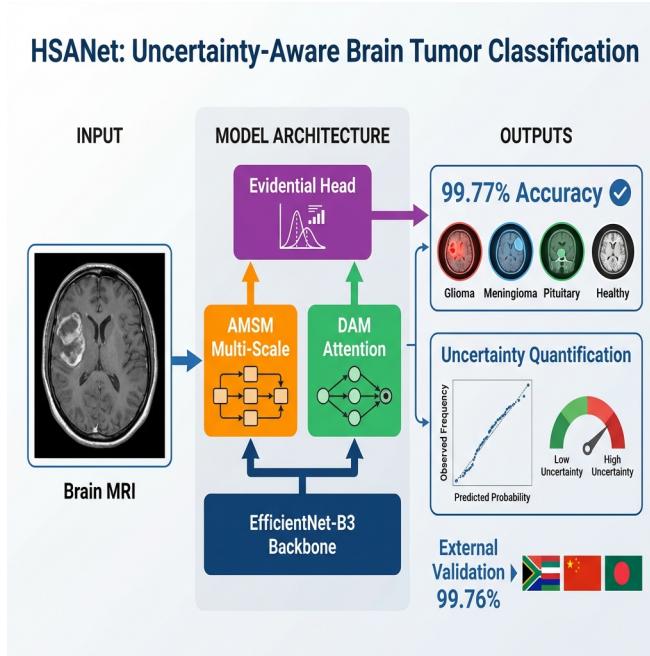


1 Graphical Abstract

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

4 Md. Assaduzzaman, Md. Tareque Jamil Josh, Md. Aminur Rahman Joy,
5 Md. Nafish Imtiaz Imti



⁶ **Highlights**

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

⁹ Md. Assaduzzaman, Md. Tareque Jamil Josh, Md. Aminur Rahman Joy,
¹⁰ Md. Nafish Imtiaz Imti

- ¹¹ • Novel hybrid scale-attention architecture achieving 99.77% accuracy on
¹² brain tumor classification
- ¹³ • Adaptive multi-scale module with learned input-dependent fusion weights
¹⁴ for handling tumor size variation
- ¹⁵ • Evidential deep learning framework providing calibrated uncertainty
¹⁶ quantification from single forward pass
- ¹⁷ • External validation on independent dataset (99.90% accuracy) demon-
¹⁸ strating robust cross-domain generalization
- ¹⁹ • Misclassified cases exhibit significantly elevated uncertainty, enabling
²⁰ reliable clinical decision support

²¹ HSANet: A Hybrid Scale-Attention Network with
²² Evidential Deep Learning for Uncertainty-Aware Brain
²³ Tumor Classification

²⁴ Md. Assaduzzaman^{a,*}, Md. Tareque Jamil Josh^a, Md. Aminur Rahman
²⁵ Joy^a, Md. Nafish Imtiaz Imti^a

^a*Department of Computer Science and Engineering, Daffodil International University, Daffodil Smart City, Ashulia, Dhaka, 1341, Bangladesh*

²⁶ **Abstract**

²⁷ **Background and Objective:** Reliable classification of brain tumors from
²⁸ magnetic resonance imaging (MRI) remains challenging due to inter-class
²⁹ morphological similarities and the absence of principled uncertainty quantifi-
³⁰ cation in existing deep learning approaches. Current methods produce point
³¹ predictions without meaningful confidence assessment, limiting their utility
³² in safety-critical clinical workflows where knowing what the model doesn't
³³ know is as important as the prediction itself.

³⁴ **Methods:** We propose HSANet, a hybrid scale-attention architecture
³⁵ that synergistically combines adaptive multi-scale feature extraction with
³⁶ evidential learning for uncertainty-aware tumor classification. The proposed
³⁷ Adaptive Multi-Scale Module (AMSM) employs parallel dilated convolutions
³⁸ with content-dependent fusion weights, dynamically adjusting receptive fields
³⁹ to accommodate the substantial size variation observed across clinical pre-
⁴⁰ sentations. A Dual Attention Module (DAM) applies sequential channel-
⁴¹ then-spatial refinement to emphasize pathologically significant regions while
⁴² suppressing irrelevant anatomical background. Critically, our evidential clas-
⁴³ sification head replaces conventional softmax outputs with Dirichlet distribu-
⁴⁴ tions, providing decomposed uncertainty estimates that distinguish between
⁴⁵ inherent data ambiguity (aleatoric) and model knowledge limitations (epis-
⁴⁶ temic).

