

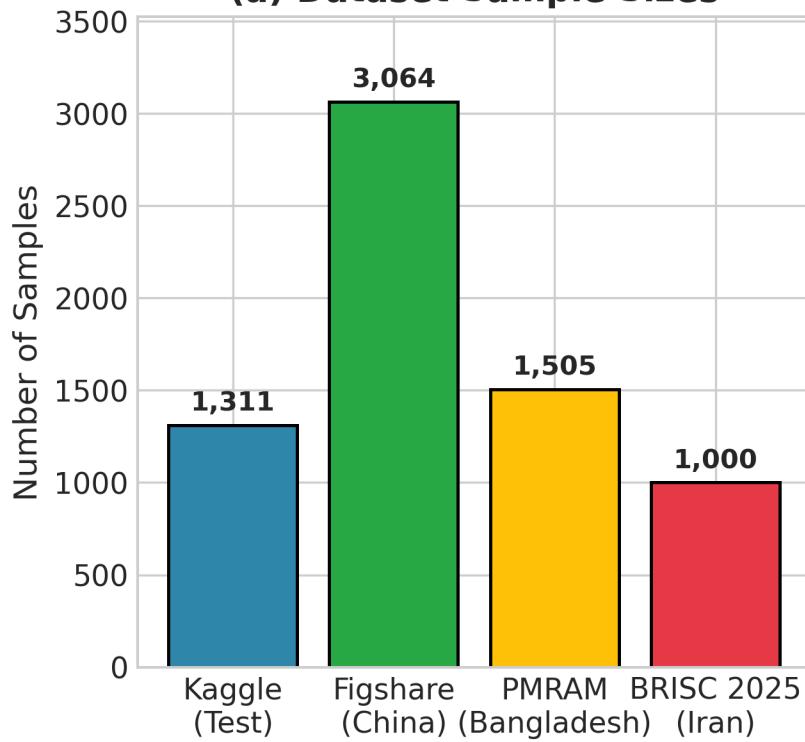
¹ Graphical Abstract

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

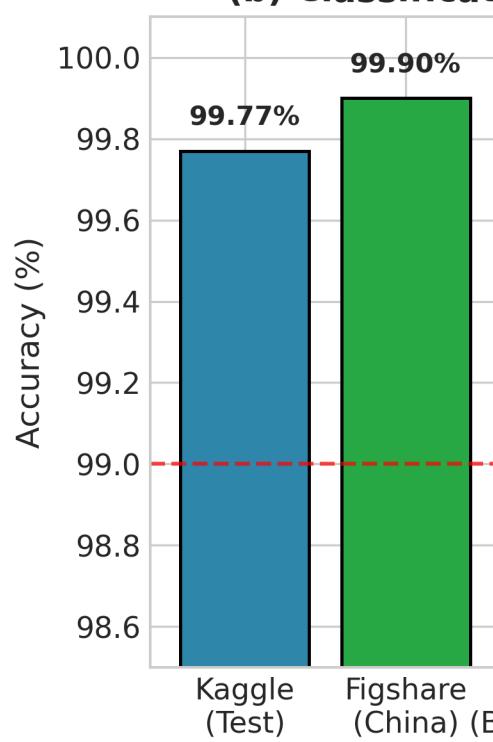
⁴ Author 1, Author 2, Author 3, Author 4

