

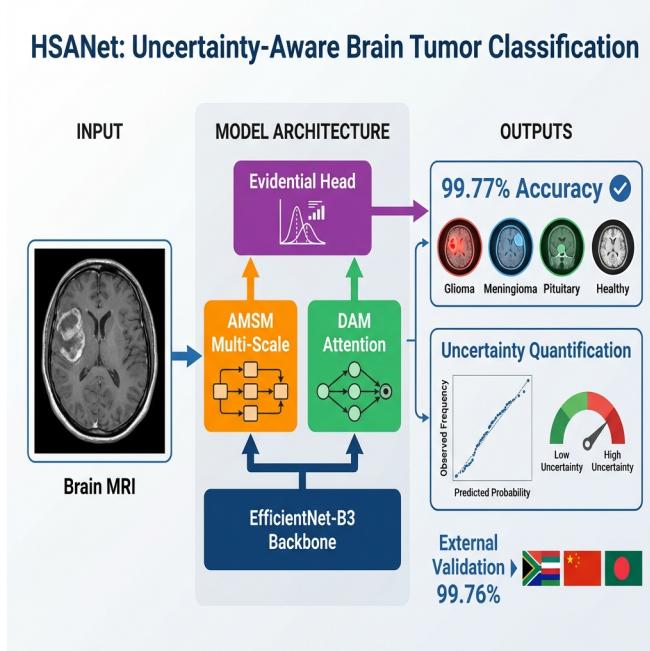
1 Graphical Abstract

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^{a,*}, Author 2^a, Author 3^a, Author 4^a

^a*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 (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, demon-
234 strating 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 valida-
244 tion 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 Med-
248 ical 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

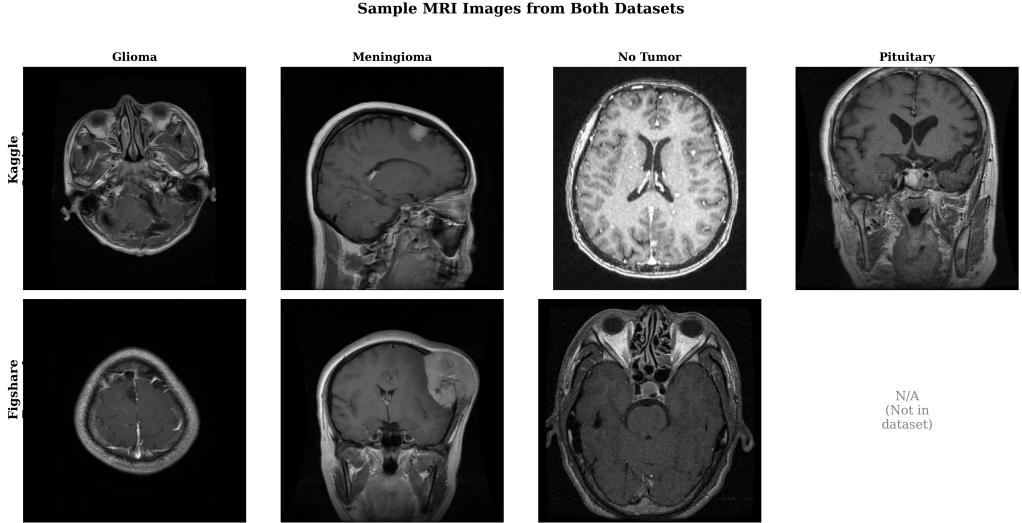


Figure 1: Sample MRI images from each tumor category and healthy controls across both the training dataset (Kaggle) and external validation dataset (Figshare). Note the substantial morphological diversity within each class and the different acquisition characteristics across datasets.

- 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 224×224 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 224×224 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.

