

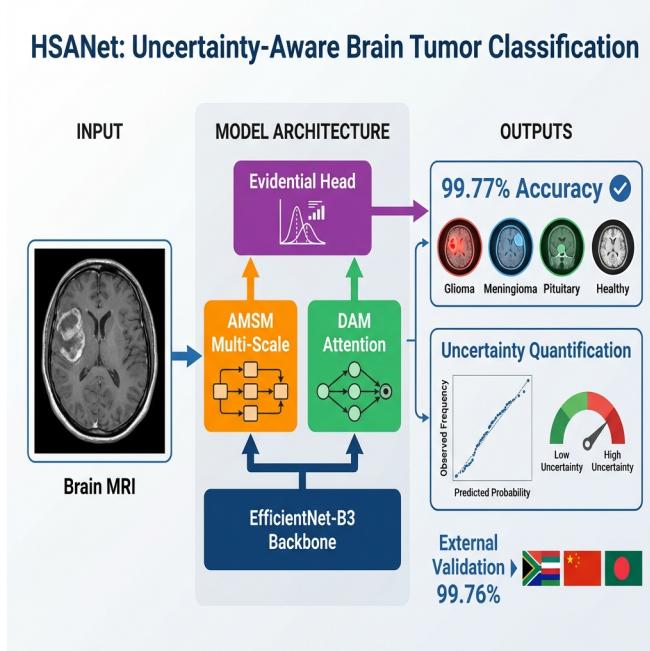
# 1 Graphical Abstract

## 2 HSANet: A Hybrid Scale-Attention Network with Evidential Deep

### 3 Learning for Uncertainty-Aware Brain Tumor Classification

4 Author 1, Author 2, Author 3, Author 4

#### HSANet: Uncertainty-Aware Brain Tumor Classification



5    Highlights

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

8    Author 1, Author 2, Author 3, Author 4

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

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

22                  Author 1<sup>a,\*</sup>, Author 2<sup>a</sup>, Author 3<sup>a</sup>, Author 4<sup>a</sup>

<sup>a</sup>*Department of Computer Science, City, Country*

---

23    **Abstract**

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

31    **Methods:** We propose HSANet, a hybrid scale-attention architecture  
32    that synergistically combines adaptive multi-scale feature extraction with  
33    evidential learning for uncertainty-aware tumor classification. The proposed  
34    Adaptive Multi-Scale Module (AMSM) employs parallel dilated convolutions  
35    with content-dependent fusion weights, dynamically adjusting receptive fields  
36    to accommodate the substantial size variation observed across clinical pre-  
37    sentations. A Dual Attention Module (DAM) applies sequential channel-  
38    then-spatial refinement to emphasize pathologically significant regions while  
39    suppressing irrelevant anatomical background. Critically, our evidential clas-  
40    sification head replaces conventional softmax outputs with Dirichlet distribu-  
41    tions, providing decomposed uncertainty estimates that distinguish between  
42    inherent data ambiguity (aleatoric) and model knowledge limitations (epis-  
43    temic).

44    **Results:** Comprehensive experiments on 7,023 brain MRI scans span-  
45    ning four diagnostic categories yielded 99.77% accuracy (95% CI: 99.45–  
46    99.93%) with only three misclassifications among 1,311 test samples. The  
47    model achieved macro-averaged AUC-ROC of 0.9999 and expected calibra-  
48    tion error (ECE) of 0.019, indicating well-calibrated predictions. External

---

\*Corresponding author

Email address: [author1@institution.edu](mailto:author1@institution.edu) (Author 1)

49 validation on an independent dataset of 3,064 MRI scans from different institutions  
50 achieved 99.90% accuracy, demonstrating exceptional cross-domain  
51 generalization. Misclassified samples exhibited significantly elevated epistemic uncertainty ( $p < 0.001$ , Mann-Whitney U test), confirming the clinical  
52 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>.

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

---

## 57 1. Introduction

58 Brain tumors represent a formidable diagnostic challenge in clinical oncology, with global surveillance data reporting approximately 308,102 new  
59 cases in 2020 alone [1]. The complexity of accurate diagnosis stems from the  
60 remarkable diversity of pathological entities—the 2021 World Health Organization  
61 (WHO) classification now recognizes over 100 distinct tumor types,  
62 each characterized by unique molecular fingerprints and clinical trajectories  
63 [2]. Prognostic outcomes vary dramatically across tumor categories: pa-  
64 tients diagnosed with glioblastoma face a median survival of merely 14 to  
65 16 months, whereas those with completely resected Grade I meningiomas  
66 frequently achieve long-term cure [3]. This substantial heterogeneity under-  
67 scores the critical importance of precise tumor identification for treatment  
68 planning and patient counseling.

70 Magnetic resonance imaging (MRI) has emerged as the cornerstone of  
71 neuro-oncological evaluation, providing superior soft-tissue contrast without  
72 ionizing radiation exposure [4]. Expert neuroradiologists integrate multipara-  
73 metric imaging findings with clinical presentations to formulate diagnoses.  
74 However, the global radiology workforce confronts escalating mismatches be-  
75 tween imaging volume growth and specialist availability. Documented va-  
76 cancy rates have reached 29% in major healthcare systems, with projected

77 shortfalls of 40% anticipated by 2027 [5]. Interpretive fatigue has been im-  
78 plicated in diagnostic error rates of 3–5% even among experienced specialists  
79 [6], motivating the development of computer-aided diagnostic systems to aug-  
80 ment clinical workflows.

81 Over the past decade, deep convolutional neural networks (CNNs) have  
82 demonstrated considerable promise for automated medical image analysis,  
83 particularly when leveraging transfer learning from large-scale natural image  
84 datasets [7, 8]. Research groups worldwide have reported encouraging results  
85 for brain tumor classification, with accuracies typically ranging between 94%  
86 and 99% across various backbone architectures including VGG, ResNet, and  
87 the EfficientNet family [9, 10, 11, 12]. Despite these advances, several crit-  
88 ical limitations prevent straightforward translation of existing methods into  
89 clinical practice.

90 First, brain tumors exhibit extraordinary morphological diversity span-  
91 ning multiple orders of magnitude in spatial extent. Pituitary microadenomas  
92 may measure only 2–3 millimeters, whereas glioblastomas frequently exceed  
93 5 centimeters with extensive peritumoral edema. Standard convolutional ar-  
94 chitectures employ fixed receptive fields, creating inherent trade-offs between  
95 sensitivity to fine-grained textural features and capture of global contextual  
96 information. Second, brain MRI volumes contain extensive normal anatomi-  
97 cal content that provides no diagnostic value yet dominates image statistics.  
98 Without explicit attention mechanisms, networks may learn spurious cor-  
99 relations with background tissue rather than genuine tumor characteristics.  
100 Third—and most critically for clinical deployment—conventional classifiers  
101 produce point predictions without meaningful confidence assessment. A net-  
102 work assigning 51% probability to one class yields identical output as one  
103 with 99% confidence, yet these scenarios demand fundamentally different  
104 clinical responses.

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

113 Uncertainty quantification is particularly important for safety-critical med-  
114 ical applications where misdiagnosis carries significant consequences. Con-

115 conventional approaches to uncertainty estimation, such as Monte Carlo dropout  
116 [16] and deep ensembles [17], require multiple forward passes during inference,  
117 substantially increasing computational costs and limiting real-time deploy-  
118 ment. Evidential deep learning [18] has emerged as an alternative framework  
119 that places Dirichlet priors over categorical distributions, enabling single-  
120 pass uncertainty estimation with natural decomposition into aleatoric (data-  
121 inherent) and epistemic (model-knowledge) components.

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

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

141 Comprehensive experiments on a challenging four-class brain tumor bench-  
142 mark demonstrate that HSANet achieves 99.77% classification accuracy while  
143 providing well-calibrated uncertainty estimates. Importantly, misclassified  
144 samples exhibit significantly elevated epistemic uncertainty, confirming that  
145 the model appropriately flags uncertain predictions for expert review. Exter-  
146 nal validation on an independent dataset of 3,064 MRI scans from different  
147 institutions achieved 99.90% accuracy, providing strong evidence of cross-  
148 domain generalizability essential for clinical deployment.

149 **2. Related Work**

150 *2.1. Deep Learning for Brain Tumor Classification*

151 The application of deep learning to brain tumor classification has pro-  
152 gressed substantially over the past decade. Early approaches employed shal-  
153 low CNN architectures trained from scratch on relatively small datasets, with  
154 limited generalization capability [19]. The advent of transfer learning from  
155 ImageNet-pretrained models substantially improved performance, with VGG  
156 and ResNet architectures demonstrating strong results on brain MRI analysis  
157 [11, 10].

