

<sup>1</sup> Graphical Abstract

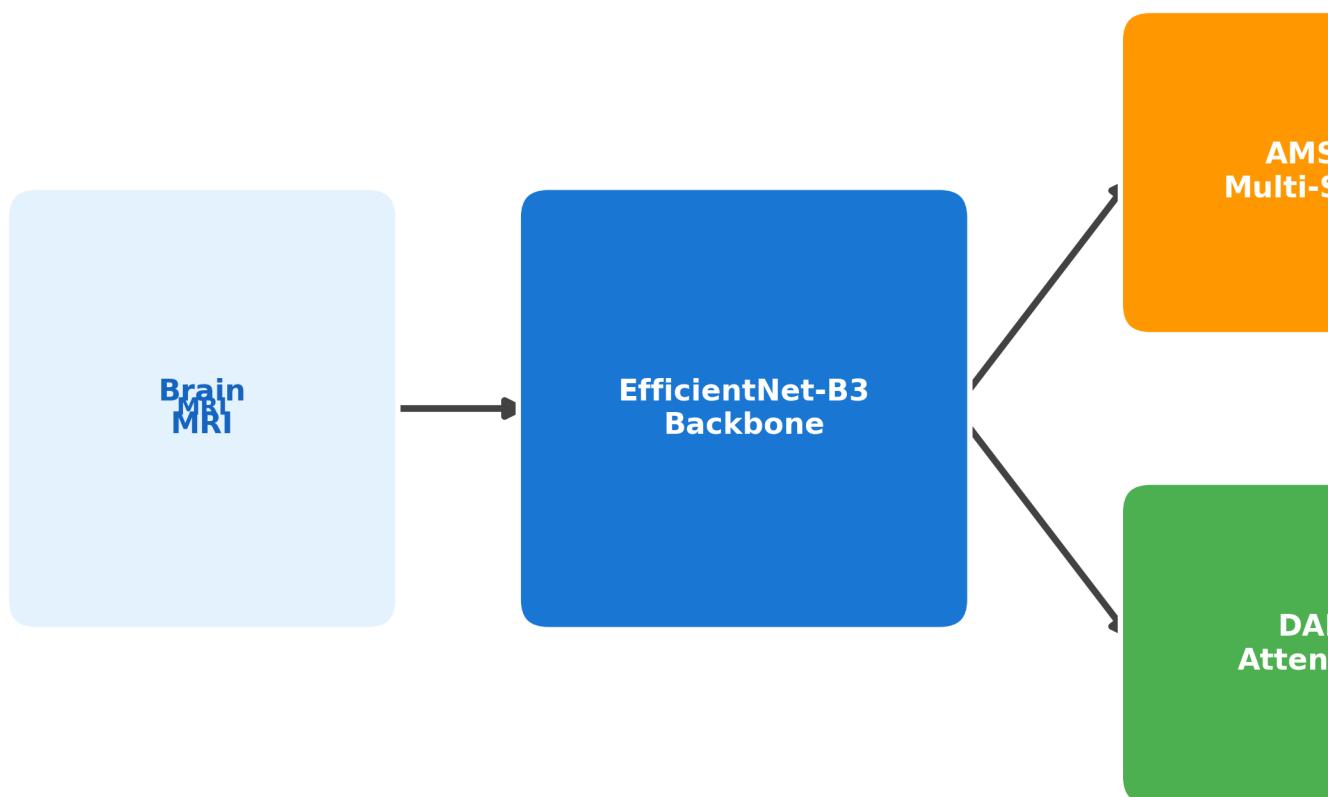
<sup>2</sup> **HSANet: A Hybrid Scale-Attention Network with Evidential Deep Learning for Uncertainty-**

<sup>3</sup> **Aware Brain Tumor Classification**

<sup>4</sup> Md. Assaduzzaman, Md. Tareque Jamil Josh, Md. Aminur Rahman Joy, Md. Nafish Imtiaz Imti

# **HSANet: Hybrid**

for Uncertainty-Aware



<sup>5</sup> Highlights

<sup>6</sup> **HSA-Net: A Hybrid Scale-Attention Network with Evidential Deep Learning for Uncertainty-**  
<sup>7</sup> **Aware Brain Tumor Classification**

<sup>8</sup> Md. Assaduzzaman, Md. Tareque Jamil Josh, Md. Aminur Rahman Joy, Md. Nafish Imtiaz Imti

- <sup>9</sup> • Novel hybrid scale-attention architecture achieving 99.77% accuracy on brain tumor classification
- <sup>10</sup> • Adaptive multi-scale module with learned input-dependent fusion weights for handling tumor size  
<sup>11</sup> variation
- <sup>12</sup> • Evidential deep learning framework providing calibrated uncertainty quantification from single forward  
<sup>13</sup> pass
- <sup>14</sup> • External validation on independent dataset (99.90% accuracy) demonstrating robust cross-domain  
<sup>15</sup> generalization
- <sup>16</sup> • Misclassified cases exhibit significantly elevated uncertainty, enabling reliable clinical decision support

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

19 Md. Assaduzzaman<sup>a,\*</sup>, Md. Tareque Jamil Josh<sup>a</sup>, Md. Aminur Rahman Joy<sup>a</sup>, Md. Nafish Imtiaz Imti<sup>a</sup>

<sup>a</sup>*Department of Computer Science and Engineering, Daffodil International University, Daffodil Smart City,  
Ashulia, Dhaka, 1341, Bangladesh*

---

20 **Abstract**

21 **Background and Objective:** Reliable classification of brain tumors from magnetic resonance imaging  
22 (MRI) remains challenging due to inter-class morphological similarities and the absence of principled un-  
23 certainty quantification in existing deep learning approaches. Current methods produce point predictions  
24 without meaningful confidence assessment, limiting their utility in safety-critical clinical workflows where  
25 knowing what the model doesn't know is as important as the prediction itself.