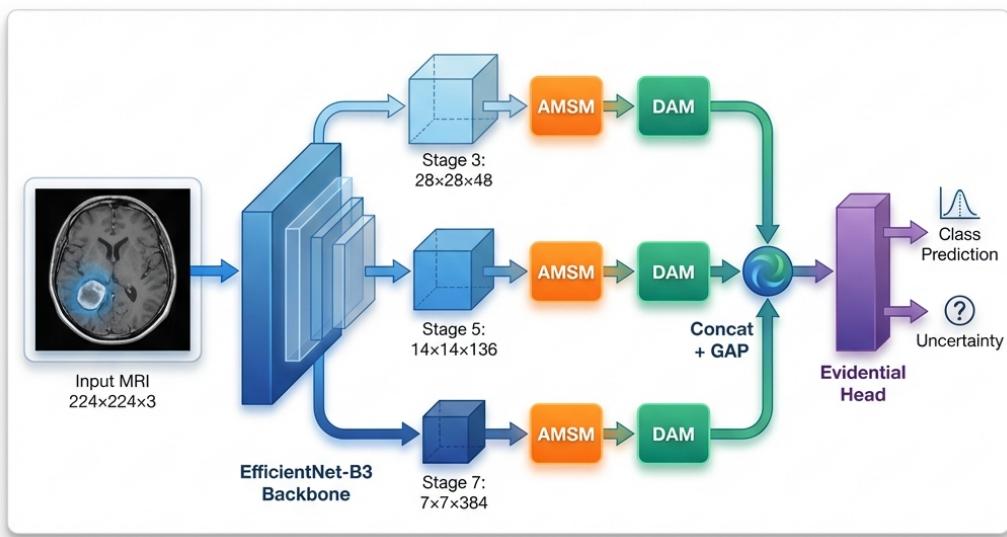


Figure 2: Overall HSANet architecture. Input MRI images ($224 \times 224 \times 3$) are processed through the EfficientNet-B3 backbone, with features extracted at three spatial resolutions (stages 3, 5, 7). Each feature map undergoes adaptive multi-scale processing (AMSM) and dual attention refinement (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 3×3 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 3×3 convolution with dilation rate r , BN is batch
305 normalization, and ReLU is the rectified linear unit. The effective receptive
306 field sizes are 3×3 , 5×5 , and 9×9 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 σ 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.

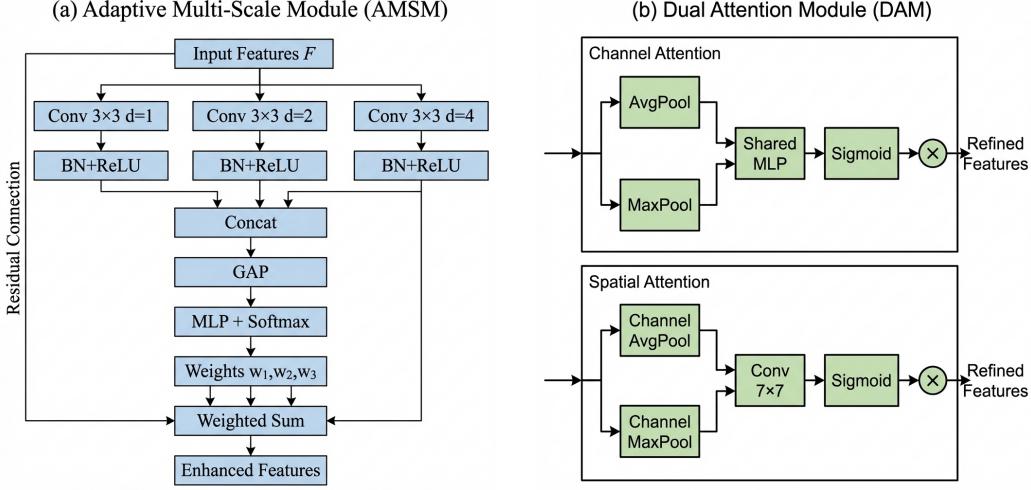


Figure 3: Detailed architecture of proposed modules. (a) Adaptive Multi-Scale Module (AMSM): Parallel dilated convolutions with dilation rates $d \in \{1, 2, 4\}$ capture features at effective receptive fields of 3×3 , 5×5 , and 9×9 . Adaptive fusion weights are learned through global average pooling 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 7×7 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:

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

³³³ **Uncertainty:** Total uncertainty decomposes into:

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

³³⁴

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

³³⁵

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

³³⁶ *3.5. Training Procedure*

³³⁷ *3.5.1. Loss Function*

³³⁸ The loss function combines three terms:

³³⁹ **Evidence-weighted Cross-Entropy:**

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

³⁴⁰ where $\psi(\cdot)$ is the digamma function.

