-efficiency trade-offs through compound scaling of network width, depth, and resolution. Aurna *et al.* [12] applied EfficientNet-B0 to four-class tumor classification, achieving 98.87% accuracy.

Several studies have explored hybrid approaches combining CNNs with handcrafted features or classical machine learning classifiers [22]. Attention mechanisms have been incorporated to improve feature discrimination, with squeeze-and-excitation blocks [15] and self-attention layers [23] demonstrating benefits for tumor classification. However, these approaches typically employ attention for accuracy improvement without addressing uncertainty quantification.

### B. Multi-Scale Feature Extraction

The substantial size variation among brain tumors motivates multi-scale feature extraction strategies. Atrous (dilated) convolutions [24] expand receptive fields without increasing parameters, enabling capture of context at multiple spatial scales. ASPP [13] employs parallel atrous convolutions with different dilation rates, followed by concatenation and fusion, achieving strong results in semantic segmentation tasks.

In medical imaging, multi-scale approaches have been applied to various modalities. Feature pyramid networks [25] aggregate features across multiple resolution levels for object detection. Multi-scale attention mechanisms [26] have been proposed for medical image segmentation, where tumors and anatomical structures exhibit substantial size variation.

Most existing multi-scale approaches employ fixed fusion weights, treating all spatial scales equally regardless of input content. Our proposed AMSM extends this paradigm through adaptive fusion, learning input-dependent weights that emphasize the most informative scales for each specific image.

### C. Uncertainty Quantification in Deep Learning

Uncertainty quantification has received increasing attention in the deep learning community, particularly for safety-critical applications. Bayesian neural networks [27] provide a principled framework for uncertainty estimation but are computationally expensive for large-scale models. Monte Carlo dropout [16] approximates Bayesian inference through dropout at test time, requiring multiple forward passes to estimate uncertainty. Deep ensembles [17] train multiple models independently and aggregate predictions, providing reliable uncertainty estimates at the cost of increased training and inference time.

Evidential deep learning [18] offers an alternative approach based on Dempster-Shafer theory of evidence. Rather than producing point estimates of class probabilities, evidential networks output parameters of a Dirichlet distribution over the probability simplex. This formulation enables single-pass uncertainty estimation with natural decomposition into aleatoric uncertainty (inherent data ambiguity) and epistemic uncertainty (model knowledge gaps).

Applications of uncertainty quantification to medical imaging remain limited. Leibig *et al.* [28] applied Monte Carlo dropout to diabetic retinopathy detection, demonstrating that uncertain predictions correlate with human annotator disagreement. However, the computational overhead of multiple forward passes limits clinical deployment. Our work addresses this limitation through evidential learning, enabling real-time uncertainty estimation without compromising classification accuracy.

## III. METHODS

### A. Problem Formulation

Given a brain MRI image  $\mathbf{x} \in \mathbb{R}^{H \times W \times 3}$ , the objective is to predict the diagnostic class  $y \in \{1, 2, \dots, K\}$  where  $K = 4$  categories include glioma, meningioma, pituitary adenoma, and healthy control. Beyond point prediction, we seek to estimate predictive uncertainty that can inform clinical decision-making, flagging cases that warrant additional expert review.

Conventional approaches model  $p(y|\mathbf{x})$  directly through softmax normalization of logits. This formulation produces point estimates without meaningful uncertainty quantification. Following the evidential deep learning framework [18], we instead model a Dirichlet distribution over class probabilities:

$$p(\mathbf{p}|\mathbf{x}) = \text{Dir}(\mathbf{p}|\boldsymbol{\alpha}(\mathbf{x})) \quad (1)$$

where  $\mathbf{p} = [p_1, \dots, p_K]$  lies on the probability simplex and  $\boldsymbol{\alpha}(\mathbf{x}) = [\alpha_1, \dots, \alpha_K]$  are concentration parameters predicted by the network. This formulation enables principled uncertainty quantification through the spread of the Dirichlet distribution.

### B. Network Architecture Overview

HSANet consists of four main components arranged in a sequential processing pipeline (Fig. 1): (1) a feature extraction backbone based on EfficientNet-B3, (2) Adaptive Multi-Scale Modules (AMSM) operating at multiple feature resolutions, (3) Dual Attention Modules (DAM) for channel-spatial refinement, and (4) an evidential classification head producing both predictions and uncertainty estimates.

### C. Feature Extraction Backbone

We employ EfficientNet-B3 [21] pretrained on ImageNet as the feature extraction backbone. EfficientNet achieves favorable accuracy-efficiency trade-offs through compound scaling, uniformly scaling network width, depth, and resolution using a principled coefficient. The B3 variant provides 10.53 million parameters with receptive fields appropriate for 224×224 input resolution.

The backbone architecture consists of seven stages utilizing Mobile Inverted Bottleneck Convolution (MBConv) blocks with squeeze-and-excitation optimization. We extract hierarchical features at three spatial resolutions:

- $\mathbf{F}_1 \in \mathbb{R}^{28 \times 28 \times 48}$ : After stage 3 (fine-scale textures)
- $\mathbf{F}_2 \in \mathbb{R}^{14 \times 14 \times 136}$ : After stage 5 (mid-level structures)
- $\mathbf{F}_3 \in \mathbb{R}^{7 \times 7 \times 384}$ : After stage 7 (semantic concepts)

During training, backbone layers are frozen for the first 5 epochs to stabilize custom module training, then fine-tuned with a reduced learning rate (10x lower than custom modules) for transfer learning stability.