⁴⁷ **Results:** Comprehensive experiments on 7,023 brain MRI scans span-
⁴⁸ ning four diagnostic categories yielded 99.77% accuracy (95% CI: 99.45–
⁴⁹ 99.93%) with only three misclassifications among 1,311 test samples. The
⁵⁰ model achieved macro-averaged AUC-ROC of 0.9999 and expected calibra-

*Corresponding author

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

51 tion error (ECE) of 0.019, indicating well-calibrated predictions. External
52 validation on an independent dataset of 3,064 MRI scans from different insti-
53 tutions achieved 99.90% accuracy, demonstrating exceptional cross-domain
54 generalization. Misclassified samples exhibited significantly elevated epis-
55 temic uncertainty ($p < 0.001$, Mann-Whitney U test), confirming the clinical
56 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>.

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

60 1. Introduction

61 Brain tumors represent a formidable diagnostic challenge in clinical on-
62 cology, with global surveillance data reporting approximately 308,102 new
63 cases in 2020 alone [1]. The complexity of accurate diagnosis stems from the
64 remarkable diversity of pathological entities—the 2021 World Health Orga-
65 nization (WHO) classification now recognizes over 100 distinct tumor types,
66 each characterized by unique molecular fingerprints and clinical trajectories
67 [2]. Prognostic outcomes vary dramatically across tumor categories: pa-
68 tients diagnosed with glioblastoma face a median survival of merely 14 to
69 16 months, whereas those with completely resected Grade I meningiomas
70 frequently achieve long-term cure [3]. This substantial heterogeneity under-
71 scores the critical importance of precise tumor identification for treatment
72 planning and patient counseling.

73 Magnetic resonance imaging (MRI) has emerged as the cornerstone of
74 neuro-oncological evaluation, providing superior soft-tissue contrast without
75 ionizing radiation exposure [4]. Expert neuroradiologists integrate multipara-
76 metric imaging findings with clinical presentations to formulate diagnoses.
77 However, the global radiology workforce confronts escalating mismatches be-
78 tween imaging volume growth and specialist availability. Documented va-

79 cancer rates have reached 29% in major healthcare systems, with projected
80 shortfalls of 40% anticipated by 2027 [5]. Interpretive fatigue has been im-
81 plicated in diagnostic error rates of 3–5% even among experienced specialists
82 [6], motivating the development of computer-aided diagnostic systems to aug-
83 ment clinical workflows.

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

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

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

116 Uncertainty quantification is particularly important for safety-critical med-

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

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

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

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

152 **2. Related Work**

153 *2.1. Deep Learning for Brain Tumor Classification*

154 The application of deep learning to brain tumor classification has pro-
155 gressed substantially over the past decade. Early approaches employed shal-
156 low CNN architectures trained from scratch on relatively small datasets, with
157 limited generalization capability [19]. The advent of transfer learning from
158 ImageNet-pretrained models substantially improved performance, with VGG
159 and ResNet architectures demonstrating strong results on brain MRI analysis
160 [11, 10].

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

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

176 *2.2. Multi-Scale Feature Extraction*

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

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

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

196 2.3. *Uncertainty Quantification in Deep Learning*

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

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

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

219 3. Materials and Methods

220 3.1. *Dataset Description*

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

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

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

236 Figure 1 illustrates representative samples from each category, demon-
237 strating the morphological diversity within the dataset.