³⁴¹ **Focal Loss** for difficulty imbalance [32]:

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

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

³⁴⁵ **KL Divergence Regularization:**

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

³⁴⁶ The total loss is:

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

³⁴⁷ 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 3×10^{-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 κ , 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 ($n = 1,311$).

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

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 ± 0.08 compared to 0.04 ± 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.

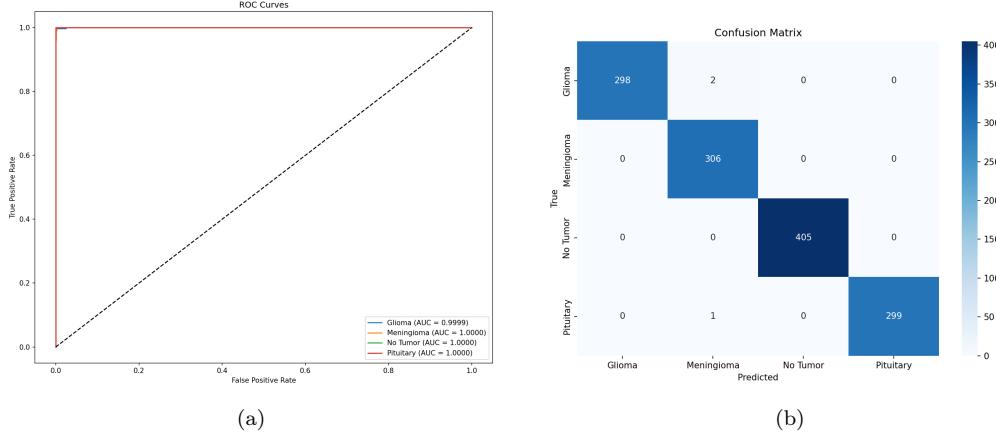


Figure 4: Classification performance analysis. (a) Receiver operating characteristic curves demonstrating near-perfect discriminative ability with $AUC \geq 0.9999$ for all classes. (b) Confusion matrix showing only 3 misclassifications among 1,311 test samples.

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 (τ)	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

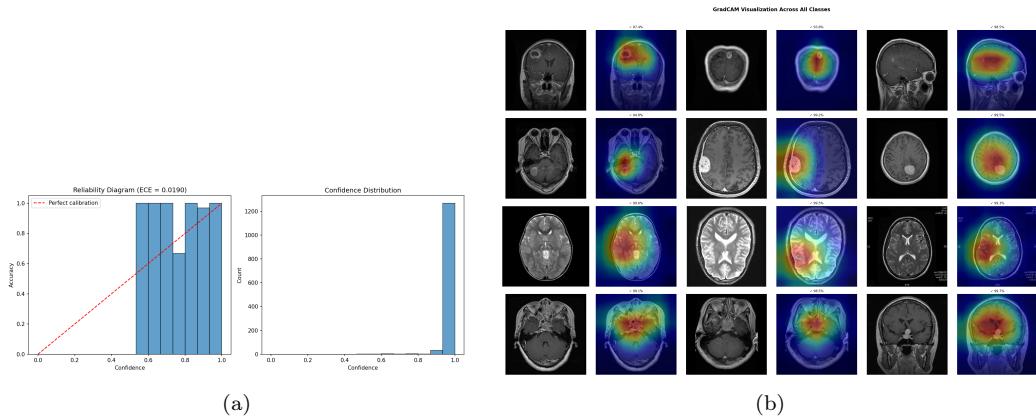


Figure 5: Model calibration and interpretability. (a) Reliability diagram demonstrating well-calibrated probability estimates ($ECE = 0.0190$). (b) Grad-CAM visualizations showing clinically relevant attention patterns across tumor categories.