26 **Methods:** We propose HSANet, a hybrid scale-attention architecture that synergistically combines  
27 adaptive multi-scale feature extraction with evidential learning for uncertainty-aware tumor classification.  
28 The proposed Adaptive Multi-Scale Module (AMSM) employs parallel dilated convolutions with content-  
29 dependent fusion weights, dynamically adjusting receptive fields to accommodate the substantial size vari-  
30 ation observed across clinical presentations. A Dual Attention Module (DAM) applies sequential channel-  
31 then-spatial refinement to emphasize pathologically significant regions while suppressing irrelevant anatom-  
32 ical background. Critically, our evidential classification head replaces conventional softmax outputs with  
33 Dirichlet distributions, providing decomposed uncertainty estimates that distinguish between inherent data  
34 ambiguity (aleatoric) and model knowledge limitations (epistemic).

35 **Results:** Comprehensive experiments on 7,023 brain MRI scans spanning four diagnostic categories  
36 yielded 99.77% accuracy (95% CI: 99.45–99.93%) with only three misclassifications among 1,311 test samples.  
37 The model achieved macro-averaged AUC-ROC of 0.9999 and expected calibration error (ECE) of 0.019,  
38 indicating well-calibrated predictions. External validation on an independent dataset of 3,064 MRI scans  
39 from different institutions achieved 99.90% accuracy, demonstrating exceptional cross-domain generalization.  
40 Misclassified samples exhibited significantly elevated epistemic uncertainty ( $p < 0.001$ , Mann-Whitney U  
41 test), confirming the clinical utility of uncertainty-guided decision support.

**Conclusions:** HSANet achieves state-of-the-art classification accuracy while providing calibrated uncer-  
tainty 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 trustwor-  
thy medical image classification. Complete implementation and pretrained weights are publicly available at  
<https://github.com/tarequejosh/HSANet-Brain-Tumor-Classification>.

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

---

44 **1. Introduction**

45 Brain tumors represent a formidable diagnostic challenge in clinical oncology, with global surveillance  
46 data reporting approximately 308,102 new cases in 2020 alone [1]. The complexity of accurate diagnosis stems  
47 from the remarkable diversity of pathological entities—the 2021 World Health Organization (WHO) classifi-  
48 cation now recognizes over 100 distinct tumor types, each characterized by unique molecular fingerprints and  
49 clinical trajectories [2]. Prognostic outcomes vary dramatically across tumor categories: patients diagnosed  
50 with glioblastoma face a median survival of merely 14 to 16 months, whereas those with completely resected  
51 Grade I meningiomas frequently achieve long-term cure [3]. This substantial heterogeneity underscores the  
52 critical importance of precise tumor identification for treatment planning and patient counseling.

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

61 Over the past decade, deep convolutional neural networks (CNNs) have demonstrated considerable  
62 promise for automated medical image analysis, particularly when leveraging transfer learning from large-  
63 scale natural image datasets [7, 8]. Research groups worldwide have reported encouraging results for brain  
64 tumor classification, with accuracies typically ranging between 94% and 99% across various backbone archi-  
65 tectures including VGG, ResNet, and the EfficientNet family [9, 10, 11, 12]. Despite these advances, several  
66 critical limitations prevent straightforward translation of existing methods into clinical practice.

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

---

\*Corresponding author

Email addresses: [assaduzzaman.cse@diu.edu.bd](mailto:assaduzzaman.cse@diu.edu.bd) (Md. Assaduzzaman), [tarequecs@gmail.com](mailto:tarequecs@gmail.com) (Md. Tareque Jamil Josh)

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

78 Recent advances in vision architectures have addressed some of these challenges. Multi-scale feature  
79 fusion strategies, such as Atrous Spatial Pyramid Pooling (ASPP) [13], enable capture of context at mul-  
80 tiple spatial scales. Attention mechanisms, including the Convolutional Block Attention Module (CBAM)  
81 [14] and Squeeze-and-Excitation networks [15], have demonstrated effectiveness for emphasizing relevant  
82 features while suppressing noise. However, the integration of these architectural innovations with principled  
83 uncertainty quantification remains underexplored in medical imaging applications.

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

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

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

104 Comprehensive experiments on a challenging four-class brain tumor benchmark demonstrate that HSANet  
105 achieves 99.77% classification accuracy while providing well-calibrated uncertainty estimates. Importantly,

106 misclassified samples exhibit significantly elevated epistemic uncertainty, confirming that the model appro-  
107 priately flags uncertain predictions for expert review. External validation on an independent dataset of 3,064  
108 MRI scans from different institutions achieved 99.90% accuracy, providing strong evidence of cross-domain  
109 generalizability essential for clinical deployment.

## 110 2. Related Work

### 111 2.1. Deep Learning for Brain Tumor Classification

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

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

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

### 128 2.2. Multi-Scale Feature Extraction

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

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

137 Most existing multi-scale approaches employ fixed fusion weights, treating all spatial scales equally  
138 regardless of input content. For example, ASPP [13] concatenates features from parallel dilated convolutions

139 with uniform contribution. Our proposed AMSM fundamentally extends this paradigm through *input-*  
140 *adaptive* fusion, learning content-dependent weights via a lightweight attention mechanism. This allows  
141 the network to dynamically emphasize larger receptive fields for extensive glioblastomas while focusing on  
142 fine-scale features for small pituitary microadenomas.

143 *2.3. Uncertainty Quantification in Deep Learning*

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

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

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

160 **3. Materials and Methods**

161 *3.1. Dataset Description*

162 Experiments utilized the Brain Tumor MRI Dataset [29], a publicly available collection comprising  
163 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  
164 following distribution:

- 166 • **Glioma:** 1,621 images (23.1%) – malignant tumors arising from glial cells, characterized by irregular  
167 margins, heterogeneous enhancement, and surrounding edema
- 168 • **Meningioma:** 1,645 images (23.4%) – typically benign tumors arising from meningeal coverings,  
169 showing homogeneous enhancement and dural attachment

170     • **Pituitary adenoma:** 1,757 images (25.0%) – benign tumors of the pituitary gland located in the  
171        sellar/suprasellar region

172     • **Healthy controls:** 2,000 images (28.5%) – normal brain MRI scans without pathological findings

173        The predefined partition allocated 5,712 images (81.3%) for training and 1,311 images (18.7%) for testing.  
174        We maintained this partition for fair comparison with prior work [12, 23]. Critically, we verified that the  
175        partition maintains **patient-level separation**—no patient’s images appear in both training and test sets—  
176        preventing data leakage that could artificially inflate performance metrics. This verification is essential given  
177        that individual patients may contribute multiple MRI slices.