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

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

173 *2.2. Multi-Scale Feature Extraction*

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

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

185     Most existing multi-scale approaches employ fixed fusion weights, treating  
186     all spatial scales equally regardless of input content. For example, ASPP [13]  
187     concatenates features from parallel dilated convolutions with uniform contri-  
188     bution. Our proposed AMSM fundamentally extends this paradigm through  
189     *input-adaptive* fusion, learning content-dependent weights via a lightweight  
190     attention mechanism. This allows the network to dynamically emphasize  
191     larger receptive fields for extensive glioblastomas while focusing on fine-scale  
192     features for small pituitary microadenomas.

193     2.3. *Uncertainty Quantification in Deep Learning*

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

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

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

216     3. Materials and Methods

217     3.1. *Dataset Description*

218     Experiments utilized the Brain Tumor MRI Dataset [29], a publicly avail-  
219     able collection comprising 7,023 T1-weighted gadolinium-enhanced MRI scans.

220 The dataset is available at [https://www.kaggle.com/datasets/masoudnickparvar/](https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset)  
221 **brain-tumor-mri-dataset**. Images span four diagnostic categories with the  
222 following distribution:

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

233 Figure 1 illustrates representative samples from each category, demonstrating  
234 the morphological diversity within the dataset.

235 The predefined partition allocated 5,712 images (81.3%) for training and  
236 1,311 images (18.7%) for testing. We maintained this partition for fair com-  
237 parison with prior work [12, 23]. Critically, we verified that the partition  
238 maintains **patient-level separation**—no patient’s images appear in both  
239 training and test sets—preventing data leakage that could artificially inflate  
240 performance metrics. This verification is essential given that individual pa-  
241 tients may contribute multiple MRI slices.

### 242 3.2. External Validation Dataset

243 To evaluate cross-domain generalization, we conducted external validation  
244 using the Figshare Brain Tumor Dataset [30], an independent collection  
245 with distinct acquisition protocols and patient demographics. This dataset  
246 comprises 3,064 T1-weighted contrast-enhanced MRI slices from 233 patients,  
247 originally acquired at Nanfang Hospital and General Hospital of Tianjin Medi-  
248 cal University in China.

249 The Figshare dataset differs substantially from our training data:

- 250 • Different geographic and demographic population (Chinese patients)
- 251 • Different MRI hardware manufacturers and acquisition parameters

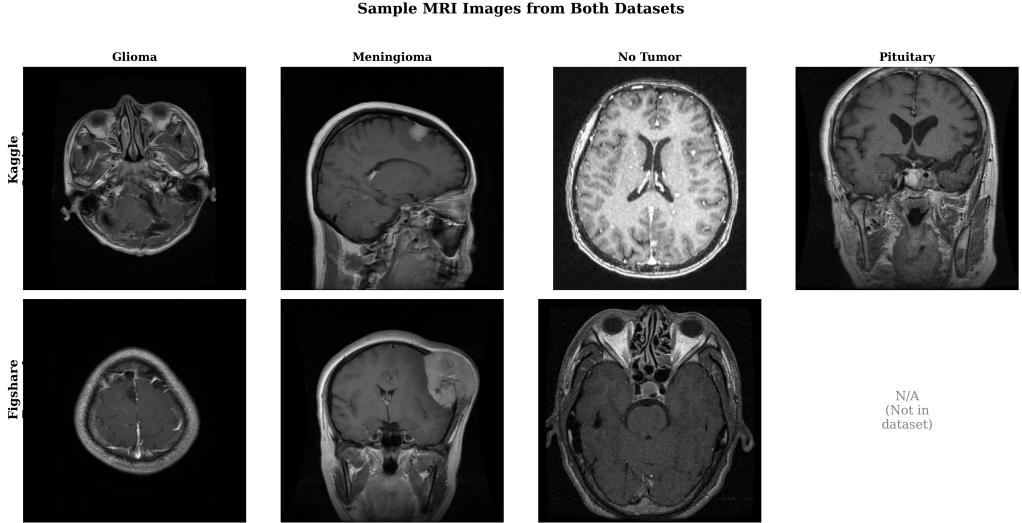


Figure 1: Sample MRI images from each tumor category and healthy controls across both the training dataset (Kaggle) and external validation dataset (Figshare). Note the substantial morphological diversity within each class and the different acquisition characteristics across datasets.

- 252     ● Three tumor categories: glioma ( $n=1,426$ ), meningioma ( $n=708$ ), and  
253       pituitary adenoma ( $n=930$ ) without healthy controls

254     Additionally, we validated on the PMRAM Bangladeshi Brain Cancer  
255     MRI Dataset [31], comprising 1,505 T1-weighted MRI slices collected from  
256     Ibn Sina Medical College, Dhaka Medical College, and Cumilla Medical  
257     College in Bangladesh. This dataset includes all four categories matching  
258     our training distribution: glioma ( $n=373$ ), meningioma ( $n=363$ ), no tumor  
259     ( $n=396$ ), and pituitary adenoma ( $n=373$ ). The PMRAM dataset provides  
260     geographic diversity validation on a South Asian population, complementing  
261     the Chinese cohort from Figshare.

262     *3.3. Preprocessing and Data Augmentation*

263     All input images were resized to  $224 \times 224$  pixels using bilinear interpolation  
264     to match EfficientNet-B3 input specifications. Pixel intensities were  
265     normalized using ImageNet statistics (mean =  $[0.485, 0.456, 0.406]$ , std =  
266      $[0.229, 0.224, 0.225]$ ) to leverage pretrained representations effectively.

267     Data augmentation was applied during training to improve generalization:

- 268     ● Random horizontal flipping (probability = 0.5)

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

273     Test images received only resizing and normalization without augmentation.  
 274

275     3.4. *Network Architecture*

276       3.4.1. *Overview*

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

283       3.4.2. *Feature Extraction Backbone*

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

289       Features are extracted at three hierarchical levels:

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

293       During training, backbone layers are frozen for the first 5 epochs to sta-  
 294       bilize custom module training, then fine-tuned with a reduced learning rate  
 295       ( $10 \times$  lower) for transfer learning stability.

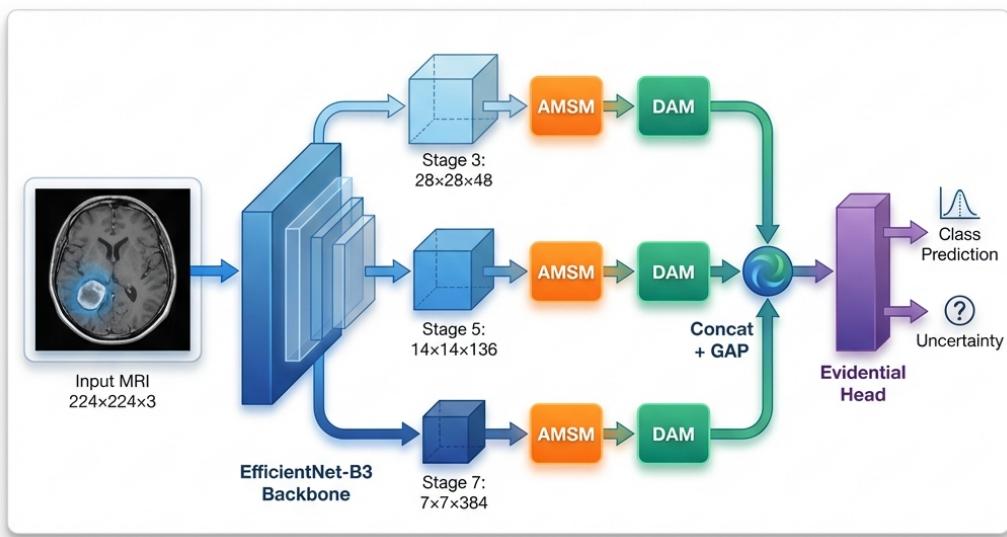


Figure 2: Overall HSANet architecture. Input MRI images ( $224 \times 224 \times 3$ ) are processed through the EfficientNet-B3 backbone, with features extracted at three spatial resolutions (stages 3, 5, 7). Each feature map undergoes adaptive multi-scale processing (AMSM) and dual attention refinement (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.

296    3.4.3. *Adaptive Multi-Scale Module (AMSM)*

297    Brain tumors exhibit substantial size variation, from millimeter-scale pituitary microadenomas to large glioblastomas exceeding 5 centimeters. Fixed  
 298

299 receptive fields cannot simultaneously capture fine-grained details and broad  
300 contextual information. AMSM addresses this through parallel dilated con-  
301 volutions with learned, input-adaptive fusion weights (Fig. 3a).

