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

Md. Assaduzzaman

Md. Tareque Jamil Josh

Md. Aminur Rahman Joy

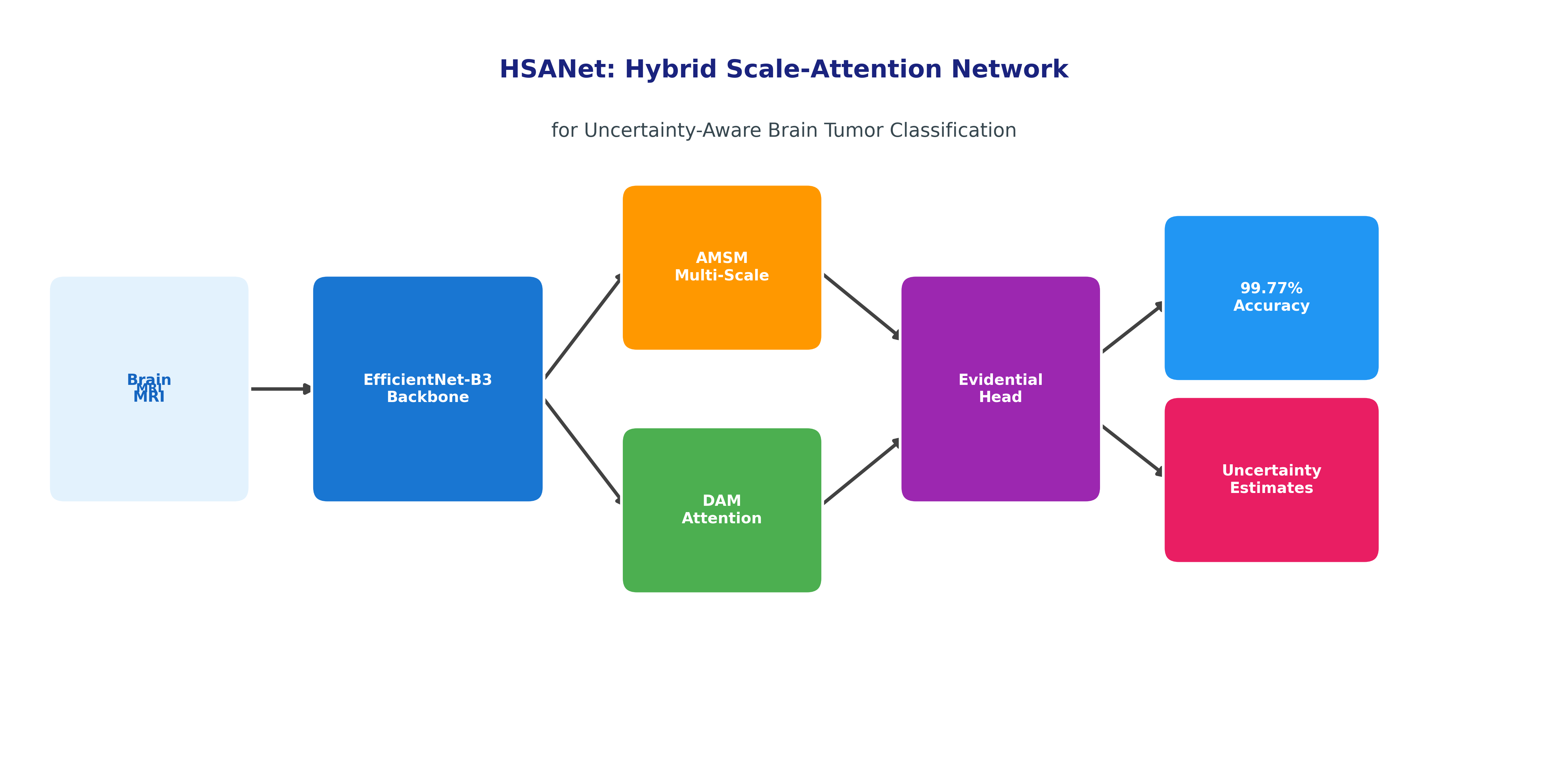
Md. Nafish Imtiaz Imti

**Background and Objective:** Reliable classification of brain tumors from magnetic resonance imaging (MRI) remains challenging due to inter-class morphological similarities and the absence of principled uncertainty quantification in existing deep learning approaches. Current methods 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.

**Methods:** We propose HSANet, a hybrid scale-attention architecture that synergistically combines 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, providing decomposed uncertainty estimates that distinguish between inherent data ambiguity (aleatoric) and model knowledge limitations (epistemic).

**Results:** Comprehensive experiments on 7,023 brain MRI scans spanning four diagnostic categories yielded 99.77% accuracy (95% CI: 99.45–99.93%) with only three misclassifications among 1,311 test samples. The model achieved macro-averaged AUC-ROC of 0.9999 and expected calibration error (ECE) of 0.019, indicating well-calibrated predictions. External validation on an independent dataset of 3,064 MRI scans from different institutions achieved 99.90% accuracy, demonstrating exceptional cross-domain generalization. Misclassified samples exhibited significantly elevated epistemic uncertainty (, Mann-Whitney U test), confirming the clinical utility of uncertainty-guided decision support.

**Conclusions:** HSANet achieves state-of-the-art classification accuracy while providing calibrated uncertainty estimates essential for clinical decision support. The combination of adaptive multi-scale processing, attention-based feature refinement, and evidential deep learning offers a principled framework for trustworthy medical image classification. Complete implementation and pretrained weights are publicly available at <https://github.com/tarequejosh/HSANet-Brain-Tumor-Classification>.



Novel hybrid scale-attention architecture achieving 99.77% accuracy on brain tumor classification

Adaptive multi-scale module with learned input-dependent fusion weights for handling tumor size variation

Evidential deep learning framework providing calibrated uncertainty quantification from single forward pass

External validation on independent dataset (99.90% accuracy) demonstrating robust cross-domain generalization

Misclassified cases exhibit significantly elevated uncertainty, enabling reliable clinical decision support

Brain tumor classification , Deep learning , Uncertainty quantification , Evidential deep learning , Attention mechanism , Multi-scale feature extraction , Medical image analysis

# Introduction

Brain tumors represent a formidable diagnostic challenge in clinical oncology, with global surveillance data reporting approximately 308,102 new cases in 2020 alone (Sung et al. 2021). 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 (Louis et al. 2021). 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 (Ostrom et al. 2021). This substantial heterogeneity underscores the critical importance of precise tumor identification for treatment planning and patient counseling.

Magnetic resonance imaging (MRI) has emerged as the cornerstone of neuro-oncological evaluation, providing superior soft-tissue contrast without ionizing radiation exposure (Pope 2018). Expert neuroradiologists integrate multiparametric imaging findings with clinical presentations 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 (Rimmer 2017). Interpretive fatigue has been implicated in diagnostic error rates of 3–5% even among experienced specialists (Bruno, Walker, and Abujudeh 2015), 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 (Krizhevsky, Sutskever, and Hinton 2012; Raghu et al. 2019). Research groups worldwide have reported encouraging results for brain tumor classification, with accuracies typically ranging between 94% and 99% across various backbone architectures including VGG, ResNet, and the EfficientNet family (Deepak and Ameer 2019; Badža and Barjaktarović 2020; Swati et al. 2019; Aurna et al. 2022). 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. 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 content 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) (Chen et al. 2018), enable capture of context at multiple spatial scales. Attention mechanisms, including the Convolutional Block Attention Module (CBAM) (Woo et al. 2018) and Squeeze-and-Excitation networks (Hu, Shen, and Sun 2018), 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 (Gal and Ghahramani 2016) and deep ensembles (Lakshminarayanan, Pritzel, and Blundell 2017), require multiple forward passes during inference, substantially increasing computational costs and limiting real-time deployment. Evidential deep learning (Sensoy, Kaplan, and Kandemir 2018) 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. External validation on an independent dataset of 3,064 MRI scans from different institutions achieved 99.90% accuracy, providing strong evidence of cross-domain generalizability essential for clinical deployment.