HSANet External Validation Results

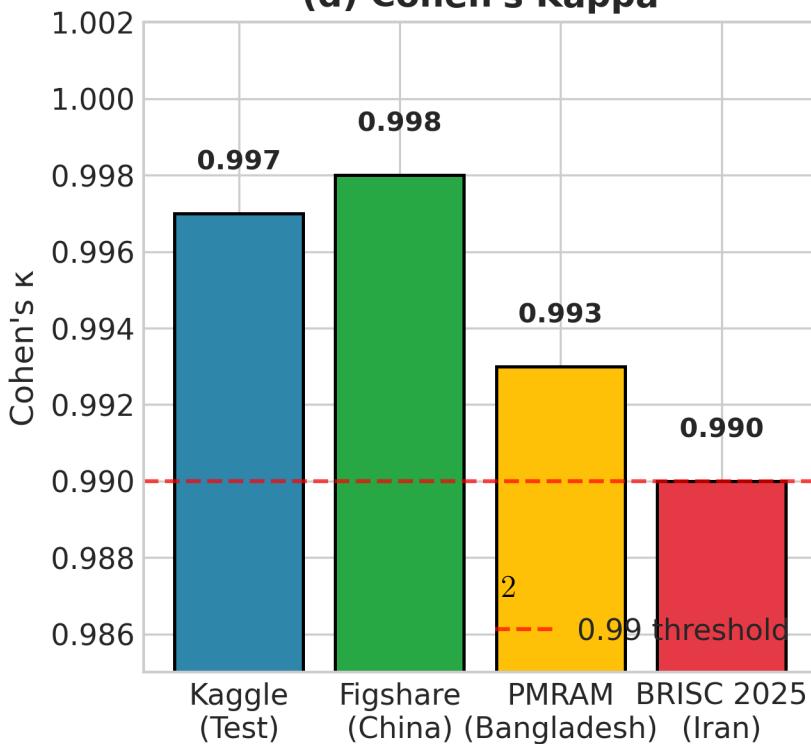
(a) Dataset Sample Sizes



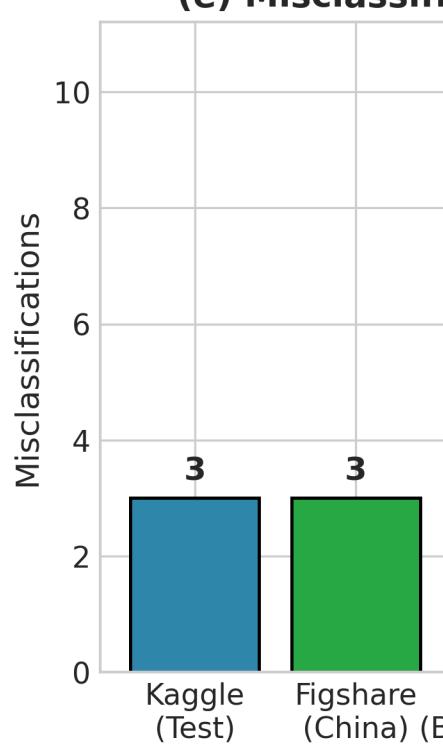
(b) Classification Accuracy (%)



(d) Cohen's Kappa



(e) Misclassifications



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 three independent datasets (5,569 samples, 99.59%
16 combined accuracy) demonstrating 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 spanning
45 four diagnostic categories yielded 99.77% accuracy (95% CI: 99.45–99.93%)
46 with only three misclassifications among 1,311 test samples. The model
47 achieved macro-averaged AUC-ROC of 0.9999 and expected calibration er-
48 ror (ECE) of 0.019, indicating well-calibrated predictions. External valida-

*Corresponding author

Email address: `author1@institution.edu` (Author 1)

49 tion on three independent datasets—Figshare (n=3,064; 99.90% accuracy),
50 PMRAM (n=1,505; 99.47%), and BRISC 2025 (n=1,000; 99.30%)—totaling
51 5,569 samples from China, Bangladesh, and Iran, demonstrated exceptional
52 cross-domain generalization with combined accuracy of 99.59%. Misclassi-
53 fied samples exhibited significantly elevated epistemic uncertainty ($p < 0.001$,
54 Mann-Whitney U test), confirming the clinical utility of uncertainty-guided
55 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>.

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

59 1. Introduction

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

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

77 tween imaging volume growth and specialist availability. Documented va-
78 cancy rates have reached 29% in major healthcare systems, with projected
79 shortfalls of 40% anticipated by 2027 [5]. Interpretive fatigue has been im-
80 plicated in diagnostic error rates of 3–5% even among experienced specialists
81 [6], motivating the development of computer-aided diagnostic systems to aug-
82 ment clinical workflows.

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

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

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

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

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

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

143 Comprehensive experiments on a challenging four-class brain tumor bench-
144 mark demonstrate that HSANet achieves 99.77% classification accuracy while
145 providing well-calibrated uncertainty estimates. Importantly, misclassified
146 samples exhibit significantly elevated epistemic uncertainty, confirming that
147 the model appropriately flags uncertain predictions for expert review. Exter-
148 nal validation on three independent datasets—Figshare ($n=3,064$; 99.90%),
149 PMRAM ($n=1,505$; 99.47%), and BRISC 2025 ($n=1,000$; 99.30%)—totaling
150 5,569 samples from multiple countries achieved 99.59% combined accuracy,

151 providing strong evidence of cross-domain generalizability essential for clinical
152 deployment.

153 **2. Related Work**

154 *2.1. Deep Learning for Brain Tumor Classification*

155 The application of deep learning to brain tumor classification has pro-
156 gressed substantially over the past decade. Early approaches employed shal-
157 low CNN architectures trained from scratch on relatively small datasets, with
158 limited generalization capability [19]. The advent of transfer learning from
159 ImageNet-pretrained models substantially improved performance, with VGG
160 and ResNet architectures demonstrating strong results on brain MRI analysis
161 [11, 10].

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

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

177 *2.2. Multi-Scale Feature Extraction*

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

184 In medical imaging, multi-scale approaches have been applied to various
185 modalities. Feature pyramid networks [25] aggregate features across multiple

186 resolution levels. Multi-scale attention mechanisms [26] have been proposed
187 for medical image segmentation, where tumors and anatomical structures
188 exhibit substantial size variation.