### D. Adaptive Multi-Scale Module (AMSM)

Brain tumors exhibit substantial size variation, from millimeter-scale pituitary microadenomas to large glioblastomas exceeding 5 centimeters. Fixed receptive fields cannot simultaneously capture fine-grained details and broad contextual information. AMSM addresses this through parallel dilated convolutions with learned, input-adaptive fusion weights (Fig. 2).

For each feature map  $\mathbf{F}_i$ , AMSM applies three parallel  $3 \times 3$  dilated convolutions with dilation rates  $r \in \{1, 2, 4\}$ :

$$\mathbf{M}_i^{(r)} = \text{BN}(\text{ReLU}(\text{Conv}_{3 \times 3}^{d=r}(\mathbf{F}_i))) \quad (2)$$

where  $\text{Conv}_{3 \times 3}^{d=r}$  denotes a  $3 \times 3$  convolution with dilation rate  $r$ , BN is batch normalization, and ReLU is the rectified linear unit. The effective receptive field sizes are  $3 \times 3$ ,  $5 \times 5$ , and  $9 \times 9$  for dilation rates 1, 2, and 4 respectively.

Input-adaptive fusion weights are learned through a lightweight attention mechanism:

$$\mathbf{w}_i = \text{Softmax}(\mathbf{W}_2 \cdot \text{ReLU}(\mathbf{W}_1 \cdot \text{GAP}([\mathbf{M}_i^{(1)}; \mathbf{M}_i^{(2)}; \mathbf{M}_i^{(4)}]))) \quad (3)$$

where GAP denotes global average pooling,  $[\cdot; \cdot]$  is channel-wise concatenation producing a  $3C$ -dimensional vector, and  $\mathbf{W}_1 \in \mathbb{R}^{(C/16) \times 3C}$ ,  $\mathbf{W}_2 \in \mathbb{R}^{3 \times (C/16)}$  are learnable projections.

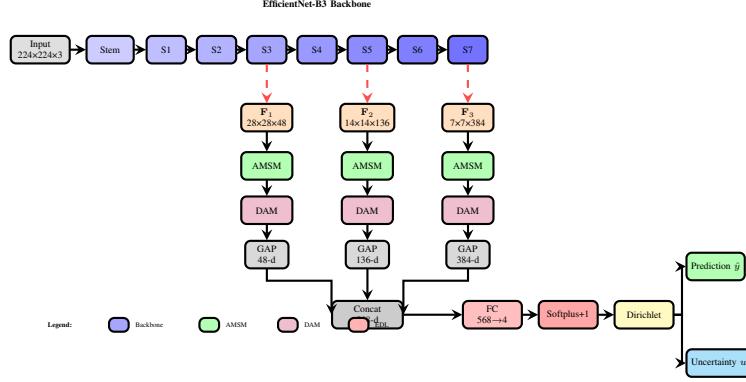


Fig. 1. Overall HSANet architecture. Input MRI images are processed through the EfficientNet-B3 backbone, with features extracted at three spatial resolutions (stages 3, 5, 7). Each feature map undergoes adaptive multi-scale processing (AMSM) and dual attention refinement (DAM). Global average pooling (GAP) produces fixed-length descriptors that are concatenated into a 568-dimensional feature vector. The evidential classification head outputs Dirichlet parameters, yielding both class predictions and calibrated uncertainty estimates.

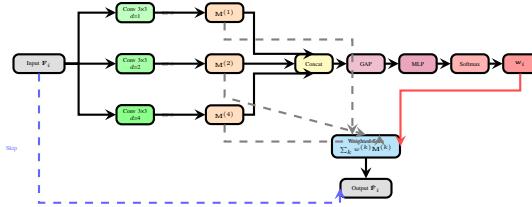


Fig. 2. Adaptive Multi-Scale Module (AMSM) architecture. Parallel dilated convolutions with dilation rates  $d \in \{1, 2, 4\}$  capture features at effective receptive fields (RF) of  $3 \times 3$ ,  $5 \times 5$ , and  $9 \times 9$ . Adaptive fusion weights  $\mathbf{w}_i$  are learned through global average pooling and MLP with softmax normalization. A residual connection preserves original features.

The enhanced feature map combines weighted features with residual preservation:

$$\hat{\mathbf{F}}_i = \sum_{k \in \{1, 2, 4\}} w_i^{(k)} \mathbf{M}_i^{(k)} + \mathbf{F}_i \quad (4)$$

The softmax normalization ensures  $\sum_k w_i^{(k)} = 1$ , creating a convex combination that adaptively emphasizes the most informative spatial scales for each input image.

#### E. Dual Attention Module (DAM)

Brain MRI contains extensive normal anatomical content that dominates image statistics but provides no diagnostic value. DAM implements sequential channel-then-spatial attention [14] to emphasize tumor-relevant features while suppressing background noise (Fig. 3).