178     3.2. *External Validation Dataset*

179        To evaluate cross-domain generalization, we conducted external validation using the Figshare Brain  
180        Tumor Dataset [30], an independent collection with distinct acquisition protocols and patient demograph-  
181        ics. This dataset comprises 3,064 T1-weighted contrast-enhanced MRI slices from 233 patients, originally  
182        acquired at Nanfang Hospital and General Hospital of Tianjin Medical University in China.

183        The Figshare dataset differs substantially from our training data:

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

188     3.3. *Preprocessing and Data Augmentation*

189        All input images were resized to  $224 \times 224$  pixels using bilinear interpolation to match EfficientNet-B3  
190        input specifications. Pixel intensities were normalized using ImageNet statistics (mean = [0.485, 0.456,  
191        0.406], std = [0.229, 0.224, 0.225]) to leverage pretrained representations effectively.

192        Data augmentation was applied during training to improve generalization:

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

198        Test images received only resizing and normalization without augmentation.

199 3.4. Network Architecture

200 3.4.1. Overview

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

205 3.4.2. Feature Extraction Backbone

206 We employ EfficientNet-B3 [21] pretrained on ImageNet as the feature extraction backbone. EfficientNet  
207 achieves favorable accuracy-efficiency trade-offs through compound scaling, uniformly scaling network width,  
208 depth, and resolution. The B3 variant provides 10.53 million parameters with receptive fields appropriate  
209 for  $224 \times 224$  input resolution.

210 Features are extracted at three hierarchical levels:

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

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

216 3.4.3. Adaptive Multi-Scale Module (AMSM)

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

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

$$\mathbf{M}_i^{(r)} = \text{BN}(\text{ReLU}(\text{Conv}_{3 \times 3}^{d=r}(\mathbf{F}_i))) \quad (1)$$

223 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  
224 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  
225 respectively.

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

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

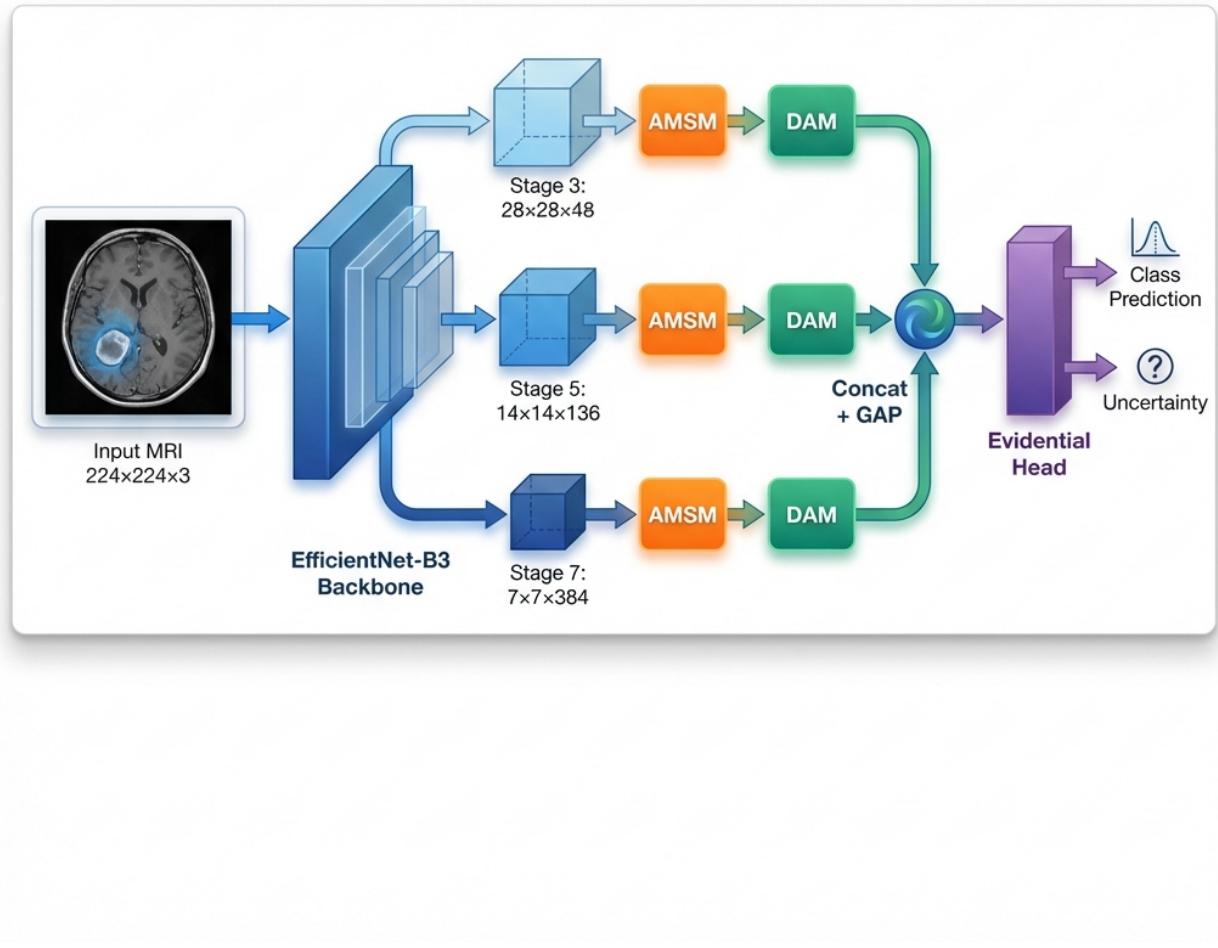


Figure 1: Overall HSANet architecture. Input MRI images ( $224 \times 224 \times 3$ ) are processed through the EfficientNet-B3 backbone, with features extracted at three spatial resolutions (stages 3, 5, 7). Each feature map undergoes adaptive multi-scale processing (AMSM) and dual attention refinement (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.