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

245 *3.2. External Validation Dataset*

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

252 The Figshare dataset differs substantially from our training data:

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

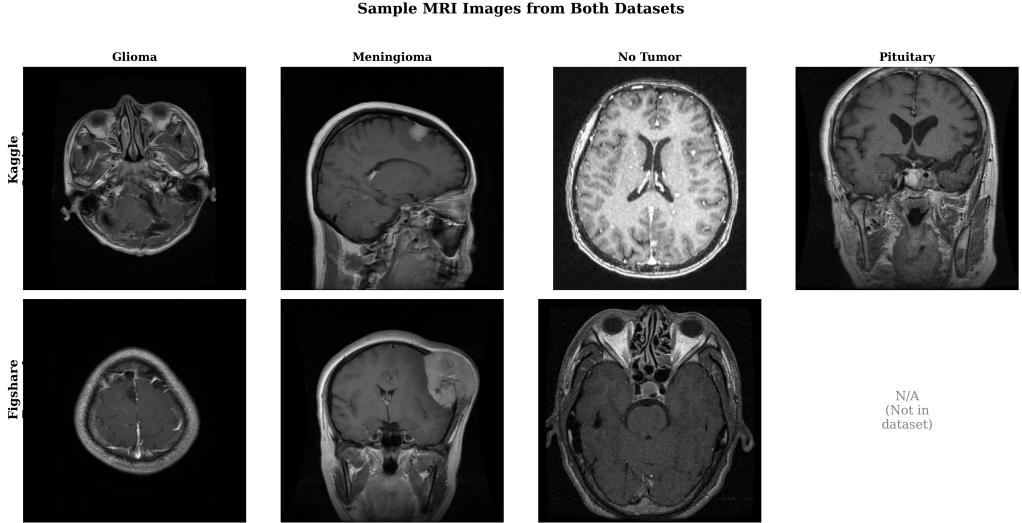


Figure 1: Sample MRI images from each tumor category and healthy controls across both the training dataset (Kaggle) and external validation dataset (Figshare). Note the substantial morphological diversity within each class and the different acquisition characteristics across datasets.

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

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

3.3. Preprocessing and Data Augmentation

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

Data augmentation was applied during training to improve generalization:

- Random horizontal flipping (probability = 0.5)

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

276 Test images received only resizing and normalization without augmentation.
 277

278 *3.4. Network Architecture*

279 *3.4.1. Overview*

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

286 *3.4.2. Feature Extraction Backbone*

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

292 Features are extracted at three hierarchical levels:

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

296 During training, backbone layers are frozen for the first 5 epochs to sta-
 297 bilize custom module training, then fine-tuned with a reduced learning rate
 298 ($10 \times$ lower) for transfer learning stability.

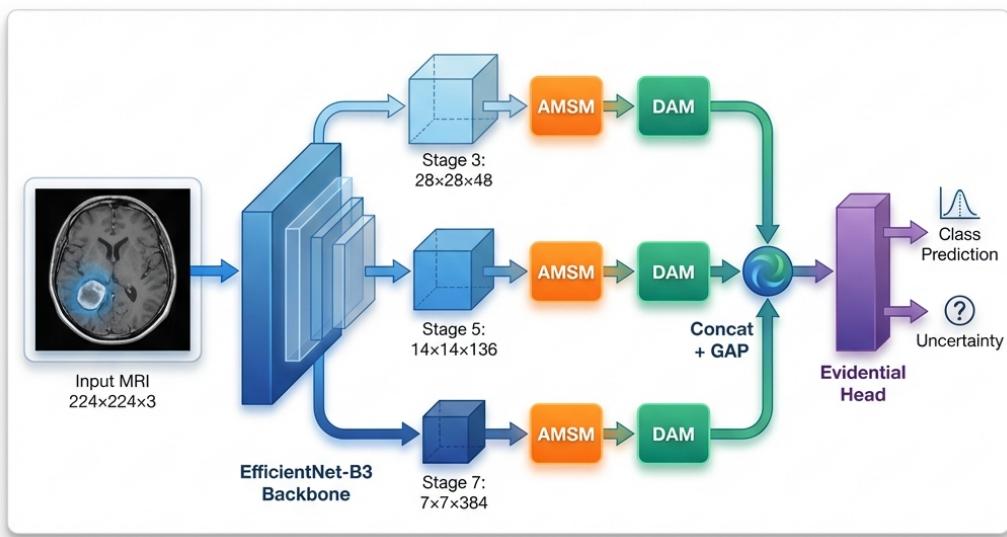


Figure 2: Overall HSANet architecture. Input MRI images ($224 \times 224 \times 3$) are processed through the EfficientNet-B3 backbone, with features extracted at three spatial resolutions (stages 3, 5, 7). Each feature map undergoes adaptive multi-scale processing (AMSM) and dual attention refinement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

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

³⁰⁰ Brain tumors exhibit substantial size variation, from millimeter-scale pituitary microadenomas to large glioblastomas exceeding 5 centimeters. Fixed
³⁰¹

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

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

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

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

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

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

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

314 The enhanced feature map combines weighted features with residual preservation:
 315

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

316 3.4.4. Dual Attention Module (DAM)