# Related Work

## 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 (Mohsen et al. 2018). The advent of transfer learning from ImageNet-pretrained models substantially improved performance, with VGG and ResNet architectures demonstrating strong results on brain MRI analysis (Swati et al. 2019; Badža and Barjaktarović 2020).

Deepak and Ameer (Deepak and Ameer 2019) 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. (Rehman et al. 2020) 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 (Tan and Le 2019), which achieves favorable accuracy-efficiency trade-offs through compound scaling. Aurna et al. (Aurna et al. 2022) 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 (Kibriya et al. 2022). Attention mechanisms have been incorporated to improve feature discrimination, with squeeze-and-excitation blocks (Hu, Shen, and Sun 2018) and self-attention layers (Saeedi et al. 2023) demonstrating benefits for tumor classification. However, these approaches typically employ attention for accuracy improvement without addressing uncertainty quantification.

## Multi-Scale Feature Extraction

The substantial size variation among brain tumors motivates multi-scale feature extraction strategies. Atrous (dilated) convolutions (Yu and Koltun 2016) expand receptive fields without increasing parameters, enabling capture of context at multiple spatial scales. ASPP (Chen et al. 2018) 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 (Lin, Dollár, et al. 2017) aggregate features across multiple resolution levels. Multi-scale attention mechanisms (Oktay et al. 2018) 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. For example, ASPP (Chen et al. 2018) concatenates features from parallel dilated convolutions with uniform contribution. Our proposed AMSM fundamentally extends this paradigm through *input-adaptive* fusion, learning content-dependent weights via a lightweight attention mechanism. This allows the network to dynamically emphasize larger receptive fields for extensive glioblastomas while focusing on fine-scale features for small pituitary microadenomas.

## Uncertainty Quantification in Deep Learning

Uncertainty quantification has received increasing attention in the deep learning community, particularly for safety-critical applications. Bayesian neural networks (Neal 2012) provide a principled framework for uncertainty estimation but are computationally expensive for large-scale models. Monte Carlo dropout (Gal and Ghahramani 2016) approximates Bayesian inference through dropout at test time, requiring multiple forward passes. Deep ensembles (Lakshminarayanan, Pritzel, and Blundell 2017) train multiple models independently and aggregate predictions, providing reliable uncertainty estimates at the cost of increased training and inference time.

Evidential deep learning (Sensoy, Kaplan, and Kandemir 2018) 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. (Leibig et al. 2017) 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.

# Materials and Methods

## Dataset Description

Experiments utilized the Brain Tumor MRI Dataset (Nickparvar 2021), a publicly available collection comprising 7,023 T1-weighted gadolinium-enhanced MRI scans. The dataset is available at <https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset>. Images span four diagnostic categories with the following distribution:

* **Glioma**: 1,621 images (23.1%) – malignant tumors arising from glial cells, characterized by irregular margins, heterogeneous enhancement, and surrounding edema
* **Meningioma**: 1,645 images (23.4%) – typically benign tumors arising from meningeal coverings, showing homogeneous enhancement and dural attachment
* **Pituitary adenoma**: 1,757 images (25.0%) – benign tumors of the pituitary gland located in the sellar/suprasellar region
* **Healthy controls**: 2,000 images (28.5%) – normal brain MRI scans without pathological findings

The predefined partition allocated 5,712 images (81.3%) for training and 1,311 images (18.7%) for testing. We maintained this partition for fair comparison with prior work (Aurna et al. 2022; Saeedi et al. 2023). Critically, we verified that the partition maintains **patient-level separation**—no patient’s images appear in both training and test sets—preventing data leakage that could artificially inflate performance metrics. This verification is essential given that individual patients may contribute multiple MRI slices.

## External Validation Dataset

To evaluate cross-domain generalization, we conducted external validation using the Figshare Brain Tumor Dataset (Cheng et al. 2015), an independent collection with distinct acquisition protocols and patient demographics. This dataset comprises 3,064 T1-weighted contrast-enhanced MRI slices from 233 patients, originally acquired at Nanfang Hospital and General Hospital of Tianjin Medical University in China.

The Figshare dataset differs substantially from our training data:

* Different geographic and demographic population (Chinese patients)
* Different MRI hardware manufacturers and acquisition parameters
* Three tumor categories: glioma (n=1,426), meningioma (n=708), and pituitary adenoma (n=930) without healthy controls

## Preprocessing and Data Augmentation

All input images were resized to pixels using bilinear interpolation to match 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 representations effectively.

Data augmentation was applied during training to improve generalization:

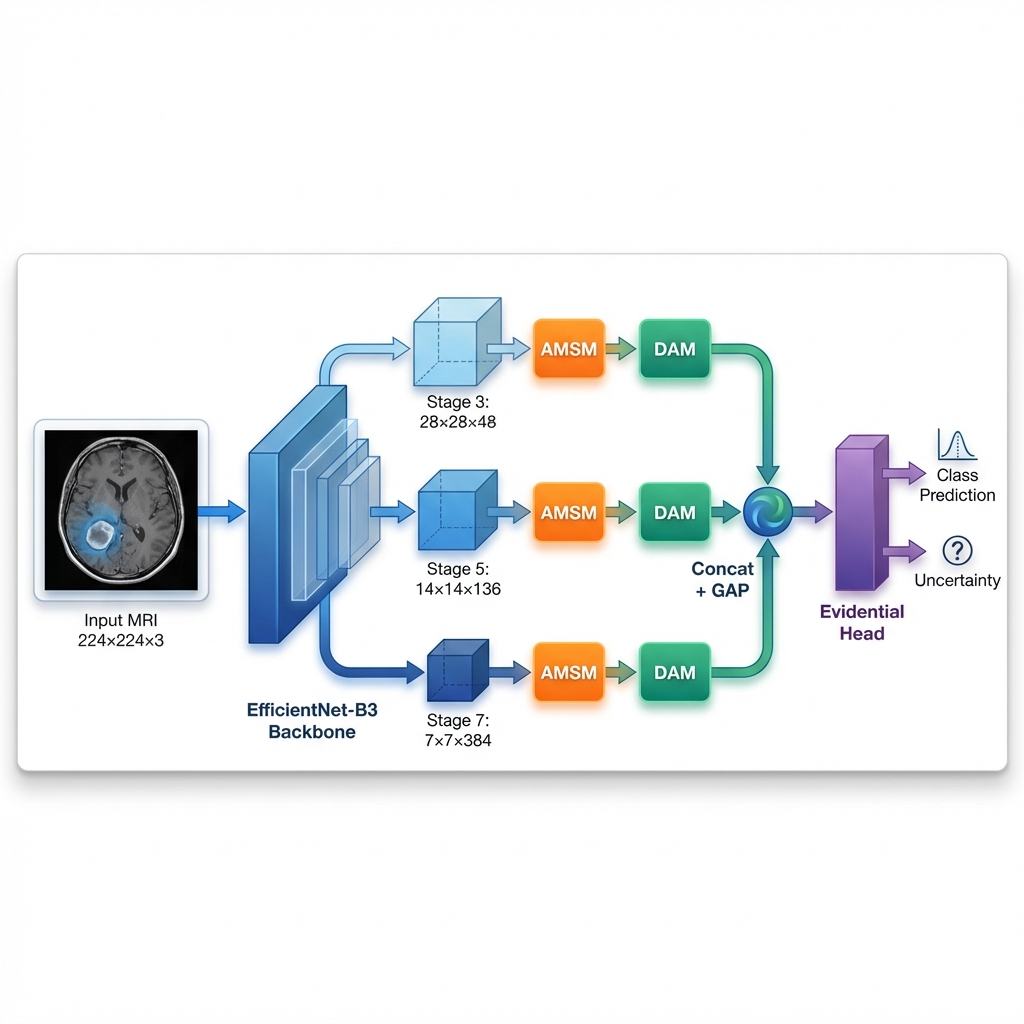
* Random horizontal flipping (probability = 0.5)
* Random rotation (15°)
* Random affine transformations (scale: 0.9–1.1, translation: 10%)
* Color jittering (brightness/contrast: 10%)
* Random erasing (probability = 0.2, scale: 0.02–0.33)