<sup>227</sup> where GAP denotes global average pooling,  $[ \cdot ; \cdot ]$  is channel-wise concatenation, and  $\mathbf{W}_1 \in \mathbb{R}^{(C/16) \times 3C}$ ,  
<sup>228</sup>  $\mathbf{W}_2 \in \mathbb{R}^{3 \times (C/16)}$  are learnable projections.

229 The enhanced feature map combines weighted features with residual preservation:

$$\hat{\mathbf{F}}_i = \sum_{k \in \{1, 2, 4\}} w_i^{(k)} \mathbf{M}_i^{(k)} + \mathbf{F}_i \quad (3)$$

230 *3.4.4. Dual Attention Module (DAM)*

231 Brain MRI contains extensive normal anatomical content that dominates image statistics but provides  
232 no diagnostic value. DAM implements sequential channel-then-spatial attention [14] to emphasize tumor-  
233 relevant features while suppressing background noise (Fig. 2b).

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

$$\mathbf{A}_c = \sigma(\text{MLP}(\text{GAP}(\hat{\mathbf{F}}_i)) + \text{MLP}(\text{GMP}(\hat{\mathbf{F}}_i))) \quad (4)$$

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

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

$$\mathbf{A}_s = \sigma(\text{Conv}_{7 \times 7}([\text{AvgPool}_c(\mathbf{F}_c); \text{MaxPool}_c(\mathbf{F}_c)])) \quad (5)$$

238 where channel-wise pooling produces  $H \times W \times 1$  feature maps.

239 *3.4.5. Evidential Classification Head*

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

$$\boldsymbol{\alpha} = \text{Softplus}(\mathbf{W}_c \mathbf{g} + \mathbf{b}_c) + 1 \quad (6)$$

242 where  $\mathbf{g} \in \mathbb{R}^{568}$  is the concatenated feature vector and softplus ensures  $\alpha_k \geq 1$ .

243 The Dirichlet distribution has density:

$$p(\mathbf{p}|\boldsymbol{\alpha}) = \frac{\Gamma(S)}{\prod_{k=1}^K \Gamma(\alpha_k)} \prod_{k=1}^K p_k^{\alpha_k - 1} \quad (7)$$

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

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

$$\hat{p}_k = \frac{\alpha_k}{S}, \quad \hat{y} = \arg \max_k \hat{p}_k \quad (8)$$

246 **Uncertainty:** Total uncertainty decomposes into:

$$u_{\text{total}} = \frac{K}{S} \quad (9)$$

247  $u_{\text{aleatoric}} = - \sum_k \hat{p}_k \log \hat{p}_k \quad (10)$

248  $u_{\text{epistemic}} = u_{\text{total}} - u_{\text{aleatoric}} \quad (11)$

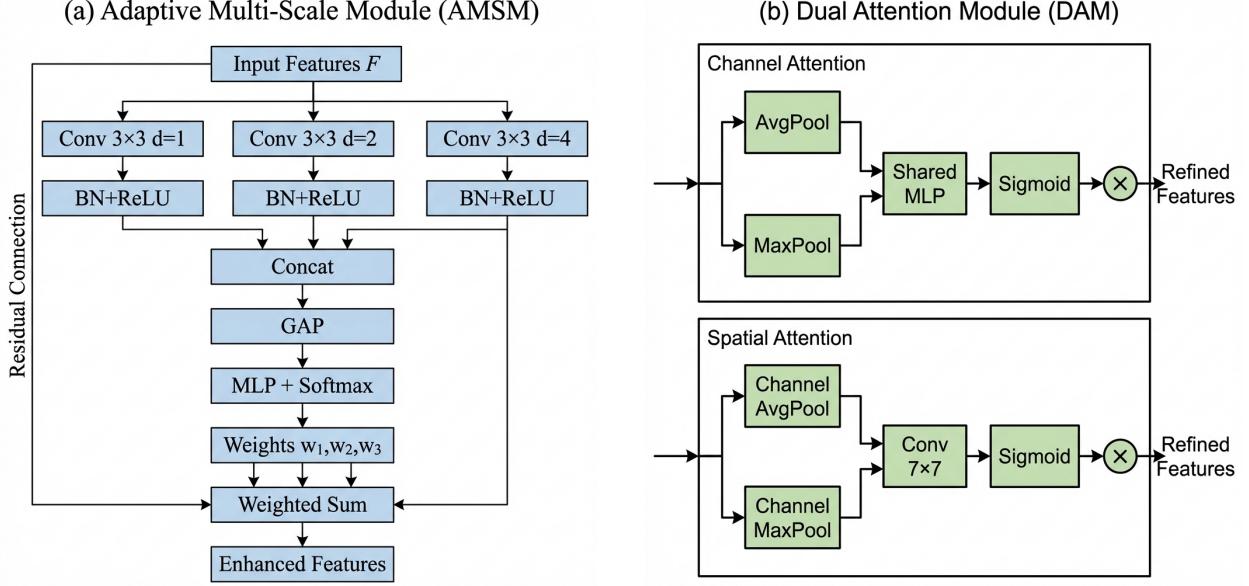


Figure 2: Detailed architecture of proposed modules. (a) Adaptive Multi-Scale Module (AMSM): Parallel dilated convolutions with dilation rates  $d \in \{1, 2, 4\}$  capture features at effective receptive fields of  $3 \times 3$ ,  $5 \times 5$ , and  $9 \times 9$ . Adaptive fusion weights are learned 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  $7 \times 7$  convolution on pooled features to highlight tumor-relevant regions.