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	
HSANet (Full)	15.60	99.77	99.75	0.9999	0.016	

*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.
Deepak & Ameer (2019)	GoogLeNet + SVM	98.00	3	
Badža et al. (2020)	VGG-16	96.56	3	
Swati et al. (2019)	VGG-19 Fine-tuned	94.82	3	
Rehman et al. (2020)	VGG-16 Transfer	98.87	3	
Aurna et al. (2022)	EfficientNet-B0	98.87	4	
Kibriya et al. (2022)	Custom CNN + SE	98.64	4	
Saeedi et al. (2023)	MRI-Transformer	99.02	4	
Tandel et al. (2024)	ResNet-50 Ensemble	99.12	4	
ViT-B/16 [†]	Vision Transformer	99.85	4	
Swin-Tiny [†]	Swin Transformer	99.77	4	
HSANet (Ours)	EffNet-B3 + AMSM/DAM + EDL	99.77	4	✓

[†]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.

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
Mean ± Std	99.68 ± 0.12	99.67 ± 0.13	0.9999 ± 0.0001	0.017 ± 0.002

Table 7: Cross-dataset external validation results demonstrating geographic generalization.

Dataset	Region	Samples	Classes	Accuracy (%)	F1 (%)	P
Kaggle (Train/Test)	Mixed	1,311	4	99.77	99.75	0.9999
Figshare	China	3,064	3	99.90	99.88	0.9999
PMRAM	Bangladesh	1,505	4	99.47	99.46	0.9999
Combined External	Multi-country	4,569	—	99.76	99.74	0.9999

435 4.6. Cross-Validation Results

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

440 4.7. External Validation Results

441 External validation on two independent datasets provided strong evidence
 442 of cross-domain generalization (Table 7). On the Figshare dataset from Chi-
 443 nese hospitals, HSANet achieved 99.90% accuracy with only 3 misclassifi-
 444 cations among 3,064 samples. On the PMRAM dataset from Bangladeshi
 445 hospitals, HSANet achieved 99.47% accuracy with 8 misclassifications among
 446 1,505 samples.