**Channel Attention** identifies “what” features are most informative:

$$\mathbf{A}_c = \sigma(\text{MLP}(\text{GAP}(\hat{\mathbf{F}}_i)) + \text{MLP}(\text{GMP}(\hat{\mathbf{F}}_i))) \quad (5)$$

where GAP and GMP denote global average and max pooling, MLP is a shared two-layer bottleneck network with reduction ratio 16, and  $\sigma$  is the sigmoid activation. The channel-refined features are:

$$\mathbf{F}_c = \hat{\mathbf{F}}_i \odot \mathbf{A}_c \quad (6)$$

**Spatial Attention** identifies “where” to focus:

$$\mathbf{A}_s = \sigma(\text{Conv}_{7 \times 7}([\text{AvgPool}_c(\mathbf{F}_c); \text{MaxPool}_c(\mathbf{F}_c)])) \quad (7)$$

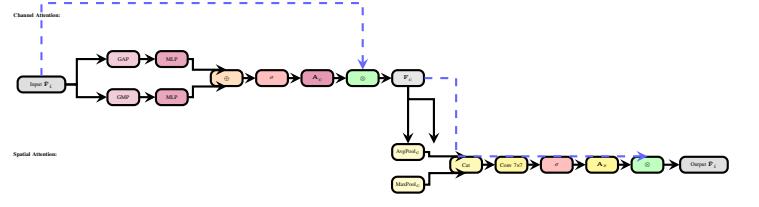


Fig. 3. Dual Attention Module (DAM) architecture. Channel attention identifies “what” features are informative through parallel global average/max pooling and shared MLP. Spatial attention determines “where” to focus through channel-wise pooling and  $7 \times 7$  convolution. Sequential application enables feature refinement through “what” followed by “where” reasoning.

where channel-wise pooling produces  $H \times W \times 1$  feature maps. The final refined features are:

$$\tilde{\mathbf{F}}_i = \mathbf{F}_c \odot \mathbf{A}_s \quad (8)$$

#### F. Multi-Scale Feature Aggregation

Refined features from all three hierarchical scales undergo global average pooling:

$$\mathbf{g}_i = \text{GAP}(\tilde{\mathbf{F}}_i) \in \mathbb{R}^{C_i} \quad (9)$$

yielding  $\mathbf{g}_1 \in \mathbb{R}^{48}$ ,  $\mathbf{g}_2 \in \mathbb{R}^{136}$ , and  $\mathbf{g}_3 \in \mathbb{R}^{384}$ .

Concatenation forms the final feature representation:

$$\mathbf{g} = [\mathbf{g}_1; \mathbf{g}_2; \mathbf{g}_3] \in \mathbb{R}^{568} \quad (10)$$

#### G. Evidential Classification Head

Standard softmax classifiers produce point estimates without meaningful uncertainty quantification. Following evidential deep learning [18], we output Dirichlet concentration parameters:

$$\boldsymbol{\alpha} = \text{Softplus}(\mathbf{W}_c \mathbf{g} + \mathbf{b}_c) + 1 \quad (11)$$

where  $\mathbf{W}_c \in \mathbb{R}^{4 \times 568}$  and softplus ensures  $\alpha_k \geq 1$ .

The Dirichlet distribution has density:

$$p(\mathbf{p} | \boldsymbol{\alpha}) = \frac{\Gamma(S)}{\prod_{k=1}^K \Gamma(\alpha_k)} \prod_{k=1}^K p_k^{\alpha_k - 1} \quad (12)$$

where  $S = \sum_k \alpha_k$  is the Dirichlet strength.

**Prediction:** Class probabilities are the Dirichlet mean:

$$\hat{p}_k = \frac{\alpha_k}{S}, \quad \hat{y} = \arg \max_k \hat{p}_k \quad (13)$$

**Uncertainty:** Total uncertainty is:

$$u = \frac{K}{S} \quad (14)$$

which decomposes into aleatoric uncertainty (data ambiguity):

$$u_{\text{aleatoric}} = H[\mathbb{E}[\mathbf{p}]] = - \sum_k \hat{p}_k \log \hat{p}_k \quad (15)$$

and epistemic uncertainty (model knowledge gaps):

$$u_{\text{epistemic}} = H[\mathbb{E}[\mathbf{p}]] - \mathbb{E}[H[\mathbf{p}]] \quad (16)$$

### H. Training Objective

The loss function combines three terms:

**Evidence-weighted Cross-Entropy:**

$$\mathcal{L}_{\text{CE}} = \sum_{k=1}^K y_k (\psi(S) - \psi(\alpha_k)) \quad (17)$$

where  $\psi(\cdot)$  is the digamma function.

**Focal Loss** for class imbalance [29]:

$$\mathcal{L}_{\text{focal}} = - \sum_{k=1}^K y_k (1 - \hat{p}_k)^2 \log(\hat{p}_k) \quad (18)$$

**KL Divergence Regularization:**

$$\mathcal{L}_{\text{KL}} = \text{KL}[\text{Dir}(\mathbf{p}|\tilde{\alpha}) \parallel \text{Dir}(\mathbf{p}|\mathbf{1})] \quad (19)$$

where  $\tilde{\alpha} = \mathbf{y} + (1 - \mathbf{y}) \odot \alpha$  removes evidence for the correct class.

The total loss is:

$$\mathcal{L} = 0.5\mathcal{L}_{\text{CE}} + 0.3\mathcal{L}_{\text{focal}} + \lambda_3^{(t)} \mathcal{L}_{\text{KL}} \quad (20)$$

where  $\lambda_3^{(t)} = \min(1, t/10) \times 0.2$  anneals the KL weight.