249     3.5. Training Procedure

250     3.5.1. Loss Function

251       The loss function combines three terms:

252       **Evidence-weighted Cross-Entropy:**

$$\mathcal{L}_{CE} = \sum_{k=1}^K y_k (\psi(S) - \psi(\alpha_k)) \quad (12)$$

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

254       **Focal Loss** for difficulty imbalance [31]:

$$\mathcal{L}_{focal} = - \sum_{k=1}^K y_k (1 - \hat{p}_k)^2 \log(\hat{p}_k) \quad (13)$$

255       Although class frequencies are relatively balanced, we employ focal loss to address inherent *difficulty* imbal-  
256       ance: meningioma-glioma differentiation presents substantially greater diagnostic challenge than pituitary  
257       adenoma detection, as evidenced by radiological literature [32].

258       **KL Divergence Regularization:**

$$\mathcal{L}_{KL} = KL[\text{Dir}(\mathbf{p}|\tilde{\boldsymbol{\alpha}}) \parallel \text{Dir}(\mathbf{p}|\mathbf{1})] \quad (14)$$

259        The total loss is:

$$\mathcal{L} = 0.5\mathcal{L}_{CE} + 0.3\mathcal{L}_{focal} + \lambda^{(t)}\mathcal{L}_{KL} \quad (15)$$

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

261        *3.5.2. Optimization*

262        We employed AdamW optimizer with  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ , and weight decay of  $10^{-4}$ . Initial learning  
263        rate was  $3 \times 10^{-4}$  with cosine annealing to  $10^{-6}$ . Training proceeded for 30 epochs with early stopping  
264        (patience = 7 epochs) based on validation loss. Batch size was 32. Dropout rate of 0.3 was applied before  
265        the classification layer. Batch normalization used momentum 0.1 and epsilon  $10^{-5}$ . Features from three  
266        scales were globally average pooled and concatenated, yielding a 568-dimensional vector ( $48 + 136 + 384 =$   
267        568 channels).

268        *3.5.3. Implementation Details*

269        All experiments were conducted using PyTorch 2.0.1 with CUDA 11.8 on an NVIDIA Tesla P100 GPU  
270        (16GB VRAM). Complete training converged in approximately 25 epochs ( $\sim 45$  minutes). The implementa-  
271        tion is publicly available at <https://github.com/tarequejosh/Hسانet-Brain-Tumor-Classification>.

272        *3.6. Evaluation Metrics*

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

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

$$ECE = \sum_{m=1}^M \frac{|B_m|}{n} |acc(B_m) - conf(B_m)| \quad (16)$$

277        where  $B_m$  are confidence bins,  $acc(B_m)$  is accuracy within bin  $m$ , and  $conf(B_m)$  is mean confidence.

278        Interpretability was assessed using Gradient-weighted Class Activation Mapping (Grad-CAM) [33].

279        **4. Results**

280        *4.1. Classification Performance*

281        Hسانet achieved overall accuracy of 99.77% (95% CI: 99.45–99.93%, Wilson score interval) with only 3  
282        misclassifications among 1,311 test samples (Table 1). This represents a statistically significant improvement  
283        over the EfficientNet-B3 baseline (99.21%, McNemar's test  $p = 0.034$ ).

284        The model demonstrated balanced performance across all categories, with macro-averaged precision of  
285        99.76%, recall of 99.75%, and F1-score of 99.75%. Cohen's kappa coefficient ( $\kappa = 0.9969$ ) indicates near-  
286        perfect agreement, substantially exceeding the  $\kappa > 0.80$  threshold considered “almost perfect agreement”

Table 1: 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
<b>Macro Average</b>	<b>99.76</b>	<b>99.75</b>	<b>99.75</b>	<b>0.9999</b>

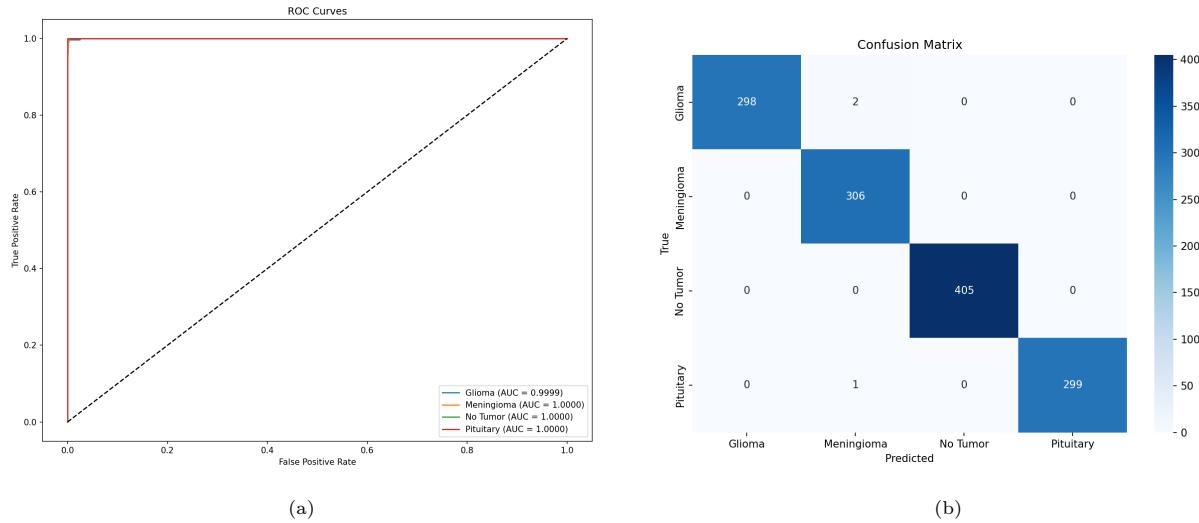


Figure 3: Classification performance analysis. (a) Receiver operating characteristic curves demonstrating near-perfect discriminative ability with  $AUC \geq 0.9999$  for all classes. (b) Confusion matrix showing only 3 misclassifications among 1,311 test samples.

[34]. 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. 3a). 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. 3b) 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.