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

197 2.3. Uncertainty Quantification in Deep Learning

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

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

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

220 **3. Materials and Methods**

221 *3.1. Dataset Description*

222 Experiments utilized the Brain Tumor MRI Dataset [29], a publicly avail-
223 able collection comprising 7,023 T1-weighted gadolinium-enhanced MRI scans.
224 The dataset is available at [https://www.kaggle.com/datasets/masoudnickparvar/](https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset)
225 **brain-tumor-mri-dataset**. Images span four diagnostic categories with the
226 following distribution:

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

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

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

246 *3.2. External Validation Dataset*

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

253 The Figshare dataset differs substantially from our training data:

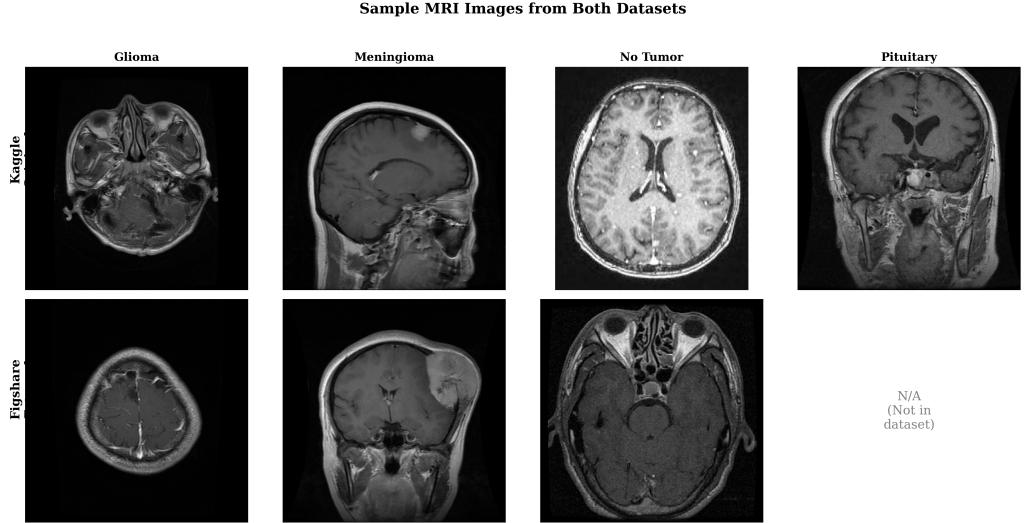


Figure 1: Sample MRI images from each tumor category and healthy controls across both the training dataset (Kaggle) and external validation dataset (Figshare). Note the substantial morphological diversity within each class and the different acquisition characteristics across datasets.

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

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

266 Furthermore, we validated on the BRISC 2025 dataset [32], a recent
 267 expert-annotated collection of 6,000 T1-weighted brain MRI slices from Ira-
 268 nian institutions. We used the official test split of 1,000 images comprising all
 269 four categories: glioma ($n=254$), meningioma ($n=306$), no tumor ($n=140$),
 270 and pituitary adenoma ($n=300$). BRISC 2025 provides additional validation

271 on a Middle Eastern population with physician-reviewed annotations, further
272 extending the geographic diversity of our external validation.

273 *3.3. Preprocessing and Data Augmentation*

274 All input images were resized to 224×224 pixels using bilinear interpolation
275 to match EfficientNet-B3 input specifications. Pixel intensities were
276 normalized using ImageNet statistics (mean = [0.485, 0.456, 0.406], std =
277 [0.229, 0.224, 0.225]) to leverage pretrained representations effectively.

278 Data augmentation was applied during training to improve generalization:

- 279 • Random horizontal flipping (probability = 0.5)
- 280 • Random rotation ($\pm 15^\circ$)
- 281 • Random affine transformations (scale: 0.9–1.1, translation: $\pm 10\%$)
- 282 • Color jittering (brightness/contrast: $\pm 10\%$)
- 283 • Random erasing (probability = 0.2, scale: 0.02–0.33)