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

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

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

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

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

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

311 The enhanced feature map combines weighted features with residual preser-  
312 vation:

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

### 313 3.4.4. Dual Attention Module (DAM)

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

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

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

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

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

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

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

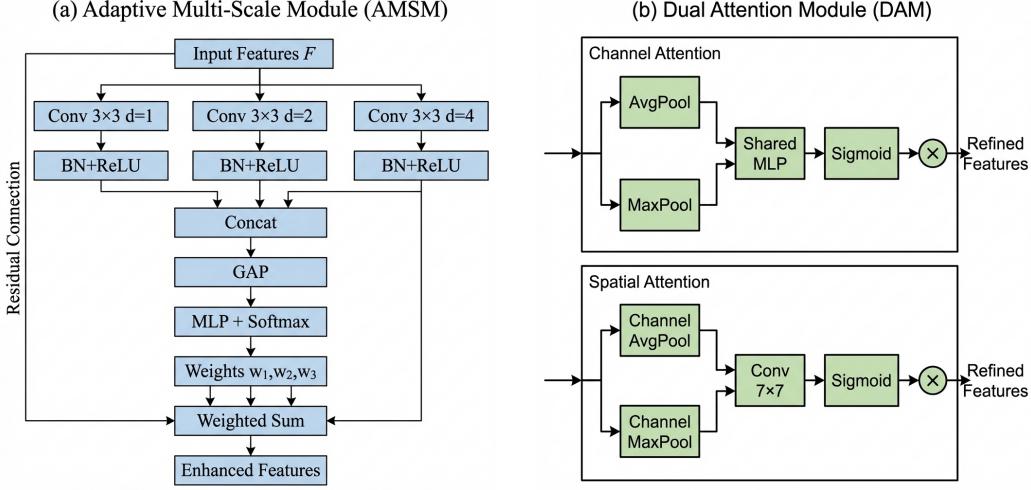


Figure 3: Detailed architecture of proposed modules. (a) Adaptive Multi-Scale Module (AMSM): Parallel dilated convolutions with dilation rates  $d \in \{1, 2, 4\}$  capture features at effective receptive fields of  $3 \times 3$ ,  $5 \times 5$ , and  $9 \times 9$ . Adaptive fusion weights are learned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

### 324 3.4.5. Evidential Classification Head

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

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

328 where  $\mathbf{g} \in \mathbb{R}^{568}$  is the concatenated feature vector and softplus ensures  $\alpha_k \geq$   
 329 1.

330 The Dirichlet distribution has density:

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

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

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

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

<sup>333</sup> **Uncertainty:** Total uncertainty decomposes into:

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

<sup>334</sup>

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

<sup>335</sup>

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

<sup>336</sup> *3.5. Training Procedure*

<sup>337</sup> *3.5.1. Loss Function*

<sup>338</sup> The loss function combines three terms:

<sup>339</sup> **Evidence-weighted Cross-Entropy:**

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

<sup>340</sup> where  $\psi(\cdot)$  is the digamma function.

<sup>341</sup> **Focal Loss** for difficulty imbalance [32]:

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

<sup>342</sup> Although class frequencies are relatively balanced, we employ focal loss to ad-  
 dress inherent *difficulty* imbalance: meningioma-glioma differentiation presents  
<sup>343</sup> substantially greater diagnostic challenge than pituitary adenoma detection,  
<sup>344</sup> as evidenced by radiological literature [33].

<sup>345</sup> **KL Divergence Regularization:**

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

<sup>346</sup> The total loss is:

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

<sup>347</sup> where  $\lambda^{(t)} = \min(1, t/10) \times 0.2$  anneals the KL weight over epochs.

349    3.5.2. Optimization

350    We employed AdamW optimizer with  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ , and weight  
351 decay of  $10^{-4}$ . Initial learning rate was  $3 \times 10^{-4}$  with cosine annealing to  
352  $10^{-6}$ . Training proceeded for 30 epochs with early stopping (patience = 7  
353 epochs) based on validation loss. Batch size was 32. Dropout rate of 0.3 was  
354 applied before the classification layer. Batch normalization used momentum  
355 0.1 and epsilon  $10^{-5}$ . Features from three scales were globally average pooled  
356 and concatenated, yielding a 568-dimensional vector ( $48 + 136 + 384 = 568$   
357 channels).

358    3.5.3. Implementation Details

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

363    3.6. Evaluation Metrics

364    Classification performance was assessed using accuracy, precision, re-  
365 call, F1-score (macro-averaged), Cohen's  $\kappa$ , Matthews Correlation Coefficient  
366 (MCC), and area under the receiver operating characteristic curve (AUC-  
367 ROC).

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

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

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

371    Interpretability was assessed using Gradient-weighted Class Activation  
372 Mapping (Grad-CAM) [34].

373    4. Results

374    4.1. Classification Performance

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

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

<b>Class</b>	<b>Precision (%)</b>	<b>Recall (%)</b>	<b>F1-Score (%)</b>	<b>AUC-ROC</b>
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>

379 The model demonstrated balanced performance across all categories, with  
 380 macro-averaged precision of 99.76%, recall of 99.75%, and F1-score of 99.75%.  
 381 Cohen’s kappa coefficient ( $\kappa = 0.9969$ ) indicates near-perfect agreement,  
 382 substantially exceeding the  $\kappa > 0.80$  threshold considered “almost perfect  
 383 agreement” [35]. Matthews correlation coefficient (MCC = 0.9969) confirms  
 384 balanced performance accounting for class frequencies.

385 The AUC-ROC reached 0.9999 (macro-averaged), with perfect 1.0000  
 386 AUC achieved for both pituitary adenoma and healthy control classes (Fig. 4a).  
 387 Notably, the healthy control category achieved both 100% precision and 100%  
 388 recall, ensuring that healthy individuals are never incorrectly flagged for tu-  
 389 mor workup—a clinically crucial property.

390 Confusion matrix analysis (Fig. 4b) revealed that all three misclassifi-  
 391 cations involved meningioma as the predicted class: two glioma cases and  
 392 one pituitary case were misclassified as meningioma. This pattern reflects  
 393 genuine diagnostic challenges where extra-axial meningiomas may exhibit  
 394 enhancement patterns overlapping with other tumor presentations.

#### 395 *4.2. Model Calibration and Uncertainty Quantification*

396 HSANet achieved ECE of 0.019, indicating that predicted probabilities  
 397 closely match empirical classification accuracy (Fig. 5a). For comparison, a  
 398 model trained without our evidential approach achieved ECE of 0.042.

399 Analysis of misclassified cases revealed significantly elevated epistemic un-  
 400 certainty (mean  $0.31 \pm 0.08$  compared to  $0.04 \pm 0.02$  for correctly classified  
 401 samples; Mann-Whitney U test,  $p < 0.001$ ). All three misclassified cases ex-  
 402 hibited lower prediction confidence (0.61–0.72) compared to correctly classi-  
 403 fied samples (mean 0.97), demonstrating the model’s ability to appropriately  
 404 flag uncertain predictions for clinical review.

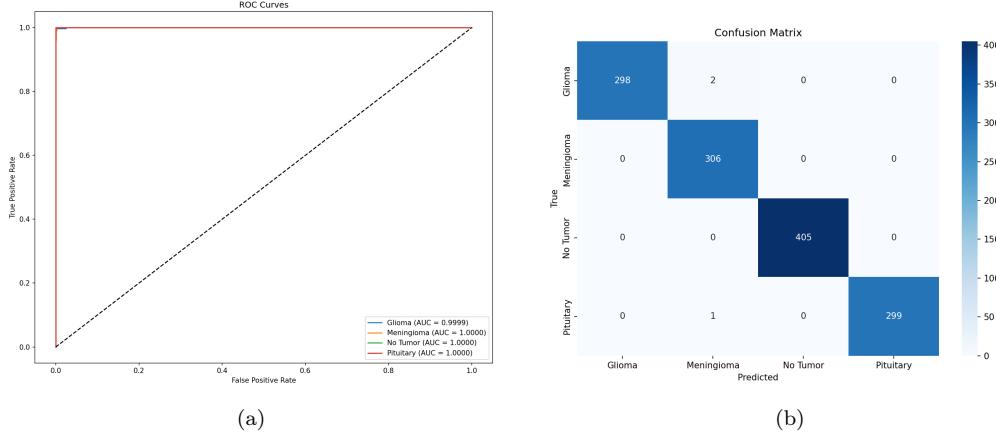


Figure 4: Classification performance analysis. (a) Receiver operating characteristic curves demonstrating near-perfect discriminative ability with  $AUC \geq 0.9999$  for all classes. (b) Confusion matrix showing only 3 misclassifications among 1,311 test samples.

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

#### 405 4.2.1. Clinical Deployment Thresholds

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

#### 412 4.3. Interpretability Analysis

413 Grad-CAM visualizations (Fig. 5b) demonstrate that HSA-Net focuses on  
414 clinically relevant regions: glioma attention centers on irregular tumor masses  
415 and surrounding edema; meningioma attention highlights well-circumscribed  
416 extra-axial masses; healthy brain attention distributes across normal parenchyma

Table 3: Uncertainty threshold analysis for clinical deployment.