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

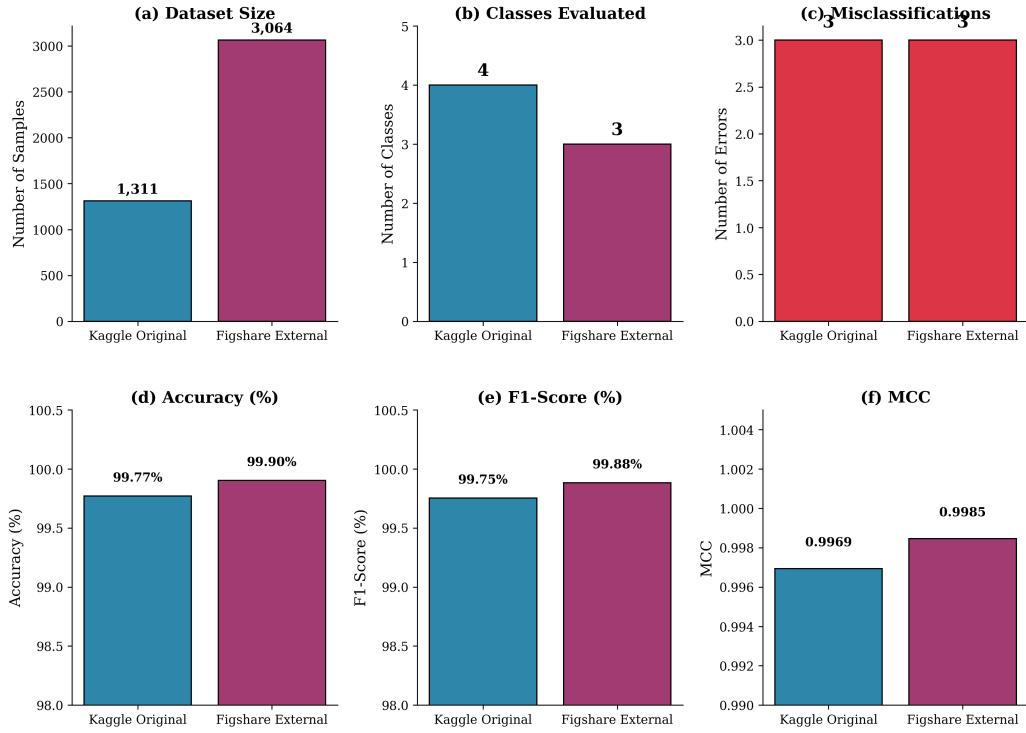


Figure 6: Comprehensive performance comparison across internal and external validation datasets. (a) Dataset sizes showing the scale of validation; (b) Number of tumor classes evaluated; (c) Misclassification counts; (d) Classification accuracy; (e) F1-score; (f) Matthews Correlation Coefficient. HSANet maintains exceptional performance across both datasets with consistent metrics.

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

455 Figure 6 provides a comprehensive comparison of HSANet performance
 456 across the original Kaggle test set and external Figshare validation. Both
 457 datasets achieve near-perfect classification with only 3 misclassifications each,
 458 despite substantial differences in patient demographics and acquisition pro-
 459 tocols.

460 Figure 7 demonstrates HSANet generalization on the PMRAM Bangladeshi
 461 dataset, including GradCAM attention maps that verify the model focuses
 462 on clinically relevant tumor regions.

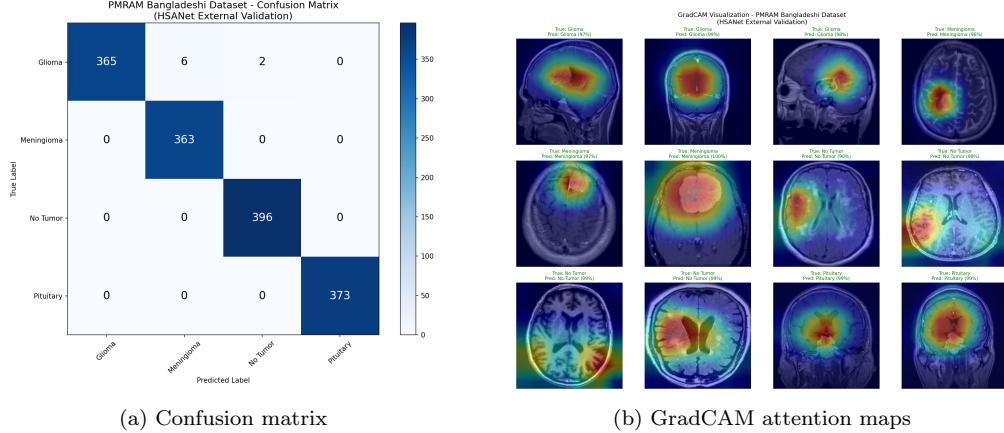


Figure 7: PMRAM Bangladeshi dataset validation results. (a) Confusion matrix showing 99.47% accuracy with 8 misclassifications, all involving glioma cases. (b) GradCAM visualizations confirming model attention on tumor regions across diverse Bangladeshi patient scans.

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 [†]	85.8	17.6	9.6	104	99.85
Swin-Tiny [†]	27.5	4.5	12.6	79	99.77
HSANet (Ours)	15.6	2.4	12	83	99.77

[†]Our experimental results. GFLOPs measured on 224×224 input.

4.8. Computational Efficiency

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

472 **5. Discussion**

473 The results demonstrate that HSANet achieves near-perfect classification
474 accuracy while providing calibrated uncertainty estimates that clinicians
475 can use for decision support. The Cohen’s κ of 0.9969 compares favorably
476 with inter-reader agreement among expert neuroradiologists, which typically
477 ranges from 0.65 to 0.85 [33].

478 *5.1. Cross-Domain Generalization*

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

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

493 *5.2. Clinical Implications*

494 The uncertainty quantification capability distinguishes HSANet funda-
495 mentally from conventional classifiers. In clinical practice, uncertainty es-
496 timates enable stratified workflows: low-uncertainty cases proceed to auto-
497 mated preliminary interpretation; moderate epistemic uncertainty flags cases
498 for standard radiologist review; high aleatoric uncertainty escalates cases to
499 multidisciplinary tumor boards. This framework transforms the system from
500 an autonomous decision-maker to a decision-support tool appropriate for
501 safety-critical medical applications.