284 Test images received only resizing and normalization without augmentation.
285

286 *3.4. Network Architecture*

287 *3.4.1. Overview*

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

294 *3.4.2. Feature Extraction Backbone*

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

300 Features are extracted at three hierarchical levels:

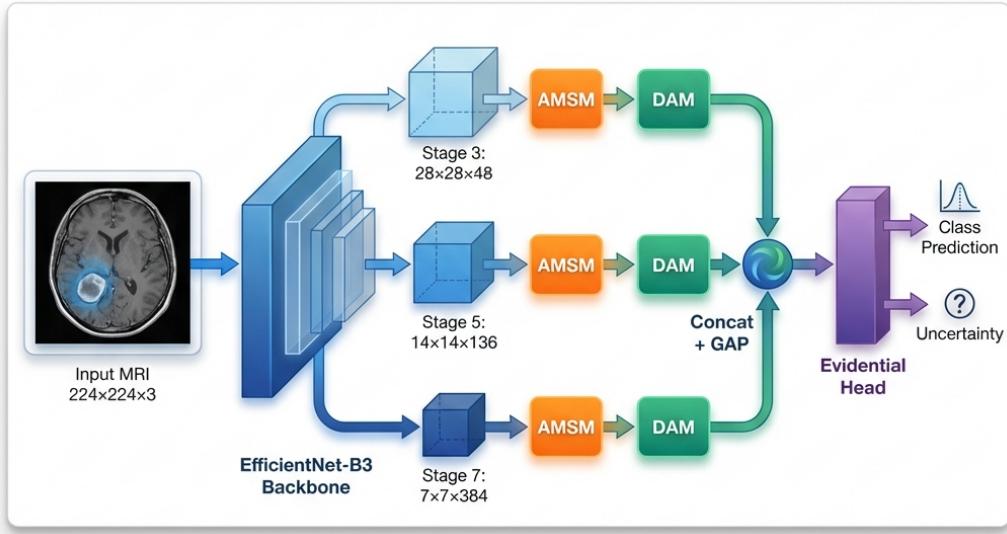


Figure 2: Overall HSANet architecture. Input MRI images ($224 \times 224 \times 3$) are processed through the EfficientNet-B3 backbone, with features extracted at three spatial resolutions (stages 3, 5, 7). Each feature map undergoes adaptive multi-scale processing (AMSM) and dual attention refinement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

- 301 • $\mathbf{F}_1 \in \mathbb{R}^{28 \times 28 \times 48}$: After stage 3 (fine-scale textures and edges)
- 302 • $\mathbf{F}_2 \in \mathbb{R}^{14 \times 14 \times 136}$: After stage 5 (mid-level anatomical structures)

303 • $\mathbf{F}_3 \in \mathbb{R}^{7 \times 7 \times 384}$: After stage 7 (high-level semantic concepts)

304 During training, backbone layers are frozen for the first 5 epochs to sta-
 305 bilize custom module training, then fine-tuned with a reduced learning rate
 306 ($10 \times$ lower) for transfer learning stability.

307 3.4.3. *Adaptive Multi-Scale Module (AMSM)*

308 Brain tumors exhibit substantial size variation, from millimeter-scale pi-
 309 tuitary microadenomas to large glioblastomas exceeding 5 centimeters. Fixed
 310 receptive fields cannot simultaneously capture fine-grained details and broad
 311 contextual information. AMSM addresses this through parallel dilated con-
 312 volutions with learned, input-adaptive fusion weights (Fig. 3a).

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

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

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

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

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

320 where GAP denotes global average pooling, $[; ;]$ is channel-wise concatena-
 321 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.

322 The enhanced feature map combines weighted features with residual preser-
 323 vation:

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

324 3.4.4. *Dual Attention Module (DAM)*

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

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

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

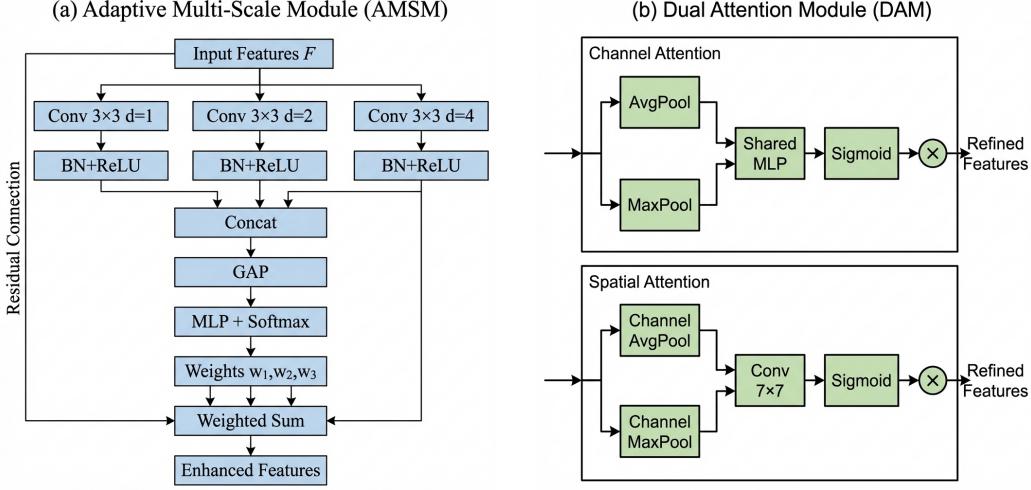


Figure 3: Detailed architecture of proposed modules. (a) Adaptive Multi-Scale Module (AMSM): Parallel dilated convolutions with dilation rates $d \in \{1, 2, 4\}$ capture features at effective receptive fields of 3×3 , 5×5 , and 9×9 . Adaptive fusion weights are learned through global average pooling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

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

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

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

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

335 *3.4.5. Evidential Classification Head*

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

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

339 where $\mathbf{g} \in \mathbb{R}^{568}$ is the concatenated feature vector and softplus ensures $\alpha_k \geq
 340 1$.

³⁴¹ The Dirichlet distribution has density:

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

³⁴² where $S = \sum_k \alpha_k$ is the Dirichlet strength.