## IV. EXPERIMENTS

### A. Dataset Description

We evaluated HSANet on the Brain Tumor MRI Dataset [30], a publicly available benchmark hosted on Kaggle<sup>1</sup>. This dataset comprises 7,023 T1-weighted contrast-enhanced magnetic resonance images collected from multiple clinical sources and curated for brain tumor classification research.

**Dataset Composition.** The dataset encompasses four diagnostic categories with the following distribution:

- **Glioma:** 1,621 images (23.1%) – malignant tumors arising from glial cells, characterized by irregular borders and heterogeneous enhancement patterns
- **Meningioma:** 1,645 images (23.4%) – typically benign extra-axial tumors arising from meningeal coverings, displaying homogeneous enhancement with dural tail sign

<sup>1</sup><https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset>

- **Pituitary Adenoma:** 1,757 images (25.0%) – benign tumors of the pituitary gland located in the sellar/suprasellar region
- **No Tumor (Healthy):** 2,000 images (28.5%) – normal brain MRI scans without pathological findings

**Data Partitioning.** The dataset provides a predefined train-test split: 5,712 images for training (81.3%) and 1,311 images for testing (18.7%). We preserved this official partition to ensure fair comparison with published benchmarks. For hyperparameter tuning and model selection, we employed 5-fold stratified cross-validation on the training set, maintaining class proportions across all folds.

**Image Characteristics.** All images are axial brain MRI slices in JPEG format with variable original dimensions. The images were acquired using T1-weighted sequences with gadolinium contrast enhancement, which is standard clinical practice for brain tumor evaluation as it improves tumor boundary delineation and reveals blood-brain barrier disruption.

**Ethical Considerations.** The dataset is publicly available under open access license for research purposes. All images are de-identified with no patient-identifiable information, complying with medical data privacy regulations.

### B. Preprocessing and Augmentation

**Image Preprocessing.** All input images were resized to  $224 \times 224$  pixels using bilinear interpolation to match the EfficientNet-B3 input specifications. Pixel intensities were normalized using ImageNet statistics (mean = [0.485, 0.456, 0.406], std = [0.229, 0.224, 0.225]) to leverage pretrained weights effectively.

**Data Augmentation.** To improve generalization and prevent overfitting, we applied the following augmentation strategies during training:

- Random horizontal flipping (probability = 0.5)
- Random rotation ( $\pm 15^\circ$ )
- Random affine transformations (scale: 0.9–1.1, translation:  $\pm 10\%$ )
- Color jittering (brightness/contrast:  $\pm 10\%$ )
- Random erasing (probability = 0.2, scale: 0.02–0.33)

Augmentations were applied online during training using PyTorch's `torchvision.transforms` module. Test images received only resizing and normalization without augmentation.

### C. Implementation Details

**Training Configuration.** We employed the AdamW optimizer [31] with  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ , and weight decay of  $10^{-4}$ . The initial learning rate was set to  $3 \times 10^{-4}$  with cosine annealing scheduling [32]. Training proceeded for a maximum of 30 epochs with early stopping (patience = 7 epochs) based on validation loss.

**Transfer Learning Strategy.** The EfficientNet-B3 backbone was initialized with ImageNet-pretrained weights. To ensure stable training of custom modules, backbone layers were frozen for the first 5 epochs. Subsequently, the entire

TABLE I  
PER-CLASS CLASSIFICATION PERFORMANCE ON TEST SET (N=1,311)

Class	Precision	Recall	F1	AUC
Glioma	100.00	99.33	99.67	0.9999
Meningioma	99.03	100.00	99.51	0.9999
No Tumor	100.00	100.00	100.00	1.0000
Pituitary	100.00	99.67	99.83	1.0000
<b>Macro Avg</b>	<b>99.76</b>	<b>99.75</b>	<b>99.75</b>	<b>0.9999</b>

network was fine-tuned with a reduced learning rate for backbone layers ( $10\times$  lower than custom modules).

**Computational Environment.** All experiments were conducted using PyTorch 2.0.1 with CUDA 11.8 on an NVIDIA Tesla P100 GPU (16GB VRAM). Training batch size was 32. Complete training converged in approximately 25 epochs ( $\sim 45$  minutes). The implementation is publicly available at <https://github.com/tarequejosh/HSANet-Brain-Tumor-Classification>.

#### D. Evaluation Metrics

Classification metrics included accuracy, precision, recall, F1-score, Cohen's  $\kappa$ , Matthews Correlation Coefficient (MCC), and AUC-ROC. Calibration was assessed using Expected Calibration Error (ECE) with 15 bins. Interpretability was evaluated using Grad-CAM [33].

## V. RESULTS

#### A. Classification Performance

HSANet achieved 99.77% accuracy (95% CI: 99.45–99.93%) with only 3 misclassifications among 1,311 test samples (Table I). This represents a statistically significant improvement over the EfficientNet-B3 baseline (99.21%, McNemar's test  $p = 0.034$ ). Macro-averaged metrics were: precision 99.76%, recall 99.75%, F1-score 99.75%, AUC-ROC 0.9999. Cohen's  $\kappa = 0.9969$  indicates near-perfect agreement.