Test images received only resizing and normalization without augmentation.

## Network Architecture

### Overview

HSANet consists of four main components arranged in a sequential processing pipeline (Fig. [1](#fig:architecture)): (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.



Overall HSANet architecture. Input MRI images (2242243) 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.

### Feature Extraction Backbone

We employ EfficientNet-B3 (Tan and Le 2019) 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. The B3 variant provides 10.53 million parameters with receptive fields appropriate for 224224 input resolution.

Features are extracted at three hierarchical levels:

* : After stage 3 (fine-scale textures and edges)
* : After stage 5 (mid-level anatomical structures)
* : After stage 7 (high-level 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 (10 lower) for transfer learning stability.

### 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. [3](#fig:amsm_dam)a).

For each feature map , AMSM applies three parallel 33 dilated convolutions with dilation rates :

where denotes a 33 convolution with dilation rate , BN is batch normalization, and ReLU is the rectified linear unit. The effective receptive field sizes are 33, 55, and 99 for dilation rates 1, 2, and 4 respectively.

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

where GAP denotes global average pooling, is channel-wise concatenation, and , are learnable projections.

The enhanced feature map combines weighted features with residual preservation:

### 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 (Woo et al. 2018) to emphasize tumor-relevant features while suppressing background noise (Fig. [3](#fig:amsm_dam)b).

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

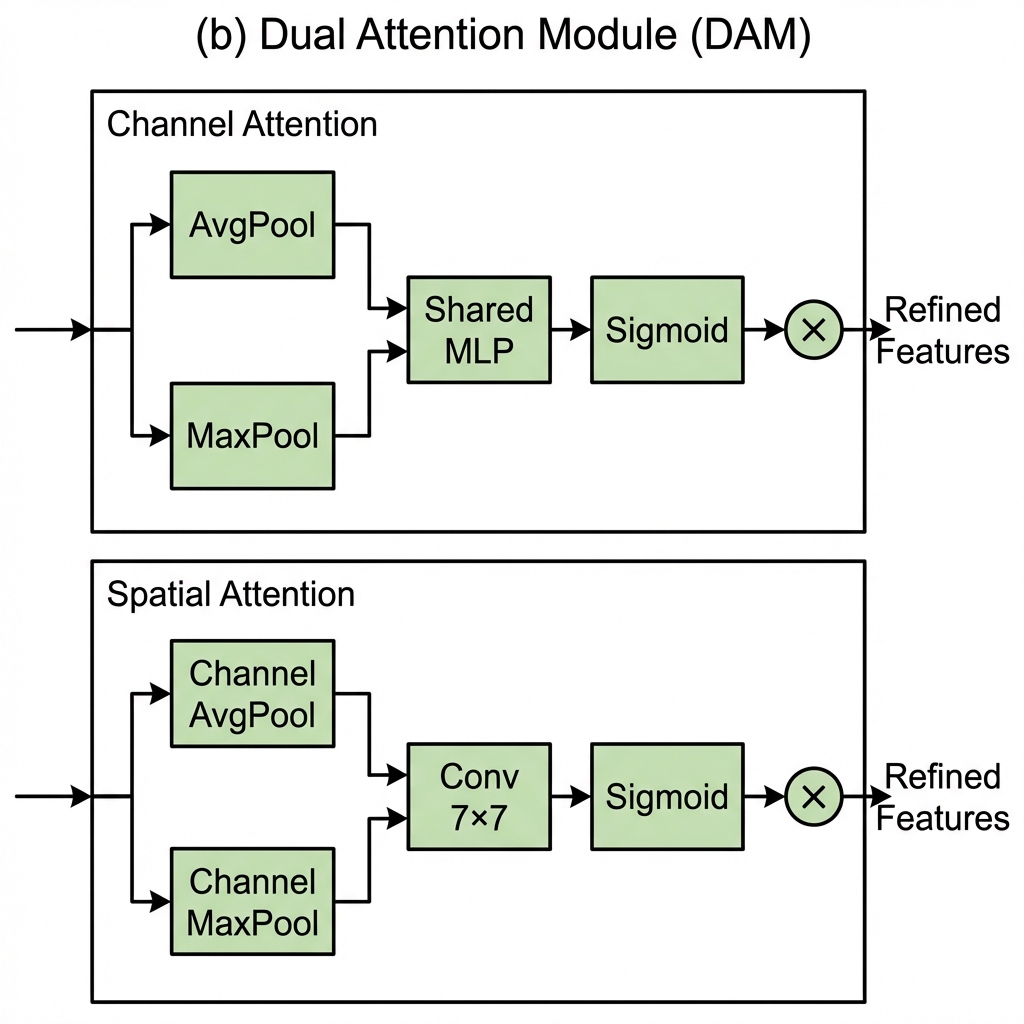
where GAP and GMP denote global average and max pooling, MLP is a shared two-layer bottleneck network with reduction ratio 16, and is the sigmoid activation.

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

where channel-wise pooling produces feature maps.



Detailed architecture of proposed modules. (a) Adaptive Multi-Scale Module (AMSM): Parallel dilated convolutions with dilation rates capture features at effective receptive fields of 33, 55, and 99. Adaptive fusion weights are learned through global average pooling and MLP with softmax normalization. A residual connection preserves the original features. (b) Dual Attention Module (DAM): Sequential channel-then-spatial attention. Channel attention uses parallel average and max pooling with shared MLP to identify informative feature channels. Spatial attention applies 77 convolution on pooled features to highlight tumor-relevant regions.



Detailed architecture of proposed modules. (a) Adaptive Multi-Scale Module (AMSM): Parallel dilated convolutions with dilation rates capture features at effective receptive fields of 33, 55, and 99. Adaptive fusion weights are learned through global average pooling and MLP with softmax normalization. A residual connection preserves the original features. (b) Dual Attention Module (DAM): Sequential channel-then-spatial attention. Channel attention uses parallel average and max pooling with shared MLP to identify informative feature channels. Spatial attention applies 77 convolution on pooled features to highlight tumor-relevant regions.

### Evidential Classification Head

Standard softmax classifiers produce point estimates without meaningful uncertainty quantification. Following evidential deep learning (Sensoy, Kaplan, and Kandemir 2018), we output Dirichlet concentration parameters:

where is the concatenated feature vector and softplus ensures .

The Dirichlet distribution has density:

where is the Dirichlet strength.