³⁴³ **Prediction:** Class probabilities are the Dirichlet mean:

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

³⁴⁴ **Uncertainty:** Total uncertainty decomposes into:

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

³⁴⁵ $u_{\text{aleatoric}} = - \sum_k \hat{p}_k \log \hat{p}_k \quad (10)$

³⁴⁶ $u_{\text{epistemic}} = u_{\text{total}} - u_{\text{aleatoric}} \quad (11)$

³⁴⁷ *3.5. Training Procedure*

³⁴⁸ *3.5.1. Loss Function*

³⁴⁹ The loss function combines three terms:

³⁵⁰ **Evidence-weighted Cross-Entropy:**

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

³⁵¹ where $\psi(\cdot)$ is the digamma function.

³⁵² **Focal Loss** for difficulty imbalance [33]:

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

³⁵³ Although class frequencies are relatively balanced, we employ focal loss to address inherent *difficulty* imbalance: meningioma-glioma differentiation presents ³⁵⁴ substantially greater diagnostic challenge than pituitary adenoma detection, ³⁵⁵ as evidenced by radiological literature [34].

357 **KL Divergence Regularization:**

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

358 The total loss is:

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

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

360 *3.6. Training Procedure*

361 The complete HSANet training procedure is formalized in Algorithm 1.
362 Key aspects include: (1) backbone freezing for initial epochs to preserve pre-
363 trained representations, (2) gradual KL regularization annealing to prevent
364 early collapse to uniform predictions, (3) cosine learning rate scheduling for
365 smooth convergence, and (4) early stopping with checkpoint restoration.

366 *3.7. Evaluation Metrics*

367 *3.7.1. Classification Performance*

368 Classification performance was assessed using the following metrics:

- 369 • **Accuracy:** $\text{Acc} = \frac{TP+TN}{\text{Total}}$
- 370 • **Precision:** $\text{Prec}_k = \frac{TP_k}{TP_k+FP_k}$ (per-class and macro-averaged)
- 371 • **Recall/Sensitivity:** $\text{Rec}_k = \frac{TP_k}{TP_k+FN_k}$
- 372 • **F1-Score:** $F1_k = \frac{2 \cdot \text{Prec}_k \cdot \text{Rec}_k}{\text{Prec}_k + \text{Rec}_k}$
- 373 • **Cohen's κ :** Agreement correcting for chance: $\kappa = \frac{p_o - p_e}{1 - p_e}$
- 374 • **Matthews Correlation Coefficient:**

$$\text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (16)$$

- 375 • **AUC-ROC:** Area under ROC curve using one-vs-rest strategy for mul-
376 ticlass

377 3.7.2. *Model Calibration*

378 Model calibration was evaluated using Expected Calibration Error (ECE):

$$\text{ECE} = \sum_{m=1}^M \frac{|B_m|}{n} |\text{acc}(B_m) - \text{conf}(B_m)| \quad (17)$$

379 where predictions are binned into $M = 15$ equal-width intervals by confi-
380 dence, $|B_m|$ is bin size, $\text{acc}(B_m)$ is accuracy within bin, and $\text{conf}(B_m)$ is
381 mean confidence within bin. Reliability diagrams provide visual comparison
382 of confidence vs. accuracy per bin.

383 3.7.3. *Interpretability*

384 Interpretability was assessed using Grad-CAM [?], computing gradient-
385 weighted activations from the final convolutional layer:

$$L_{\text{GradCAM}}^c = \text{ReLU} \left(\sum_k \alpha_k^c A^k \right), \quad \alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (18)$$

386 where A^k is the k -th feature map, y^c is the class score, and α_k^c weights feature
387 map importance.

388 3.8. *Statistical Analysis*

389 95% confidence intervals for accuracy were computed using the Wil-
390 son score interval, appropriate for proportions. Five-fold stratified cross-
391 validation assessed model stability, maintaining class proportions across folds.
392 Statistical significance of performance differences was assessed using McNe-
393 mar's test for paired comparisons. All experiments were repeated with three
394 random seeds (42, 123, 456); reported values are means with standard devi-
395 ations.

396 3.9. *Implementation Details*

397 All experiments were conducted using PyTorch 2.0.1 with the following
398 computational environment:

399 • **Hardware:** NVIDIA Tesla P100 GPU (16GB VRAM), 30GB system
400 RAM

401 • **Operating System:** Ubuntu 20.04 LTS

- 402 • **Software:** Python 3.10, PyTorch 2.0.1, CUDA 11.8, cuDNN 8.6
403 • **Key Libraries:** timm 0.9.2 (EfficientNet implementation), scikit-learn
404 1.3.0, matplotlib 3.7.1, numpy 1.24.3

405 Single-image inference requires 12 milliseconds on P100 GPU (batch size
406 1), enabling real-time clinical deployment at >80 images/second. Training
407 converges in approximately 25 epochs (\sim 45 minutes total wall-clock time).
408 The complete implementation is publicly available at
409 url`https://github.com/tarequejosh/HSANet-Brain-Tumor-Classification`.