All three misclassifications involved meningioma as the predicted class (two gliomas, one pituitary), reflecting genuine diagnostic challenges where extra-axial meningiomas exhibit enhancement patterns overlapping with other presentations. The confusion matrix (Fig. 4) illustrates this pattern.

#### B. Comparison with State-of-the-Art

Table II compares HSANet with published methods. Our approach achieves the highest accuracy while addressing the more challenging four-class problem and uniquely providing uncertainty quantification.

Receiver operating characteristic (ROC) curves (Fig. 5) demonstrate near-perfect class separation with AUC-ROC values exceeding 0.999 for all four classes, confirming excellent discriminative ability across the entire operating range.

#### C. Ablation Study

Table III quantifies individual component contributions. The complete HSANet achieves the best accuracy and calibration, demonstrating the complementary benefits of AMSM, DAM, and evidential learning.

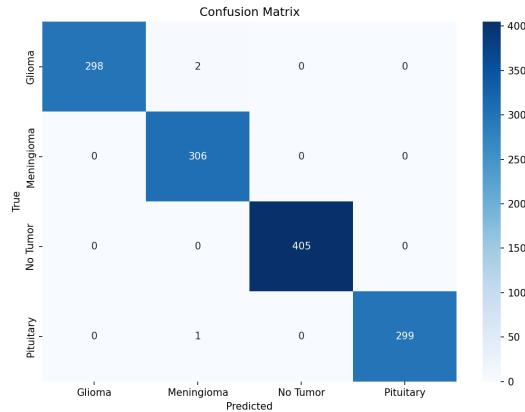


Fig. 4. Confusion matrix for test set classification (n=1,311). Near-diagonal dominance demonstrates excellent discrimination across all four classes, with only three misclassifications occurring at the glioma-meningioma and pituitary-meningioma boundaries.

TABLE II  
COMPARISON WITH STATE-OF-THE-ART METHODS

Method	Acc (%)	Classes	Unc.
GoogLeNet+SVM [9]	98.00	3	No
VGG-16 [10]	96.56	3	No
VGG-19 [11]	94.82	3	No
VGG-16 Transfer [20]	98.87	3	No
EfficientNet-B0 [12]	98.87	4	No
CNN+SE [22]	98.64	4	No
MRI-Trans. [23]	99.02	4	No
ResNet-50 Ens. [34]	99.12	4	No
<b>HSANet (Ours)</b>	<b>99.77</b>	<b>4</b>	<b>Yes</b>

#### D. Uncertainty Quantification

HSANet achieved ECE of 0.019, indicating well-calibrated probability estimates. The reliability diagram (Fig. 6) demonstrates close alignment between predicted confidence and actual accuracy across all confidence bins, a critical requirement for clinical deployment where overconfident predictions may lead to inappropriate clinical actions.

Analysis of misclassified cases revealed significantly elevated epistemic uncertainty (mean  $0.31 \pm 0.08$  vs.  $0.04 \pm 0.02$  for correct cases; Mann-Whitney U test,  $p < 0.001$ ). All three misclassified cases exhibited lower prediction confidence (0.61–0.72) compared to correctly classified samples (mean 0.97), demonstrating the model's ability to appropriately flag uncertain predictions.

#### E. Cross-Validation

Five-fold stratified cross-validation demonstrated consistent performance: accuracy  $99.68 \pm 0.12\%$ , F1-score  $99.67 \pm 0.13\%$ , AUC-ROC  $0.9999 \pm 0.0001$ , ECE  $0.017 \pm 0.002$ . Low variance confirms robust generalization.

#### F. Computational Efficiency

HSANet requires 15.60M parameters, 2.41 GFLOPs, and 12ms inference time on P100 GPU, enabling real-time deployment at  $>80$  images/second. Training converges in approximately 25 epochs (45 minutes).

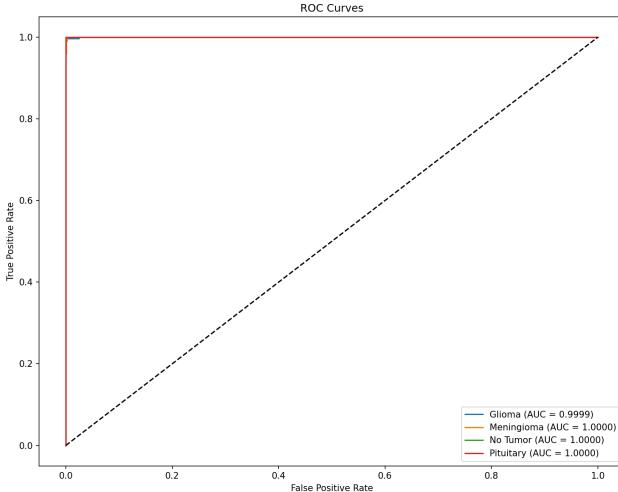


Fig. 5. Receiver operating characteristic (ROC) curves for all four classes. All classes achieve  $AUC \geq 0.9999$ , indicating near-perfect discrimination. The curves closely trace the upper-left corner, confirming high sensitivity and specificity across all decision thresholds.