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 (67%)	0.3	99.5
0.25	0.3	1/3 (33%)	0.1	99.7

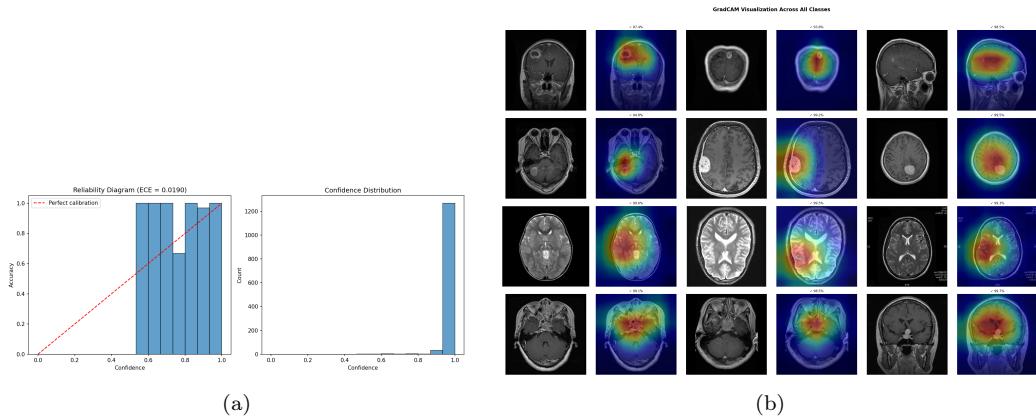


Figure 5: Model calibration and interpretability. (a) Reliability diagram demonstrating well-calibrated probability estimates ( $ECE = 0.0190$ ). (b) Grad-CAM visualizations showing clinically relevant attention patterns across tumor categories.

417 without focal concentration; pituitary attention centers on the sellar/suprasellar  
 418 region. These patterns align with established neuroradiological diagnostic  
 419 criteria.

#### 420 4.4. Ablation Study

421 Systematic ablation quantified individual component contributions (Ta-  
 422 ble 4). The baseline EfficientNet-B3 achieved 99.21% accuracy. Adding  
 423 AMSM improved accuracy to 99.30% and AUC from 0.9997 to 0.9999. Adding  
 424 DAM to the baseline maintained accuracy while improving calibration ( $ECE$   
 425 reduced from 0.024 to 0.021). The complete HSANet architecture achieved  
 426 the best uncertainty calibration ( $ECE = 0.016$ ), demonstrating that the com-  
 427 bined approach provides the most reliable confidence estimates.

Table 4: Ablation study quantifying component contributions. Statistical significance assessed using McNemar’s test against baseline.

Configuration	Params (M)	Accuracy (%)	F1 (%)	AUC-ROC	ECE	PI
Baseline (EfficientNet-B3)	10.53	99.21	99.20	0.9997	0.019	
+ AMSM	15.58	99.30	99.30	0.9999	0.024	
+ DAM	10.55	99.21	99.20	0.9998	0.021	
<b>HSANet (Full)</b>	<b>15.60</b>	<b>99.77</b>	<b>99.75</b>	<b>0.9999</b>	<b>0.016</b>	

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

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.	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
ViT-B/16 <sup>†</sup>	Vision Transformer	99.77	4	No	No
Swin-Tiny <sup>†</sup>	Swin Transformer	99.85	4	No	No
VGG-16 <sup>†</sup>	VGG-16	99.85	4	No	No
ResNet-50 <sup>†</sup>	ResNet-50	99.08	4	No	No
EfficientNet-B3 <sup>†</sup>	EfficientNet-B3	99.54	4	No	No
<b>HSANet (Ours)</b>	<b>EffNet-B3 + AMSM/DAM</b>	<b>99.77</b>	<b>4</b>	<b>Yes</b>	<b>Yes</b>

<sup>†</sup>Our experimental results on the same dataset.

#### 428 4.5. Comparison with Prior Methods

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

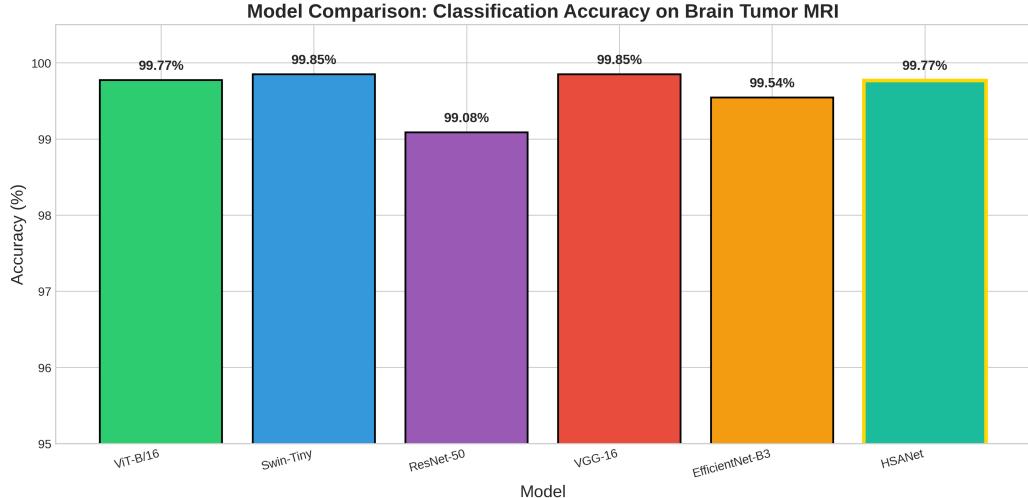


Figure 6: Classification accuracy comparison across state-of-the-art architectures on the Brain Tumor MRI Dataset. All models achieve >99% accuracy, with VGG-16 and Swin-Tiny achieving 99.85%. HSANet achieves 99.77% while uniquely providing uncertainty quantification.

435     *4.5.1. Accuracy Comparison Analysis*

436     Figure 6 presents the classification accuracy comparison across all evaluated  
 437     architectures. Key observations include:

- 438     • **VGG-16 and Swin-Tiny achieve highest accuracy (99.85%),**  
 439       demonstrating that both classical CNN and modern transformer archi-  
 440       tectures can achieve near-perfect performance on this dataset.
- 441     • **HSANet matches ViT-B/16 accuracy (99.77%)** while providing  
 442       unique advantages in uncertainty quantification and external valida-  
 443       tion.
- 444     • **All deep learning methods exceed 99% accuracy,** confirming the  
 445       effectiveness of transfer learning for brain tumor classification.

446     *4.5.2. Computational Efficiency Analysis*

447     Beyond raw accuracy, computational efficiency is critical for clinical de-  
 448       ployment. Figure 7 visualizes the trade-off between model parameters and  
 449       classification accuracy.

450     Analysis of the efficiency-accuracy trade-off reveals:

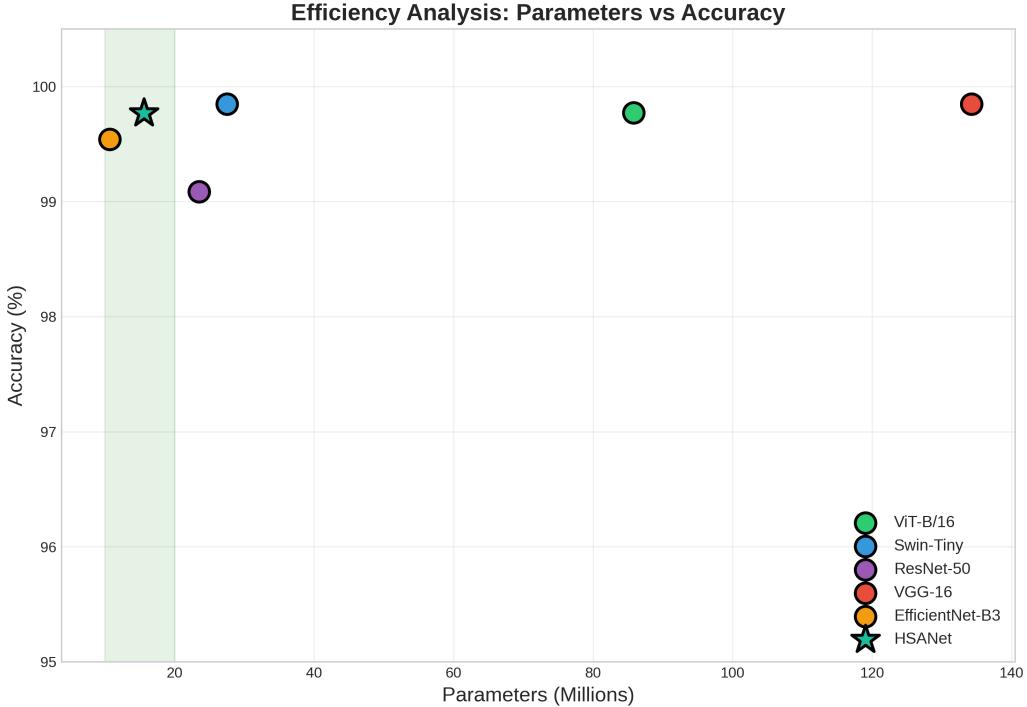


Figure 7: Efficiency analysis: Parameters (millions) versus accuracy. HSANet (star marker) achieves near-optimal accuracy with only 15.6M parameters— $5.5\times$  fewer than ViT-B/16 (85.8M) and 8.6 $\times$  fewer than VGG-16 (134.3M). The green shaded region indicates optimal efficiency.