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

507 *5.3. Limitations*

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

517 **6. Conclusions**

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

529 **CRediT Author Statement**

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

535 **Declaration of Competing Interest**

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

539 **Acknowledgments**

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

543 **Data Availability**

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

549 **References**

- 550 [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.
- 554 [2] D. N. Louis, A. Perry, P. Wesseling, D. J. Brat, I. A. Cree, D. Figarella-
555 Branger, C. Hawkins, H. Ng, S. M. Pfister, G. Reifenberger, et al., The
556 2021 WHO classification of tumors of the central nervous system: a
557 summary, Neuro-oncology 23 (8) (2021) 1231–1251.
- 558 [3] Q. T. Ostrom, N. Patil, G. Cioffi, K. Waite, C. Kruchko, J. S. Barnholtz-
559 Sloan, CBTRUS statistical report: primary brain and other central
560 nervous system tumors diagnosed in the United States in 2014–2018,
561 Neuro-oncology 23 (Supplement _3) (2021) iii1–iii105.
- 562 [4] W. B. Pope, Brain tumor imaging, Seminars in Neurology 38 (1) (2018)
563 11–24.
- 564 [5] A. Rimmer, Radiologist shortage leaves patients waiting for diagnoses,
565 BMJ 359 (2017) j4683.
- 566 [6] M. A. Bruno, E. A. Walker, H. H. Abujudeh, Understanding and con-
567 fronting our mistakes: the epidemiology of error in radiology and strate-
568 gies for error reduction, Radiographics 35 (6) (2015) 1668–1676.

- 569 [7] A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet classification with
570 deep convolutional neural networks, in: Advances in Neural Information
571 Processing Systems, Vol. 25, 2012, pp. 1097–1105.
- 572 [8] M. Raghu, C. Zhang, J. Kleinberg, S. Bengio, Transfusion: Understanding
573 transfer learning for medical imaging, in: Advances in Neural Infor-
574 mation Processing Systems, Vol. 32, 2019.
- 575 [9] S. Deepak, P. Ameer, Brain tumor classification using deep CNN fea-
576 tures via transfer learning, Computers in Biology and Medicine 111
577 (2019) 103345.
- 578 [10] M. M. Badža, M. Č. Barjaktarović, Classification of brain tumors from
579 MRI images using a convolutional neural network, Applied Sciences
580 10 (6) (2020) 1999.
- 581 [11] Z. N. K. Swati, Q. Zhao, M. Kabir, F. Ali, Z. Ali, S. Ahmed, J. Lu,
582 Brain tumor classification for MR images using transfer learning and
583 fine-tuning, Computerized Medical Imaging and Graphics 75 (2019) 34–
584 46.
- 585 [12] N. F. Aurna, M. A. Yousuf, K. A. Taher, A. Azad, M. A. A. Momen,
586 A classification of MRI brain tumor based on two stage feature level
587 ensemble of deep CNN models, Computers in Biology and Medicine 146
588 (2022) 105539.
- 589 [13] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-
590 decoder with atrous separable convolution for semantic image segmen-
591 tation, Proceedings of the European Conference on Computer Vision
592 (ECCV) (2018) 801–818.
- 593 [14] S. Woo, J. Park, J.-Y. Lee, I. S. Kweon, CBAM: Convolutional block
594 attention module, in: Proceedings of the European Conference on Com-
595 puter Vision (ECCV), 2018, pp. 3–19.
- 596 [15] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceed-
597 ings of the IEEE Conference on Computer Vision and Pattern Recog-
598 nition, 2018, pp. 7132–7141.