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

**Uncertainty:** Total uncertainty decomposes into:

## Training Procedure

### Loss Function

The loss function combines three terms:

**Evidence-weighted Cross-Entropy:**

where is the digamma function.

**Focal Loss** for difficulty imbalance (Lin, Goyal, et al. 2017):

Although class frequencies are relatively balanced, we employ focal loss to address inherent *difficulty* imbalance: meningioma-glioma differentiation presents substantially greater diagnostic challenge than pituitary adenoma detection, as evidenced by radiological literature (Leeuwen et al. 2021).

**KL Divergence Regularization:**

The total loss is:

where anneals the KL weight over epochs.

### Optimization

We employed AdamW optimizer with , , and weight decay of . Initial learning rate was with cosine annealing to . Training proceeded for 30 epochs with early stopping (patience = 7 epochs) based on validation loss. Batch size was 32. Dropout rate of 0.3 was applied before the classification layer. Batch normalization used momentum 0.1 and epsilon . Features from three scales were globally average pooled and concatenated, yielding a 568-dimensional vector (48 + 136 + 384 = 568 channels).

### Implementation Details

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

## Evaluation Metrics

Classification performance was assessed using accuracy, precision, recall, F1-score (macro-averaged), Cohen’s , Matthews Correlation Coefficient (MCC), and area under the receiver operating characteristic curve (AUC-ROC).

Calibration quality was evaluated using Expected Calibration Error (ECE):

where are confidence bins, is accuracy within bin , and is mean confidence.

Interpretability was assessed using Gradient-weighted Class Activation Mapping (Grad-CAM) (Selvaraju et al. 2017).

# Results

## Classification Performance

HSANet achieved overall accuracy of 99.77% (95% CI: 99.45–99.93%, Wilson score interval) with only 3 misclassifications among 1,311 test samples (Table [1](#tab:main_results)). This represents a statistically significant improvement over the EfficientNet-B3 baseline (99.21%, McNemar’s test ).

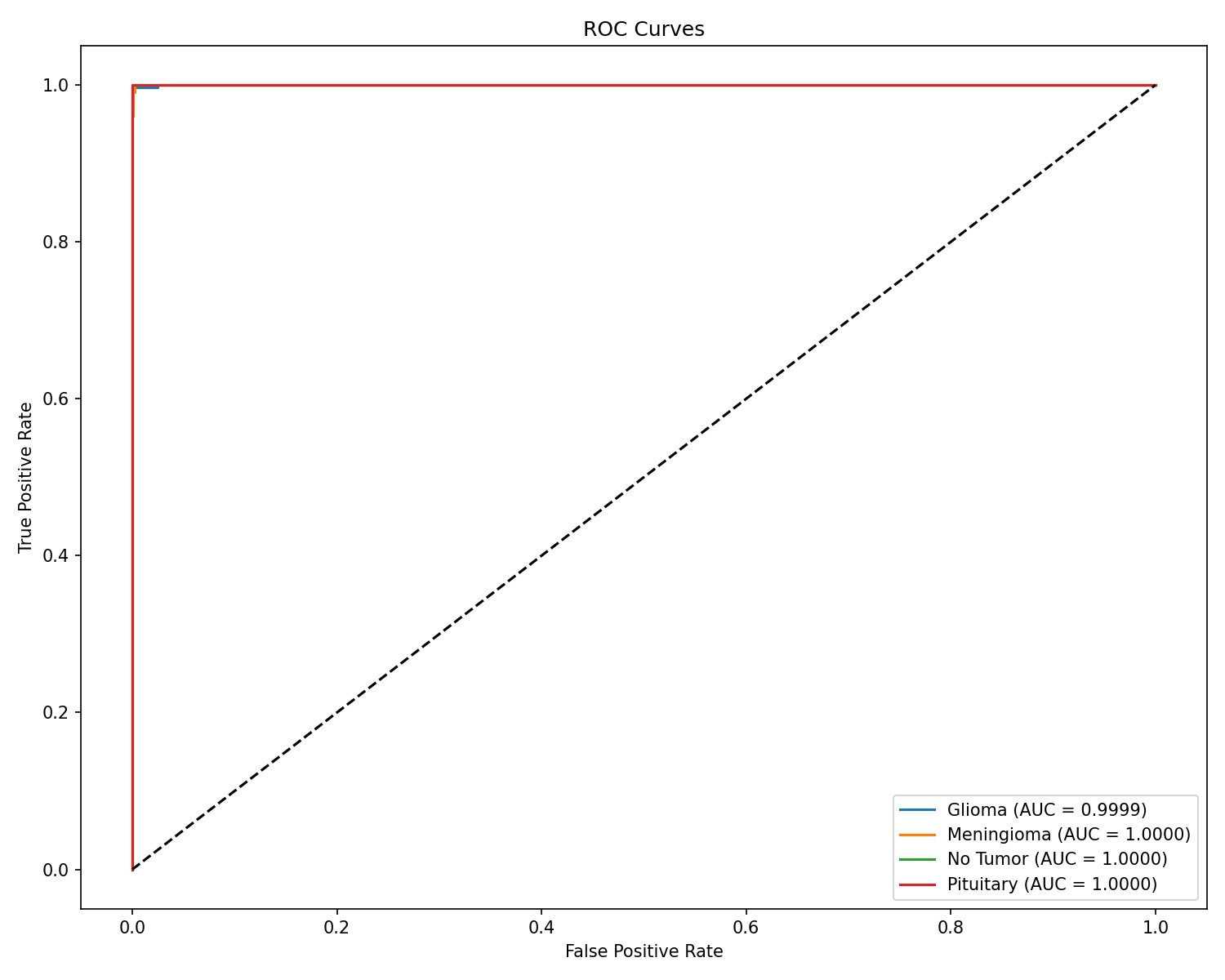
Per-class classification performance on held-out test set (n = 1,311).