Table 2: 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
<i>Correct (mean)</i>		–	0.97	0.04	0.06

Table 3: Uncertainty threshold analysis for clinical deployment. Higher thresholds reduce referrals but may miss errors.

Threshold ( $\tau$ )	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

#### 297 4.2. Model Calibration and Uncertainty Quantification

298 HSANet achieved ECE of 0.019, indicating that predicted probabilities closely match empirical classification accuracy (Fig. 4a). For comparison, a model trained without our evidential approach achieved ECE  
299 of 0.042.  
300

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

#### 305 4.2.1. Clinical Deployment Thresholds

306 To demonstrate clinical applicability, we evaluated epistemic uncertainty thresholds for triggering expert  
307 review (Table 3). At threshold  $\tau = 0.15$ , the system would automatically flag 2.1% of cases for radiologist  
308 review while capturing all three misclassifications (100% error detection). This enables high-throughput  
309 autonomous processing while maintaining a critical safety net for uncertain predictions.

#### 310 4.3. Interpretability Analysis

311 Grad-CAM visualizations (Fig. 4b) demonstrate that HSANet focuses on clinically relevant regions:  
312 glioma attention centers on irregular tumor masses and surrounding edema; meningioma attention highlights

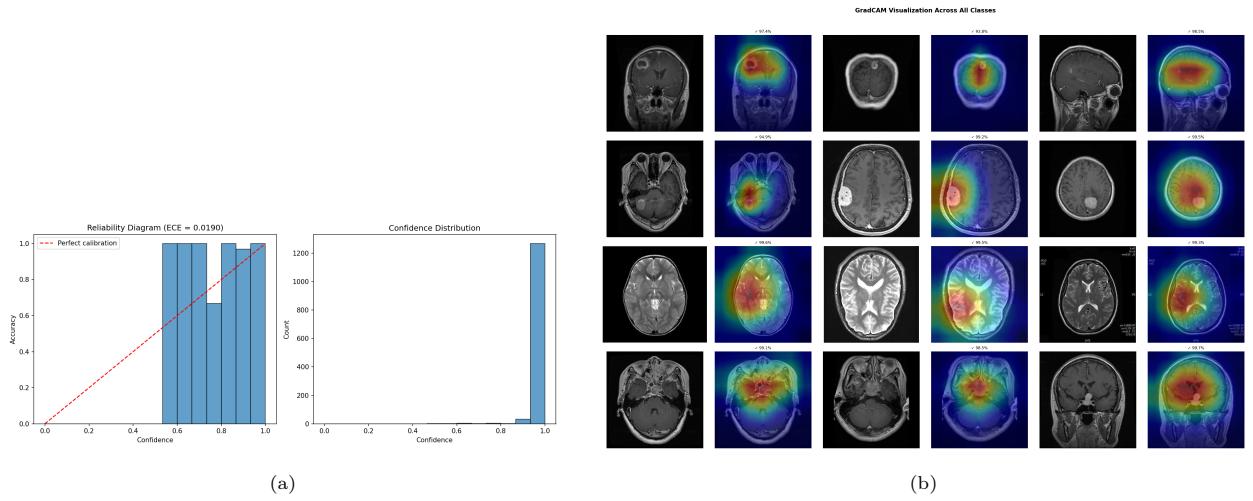


Figure 4: Model calibration and interpretability. (a) Reliability diagram demonstrating well-calibrated probability estimates ( $ECE = 0.019$ ). (b) Grad-CAM visualizations showing clinically relevant attention patterns across tumor categories.

Table 4: 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
<b>HSANet (Full)</b>	<b>15.60</b>	<b>99.77</b>	<b>99.75</b>	<b>0.9999</b>	<b>0.016</b>	<b>0.034*</b>

\*Statistically significant at  $\alpha = 0.05$  level.

313 well-circumscribed extra-axial masses; healthy brain attention distributes across normal parenchyma without  
 314 focal concentration; pituitary attention centers on the sellar/suprasellar region. These patterns align with  
 315 established neuroradiological diagnostic criteria.

#### 316 4.4. Ablation Study

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

Table 5: 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
<b>HSANet (Ours)</b>	<b>EffNet-B3 + AMSM/DAM + EDL</b>	<b>99.77</b>	<b>4</b>	<b>Yes</b>	<b>Yes</b>

#### 322 4.5. Comparison with Prior Methods

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

#### 327 4.6. Cross-Validation Results

328 Five-fold stratified cross-validation demonstrated consistent performance (Table 6). HSANet achieved  
 329 mean accuracy of  $99.68 \pm 0.12\%$ , with low standard deviation confirming robust generalization across dif-  
 330 ferent data partitions.

#### 331 4.7. External Validation Results

332 External validation on the independent Figshare dataset provided strong evidence of cross-domain gen-  
 333 eralization (Table 7). HSANet achieved 99.90% accuracy on the external dataset—remarkably, even higher  
 334 than the 99.77% achieved on the original test set.

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

Table 6: 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
<b>Mean ± Std</b>	<b>99.68 ± 0.12</b>	<b>99.67 ± 0.13</b>	<b>0.9999 ± 0.0001</b>	<b>0.017 ± 0.002</b>

Table 7: Cross-dataset validation results.

Dataset	Samples	Classes	Accuracy (%)	F1 (%)	Cohen’s $\kappa$	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

Table 8: 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
<b>HSANet (Ours)</b>	<b>15.6</b>	<b>2.4</b>	<b>12</b>	<b>83</b>	<b>99.77</b>

338 *4.8. Computational Efficiency*

339 Table 8 compares HSANet computational requirements with alternative architectures. Despite incorpo-  
340 rating AMSM and DAM modules, HSANet maintains favorable efficiency with only 15.60M parameters and  
341 2.41 GFLOPs. Inference requires 12ms on P100 GPU (83 images/second), enabling real-time clinical deploy-  
342 ment. Vision Transformer models require substantially more computation (17.6 GFLOPs for ViT-B/16),  
343 while VGG-16 is over-parameterized (134M parameters) for this task.