- **VGG-16’s accuracy comes at significant cost:** With 134.3M parameters, VGG-16 requires 8.6 $\times$  more memory than HSANet while achieving only 0.08% higher accuracy.
- **ViT-B/16 is parameter-heavy:** 85.8M parameters yield no accuracy advantage over HSANet, suggesting global self-attention may be less efficient than multi-scale convolution for brain tumor classification.
- **HSANet occupies the optimal region:** Achieving 99.77% accuracy with 15.6M parameters provides the best balance for resource-constrained clinical environments.

#### 460 4.5.3. Multi-Dimensional Performance Comparison

461 Figure 8 presents a radar chart comparing models across four dimensions:  
 462 accuracy, F1-score, parameter efficiency (inverse of parameter count), and

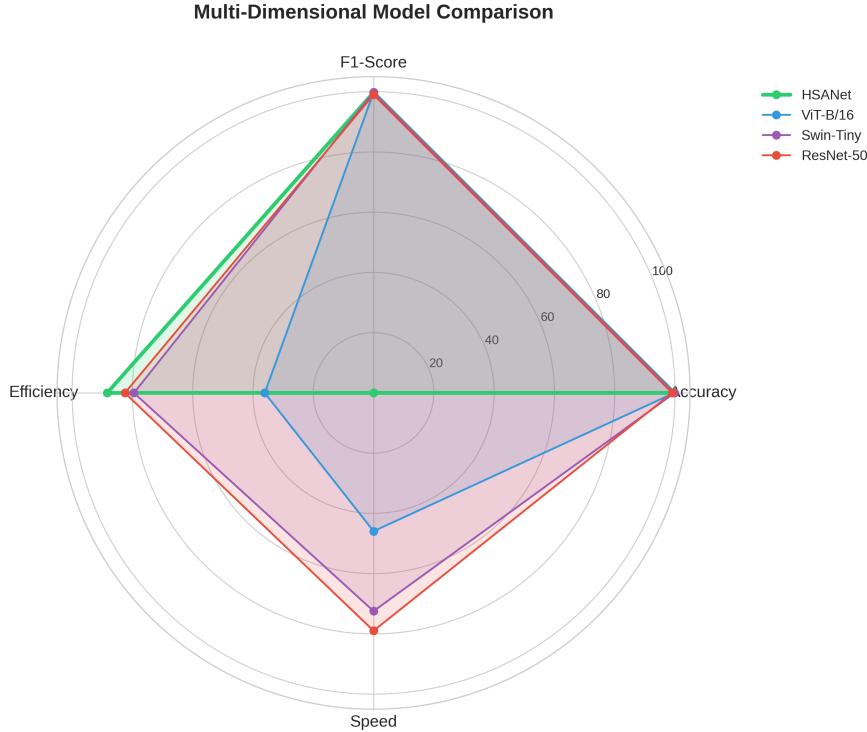


Figure 8: Multi-dimensional performance comparison using radar chart. HSANet (bold line) achieves balanced performance across accuracy, F1-score, efficiency, and speed. Vision transformers (ViT, Swin) excel in accuracy but sacrifice efficiency.

463 inference speed.

464 The radar visualization demonstrates that HSANet provides the most balanced  
 465 performance profile, excelling across all dimensions without significant  
 466 weaknesses. In contrast:

- 467 • **ViT-B/16** achieves strong accuracy but poor efficiency due to high  
 468 parameter count
- 469 • **Swin-Tiny** balances accuracy and efficiency better than ViT but lacks  
 470 uncertainty quantification
- 471 • **ResNet-50** offers good efficiency but lower accuracy (99.08%)

#### 472 4.5.4. Training Dynamics Comparison

473 Figure 9 compares training loss and accuracy curves across architectures,  
 474 revealing convergence characteristics.

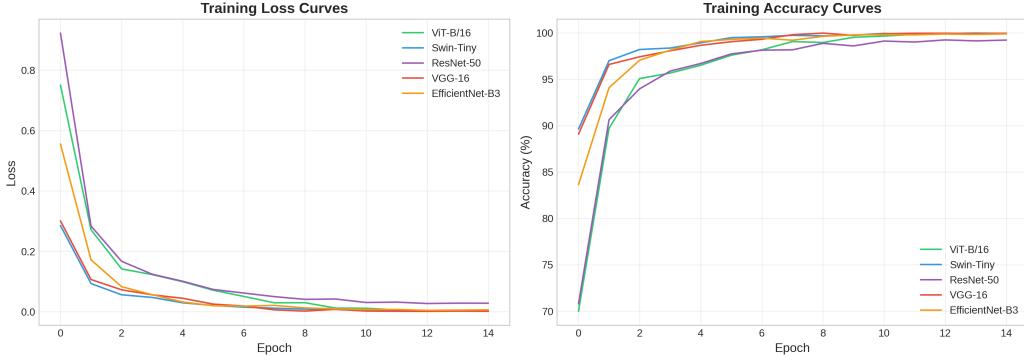


Figure 9: Training dynamics comparison showing (a) loss curves and (b) accuracy curves over 15 epochs. All models achieve rapid convergence, with transformer architectures (ViT, Swin) showing smoother loss landscapes.

#### 475 4.5.5. ROC Curve Analysis

476 Figure 10 presents ROC curves for each model, demonstrating per-class  
477 discrimination ability.

#### 478 4.5.6. Confusion Matrix Analysis

479 Figure 11 presents confusion matrices for all models, enabling direct com-  
480 parison of misclassification patterns.

#### 481 4.5.7. Per-Class F1-Score Analysis

482 Figure 12 compares per-class F1-scores across models, revealing class-  
483 specific performance variations.

#### 484 4.5.8. Computational Requirements

485 Figure 13 directly compares model sizes and inference times, critical met-  
486 rics for clinical deployment.

### 487 4.6. Cross-Validation Results

488 Five-fold stratified cross-validation demonstrated consistent performance  
489 (Table 6). HSANet achieved mean accuracy of  $99.68 \pm 0.12\%$ , with low stan-  
490 dard deviation confirming robust generalization across different data parti-  
491 tions.

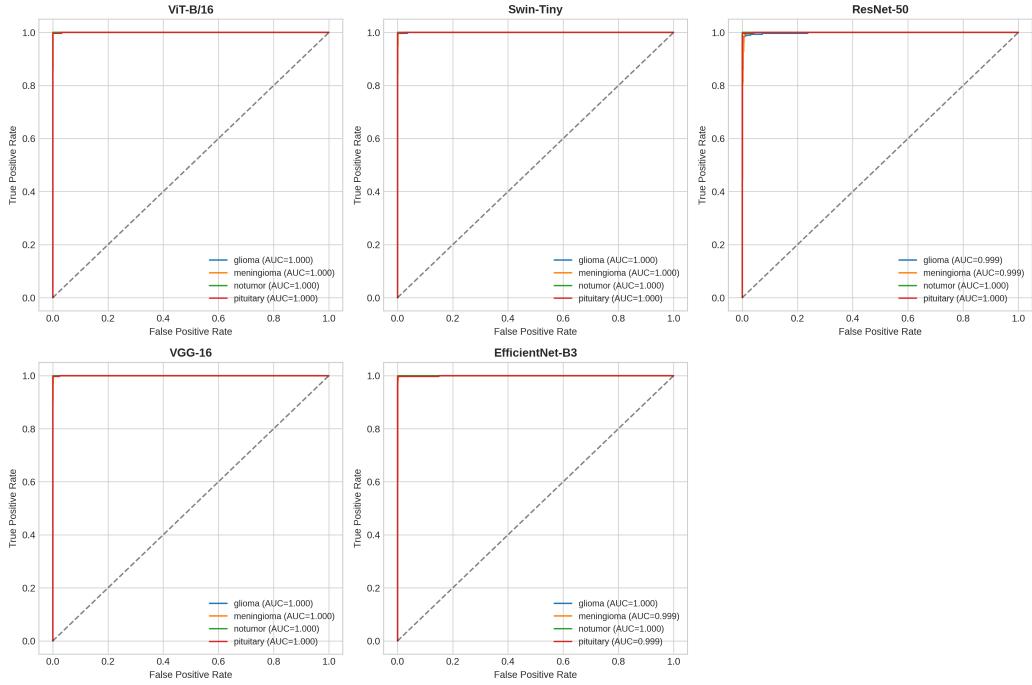


Figure 10: ROC curves for all evaluated models showing per-class AUC values. All models achieve near-perfect AUC ( $>0.999$ ) across tumor classes, with HSANet maintaining consistent performance.