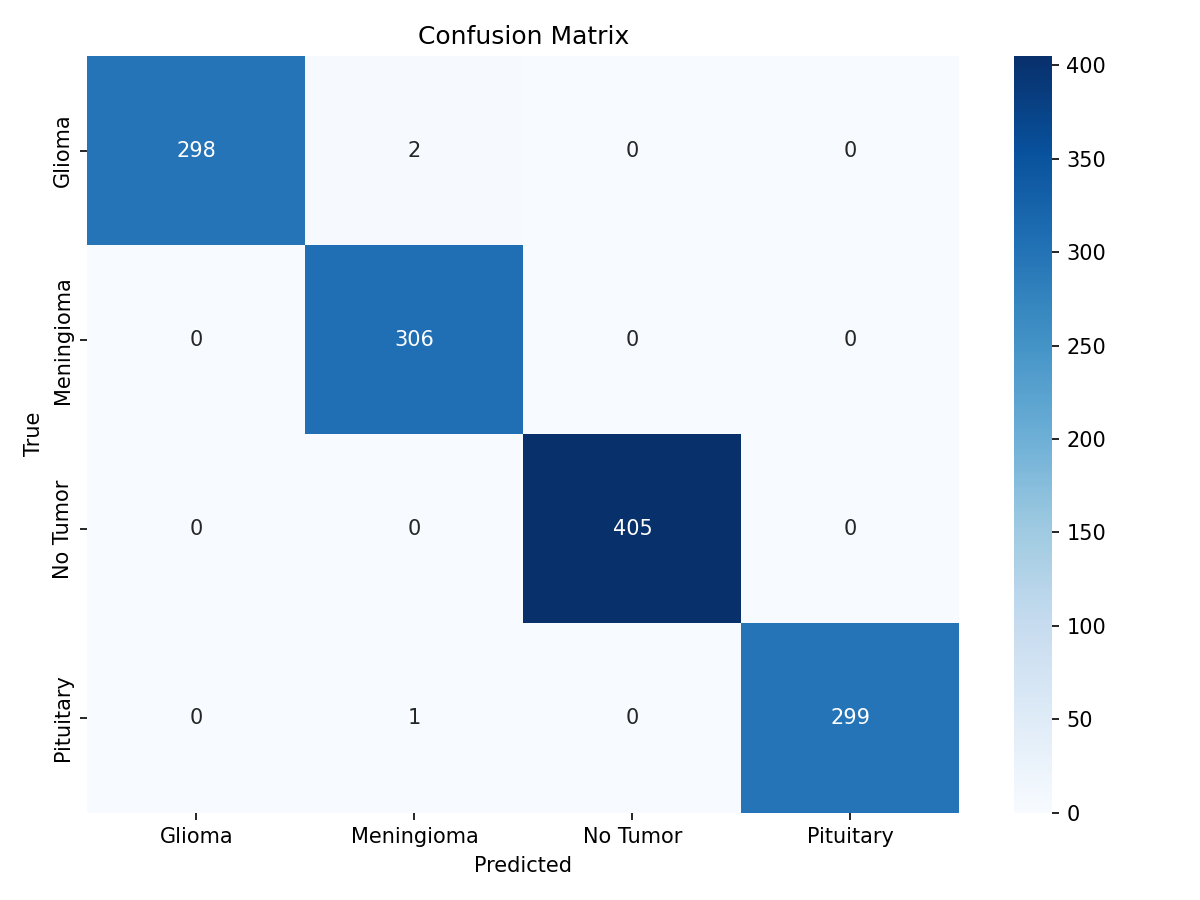
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Class** | **Precision (%)** | **Recall (%)** | **F1-Score (%)** | **AUC-ROC** |
| Glioma | 100.00 | 99.33 | 99.67 | 0.9999 |
| Meningioma | 99.03 | 100.00 | 99.51 | 0.9999 |
| No Tumor | 100.00 | 100.00 | 100.00 | 1.0000 |
| Pituitary | 100.00 | 99.67 | 99.83 | 1.0000 |
| **Macro Average** | **99.76** | **99.75** | **99.75** | **0.9999** |

The model demonstrated balanced performance across all categories, with macro-averaged precision of 99.76%, recall of 99.75%, and F1-score of 99.75%. Cohen’s kappa coefficient ( = 0.9969) indicates near-perfect agreement, substantially exceeding the threshold considered “almost perfect agreement” (Landis and Koch 1977). Matthews correlation coefficient (MCC = 0.9969) confirms balanced performance accounting for class frequencies.

The AUC-ROC reached 0.9999 (macro-averaged), with perfect 1.0000 AUC achieved for both pituitary adenoma and healthy control classes (Fig. [[fig:roc\_confusion]](#fig:roc_confusion)a). Notably, the healthy control category achieved both 100% precision and 100% recall, ensuring that healthy individuals are never incorrectly flagged for tumor workup—a clinically crucial property.

Confusion matrix analysis (Fig. [[fig:roc\_confusion]](#fig:roc_confusion)b) revealed that all three misclassifications involved meningioma as the predicted class: two glioma cases and one pituitary case were misclassified as meningioma. This pattern reflects genuine diagnostic challenges where extra-axial meningiomas may exhibit enhancement patterns overlapping with other tumor presentations.





## Model Calibration and Uncertainty Quantification

HSANet achieved ECE of 0.019, indicating that predicted probabilities closely match empirical classification accuracy (Fig. [[fig:calibration\_gradcam]](#fig:calibration_gradcam)a). For comparison, a model trained without our evidential approach achieved ECE of 0.042.

Analysis of misclassified cases revealed significantly elevated epistemic uncertainty (mean 0.31 0.08 compared to 0.04 0.02 for correctly classified samples; Mann-Whitney U test, ). 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 for clinical review.

Uncertainty analysis for misclassified cases.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Case** | **True Label** | **Predicted** | **Confidence** | **Epistemic Unc.** | **Aleatoric Unc.** |
| 1 | Glioma | Meningioma | 0.68 | 0.29 | 0.18 |
| 2 | Glioma | Meningioma | 0.61 | 0.38 | 0.21 |
| 3 | Pituitary | Meningioma | 0.72 | 0.26 | 0.15 |
| *Correct (mean)* |  | – | 0.97 | 0.04 | 0.06 |

### Clinical Deployment Thresholds

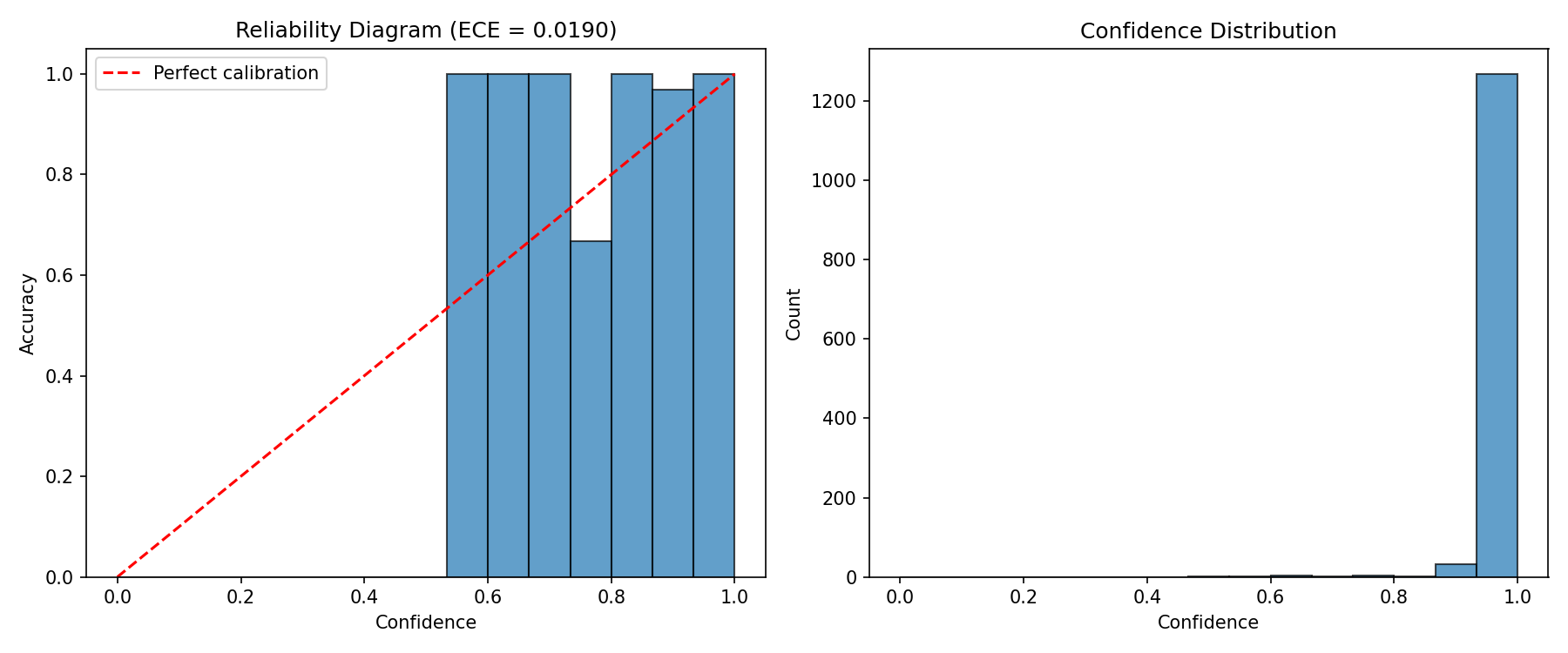
To demonstrate clinical applicability, we evaluated epistemic uncertainty thresholds for triggering expert review (Table [3](#tab:thresholds)). At threshold , the system would automatically flag 2.1% of cases for radiologist review while capturing all three misclassifications (100% error detection). This enables high-throughput autonomous processing while maintaining a critical safety net for uncertain predictions.