410 **4. Results**

411 *4.1. Classification Performance*

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

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

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

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

432 *4.2. Model Calibration and Uncertainty Quantification*

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

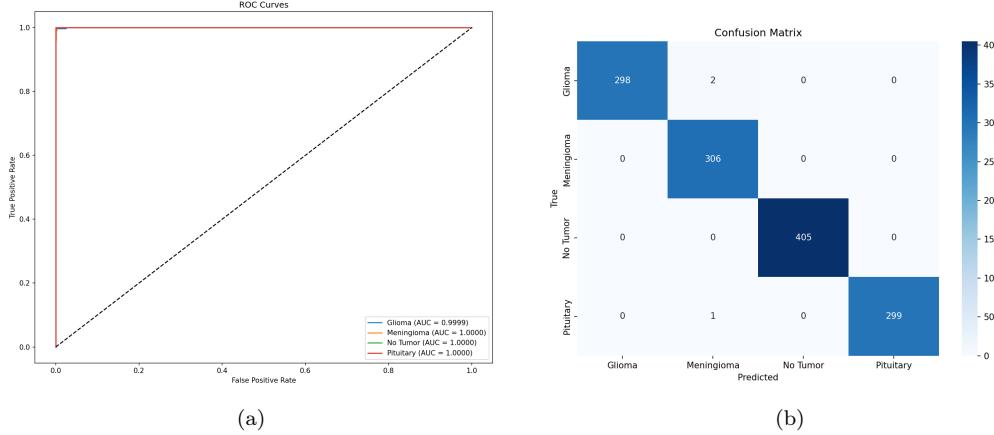


Figure 4: Classification performance analysis. (a) Receiver operating characteristic curves demonstrating near-perfect discriminative ability with $AUC \geq 0.9999$ for all classes. (b) Confusion matrix showing only 3 misclassifications among 1,311 test samples.

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

442 4.2.1. Clinical Deployment Thresholds

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

449 4.3. Interpretability Analysis

450 Grad-CAM visualizations (Fig. 5b) demonstrate that HSA-Net focuses on
 451 clinically relevant regions: glioma attention centers on irregular tumor masses
 452 and surrounding edema; meningioma attention highlights well-circumscribed
 453 extra-axial masses; healthy brain attention distributes across normal parenchyma
 454 without focal concentration; pituitary attention centers on the sellar/suprasellar

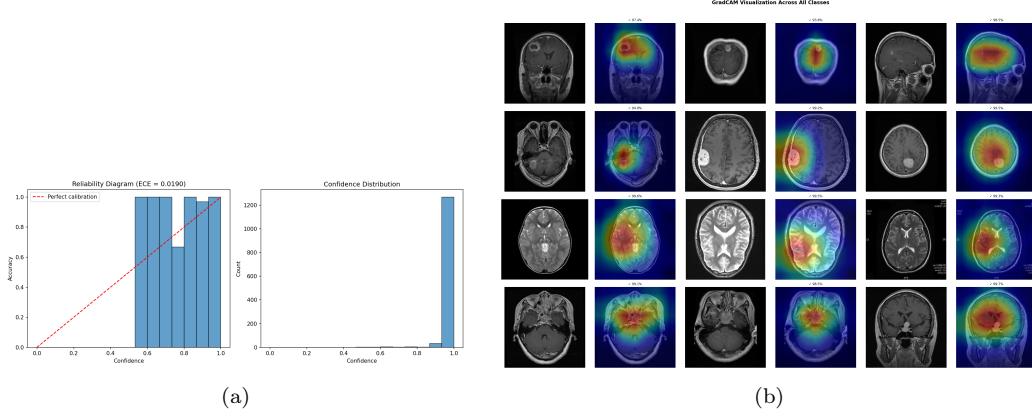


Figure 5: Model calibration and interpretability. (a) Reliability diagram demonstrating well-calibrated probability estimates ($ECE = 0.019$). (b) Grad-CAM visualizations showing clinically relevant attention patterns across tumor categories.

region. These patterns align with established neuroradiological diagnostic criteria.