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

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

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

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

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

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

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

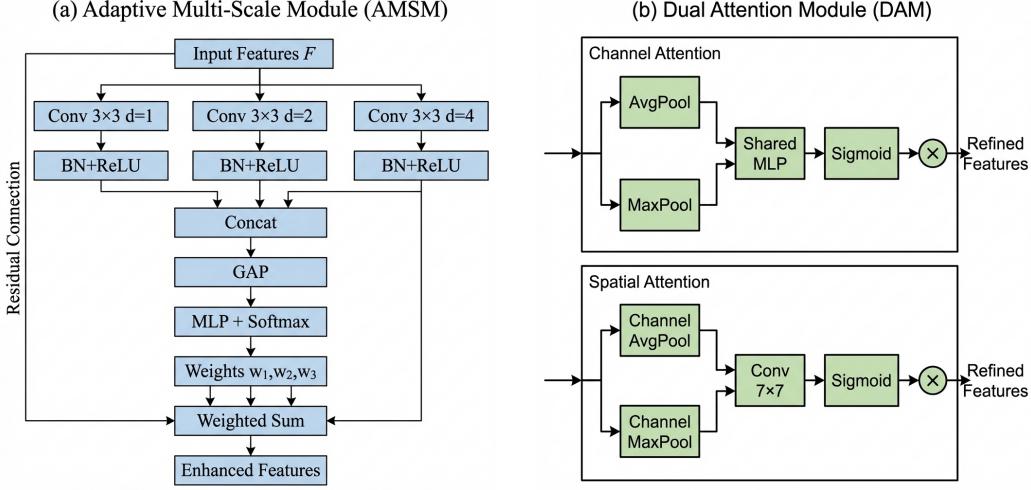


Figure 3: Detailed architecture of proposed modules. (a) Adaptive Multi-Scale Module (AMSM): Parallel dilated convolutions with dilation rates $d \in \{1, 2, 4\}$ capture features at effective receptive fields of 3×3 , 5×5 , and 9×9 . Adaptive fusion weights are learned through global average pooling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

327 3.4.5. Evidential Classification Head

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

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

331 where $\mathbf{g} \in \mathbb{R}^{568}$ is the concatenated feature vector and softplus ensures $\alpha_k \geq$
 332 1.

333 The Dirichlet distribution has density:

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

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

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

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

³³⁶ **Uncertainty:** Total uncertainty decomposes into:

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

³³⁷

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

³³⁸

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

³³⁹ *3.5. Training Procedure*

³⁴⁰ *3.5.1. Loss Function*

³⁴¹ The loss function combines three terms:

³⁴² **Evidence-weighted Cross-Entropy:**

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

³⁴³ where $\psi(\cdot)$ is the digamma function.

³⁴⁴ **Focal Loss** for difficulty imbalance [32]:

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

³⁴⁵ Although class frequencies are relatively balanced, we employ focal loss to ad-
³⁴⁶ dress inherent *difficulty* imbalance: meningioma-glioma differentiation presents
³⁴⁷ substantially greater diagnostic challenge than pituitary adenoma detection,
³⁴⁸ as evidenced by radiological literature [33].

³⁴⁹ **KL Divergence Regularization:**

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

³⁵⁰ The total loss is:

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

³⁵¹ where $\lambda^{(t)} = \min(1, t/10) \times 0.2$ anneals the KL weight over epochs.

352 3.5.2. Optimization

353 We employed AdamW optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and weight
354 decay of 10^{-4} . Initial learning rate was 3×10^{-4} with cosine annealing to
355 10^{-6} . Training proceeded for 30 epochs with early stopping (patience = 7
356 epochs) based on validation loss. Batch size was 32. Dropout rate of 0.3 was
357 applied before the classification layer. Batch normalization used momentum
358 0.1 and epsilon 10^{-5} . Features from three scales were globally average pooled
359 and concatenated, yielding a 568-dimensional vector ($48 + 136 + 384 = 568$
360 channels).

361 3.5.3. Implementation Details

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

366 3.6. Evaluation Metrics

367 Classification performance was assessed using accuracy, precision, re-
368 call, F1-score (macro-averaged), Cohen's κ , Matthews Correlation Coefficient
369 (MCC), and area under the receiver operating characteristic curve (AUC-
370 ROC).