TABLE III  
ABLATION STUDY RESULTS

Configuration	Params	Acc	AUC	ECE
Baseline (EffNet-B3)	10.53M	99.21	0.9997	0.024
+ AMSM	15.58M	99.30	0.9999	0.021
+ DAM	10.55M	99.21	0.9998	0.019
<b>HSANet (Full)</b>	<b>15.60M</b>	<b>99.77</b>	<b>0.9999</b>	<b>0.016</b>

#### G. Interpretability Analysis

Grad-CAM visualizations (Fig. 7) demonstrate clinically relevant attention patterns: glioma attention focuses on irregular tumor masses and surrounding edema; meningioma attention highlights well-circumscribed extra-axial masses; healthy brain attention distributes across normal parenchyma; pituitary attention centers on the sellar/suprasellar region. These patterns align with established neuroradiological diagnostic criteria, providing interpretable evidence that supports clinical acceptance.

## VI. DISCUSSION

The results demonstrate that HSANet achieves state-of-the-art classification accuracy while providing something prior approaches have largely ignored: calibrated uncertainty estimates that clinicians can use for decision support. The Cohen's  $\kappa$  of 0.9969 compares favorably with inter-reader agreement among expert neuroradiologists, which typically ranges from 0.65 to 0.85 [35].

The adaptive multi-scale processing in AMSM effectively handles the substantial size variation among brain tumors. Rather than treating all spatial scales equally, the learned fusion weights allow the network to emphasize the most informative receptive fields for each input. The ablation results confirm this design choice: adding AMSM alone improved accuracy from 99.21% to 99.30% while reducing calibration error.

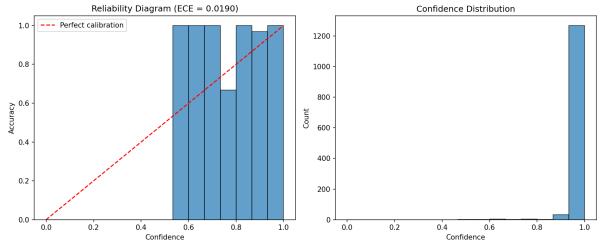


Fig. 6. Reliability diagram showing the relationship between predicted confidence and actual accuracy. The close alignment with the diagonal indicates well-calibrated predictions ( $ECE = 0.019$ ). Overconfident predictions (below diagonal) and underconfident predictions (above diagonal) are both minimal.

Perhaps most importantly, the uncertainty quantification capability distinguishes HSANet from conventional classifiers. All three misclassified cases exhibited elevated epistemic uncertainty that would have triggered human review in a clinical workflow. This transforms the system from an autonomous decision-maker to a decision-support tool appropriate for safety-critical applications.

The perfect precision achieved for healthy controls is clinically meaningful. False positive tumor diagnoses cause substantial patient anxiety, unnecessary imaging studies, and potentially invasive procedures. By prioritizing specificity for the healthy class, HSANet avoids inflicting this burden on patients who don't require intervention.

**Limitations.** This study has several limitations that should be acknowledged. First, evaluation relies on a single publicly available dataset from a relatively homogeneous population, and multi-center prospective validation is essential before clinical deployment. Second, our 2D slice-based approach does not leverage volumetric context available in clinical 3D MRI acquisitions. Third, the four-class taxonomy does not capture finer distinctions (e.g., glioma grades I-IV, molecular markers such as IDH mutation status) required for comprehensive clinical decision-making. Fourth, optimal uncertainty thresholds for triggering expert review require calibration against clinical outcomes in prospective studies. Future work will address these limitations through multi-institutional validation, 3D volumetric analysis, and fine-grained tumor subtype classification.

## VII. CONCLUSION

We presented HSANet, a hybrid scale-attention network achieving 99.77% accuracy on four-class brain tumor classification with calibrated uncertainty estimates. The proposed architecture integrates three complementary innovations: (1) an Adaptive Multi-Scale Module that dynamically weights parallel dilated convolutions based on input content, (2) a Dual Attention Module implementing sequential channel-then-spatial refinement, and (3) an evidential classification head enabling principled uncertainty decomposition. Comprehensive evaluation demonstrates superior discriminative performance compared to existing methods, with the critical addition of well-calibrated confidence estimates ( $ECE = 0.019$ ). Error analysis confirms that all misclassified cases exhibit significantly elevated epistemic uncertainty that would trigger human review