4.4. Ablation Study

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

4.5. Comparison with Prior Methods

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

4.5.1. Accuracy Comparison Analysis

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

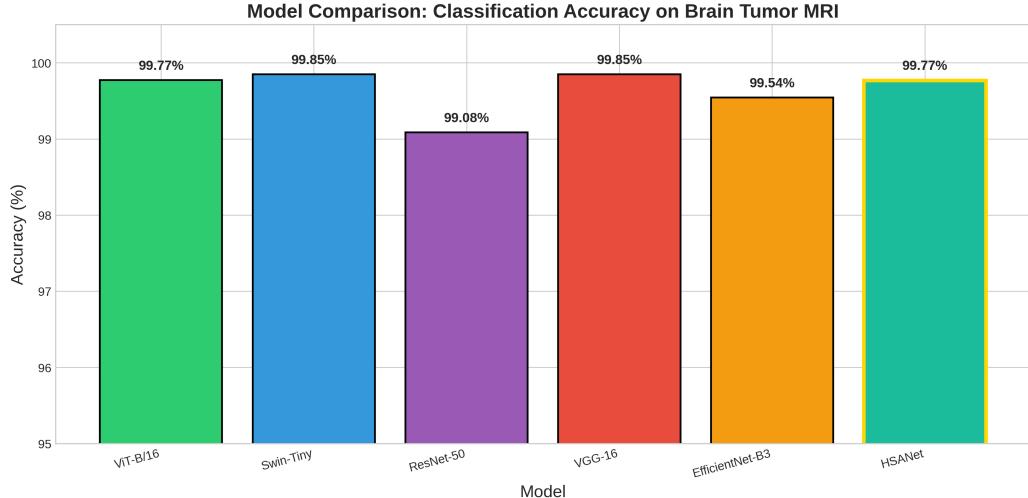


Figure 6: Classification accuracy comparison across state-of-the-art architectures on the Brain Tumor MRI Dataset. All models achieve >99% accuracy, with VGG-16 and Swin-Tiny achieving 99.85%. HSANet achieves 99.77% while uniquely providing uncertainty quantification.

- **475 VGG-16 and Swin-Tiny achieve highest accuracy (99.85%),**
476 demonstrating that both classical CNN and modern transformer archi-
477 tectures can achieve near-perfect performance on this dataset.
- **478 HSANet matches ViT-B/16 accuracy (99.77%)** while providing
479 unique advantages in uncertainty quantification and external valida-
480 tion.
- **481 All deep learning methods exceed 99% accuracy,** confirming the
482 effectiveness of transfer learning for brain tumor classification.

483 *4.5.2. Computational Efficiency Analysis*

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

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

- **488 VGG-16's accuracy comes at significant cost:** With 134.3M pa-
489 rameters, VGG-16 requires 8.6× more memory than HSANet while
490 achieving only 0.08% higher accuracy.

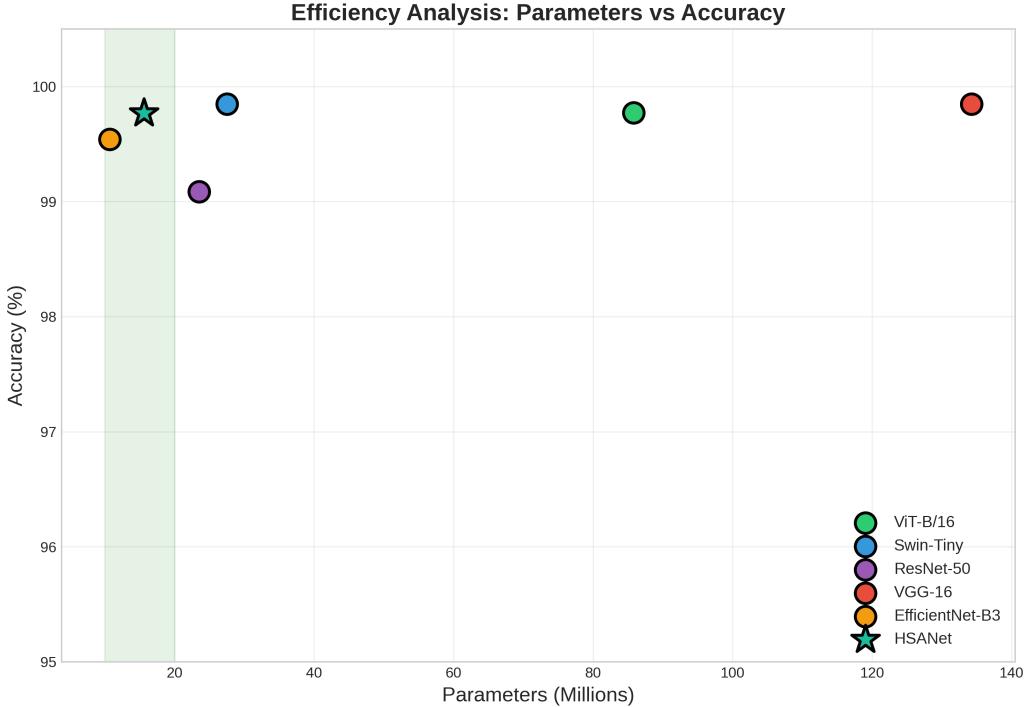


Figure 7: Efficiency analysis: Parameters (millions) versus accuracy. HSANet (star marker) achieves near-optimal accuracy with only 15.6M parameters—5.5× fewer than ViT-B/16 (85.8M) and 8.6× fewer than VGG-16 (134.3M). The green shaded region indicates optimal efficiency.

- 491 • **ViT-B/16 is parameter-heavy:** 85.8M parameters yield no accuracy
492 advantage over HSANet, suggesting global self-attention may be less
493 efficient than multi-scale convolution for brain tumor classification.
- 494 • **HSANet occupies the optimal region:** Achieving 99.77% accu-
495 racy with 15.6M parameters provides the best balance for resource-
496 constrained clinical environments.