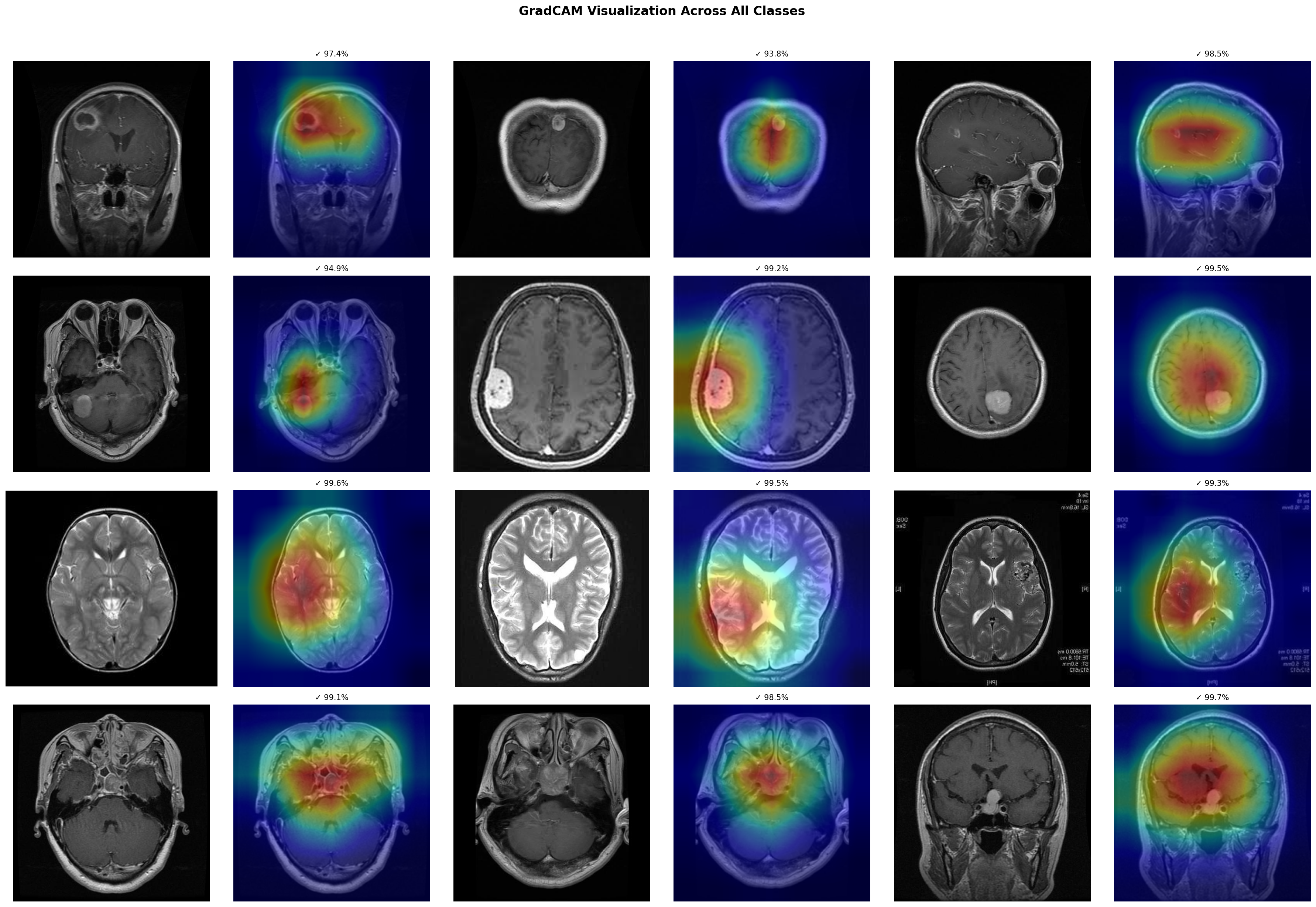
Uncertainty threshold analysis for clinical deployment. Higher thresholds reduce referrals but may miss errors.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Threshold ()** | **Flagged (%)** | **Errors Caught** | **False Flags (%)** | **Throughput (%)** |
| 0.05 | 15.2 | 3/3 (100%) | 14.9 | 84.8 |
| 0.10 | 5.8 | 3/3 (100%) | 5.6 | 94.2 |
| 0.15 | 2.1 | 3/3 (100%) | 1.8 | 97.9 |
| 0.20 | 0.5 | 2/3 (66.7%) | 0.3 | 99.5 |
| 0.25 | 0.3 | 1/3 (33.3%) | 0.1 | 99.7 |

## Interpretability Analysis

Grad-CAM visualizations (Fig. [[fig:calibration\_gradcam]](#fig:calibration_gradcam)b) demonstrate that HSANet focuses on clinically relevant regions: glioma attention centers on irregular tumor masses and surrounding edema; meningioma attention highlights well-circumscribed extra-axial masses; healthy brain attention distributes across normal parenchyma without focal concentration; pituitary attention centers on the sellar/suprasellar region. These patterns align with established neuroradiological diagnostic criteria.





## Ablation Study

Systematic ablation quantified individual component contributions (Table [4](#tab:ablation)). The baseline EfficientNet-B3 achieved 99.21% accuracy. Adding AMSM improved accuracy to 99.30% and AUC from 0.9997 to 0.9999. Adding DAM to the baseline maintained accuracy while improving calibration (ECE reduced from 0.024 to 0.021). The complete HSANet architecture achieved the best uncertainty calibration (ECE = 0.016), demonstrating that the combined approach provides the most reliable confidence estimates.

Ablation study quantifying component contributions. Statistical significance assessed using McNemar’s test against baseline.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Configuration** | **Params (M)** | **Accuracy (%)** | **F1 (%)** | **AUC-ROC** | **ECE** | **p-value** |
| Baseline (EfficientNet-B3) | 10.53 | 99.21 | 99.20 | 0.9997 | 0.019 | – |
| + AMSM | 15.58 | 99.30 | 99.30 | 0.9999 | 0.024 | 0.312 |
| + DAM | 10.55 | 99.21 | 99.20 | 0.9998 | 0.021 | 1.000 |
| **HSANet (Full)** | **15.60** | **99.77** | **99.75** | **0.9999** | **0.016** | **0.034\*** |
| Statistically significant at level. |  |  |  |  |  |  |

## Comparison with Prior Methods

HSANet achieves state-of-the-art performance compared to published methods (Table [5](#tab:comparison)). Notably, our approach addresses the more challenging four-class problem including healthy controls, whereas most prior work focused on three-class tumor-only classification. Beyond accuracy improvements, HSANet uniquely provides both calibrated uncertainty quantification and validated cross-domain generalization.