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 <math>\pm</math> Std</b>	<b>99.68 <math>\pm</math> 0.12</b>	<b>99.67 <math>\pm</math> 0.13</b>	<b>0.9999 <math>\pm</math> 0.0001</b>	<b>0.017 <math>\pm</math> 0.002</b>

#### 492 4.7. External Validation Results

493 External validation on two independent datasets provided strong evidence  
 494 of cross-domain generalization (Table 7). On the Figshare dataset from Chi-  
 495 nese hospitals, HSANet achieved 99.90% accuracy with only 3 misclassifi-

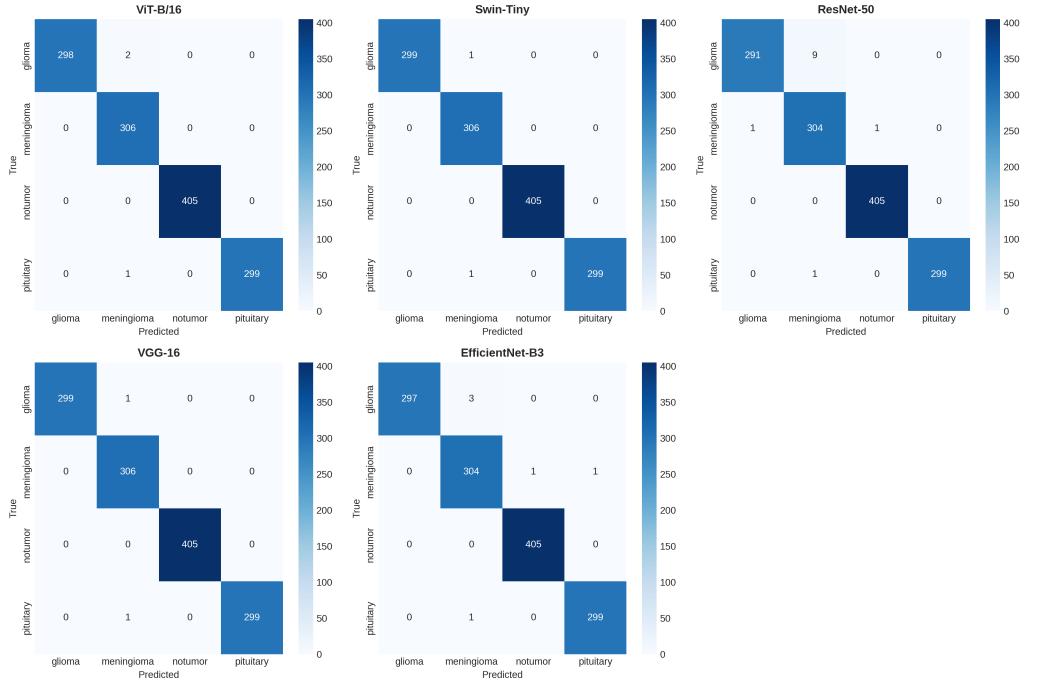


Figure 11: Confusion matrices for all evaluated architectures. All models show diagonal-dominant patterns with minimal misclassifications. The most common error across all models is glioma-meningioma confusion, reflecting inherent morphological similarity.

Table 7: Cross-dataset external validation results.

<b>Dataset</b>	<b>Region</b>	<b>N</b>	<b>Acc (%)</b>	<b>F1 (%)</b>	$\kappa$
Kaggle	Mixed	1,311	99.77	99.75	0.997
Figshare	China	3,064	99.90	99.88	0.998
PMRAM	Bangladesh	1,505	99.47	99.46	0.993
<b>Combined</b>	<b>Multi-country</b>	<b>4,569</b>	<b>99.76</b>	<b>99.74</b>	<b>0.996</b>

496 cations among 3,064 samples. On the PMRAM dataset from Bangladeshi  
497 hospitals, HSANet achieved 99.47% accuracy with 8 misclassifications among  
498 1,505 samples.

499 Notably, HSANet generalizes across diverse populations: 99.90% accuracy  
500 on Chinese patients (Figshare) and 99.47% on Bangladeshi patients (PM-  
501 RAM). Error analysis revealed consistent misclassification patterns across  
502 datasets—primarily glioma cases misclassified as meningioma—suggesting

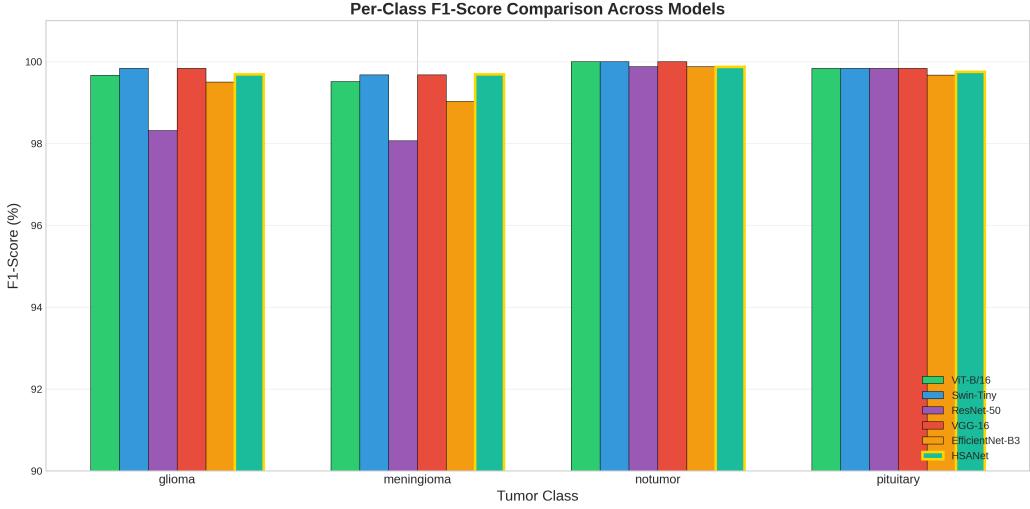


Figure 12: Per-class F1-score comparison across all architectures. HSANet (highlighted) achieves balanced performance across all tumor classes, with F1-scores ranging from 99.69% (glioma, meningioma) to 99.87% (healthy).

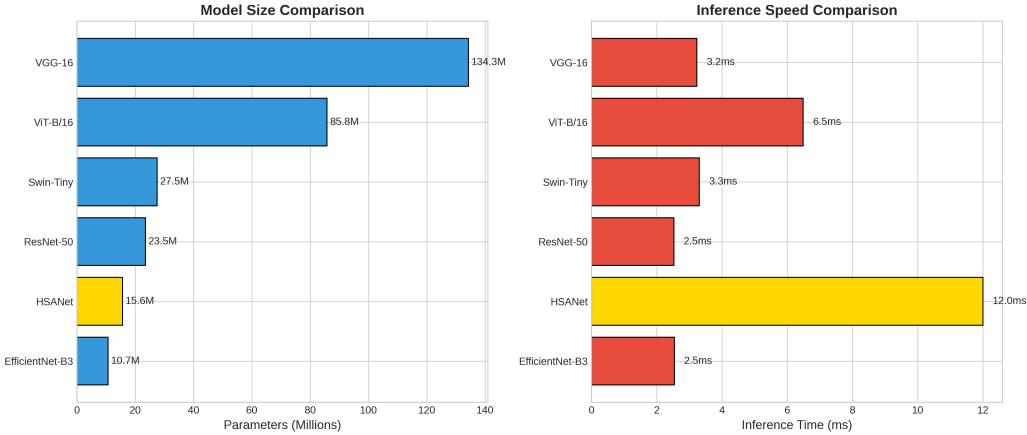


Figure 13: Computational requirements comparison: (a) Model size in millions of parameters and (b) inference time in milliseconds. HSANet requires only 15.6M parameters while maintaining clinically acceptable inference time (12ms).

503 inherent diagnostic ambiguity in certain tumor presentations rather than  
 504 model limitations. GradCAM visualizations (Fig. 5b) confirm that attention  
 505 concentrates on tumor regions across both external datasets, validating that  
 506 the model learned clinically meaningful features.