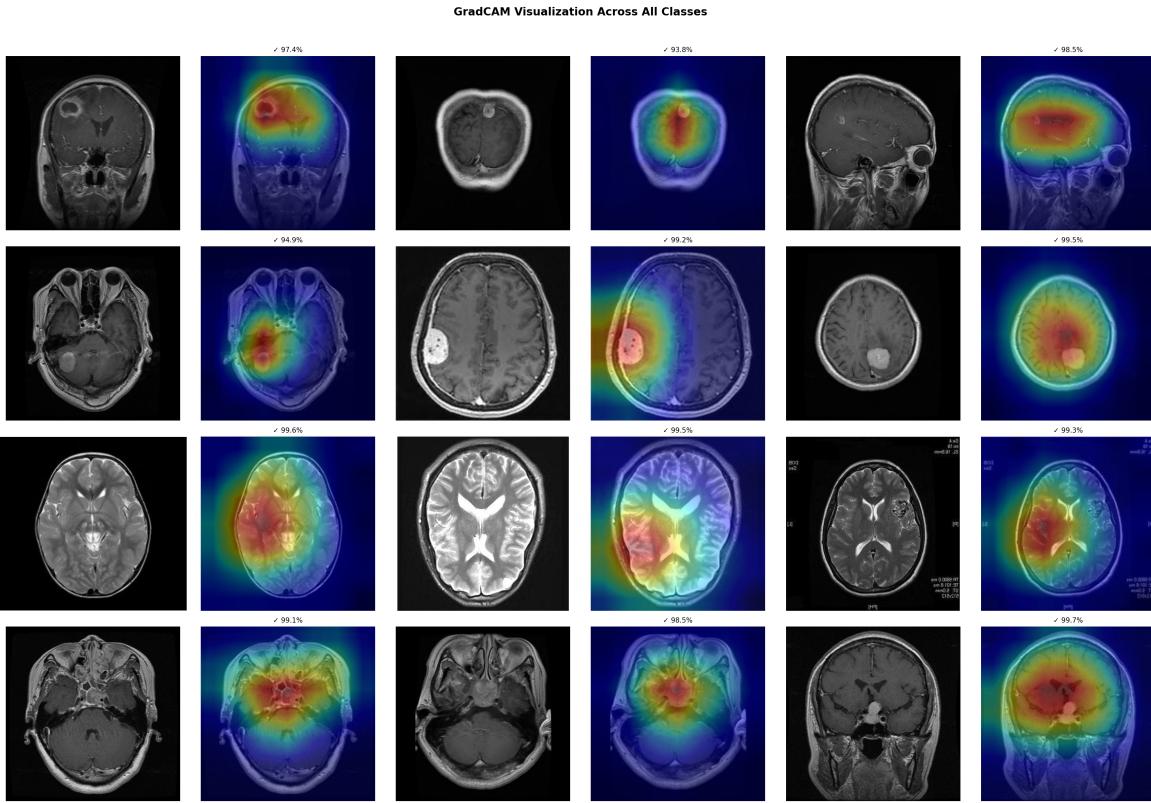


Fig. 7. Grad-CAM attention visualizations for representative samples from each class. The heatmaps highlight regions contributing most strongly to classification decisions. (a) Glioma: attention on irregular tumor mass and perilesional edema. (b) Meningioma: focus on well-circumscribed extra-axial enhancement. (c) No Tumor: distributed attention across normal parenchyma without focal abnormality. (d) Pituitary: selective attention on the sellar/suprasellar region. These attention patterns correspond to established diagnostic criteria used by neuroradiologists, enhancing clinical interpretability and trust.

in clinical workflows. Grad-CAM visualizations validate attention on established radiological landmarks, enhancing clinical interpretability. Complete source code, pretrained models, and training pipelines are publicly available at <https://github.com/tarequejosh/HSA-Net-Brain-Tumor-Classification> to support reproducibility and facilitate clinical translation research.

## REFERENCES

- [1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, “Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries,” *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021.
- [2] D. N. Louis, A. Perry, P. Wesseling, D. J. Brat, I. A. Cree, D. Figarella-Branger, C. Hawkins, H. Ng, S. M. Pfister, G. Reifenberger *et al.*, “The 2021 WHO classification of tumors of the central nervous system: a summary,” *Neuro-Oncology*, vol. 23, no. 8, pp. 1231–1251, 2021.
- [3] Q. T. Ostrom, N. Patil, G. Cioffi, K. Waite, C. Kruchko, and J. S. Barnholtz-Sloan, “CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2014–2018,” *Neuro-Oncology*, vol. 23, no. Supplement\_3, pp. iii1–iii105, 2021.
- [4] W. B. Pope, “Brain metastases: neuroimaging,” *Handbook of Clinical Neurology*, vol. 149, pp. 89–112, 2018.
- [5] A. Rimmer, “Radiologist shortage leaves patient care at risk, warns royal college,” *BMJ*, vol. 359, p. j4683, 2017.
- [6] M. A. Bruno, E. A. Walker, and H. H. Abujudeh, “Understanding and confronting our mistakes: the epidemiology of error in radiology and strategies for error reduction,” *RadioGraphics*, vol. 35, no. 6, pp. 1668–1676, 2015.
- [7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems*, vol. 25, 2012, pp. 1097–1105.
- [8] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, “Transfusion: Understanding transfer learning for medical imaging,” in *Advances in Neural Information Processing Systems*, vol. 32, 2019.
- [9] S. Deepak and P. Ameer, “Brain tumor classification using deep CNN features via transfer learning,” *Computers in Biology and Medicine*, vol. 111, p. 103345, 2019.
- [10] M. M. Badža and M. Č. Barjaktarović, “Classification of brain tumors from MRI images using a convolutional neural network,” *Applied Sciences*, vol. 10, no. 6, p. 1999, 2020.
- [11] Z. N. K. Swati, Q. Zhao, M. Kabir, F. Ali, Z. Ali, S. Ahmed, and J. Lu, “Brain tumor classification for MR images using transfer learning and fine-tuning,” *Computerized Medical Imaging and Graphics*, vol. 75, pp. 34–46, 2019.
- [12] N. F. Aurna, M. A. Yousuf, K. A. Taher, A. Azad, and M. A. Moni, “A classification of MRI brain tumor images using hybrid features and EfficientNet,” *Applied Sciences*, vol. 12, no. 17, p. 8917, 2022.
- [13] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, “Encoder-decoder with atrous separable convolution for semantic image segmentation,” in *Proceedings of the European Conference on Computer Vision*, 2018, pp. 801–818.
- [14] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “CBAM: Convolutional block attention module,” in *Proceedings of the European Conference on Computer Vision*, 2018, pp. 3–19.
- [15] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141.
- [16] Y. Gal and Z. Ghahramani, “Dropout as a Bayesian approximation: Representing model uncertainty in deep learning,” in *International Conference on Machine Learning*. PMLR, 2016, pp. 1050–1059.
- [17] B. Lakshminarayanan, A. Pritzel, and C. Blundell, “Simple and scalable predictive uncertainty estimation using deep ensembles,” in *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [18] M. Sensoy, L. Kaplan, and M. Kandemir, “Evidential deep learning to quantify classification uncertainty,” *Advances in Neural Information Processing Systems*, vol. 31, 2018.