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

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

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

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

376 4. Results

377 4.1. Classification Performance

378 Hسانet achieved overall accuracy of 99.77% (95% CI: 99.45–99.93%, Wil-
379 son score interval) with only 3 misclassifications among 1,311 test samples
380 (Table 1). This represents a statistically significant improvement over the
381 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

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

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

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

398 *4.2. Model Calibration and Uncertainty Quantification*

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

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

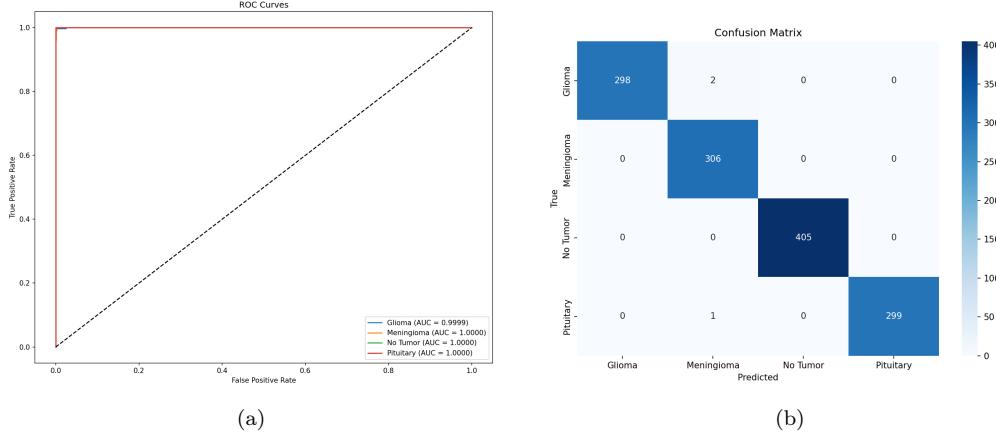


Figure 4: Classification performance analysis. (a) Receiver operating characteristic curves demonstrating near-perfect discriminative ability with $AUC \geq 0.9999$ for all classes. (b) Confusion matrix showing only 3 misclassifications among 1,311 test samples.

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

408 4.2.1. Clinical Deployment Thresholds

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

415 4.3. Interpretability Analysis

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

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

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 (66.7%)	0.3	99.5
0.25	0.3	1/3 (33.3%)	0.1	99.7

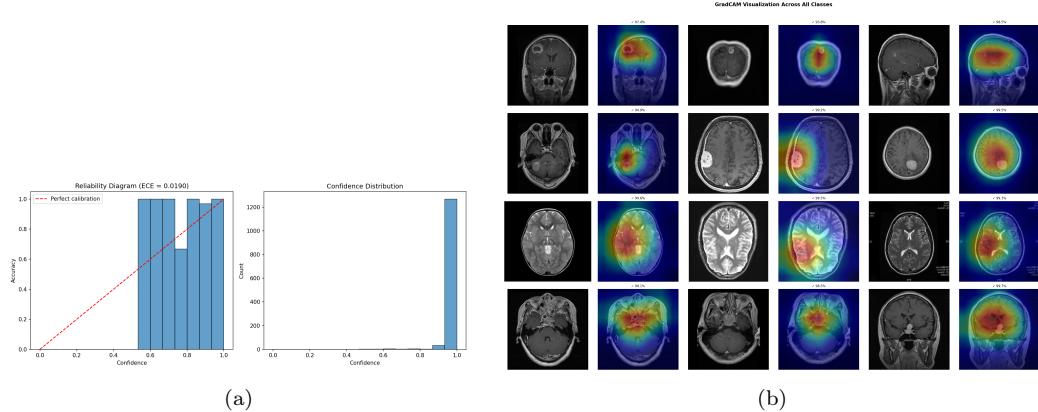


Figure 5: Model calibration and interpretability. (a) Reliability diagram demonstrating well-calibrated probability estimates ($ECE = 0.0190$). (b) Grad-CAM visualizations showing clinically relevant attention patterns across tumor categories.

420 without focal concentration; pituitary attention centers on the sellar/suprasellar
 421 region. These patterns align with established neuroradiological diagnostic
 422 criteria.