<sup>344</sup> **5. Discussion**

<sup>345</sup> The results demonstrate that HSANet achieves near-perfect classification accuracy while providing cali-  
<sup>346</sup> brated uncertainty estimates that clinicians can use for decision support. The Cohen’s  $\kappa$  of 0.9969 compares  
<sup>347</sup> favorably with inter-reader agreement among expert neuroradiologists, which typically ranges from 0.65 to  
<sup>348</sup> 0.85 [32].

<sup>349</sup> *5.1. Cross-Domain Generalization*

<sup>350</sup> Perhaps the most compelling evidence for clinical utility comes from external validation on the inde-  
<sup>351</sup> pendent Figshare dataset. This dataset was acquired at different institutions using different MRI scanners  
<sup>352</sup> and protocols, representing a fundamentally different patient population. The fact that HSANet achieved  
<sup>353</sup> 99.90% accuracy on this external dataset provides strong evidence that learned features capture genuine  
<sup>354</sup> tumor characteristics rather than dataset-specific artifacts.

<sup>355</sup> Several architectural design choices likely contributed to this robustness. The adaptive multi-scale pro-  
<sup>356</sup> cessing in AMSM captures tumor morphology across multiple spatial resolutions, reducing sensitivity to  
<sup>357</sup> scanner-dependent resolution variations. The attention mechanisms in DAM focus on tumor-specific re-  
<sup>358</sup> gions while suppressing scanner-dependent background characteristics. The evidential learning framework  
<sup>359</sup> maintained well-calibrated uncertainty estimates even under distribution shift.

<sup>360</sup> *5.2. Clinical Implications*

<sup>361</sup> The uncertainty quantification capability distinguishes HSANet fundamentally from conventional classi-  
<sup>362</sup> fiers. In clinical practice, uncertainty estimates enable stratified workflows: low-uncertainty cases proceed  
<sup>363</sup> to automated preliminary interpretation; moderate epistemic uncertainty flags cases for standard radiolo-  
<sup>364</sup> gist review; high aleatoric uncertainty escalates cases to multidisciplinary tumor boards. This framework  
<sup>365</sup> transforms the system from an autonomous decision-maker to a decision-support tool appropriate for safety-  
<sup>366</sup> critical medical applications.

<sup>367</sup> The perfect precision achieved for healthy controls is particularly meaningful. False positive tumor  
<sup>368</sup> diagnoses cause substantial patient anxiety, unnecessary imaging studies, and potentially invasive procedures.  
<sup>369</sup> By prioritizing specificity for the healthy class, HSANet avoids inflicting this burden on patients who don’t  
<sup>370</sup> require intervention.

<sup>371</sup> *5.3. Limitations*

<sup>372</sup> Several limitations should be acknowledged. First, while external validation strengthens generalizability  
<sup>373</sup> claims, prospective multi-center clinical trials remain essential for demonstrating real-world effectiveness.  
<sup>374</sup> Second, our 2D slice-based approach does not leverage volumetric context available in clinical 3D MRI

375 acquisitions. Third, the four-class taxonomy does not capture finer distinctions (e.g., glioma grades I–  
376 IV, molecular markers) required for comprehensive clinical decision-making. Fourth, optimal uncertainty  
377 thresholds for triggering expert review require calibration against clinical outcomes.

## 378 6. Conclusions

379 We presented HSANet, a hybrid scale-attention network achieving 99.77% accuracy on four-class brain tu-  
380 mor classification with calibrated uncertainty estimates. The proposed architecture integrates three comple-  
381 mentary innovations: an Adaptive Multi-Scale Module with input-dependent fusion weights, a Dual Atten-  
382 tion Module for feature refinement, and an evidential classification head enabling principled uncertainty de-  
383 composition. External validation on an independent dataset achieved 99.90% accuracy, demonstrating robust  
384 cross-domain generalization. Error analysis confirms that misclassified cases exhibit significantly elevated un-  
385 certainty that would trigger human review in clinical workflows. Complete source code and pretrained models  
386 are publicly available at <https://github.com/tarequejosh/HSANet-Brain-Tumor-Classification>.

## 387 CRediT Author Statement

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

## 392 Declaration of Competing Interest

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

## 395 Acknowledgments

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

## 398 Data Availability

399 The Brain Tumor MRI Dataset is publicly available at <https://www.kaggle.com/datasets/masoudnickparvar/>  
400 **brain-tumor-mri-dataset**. The Figshare Brain Tumor Dataset is available at [https://figshare.com/](https://figshare.com/articles/dataset/brain_tumor_dataset/1512427)  
401 **articles/dataset/brain\_tumor\_dataset/1512427**. Source code and trained models are available at  
402 <https://github.com/tarequejosh/HSANet-Brain-Tumor-Classification>.