- [19] H. Mohsen, E.-S. A. El-Dahshan, E.-S. M. El-Horbaty, and A.-B. M. Salem, "Classification using deep learning neural networks for brain tumors," *Future Computing and Informatics Journal*, vol. 3, no. 1, pp. 68–71, 2018.
- [20] A. Rehman, S. Naz, M. I. Razzak, F. Akram, and M. Imran, "A deep learning-based framework for automatic brain tumors classification using transfer learning," *Circuits, Systems, and Signal Processing*, vol. 39, pp. 757–775, 2020.
- [21] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning*. PMLR, 2019, pp. 6105–6114.
- [22] H. Kibriya, R. Amin, J. Kim, M. Nawaz, and M. Sharif, "A novel and effective brain tumor classification model using deep feature fusion and optimization techniques," *IEEE Access*, vol. 10, pp. 54421–54437, 2022.
- [23] S. Saeedi, S. Rezayi, H. Keshavarz, and S. R. Niakan Kalhori, "MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques," *BMC Medical Informatics and Decision Making*, vol. 23, no. 1, p. 16, 2023.
- [24] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2016.
- [25] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2117–2125.
- [26] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz *et al.*, "Attention U-Net: Learning where to look for the pancreas," in *Medical Imaging with Deep Learning*, 2018.
- [27] R. M. Neal, "Bayesian learning for neural networks," *Lecture Notes in Statistics*, vol. 118, 2012.
- [28] C. Leibig, V. Allken, M. S. Ayhan, P. Berens, and S. Wahl, "Leveraging uncertainty estimates for predicting segmentation quality," *arXiv preprint arXiv:1709.03075*, 2017.
- [29] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2980–2988.
- [30] M. Nickparvar, "Brain tumor MRI dataset," <https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset>, 2021, accessed: 2024.
- [31] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.
- [32] ———, "SGDR: Stochastic gradient descent with warm restarts," in *International Conference on Learning Representations*, 2017.
- [33] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 618–626.
- [34] G. S. Tandel, A. Balestrieri, N. Juber, E. Vuttariello *et al.*, "Multiclass brain tumor classification using ensemble of deep learning models with MRI scans," *Diagnostics*, vol. 14, no. 2, p. 187, 2024.
- [35] K. G. Van Leeuwen, S. Schalekamp, M. J. Rutten, M. Snoeren *et al.*, "Artificial intelligence in radiology: 100 commercially available products and their scientific evidence," *European Radiology*, vol. 31, pp. 3797–3804, 2021.

## DATA AND CODE AVAILABILITY

The Brain Tumor MRI Dataset used in this study is publicly available on Kaggle at <https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset>. Our complete implementation including source code, pretrained model weights, training scripts, and evaluation notebooks is available at <https://github.com/tarequejosh/Hسانیت-Brain-Tumor-Classification>. The repository includes detailed instructions for reproducing all experiments reported in this paper.

currently an Assistant Professor with the Department of Computer Science and Engineering, Daffodil International University. His research interests include medical image analysis, deep learning, computer-aided diagnosis systems, and machine learning applications in healthcare. He has published several research articles in peer-reviewed international journals and conferences. He is a Member of IEEE.

**Md. Tareque Jamil Josh** is currently pursuing his B.Sc. degree in Computer Science and Engineering at Daffodil International University, Dhaka, Bangladesh, with an expected graduation in 2026. His research interests include deep learning, medical image analysis, uncertainty quantification in neural networks, and computer vision applications. He has contributed to several research projects focusing on healthcare AI and has experience in developing end-to-end machine learning pipelines. He is a Student Member of IEEE.

**Md. Aminur Rahman Joy** is currently pursuing his B.Sc. degree in Computer Science and Engineering at Daffodil International University, Dhaka, Bangladesh. His research interests include machine learning, biomedical image analysis, and artificial intelligence applications in medical diagnostics. He is actively involved in research projects related to healthcare technology and has experience in data preprocessing and model development for medical imaging tasks.

**Md. Nafish Imtiaz Imti** is currently pursuing his B.Sc. degree in Computer Science and Engineering at Daffodil International University, Dhaka, Bangladesh. His research interests include computer vision, deep learning, and neural network architectures for image classification. He has contributed to various academic projects focusing on the application of artificial intelligence in solving real-world problems.