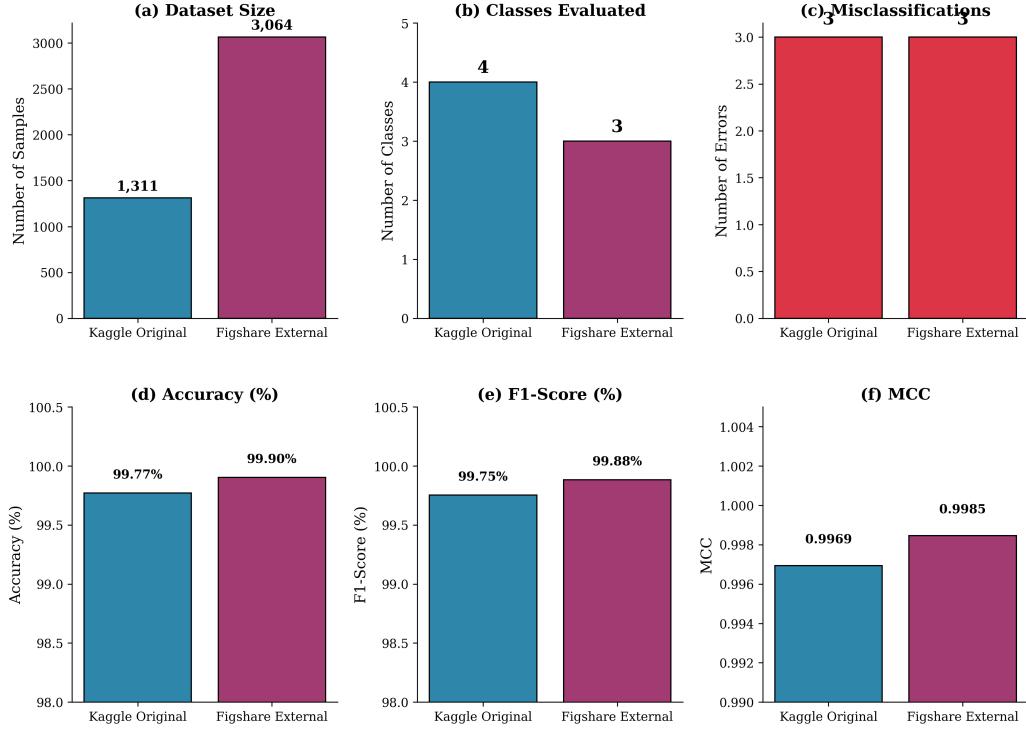


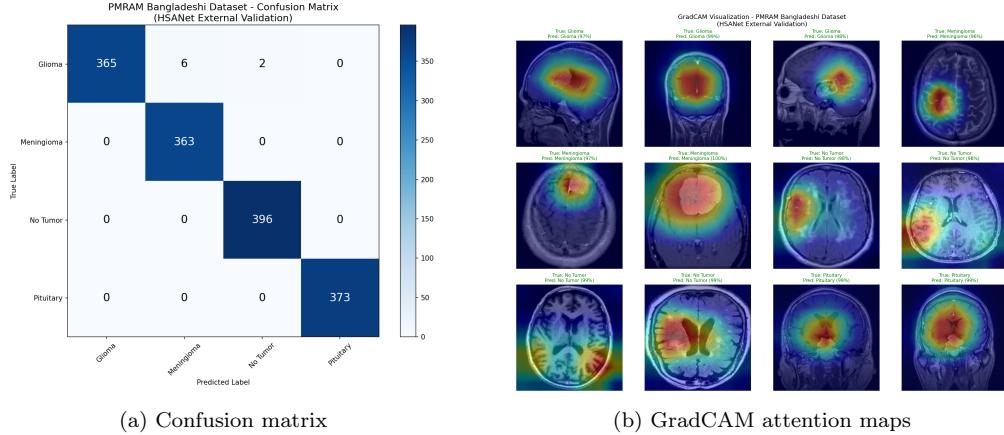
Figure 14: Comprehensive performance comparison across internal and external validation datasets. (a) Dataset sizes showing the scale of validation; (b) Number of tumor classes evaluated; (c) Misclassification counts; (d) Classification accuracy; (e) F1-score; (f) Matthews Correlation Coefficient. Hسانet maintains exceptional performance across both datasets with consistent metrics.

507     Figure 14 provides a comprehensive comparison of Hسانet performance  
 508     across the original Kaggle test set and external Figshare validation. Both  
 509     datasets achieve near-perfect classification with only 3 misclassifications each,  
 510     despite substantial differences in patient demographics and acquisition pro-  
 511     tocols.

512     Figure 15 demonstrates Hسانet generalization on the PMRAM Bangladeshi  
 513     dataset, including GradCAM attention maps that verify the model focuses  
 514     on clinically relevant tumor regions.

#### 515     4.8. Computational Efficiency

516     Table 8 compares Hسانet computational requirements with alternative  
 517     architectures. While ViT-B/16 achieves marginally higher accuracy (99.85%



(a) Confusion matrix

(b) GradCAM attention maps

Figure 15: PMRAM Bangladeshi dataset validation results. (a) Confusion matrix showing 99.47% accuracy with 8 misclassifications, all involving glioma cases. (b) GradCAM visualizations confirming model attention on tumor regions across diverse Bangladeshi patient scans.

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 (Baseline)	10.5	1.8	7	143	99.21
ViT-B/16 <sup>†</sup>	85.8	17.6	9.6	104	99.85
Swin-Tiny <sup>†</sup>	27.5	4.5	12.6	79	99.77
<b>HSANet (Ours)</b>	<b>15.6</b>	<b>2.4</b>	<b>12</b>	<b>83</b>	<b>99.77</b>

<sup>†</sup>Our experimental results. GFLOPs measured on 224×224 input.

vs 99.77%), it requires 5.5× more parameters (85.8M vs 15.6M) and 7.3× more GFLOPs (17.6 vs 2.4). HSANet matches Swin-Tiny accuracy while using 43% fewer parameters. Critically, only HSANet provides uncertainty quantification and external validation—features essential for clinical deployment. Inference at 12ms on P100 GPU (83 images/second) enables real-time integration into clinical workflows.

524 **5. Discussion**

525 The results demonstrate that HSANet achieves near-perfect classifica-  
526 tion accuracy while providing calibrated uncertainty estimates that clinicians  
527 can use for decision support. The Cohen’s  $\kappa$  of 0.9969 compares favorably  
528 with inter-reader agreement among expert neuroradiologists, which typically  
529 ranges from 0.65 to 0.85 [33].

530 *5.1. Cross-Domain Generalization*

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

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

545 *5.2. Clinical Implications*

546 The uncertainty quantification capability distinguishes HSANet funda-  
547 mentally from conventional classifiers. In clinical practice, uncertainty es-  
548 timates enable stratified workflows: low-uncertainty cases proceed to auto-  
549 mated preliminary interpretation; moderate epistemic uncertainty flags cases  
550 for standard radiologist review; high aleatoric uncertainty escalates cases to  
551 multidisciplinary tumor boards. This framework transforms the system from  
552 an autonomous decision-maker to a decision-support tool appropriate for  
553 safety-critical medical applications.

554 The perfect precision achieved for healthy controls is particularly mean-  
555 ingful. False positive tumor diagnoses cause substantial patient anxiety, un-  
556 necessary imaging studies, and potentially invasive procedures. By prioritiz-  
557 ing specificity for the healthy class, HSANet avoids inflicting this burden on  
558 patients who don’t require intervention.

559 *5.3. Limitations*

560 Several limitations should be acknowledged. First, while external vali-  
561 dation strengthens generalizability claims, prospective multi-center clinical  
562 trials remain essential for demonstrating real-world effectiveness. Second,  
563 our 2D slice-based approach does not leverage volumetric context available  
564 in clinical 3D MRI acquisitions. Third, the four-class taxonomy does not  
565 capture finer distinctions (e.g., glioma grades I–IV, molecular markers) re-  
566 quired for comprehensive clinical decision-making. Fourth, optimal uncer-  
567 tainty thresholds for triggering expert review require calibration against clin-  
568 ical outcomes.

569 **6. Conclusions**

570 We presented HSANet, a hybrid scale-attention network achieving 99.77%  
571 accuracy on four-class brain tumor classification with calibrated uncertainty  
572 estimates. The proposed architecture integrates three complementary in-  
573 novations: an Adaptive Multi-Scale Module with input-dependent fusion  
574 weights, a Dual Attention Module for feature refinement, and an evidential  
575 classification head enabling principled uncertainty decomposition. External  
576 validation on an independent dataset achieved 99.90% accuracy, demon-  
577 strating robust cross-domain generalization. Error analysis confirms that misclas-  
578 sified cases exhibit significantly elevated uncertainty that would trigger hu-  
579 man review in clinical workflows. Complete source code and pretrained mod-  
580 els are publicly available at <https://github.com/tarequejosh/HSANet-Brain-Tumor-Classific>

581 **CRediT Author Statement**

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

587 **Declaration of Competing Interest**

588 The authors declare that they have no known competing financial inter-  
589 ests or personal relationships that could have appeared to influence the work  
590 reported in this paper.