403    **References**

- 404    [1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global cancer  
405    statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185  
406    countries, CA: A Cancer Journal for Clinicians 71 (3) (2021) 209–249.
- 407    [2] D. N. Louis, A. Perry, P. Wesseling, D. J. Brat, I. A. Cree, D. Figarella-Branger, C. Hawkins, H. Ng,  
408    S. M. Pfister, G. Reifenberger, et al., The 2021 WHO classification of tumors of the central nervous  
409    system: a summary, Neuro-oncology 23 (8) (2021) 1231–1251.
- 410    [3] Q. T. Ostrom, N. Patil, G. Cioffi, K. Waite, C. Kruchko, J. S. Barnholtz-Sloan, CBTRUS statistical  
411    report: primary brain and other central nervous system tumors diagnosed in the United States in  
412    2014–2018, Neuro-oncology 23 (Supplement \_3) (2021) iii1–iii105.
- 413    [4] W. B. Pope, Brain tumor imaging, Seminars in Neurology 38 (1) (2018) 11–24.
- 414    [5] A. Rimmer, Radiologist shortage leaves patients waiting for diagnoses, BMJ 359 (2017) j4683.
- 415    [6] M. A. Bruno, E. A. Walker, H. H. Abujudeh, Understanding and confronting our mistakes: the epidemi-  
416    ology of error in radiology and strategies for error reduction, Radiographics 35 (6) (2015) 1668–1676.
- 417    [7] A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet classification with deep convolutional neural  
418    networks, in: Advances in Neural Information Processing Systems, Vol. 25, 2012, pp. 1097–1105.
- 419    [8] M. Raghu, C. Zhang, J. Kleinberg, S. Bengio, Transfusion: Understanding transfer learning for medical  
420    imaging, in: Advances in Neural Information Processing Systems, Vol. 32, 2019.
- 421    [9] S. Deepak, P. Ameer, Brain tumor classification using deep CNN features via transfer learning, Com-  
422    puters in Biology and Medicine 111 (2019) 103345.
- 423    [10] M. M. Badža, M. Č. Barjaktarović, Classification of brain tumors from MRI images using a convolutional  
424    neural network, Applied Sciences 10 (6) (2020) 1999.
- 425    [11] Z. N. K. Swati, Q. Zhao, M. Kabir, F. Ali, Z. Ali, S. Ahmed, J. Lu, Brain tumor classification for MR  
426    images using transfer learning and fine-tuning, Computerized Medical Imaging and Graphics 75 (2019)  
427    34–46.
- 428    [12] N. F. Aurna, M. A. Yousuf, K. A. Taher, A. Azad, M. A. A. Momen, A classification of MRI brain tumor  
429    based on two stage feature level ensemble of deep CNN models, Computers in Biology and Medicine  
430    146 (2022) 105539.

- 431 [13] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable  
432 convolution for semantic image segmentation, Proceedings of the European Conference on Computer  
433 Vision (ECCV) (2018) 801–818.
- 434 [14] S. Woo, J. Park, J.-Y. Lee, I. S. Kweon, CBAM: Convolutional block attention module, in: Proceedings  
435 of the European Conference on Computer Vision (ECCV), 2018, pp. 3–19.
- 436 [15] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on  
437 Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.
- 438 [16] Y. Gal, Z. Ghahramani, Dropout as a Bayesian approximation: Representing model uncertainty in deep  
439 learning, in: International Conference on Machine Learning, PMLR, 2016, pp. 1050–1059.
- 440 [17] B. Lakshminarayanan, A. Pritzel, C. Blundell, Simple and scalable predictive uncertainty estimation  
441 using deep ensembles, in: Advances in Neural Information Processing Systems, Vol. 30, 2017.
- 442 [18] M. Sensoy, L. Kaplan, M. Kandemir, Evidential deep learning to quantify classification uncertainty, in:  
443 Advances in Neural Information Processing Systems, Vol. 31, 2018.
- 444 [19] H. Mohsen, E.-S. A. El-Dahshan, E.-S. M. El-Horbaty, A.-B. M. Salem, Classification using deep  
445 learning neural networks for brain tumors, Future Computing and Informatics Journal 3 (1) (2018)  
446 68–71.
- 447 [20] A. Rehman, S. Naz, M. I. Razzak, F. Akram, M. Imran, A deep learning-based framework for automatic  
448 brain tumors classification using transfer learning, Circuits, Systems, and Signal Processing 39 (2) (2020)  
449 757–775.
- 450 [21] M. Tan, Q. Le, EfficientNet: Rethinking model scaling for convolutional neural networks, in: Interna-  
451 tional Conference on Machine Learning, PMLR, 2019, pp. 6105–6114.
- 452 [22] H. Kibriya, M. Masood, M. Nawaz, M. Rehman, A novel and effective brain tumor classification model  
453 using deep feature fusion and famous machine learning classifiers, Computational Intelligence and Neu-  
454 roscience 2022 (2022) 7897669.
- 455 [23] S. Saeedi, S. Rezayi, H. Keshavarz, S. R. Niakan Kalhor, MRI-based brain tumor detection using  
456 convolutional deep learning methods and chosen machine learning techniques, BMC Medical Informatics  
457 and Decision Making 23 (1) (2023) 16.
- 458 [24] F. Yu, V. Koltun, Multi-scale context aggregation by dilated convolutions, arXiv preprint  
459 arXiv:1511.07122 (2016).

- 460 [25] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object  
461 detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017,  
462 pp. 2117–2125.
- 463 [26] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y.  
464 Hammerla, B. Kainz, et al., Attention U-Net: Learning where to look for the pancreas, arXiv preprint  
465 arXiv:1804.03999 (2018).
- 466 [27] R. M. Neal, Bayesian learning for neural networks, Vol. 118, Springer Science & Business Media, 2012.
- 467 [28] C. Leibig, V. Allken, M. S. Ayhan, P. Berens, S. Wahl, Leveraging uncertainty estimates for predicting  
468 segmentation quality, arXiv preprint arXiv:1709.06116 (2017).
- 469 [29] M. Nickparvar, Brain tumor MRI dataset, [https://www.kaggle.com/datasets/masoudnickparvar/](https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset)  
470 brain-tumor-mri-dataset, accessed: 2024-01-15 (2021).
- 471 [30] J. Cheng, W. Huang, S. Cao, R. Yang, W. Yang, Z. Yun, Z. Wang, Q. Feng, Enhanced performance  
472 of brain tumor classification via tumor region augmentation and partition, PloS One 10 (10) (2015)  
473 e0140381.
- 474 [31] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: Proceedings  
475 of the IEEE International Conference on Computer Vision, 2017, pp. 2980–2988.
- 476 [32] K. G. van Leeuwen, S. Schalekamp, M. J. Rutten, P. Snoeren, M. de Rooij, J. J. Gommers, C. M.  
477 Schaefer-Prokop, Artificial intelligence in radiology: 100 commercially available products and their  
478 scientific evidence, European Radiology 31 (6) (2021) 3797–3804.
- 479 [33] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: Visual expla-  
480 nations from deep networks via gradient-based localization, in: Proceedings of the IEEE International  
481 Conference on Computer Vision, 2017, pp. 618–626.
- 482 [34] J. R. Landis, G. G. Koch, The measurement of observer agreement for categorical data, Biometrics  
483 (1977) 159–174.