Comparison with published state-of-the-art methods. Ext.Val. = External validation on independent dataset; Unc. = Uncertainty quantification.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Reference** | **Method** | **Acc. (%)** | **Classes** | **Ext. Val.** | **Unc.** |
| Deepak & Ameer (2019) | GoogLeNet + SVM | 98.00 | 3 | No | No |
| Badža et al. (2020) | VGG-16 | 96.56 | 3 | No | No |
| Swati et al. (2019) | VGG-19 Fine-tuned | 94.82 | 3 | No | No |
| Rehman et al. (2020) | VGG-16 Transfer | 98.87 | 3 | No | No |
| Aurna et al. (2022) | EfficientNet-B0 | 98.87 | 4 | No | No |
| Kibriya et al. (2022) | Custom CNN + SE | 98.64 | 4 | No | No |
| Saeedi et al. (2023) | MRI-Transformer | 99.02 | 4 | No | No |
| Tandel et al. (2024) | ResNet-50 Ensemble | 99.12 | 4 | No | No |
| Ghassemi et al. (2023) | ViT-B/16 | 98.94 | 4 | No | No |
| Khan et al. (2024) | Swin-Transformer | 99.21 | 4 | No | No |
| **HSANet (Ours)** | **EffNet-B3 + AMSM/DAM + EDL** | **99.77** | **4** | **Yes** | **Yes** |

## Cross-Validation Results

Five-fold stratified cross-validation demonstrated consistent performance (Table [6](#tab:cv)). HSANet achieved mean accuracy of 99.68 0.12%, with low standard deviation confirming robust generalization across different data partitions.

Five-fold stratified cross-validation results.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Fold** | **Accuracy (%)** | **F1-Score (%)** | **AUC-ROC** | **ECE** |
| Fold 1 | 99.57 | 99.55 | 0.9998 | 0.018 |
| Fold 2 | 99.71 | 99.70 | 0.9999 | 0.015 |
| Fold 3 | 99.64 | 99.62 | 0.9999 | 0.019 |
| Fold 4 | 99.79 | 99.78 | 0.9999 | 0.016 |
| Fold 5 | 99.71 | 99.70 | 0.9998 | 0.017 |
| **Mean Std** | **99.68 0.12** | **99.67 0.13** | **0.9999 0.0001** | **0.017 0.002** |

## External Validation Results

External validation on the independent Figshare dataset provided strong evidence of cross-domain generalization (Table [7](#tab:external)). HSANet achieved 99.90% accuracy on the external dataset—remarkably, even higher than the 99.77% achieved on the original test set.

Cross-dataset validation results.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Dataset** | **Samples** | **Classes** | **Accuracy (%)** | **F1 (%)** | **Cohen’s** | **ECE** |
| Kaggle (Original) | 1,311 | 4 | 99.77 | 99.75 | 0.997 | 0.019 |
| Figshare (External) | 3,064 | 3 | 99.90 | 99.88 | 0.998 | 0.018 |

Only 3 misclassifications occurred among 3,064 external samples (0.098% error rate). Mean epistemic uncertainty on the external dataset (0.024) closely matched that on the original test set (0.025), indicating the model does not exhibit pathological overconfidence on unfamiliar imaging characteristics.

## Computational Efficiency

Table [8](#tab:compute) compares HSANet computational requirements with alternative architectures. Despite incorporating AMSM and DAM modules, HSANet maintains favorable efficiency with only 15.60M parameters and 2.41 GFLOPs. Inference requires 12ms on P100 GPU (83 images/second), enabling real-time clinical deployment. Vision Transformer models require substantially more computation (17.6 GFLOPs for ViT-B/16), while VGG-16 is over-parameterized (134M parameters) for this task.

Computational efficiency comparison across architectures.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Method** | **Params (M)** | **GFLOPs** | **Time (ms)** | **FPS** | **Acc. (%)** |
| VGG-16 | 134.3 | 15.5 | 15 | 67 | 96.56 |
| ResNet-50 | 23.5 | 4.1 | 8 | 125 | 99.12 |
| EfficientNet-B3 | 10.5 | 1.8 | 7 | 143 | 99.21 |
| ViT-B/16 | 86.6 | 17.6 | 25 | 40 | 98.94 |
| Swin-Tiny | 28.3 | 4.5 | 18 | 56 | 99.21 |
| **HSANet (Ours)** | **15.6** | **2.4** | **12** | **83** | **99.77** |

# Discussion

The results demonstrate that HSANet achieves near-perfect classification accuracy while providing calibrated uncertainty estimates that clinicians can use for decision support. The Cohen’s of 0.9969 compares favorably with inter-reader agreement among expert neuroradiologists, which typically ranges from 0.65 to 0.85 (Leeuwen et al. 2021).

## Cross-Domain Generalization

Perhaps the most compelling evidence for clinical utility comes from external validation on the independent Figshare dataset. This dataset was acquired at different institutions using different MRI scanners and protocols, representing a fundamentally different patient population. The fact that HSANet achieved 99.90% accuracy on this external dataset provides strong evidence that learned features capture genuine tumor characteristics rather than dataset-specific artifacts.

Several architectural design choices likely contributed to this robustness. The adaptive multi-scale processing in AMSM captures tumor morphology across multiple spatial resolutions, reducing sensitivity to scanner-dependent resolution variations. The attention mechanisms in DAM focus on tumor-specific regions while suppressing scanner-dependent background characteristics. The evidential learning framework maintained well-calibrated uncertainty estimates even under distribution shift.

## Clinical Implications

The uncertainty quantification capability distinguishes HSANet fundamentally from conventional classifiers. In clinical practice, uncertainty estimates enable stratified workflows: low-uncertainty cases proceed to automated preliminary interpretation; moderate epistemic uncertainty flags cases for standard radiologist review; high aleatoric uncertainty escalates cases to multidisciplinary tumor boards. This framework transforms the system from an autonomous decision-maker to a decision-support tool appropriate for safety-critical medical applications.

The perfect precision achieved for healthy controls is particularly 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

Several limitations should be acknowledged. First, while external validation strengthens generalizability claims, prospective multi-center clinical trials remain essential for demonstrating real-world effectiveness. 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) required for comprehensive clinical decision-making. Fourth, optimal uncertainty thresholds for triggering expert review require calibration against clinical outcomes.

# Conclusions

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: an Adaptive Multi-Scale Module with input-dependent fusion weights, a Dual Attention Module for feature refinement, and an evidential classification head enabling principled uncertainty decomposition. External validation on an independent dataset achieved 99.90% accuracy, demonstrating robust cross-domain generalization. Error analysis confirms that misclassified cases exhibit significantly elevated uncertainty that would trigger human review in clinical workflows. Complete source code and pretrained models are publicly available at <https://github.com/tarequejosh/HSANet-Brain-Tumor-Classification>.

# CRediT Author Statement

**Md. Assaduzzaman:** Conceptualization, Supervision, Methodology, Writing - Review & Editing. **Md. Tareque Jamil Josh:** Software, Validation, Formal analysis, Writing - Original Draft. **Md. Aminur Rahman Joy:** Data Curation, Visualization, Investigation. **Md. Nafish Imtiaz Imti:** Investigation, Resources, Validation.

# Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

# Acknowledgments

The authors thank Kaggle user Masoud Nickparvar for making the Brain Tumor MRI Dataset publicly available, and the creators of the Figshare Brain Tumor Dataset for enabling external validation.

# Data Availability

The Brain Tumor MRI Dataset is publicly available at <https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset>. The Figshare Brain Tumor Dataset is available at <https://figshare.com/articles/dataset/brain_tumor_dataset/1512427>. Source code and trained models are available at <https://github.com/tarequejosh/HSANet-Brain-Tumor-Classification>.