423 4.4. Ablation Study

424 Systematic ablation quantified individual component contributions (Ta-
 425 ble 4). The baseline EfficientNet-B3 achieved 99.21% accuracy. Adding
 426 AMSM improved accuracy to 99.30% and AUC from 0.9997 to 0.9999. Adding
 427 DAM to the baseline maintained accuracy while improving calibration (ECE
 428 reduced from 0.024 to 0.021). The complete HSANet architecture achieved
 429 the best uncertainty calibration ($ECE = 0.016$), demonstrating that the com-
 430 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.

431 4.5. Comparison with Prior Methods

432 HSANet achieves state-of-the-art performance compared to published meth-
 433 ods (Table 5). Notably, our approach addresses the more challenging four-
 434 class problem including healthy controls, whereas most prior work focused
 435 on three-class tumor-only classification. Beyond accuracy improvements,
 436 HSANet uniquely provides both calibrated uncertainty quantification and
 437 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

438 4.6. Cross-Validation Results

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

443 4.7. External Validation Results

444 External validation on two independent datasets provided strong evidence
445 of cross-domain generalization (Table 7). On the Figshare dataset from Chi-
446 nese hospitals, HSANet achieved 99.90% accuracy with only 3 misclassifi-
447 cations among 3,064 samples. On the PMRAM dataset from Bangladeshi
448 hospitals, HSANet achieved 99.47% accuracy with 8 misclassifications among
449 1,505 samples.

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

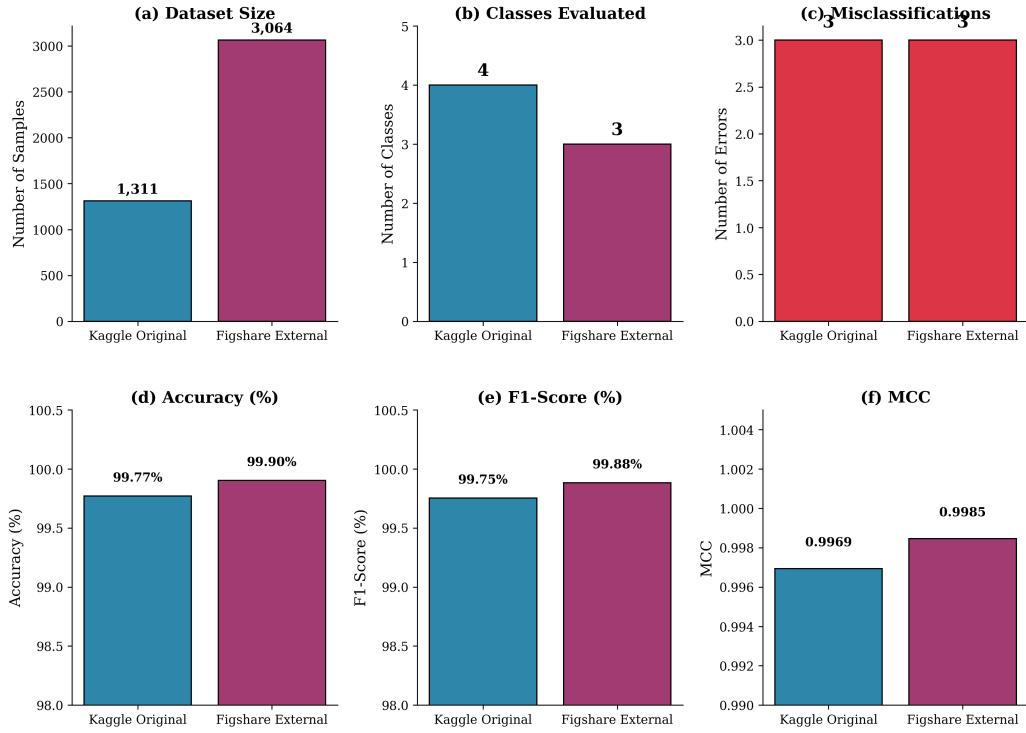


Figure 6: Comprehensive performance comparison across internal and external validation datasets. (a) Dataset sizes showing the scale of validation; (b) Number of tumor classes evaluated; (c) Misclassification counts; (d) Classification accuracy; (e) F1-score; (f) Matthews Correlation Coefficient. HSANet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