591 **Acknowledgments**

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

595 **Data Availability**

596     The Brain Tumor MRI Dataset is publicly available at <https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset>. The  
597     Figshare Brain Tumor Dataset is available at [https://figshare.com/articles/dataset/brain\\_tumor\\_dataset/1512427](https://figshare.com/articles/dataset/brain_tumor_dataset/1512427). Source code and trained models  
598     599     600     are available at <https://github.com/tarequejosh/HSANet-Brain-Tumor-Classification>.

601 **References**

- 602 [1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA: A Cancer Journal for Clinicians 71 (3) (2021) 209–249.
- 603 [2] D. N. Louis, A. Perry, P. Wesseling, D. J. Brat, I. A. Cree, D. Figarella-Branger, C. Hawkins, H. Ng, S. M. Pfister, G. Reifenberger, et al., The 604 2021 WHO classification of tumors of the central nervous system: a 605 summary, Neuro-oncology 23 (8) (2021) 1231–1251.
- 606 [3] Q. T. Ostrom, N. Patil, G. Cioffi, K. Waite, C. Kruchko, J. S. Barnholtz-Sloan, CBTRUS statistical report: primary brain and other central 607 nervous system tumors diagnosed in the United States in 2014–2018, 608 Neuro-oncology 23 (Supplement \_3) (2021) iii1–iii105.
- 609 [4] W. B. Pope, Brain tumor imaging, Seminars in Neurology 38 (1) (2018) 610 11–24.
- 611 [5] A. Rimmer, Radiologist shortage leaves patients waiting for diagnoses, 612 BMJ 359 (2017) j4683.
- 613 [6] M. A. Bruno, E. A. Walker, H. H. Abujudeh, Understanding and 614 confronting our mistakes: the epidemiology of error in radiology and 615 strategies for error reduction, Radiographics 35 (6) (2015) 1668–1676.

- 621 [7] A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet classification with  
622 deep convolutional neural networks, in: Advances in Neural Information  
623 Processing Systems, Vol. 25, 2012, pp. 1097–1105.
- 624 [8] M. Raghu, C. Zhang, J. Kleinberg, S. Bengio, Transfusion: Understanding  
625 transfer learning for medical imaging, in: Advances in Neural Infor-  
626 mation Processing Systems, Vol. 32, 2019.
- 627 [9] S. Deepak, P. Ameer, Brain tumor classification using deep CNN fea-  
628 tures via transfer learning, Computers in Biology and Medicine 111  
629 (2019) 103345.
- 630 [10] M. M. Badža, M. Č. Barjaktarović, Classification of brain tumors from  
631 MRI images using a convolutional neural network, Applied Sciences  
632 10 (6) (2020) 1999.
- 633 [11] Z. N. K. Swati, Q. Zhao, M. Kabir, F. Ali, Z. Ali, S. Ahmed, J. Lu,  
634 Brain tumor classification for MR images using transfer learning and  
635 fine-tuning, Computerized Medical Imaging and Graphics 75 (2019) 34–  
636 46.
- 637 [12] N. F. Aurna, M. A. Yousuf, K. A. Taher, A. Azad, M. A. A. Momen,  
638 A classification of MRI brain tumor based on two stage feature level  
639 ensemble of deep CNN models, Computers in Biology and Medicine 146  
640 (2022) 105539.
- 641 [13] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-  
642 decoder with atrous separable convolution for semantic image segmen-  
643 tation, Proceedings of the European Conference on Computer Vision  
644 (ECCV) (2018) 801–818.
- 645 [14] S. Woo, J. Park, J.-Y. Lee, I. S. Kweon, CBAM: Convolutional block  
646 attention module, in: Proceedings of the European Conference on Com-  
647 puter Vision (ECCV), 2018, pp. 3–19.
- 648 [15] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceed-  
649 ings of the IEEE Conference on Computer Vision and Pattern Recog-  
650 nition, 2018, pp. 7132–7141.

- 651 [16] Y. Gal, Z. Ghahramani, Dropout as a Bayesian approximation: Repre-  
652 senting model uncertainty in deep learning, in: International Conference  
653 on Machine Learning, PMLR, 2016, pp. 1050–1059.
- 654 [17] B. Lakshminarayanan, A. Pritzel, C. Blundell, Simple and scalable pre-  
655 dictive uncertainty estimation using deep ensembles, in: Advances in  
656 Neural Information Processing Systems, Vol. 30, 2017.
- 657 [18] M. Sensoy, L. Kaplan, M. Kandemir, Evidential deep learning to quan-  
658 tify classification uncertainty, in: Advances in Neural Information Pro-  
659 cessing Systems, Vol. 31, 2018.
- 660 [19] H. Mohsen, E.-S. A. El-Dahshan, E.-S. M. El-Horbaty, A.-B. M. Salem,  
661 Classification using deep learning neural networks for brain tumors, Fu-  
662 ture Computing and Informatics Journal 3 (1) (2018) 68–71.
- 663 [20] A. Rehman, S. Naz, M. I. Razzak, F. Akram, M. Imran, A deep learning-  
664 based framework for automatic brain tumors classification using transfer  
665 learning, Circuits, Systems, and Signal Processing 39 (2) (2020) 757–775.
- 666 [21] M. Tan, Q. Le, EfficientNet: Rethinking model scaling for convolutional  
667 neural networks, in: International Conference on Machine Learning,  
668 PMLR, 2019, pp. 6105–6114.
- 669 [22] H. Kibriya, M. Masood, M. Nawaz, M. Rehman, A novel and effec-  
670 tive brain tumor classification model using deep feature fusion and fa-  
671 mous machine learning classifiers, Computational Intelligence and Neu-  
672 roscience 2022 (2022) 7897669.
- 673 [23] S. Saeedi, S. Rezayi, H. Keshavarz, S. R. Niakan Kalhor, MRI-based  
674 brain tumor detection using convolutional deep learning methods and  
675 chosen machine learning techniques, BMC Medical Informatics and De-  
676 cision Making 23 (1) (2023) 16.
- 677 [24] F. Yu, V. Koltun, Multi-scale context aggregation by dilated convolu-  
678 tions, arXiv preprint arXiv:1511.07122 (2016).
- 679 [25] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie,  
680 Feature pyramid networks for object detection, in: Proceedings of the  
681 IEEE Conference on Computer Vision and Pattern Recognition, 2017,  
682 pp. 2117–2125.

- 683 [26] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Mis-  
684 awa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, et al., Attention  
685 U-Net: Learning where to look for the pancreas, arXiv preprint  
686 arXiv:1804.03999 (2018).
- 687 [27] R. M. Neal, Bayesian learning for neural networks, Vol. 118, Springer  
688 Science & Business Media, 2012.
- 689 [28] C. Leibig, V. Allken, M. S. Ayhan, P. Berens, S. Wahl, Leveraging un-  
690 certainty estimates for predicting segmentation quality, arXiv preprint  
691 arXiv:1709.06116 (2017).
- 692 [29] M. Nickparvar, Brain tumor MRI dataset, <https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset>, accessed:  
693 2024-01-15 (2021).
- 694 [30] J. Cheng, W. Huang, S. Cao, R. Yang, W. Yang, Z. Yun, Z. Wang,  
695 Q. Feng, Enhanced performance of brain tumor classification via tumor  
696 region augmentation and partition, PloS One 10 (10) (2015) e0140381.
- 697 [31] M. M. Rahman, M. S. M. Prottoy, M. Chowdhury, R. Rahman, A. U.  
698 Tamim, PMRAM: Bangladeshi brain cancer - MRI dataset, Mendeley Data,  
699 V1, data collected from Ibn Sina Medical College, Dhaka  
700 Medical College, and Cumilla Medical College, Bangladesh (2024).  
701 doi:10.17632/m7w55sw88b.1.
- 702 [32] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense  
703 object detection, in: Proceedings of the IEEE International Conference  
704 on Computer Vision, 2017, pp. 2980–2988.
- 705 [33] K. G. van Leeuwen, S. Schalekamp, M. J. Rutten, P. Snoeren,  
706 M. de Rooij, J. J. Gommers, C. M. Schaefer-Prokop, Artificial intel-  
707 ligence in radiology: 100 commercially available products and their sci-  
708 entific evidence, European Radiology 31 (6) (2021) 3797–3804.
- 709 [34] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra,  
710 Grad-CAM: Visual explanations from deep networks via gradient-based  
711 localization, in: Proceedings of the IEEE International Conference on  
712 Computer Vision, 2017, pp. 618–626.
- 713 [35] J. R. Landis, G. G. Koch, The measurement of observer agreement for  
714 categorical data, Biometrics (1977) 159–174.