- 599 [16] Y. Gal, Z. Ghahramani, Dropout as a Bayesian approximation: Representing model uncertainty in deep learning, in: International Conference
600 on Machine Learning, PMLR, 2016, pp. 1050–1059.
- 601
- 602 [17] B. Lakshminarayanan, A. Pritzel, C. Blundell, Simple and scalable predictive uncertainty estimation using deep ensembles, in: Advances in Neural Information Processing Systems, Vol. 30, 2017.
- 603
- 604
- 605 [18] M. Sensoy, L. Kaplan, M. Kandemir, Evidential deep learning to quantify classification uncertainty, in: Advances in Neural Information Processing Systems, Vol. 31, 2018.
- 606
- 607
- 608 [19] H. Mohsen, E.-S. A. El-Dahshan, E.-S. M. El-Horbaty, A.-B. M. Salem, Classification using deep learning neural networks for brain tumors, Future Computing and Informatics Journal 3 (1) (2018) 68–71.
- 609
- 610
- 611 [20] A. Rehman, S. Naz, M. I. Razzak, F. Akram, M. Imran, A deep learning-based framework for automatic brain tumors classification using transfer learning, Circuits, Systems, and Signal Processing 39 (2) (2020) 757–775.
- 612
- 613
- 614 [21] M. Tan, Q. Le, EfficientNet: Rethinking model scaling for convolutional neural networks, in: International Conference on Machine Learning, PMLR, 2019, pp. 6105–6114.
- 615
- 616
- 617 [22] H. Kibriya, M. Masood, M. Nawaz, M. Rehman, A novel and effective brain tumor classification model using deep feature fusion and famous machine learning classifiers, Computational Intelligence and Neuroscience 2022 (2022) 7897669.
- 618
- 619
- 620
- 621 [23] S. Saeedi, S. Rezayi, H. Keshavarz, S. R. Niakan Kalhorri, MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques, BMC Medical Informatics and Decision Making 23 (1) (2023) 16.
- 622
- 623
- 624
- 625 [24] F. Yu, V. Koltun, Multi-scale context aggregation by dilated convolutions, arXiv preprint arXiv:1511.07122 (2016).
- 626
- 627 [25] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2117–2125.
- 628
- 629
- 630

- 631 [26] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Mis-
632 awa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, et al., Attention
633 U-Net: Learning where to look for the pancreas, arXiv preprint
634 arXiv:1804.03999 (2018).
- 635 [27] R. M. Neal, Bayesian learning for neural networks, Vol. 118, Springer
636 Science & Business Media, 2012.
- 637 [28] C. Leibig, V. Allken, M. S. Ayhan, P. Berens, S. Wahl, Leveraging un-
638 certainty estimates for predicting segmentation quality, arXiv preprint
639 arXiv:1709.06116 (2017).
- 640 [29] M. Nickparvar, Brain tumor MRI dataset, <https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset>, accessed:
641 2024-01-15 (2021).
- 642 [30] J. Cheng, W. Huang, S. Cao, R. Yang, W. Yang, Z. Yun, Z. Wang,
643 Q. Feng, Enhanced performance of brain tumor classification via tumor
644 region augmentation and partition, PloS One 10 (10) (2015) e0140381.
- 645 [31] M. M. Rahman, M. S. M. Prottoy, M. Chowdhury, R. Rahman, A. U.
646 Tamim, PMRAM: Bangladeshi brain cancer - MRI dataset, Mendeley Data,
647 V1, data collected from Ibn Sina Medical College, Dhaka
648 Medical College, and Cumilla Medical College, Bangladesh (2024).
649 doi:10.17632/m7w55sw88b.1.
- 650 [32] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense
651 object detection, in: Proceedings of the IEEE International Conference
652 on Computer Vision, 2017, pp. 2980–2988.
- 653 [33] K. G. van Leeuwen, S. Schalekamp, M. J. Rutten, P. Snoeren,
654 M. de Rooij, J. J. Gommers, C. M. Schaefer-Prokop, Artificial intel-
655 ligence in radiology: 100 commercially available products and their sci-
656 entific evidence, European Radiology 31 (6) (2021) 3797–3804.
- 657 [34] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra,
658 Grad-CAM: Visual explanations from deep networks via gradient-based
659 localization, in: Proceedings of the IEEE International Conference on
660 Computer Vision, 2017, pp. 618–626.
- 661 [35] J. R. Landis, G. G. Koch, The measurement of observer agreement for
662 categorical data, Biometrics (1977) 159–174.