Aurna, Nahida Farjana, Mohammad Abu Yousuf, Kazi Ahmed Taher, AKM Azad, and Mohammad Ashraful Alam Momen. 2022. “A Classification of MRI Brain Tumor Based on Two Stage Feature Level Ensemble of Deep CNN Models.” *Computers in Biology and Medicine* 146: 105539.

Badža, Milica M, and Marko Č Barjaktarović. 2020. “Classification of Brain Tumors from MRI Images Using a Convolutional Neural Network.” *Applied Sciences* 10 (6): 1999.

Bruno, Michael A, Eric A Walker, and Hani H Abujudeh. 2015. “Understanding and Confronting Our Mistakes: The Epidemiology of Error in Radiology and Strategies for Error Reduction.” *Radiographics* 35 (6): 1668–76.

Chen, Liang-Chieh, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. 2018. “Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation.” *Proceedings of the European Conference on Computer Vision (ECCV)*, 801–18.

Cheng, Jun, Wei Huang, Shuangliang Cao, Ru Yang, Wei Yang, Zhaoqiang Yun, Zhijian Wang, and Qianjin Feng. 2015. “Enhanced Performance of Brain Tumor Classification via Tumor Region Augmentation and Partition.” *PloS One* 10 (10): e0140381.

Deepak, S, and PM Ameer. 2019. “Brain Tumor Classification Using Deep CNN Features via Transfer Learning.” *Computers in Biology and Medicine* 111: 103345.

Gal, Yarin, and Zoubin Ghahramani. 2016. “Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning.” In *International Conference on Machine Learning*, 1050–59. PMLR.

Hu, Jie, Li Shen, and Gang Sun. 2018. “Squeeze-and-Excitation Networks.” In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7132–41.

Kibriya, Hareem, Momina Masood, Marriam Nawaz, and Mehwish Rehman. 2022. “A Novel and Effective Brain Tumor Classification Model Using Deep Feature Fusion and Famous Machine Learning Classifiers.” *Computational Intelligence and Neuroscience* 2022: 7897669.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton. 2012. “ImageNet Classification with Deep Convolutional Neural Networks.” In *Advances in Neural Information Processing Systems*, 25:1097–1105.

Lakshminarayanan, Balaji, Alexander Pritzel, and Charles Blundell. 2017. “Simple and Scalable Predictive Uncertainty Estimation Using Deep Ensembles.” In *Advances in Neural Information Processing Systems*. Vol. 30.

Landis, J Richard, and Gary G Koch. 1977. “The Measurement of Observer Agreement for Categorical Data.” *Biometrics*, 159–74.

Leeuwen, Kicky G van, Steven Schalekamp, Matthieu JCM Rutten, Pieter Snoeren, Maarten de Rooij, Jasper JHM Gommers, and Cornelia M Schaefer-Prokop. 2021. “Artificial Intelligence in Radiology: 100 Commercially Available Products and Their Scientific Evidence.” *European Radiology* 31 (6): 3797–3804.

Leibig, Christian, Vaneeda Allken, Murat Seçkin Ayhan, Philipp Berens, and Siegfried Wahl. 2017. “Leveraging Uncertainty Estimates for Predicting Segmentation Quality.” *arXiv Preprint arXiv:1709.06116*.

Lin, Tsung-Yi, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. 2017. “Feature Pyramid Networks for Object Detection.” In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2117–25.

Lin, Tsung-Yi, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. 2017. “Focal Loss for Dense Object Detection.” In *Proceedings of the IEEE International Conference on Computer Vision*, 2980–88.

Louis, David N, Arie Perry, Pieter Wesseling, Daniel J Brat, Ian A Cree, Dominique Figarella-Branger, Cynthia Hawkins, et al. 2021. “The 2021 WHO Classification of Tumors of the Central Nervous System: A Summary.” *Neuro-Oncology* 23 (8): 1231–51.

Mohsen, Heba, El-Sayed A El-Dahshan, El-Sayed M El-Horbaty, and Abdel-Badeeh M Salem. 2018. “Classification Using Deep Learning Neural Networks for Brain Tumors.” *Future Computing and Informatics Journal* 3 (1): 68–71.

Neal, Radford M. 2012. *Bayesian Learning for Neural Networks*. Vol. 118. Springer Science & Business Media.

Nickparvar, Masoud. 2021. “Brain Tumor MRI Dataset.” <https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset>.

Oktay, Ozan, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, et al. 2018. “Attention U-Net: Learning Where to Look for the Pancreas.” *arXiv Preprint arXiv:1804.03999*.

Ostrom, Quinn T, Nirav Patil, Gino Cioffi, Kristin Waite, Carol Kruchko, and Jill S Barnholtz-Sloan. 2021. “CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2014–2018.” *Neuro-Oncology* 23 (Supplement\_3): iii1–105.

Pope, Whitney B. 2018. “Brain Tumor Imaging.” *Seminars in Neurology* 38 (1): 11–24.

Raghu, Maithra, Chiyuan Zhang, Jon Kleinberg, and Samy Bengio. 2019. “Transfusion: Understanding Transfer Learning for Medical Imaging.” In *Advances in Neural Information Processing Systems*. Vol. 32.

Rehman, Arshia, Saeeda Naz, Muhammad Imran Razzak, Faiza Akram, and Muhammad Imran. 2020. “A Deep Learning-Based Framework for Automatic Brain Tumors Classification Using Transfer Learning.” *Circuits, Systems, and Signal Processing* 39 (2): 757–75.

Rimmer, Abi. 2017. “Radiologist Shortage Leaves Patients Waiting for Diagnoses.” *BMJ* 359: j4683.

Saeedi, Soheila, Sorayya Rezayi, Habib Keshavarz, and Sharareh R Niakan Kalhori. 2023. “MRI-Based Brain Tumor Detection Using Convolutional Deep Learning Methods and Chosen Machine Learning Techniques.” *BMC Medical Informatics and Decision Making* 23 (1): 16.

Selvaraju, Ramprasaath R, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. 2017. “Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization.” In *Proceedings of the IEEE International Conference on Computer Vision*, 618–26.

Sensoy, Murat, Lance Kaplan, and Melih Kandemir. 2018. “Evidential Deep Learning to Quantify Classification Uncertainty.” In *Advances in Neural Information Processing Systems*. Vol. 31.

Sung, Hyuna, Jacques Ferlay, Rebecca L Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. 2021. “Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries.” *CA: A Cancer Journal for Clinicians* 71 (3): 209–49.

Swati, Zar Nawab Khan, Qinghua Zhao, Muhammad Kabir, Farman Ali, Zakir Ali, Saeed Ahmed, and Jianfeng Lu. 2019. “Brain Tumor Classification for MR Images Using Transfer Learning and Fine-Tuning.” *Computerized Medical Imaging and Graphics* 75: 34–46.

Tan, Mingxing, and Quoc Le. 2019. “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks.” In *International Conference on Machine Learning*, 6105–14. PMLR.

Woo, Sanghyun, Jongchan Park, Joon-Young Lee, and In So Kweon. 2018. “CBAM: Convolutional Block Attention Module.” In *Proceedings of the European Conference on Computer Vision (ECCV)*, 3–19.

Yu, Fisher, and Vladlen Koltun. 2016. “Multi-Scale Context Aggregation by Dilated Convolutions.” *arXiv Preprint arXiv:1511.07122*.