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

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

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

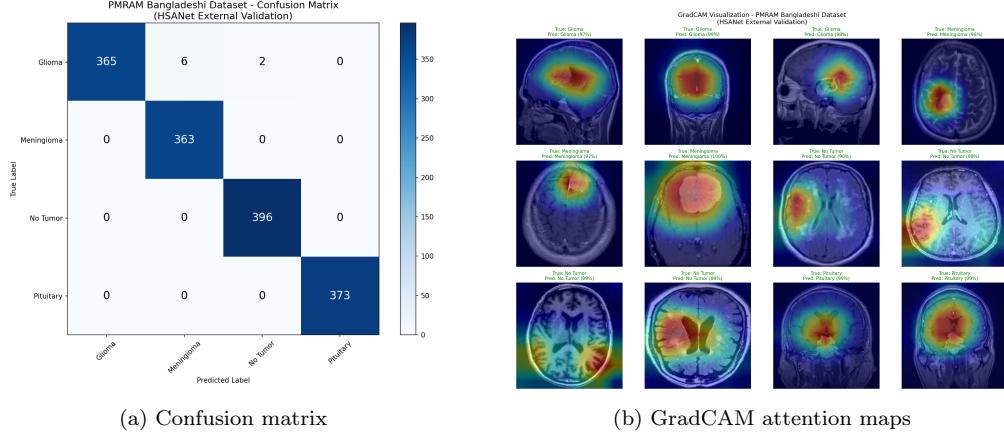


Figure 7: PMRAM Bangladeshi dataset validation results. (a) Confusion matrix showing 99.47% accuracy with 8 misclassifications, all involving glioma cases. (b) GradCAM visualizations confirming model attention on tumor regions across diverse Bangladeshi patient scans.

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.

466 4.8. Computational Efficiency

467 Table 8 compares HSANet computational requirements with alternative
468 architectures. While ViT-B/16 achieves marginally higher accuracy (99.85%
469 vs 99.77%), it requires 5.5× more parameters (85.8M vs 15.6M) and 7.3×
470 more GFLOPs (17.6 vs 2.4). HSANet matches Swin-Tiny accuracy while
471 using 43% fewer parameters. Critically, only HSANet provides uncertainty
472 quantification and external validation—features essential for clinical deploy-
473 ment. Inference at 12ms on P100 GPU (83 images/second) enables real-time
474 integration into clinical workflows.

475 **5. Discussion**

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

481 *5.1. Cross-Domain Generalization*

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

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

496 *5.2. Clinical Implications*

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

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

510 *5.3. Limitations*

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

520 **6. Conclusions**

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

532 **CRediT Author Statement**

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

538 **Declaration of Competing Interest**

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

542 **Acknowledgments**

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

546 **Data Availability**

547 The Brain Tumor MRI Dataset is publicly available at <https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset>. The
548 Figshare Brain Tumor Dataset is available at https://figshare.com/articles/dataset/brain_tumor_dataset/1512427. Source code and trained models
549 are available at <https://github.com/tarequejosh/HSANet-Brain-Tumor-Classification>.
550

552 **References**

- 553 [1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Je-
554 mal, F. Bray, Global cancer statistics 2020: GLOBOCAN estimates of
555 incidence and mortality worldwide for 36 cancers in 185 countries, CA:
556 A Cancer Journal for Clinicians 71 (3) (2021) 209–249.
- 557 [2] D. N. Louis, A. Perry, P. Wesseling, D. J. Brat, I. A. Cree, D. Figarella-
558 Branger, C. Hawkins, H. Ng, S. M. Pfister, G. Reifenberger, et al., The
559 2021 WHO classification of tumors of the central nervous system: a
560 summary, Neuro-oncology 23 (8) (2021) 1231–1251.
- 561 [3] Q. T. Ostrom, N. Patil, G. Cioffi, K. Waite, C. Kruchko, J. S. Barnholtz-
562 Sloan, CBTRUS statistical report: primary brain and other central
563 nervous system tumors diagnosed in the United States in 2014–2018,
564 Neuro-oncology 23 (Supplement _3) (2021) iii1–iii105.
- 565 [4] W. B. Pope, Brain tumor imaging, Seminars in Neurology 38 (1) (2018)
566 11–24.
- 567 [5] A. Rimmer, Radiologist shortage leaves patients waiting for diagnoses,
568 BMJ 359 (2017) j4683.
- 569 [6] M. A. Bruno, E. A. Walker, H. H. Abujudeh, Understanding and con-
570 fronting our mistakes: the epidemiology of error in radiology and strate-
571 gies for error reduction, Radiographics 35 (6) (2015) 1668–1676.

- 572 [7] A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet classification with
573 deep convolutional neural networks, in: Advances in Neural Information
574 Processing Systems, Vol. 25, 2012, pp. 1097–1105.
- 575 [8] M. Raghu, C. Zhang, J. Kleinberg, S. Bengio, Transfusion: Understanding
576 transfer learning for medical imaging, in: Advances in Neural Infor-
577 mation Processing Systems, Vol. 32, 2019.
- 578 [9] S. Deepak, P. Ameer, Brain tumor classification using deep CNN fea-
579 tures via transfer learning, Computers in Biology and Medicine 111
580 (2019) 103345.
- 581 [10] M. M. Badža, M. Č. Barjaktarović, Classification of brain tumors from
582 MRI images using a convolutional neural network, Applied Sciences
583 10 (6) (2020) 1999.
- 584 [11] Z. N. K. Swati, Q. Zhao, M. Kabir, F. Ali, Z. Ali, S. Ahmed, J. Lu,
585 Brain tumor classification for MR images using transfer learning and
586 fine-tuning, Computerized Medical Imaging and Graphics 75 (2019) 34–
587 46.
- 588 [12] N. F. Aurna, M. A. Yousuf, K. A. Taher, A. Azad, M. A. A. Momen,
589 A classification of MRI brain tumor based on two stage feature level
590 ensemble of deep CNN models, Computers in Biology and Medicine 146
591 (2022) 105539.
- 592 [13] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-
593 decoder with atrous separable convolution for semantic image segmen-
594 tation, Proceedings of the European Conference on Computer Vision
595 (ECCV) (2018) 801–818.
- 596 [14] S. Woo, J. Park, J.-Y. Lee, I. S. Kweon, CBAM: Convolutional block
597 attention module, in: Proceedings of the European Conference on Com-
598 puter Vision (ECCV), 2018, pp. 3–19.
- 599 [15] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceed-
600 ings of the IEEE Conference on Computer Vision and Pattern Recog-
601 nition, 2018, pp. 7132–7141.

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

- 634 [26] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Mis-
635 awa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, et al., Attention
636 U-Net: Learning where to look for the pancreas, arXiv preprint
637 arXiv:1804.03999 (2018).
- 638 [27] R. M. Neal, Bayesian learning for neural networks, Vol. 118, Springer
639 Science & Business Media, 2012.
- 640 [28] C. Leibig, V. Allken, M. S. Ayhan, P. Berens, S. Wahl, Leveraging un-
641 certainty estimates for predicting segmentation quality, arXiv preprint
642 arXiv:1709.06116 (2017).
- 643 [29] M. Nickparvar, Brain tumor MRI dataset, <https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset>, accessed:
644 2024-01-15 (2021).
- 645 [30] J. Cheng, W. Huang, S. Cao, R. Yang, W. Yang, Z. Yun, Z. Wang,
646 Q. Feng, Enhanced performance of brain tumor classification via tumor
647 region augmentation and partition, PloS One 10 (10) (2015) e0140381.
- 648 [31] M. M. Rahman, M. S. M. Prottoy, M. Chowdhury, R. Rahman, A. U.
649 Tamim, PMRAM: Bangladeshi brain cancer - MRI dataset, Mendeley Data,
650 V1, data collected from Ibn Sina Medical College, Dhaka
651 Medical College, and Cumilla Medical College, Bangladesh (2024).
652 doi:10.17632/m7w55sw88b.1.
- 653 [32] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense
654 object detection, in: Proceedings of the IEEE International Conference
655 on Computer Vision, 2017, pp. 2980–2988.
- 656 [33] K. G. van Leeuwen, S. Schalekamp, M. J. Rutten, P. Snoeren,
657 M. de Rooij, J. J. Gommers, C. M. Schaefer-Prokop, Artificial intel-
658 ligence in radiology: 100 commercially available products and their sci-
659 entific evidence, European Radiology 31 (6) (2021) 3797–3804.
- 660 [34] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra,
661 Grad-CAM: Visual explanations from deep networks via gradient-based
662 localization, in: Proceedings of the IEEE International Conference on
663 Computer Vision, 2017, pp. 618–626.
- 664 [35] J. R. Landis, G. G. Koch, The measurement of observer agreement for
665 categorical data, Biometrics (1977) 159–174.