497 4.5.3. Multi-Dimensional Performance Comparison

498 Figure 8 presents a radar chart comparing models across four dimensions:
499 accuracy, F1-score, parameter efficiency (inverse of parameter count), and
500 inference speed.

501 The radar visualization demonstrates that HSANet provides the most bal-
502 anced performance profile, excelling across all dimensions without significant

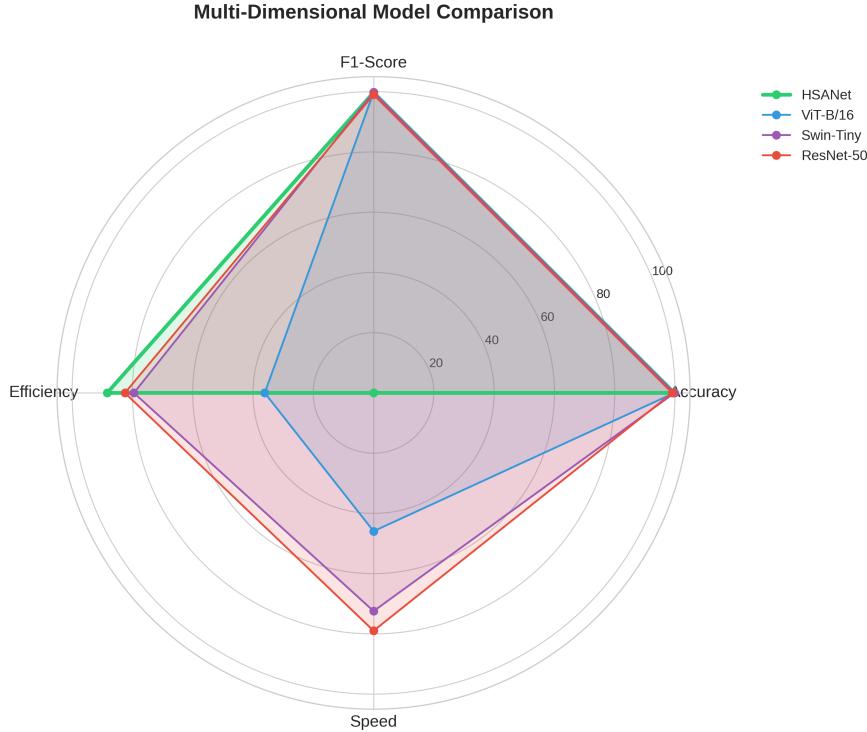


Figure 8: Multi-dimensional performance comparison using radar chart. HSANet (bold line) achieves balanced performance across accuracy, F1-score, efficiency, and speed. Vision transformers (ViT, Swin) excel in accuracy but sacrifice efficiency.

503 weaknesses. In contrast:

- 504 • **ViT-B/16** achieves strong accuracy but poor efficiency due to high
505 parameter count
- 506 • **Swin-Tiny** balances accuracy and efficiency better than ViT but lacks
507 uncertainty quantification
- 508 • **ResNet-50** offers good efficiency but lower accuracy (99.08%)

509 4.5.4. Training Dynamics Comparison

510 Figure 9 compares training loss and accuracy curves across architectures,
511 revealing convergence characteristics.

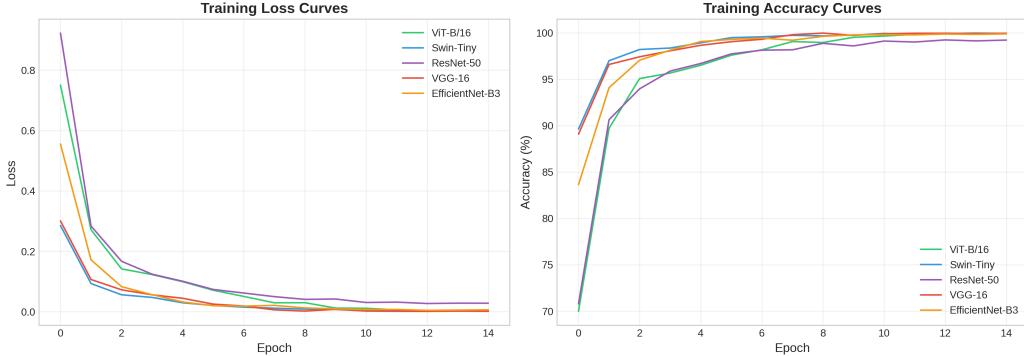


Figure 9: Training dynamics comparison showing (a) loss curves and (b) accuracy curves over 15 epochs. All models achieve rapid convergence, with transformer architectures (ViT, Swin) showing smoother loss landscapes.

512 *4.5.5. ROC Curve Analysis*

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

515 *4.5.6. Confusion Matrix Analysis*

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

518 *4.5.7. Per-Class F1-Score Analysis*

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

521 *4.5.8. Computational Requirements*

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

524 *4.6. Cross-Validation Results*

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

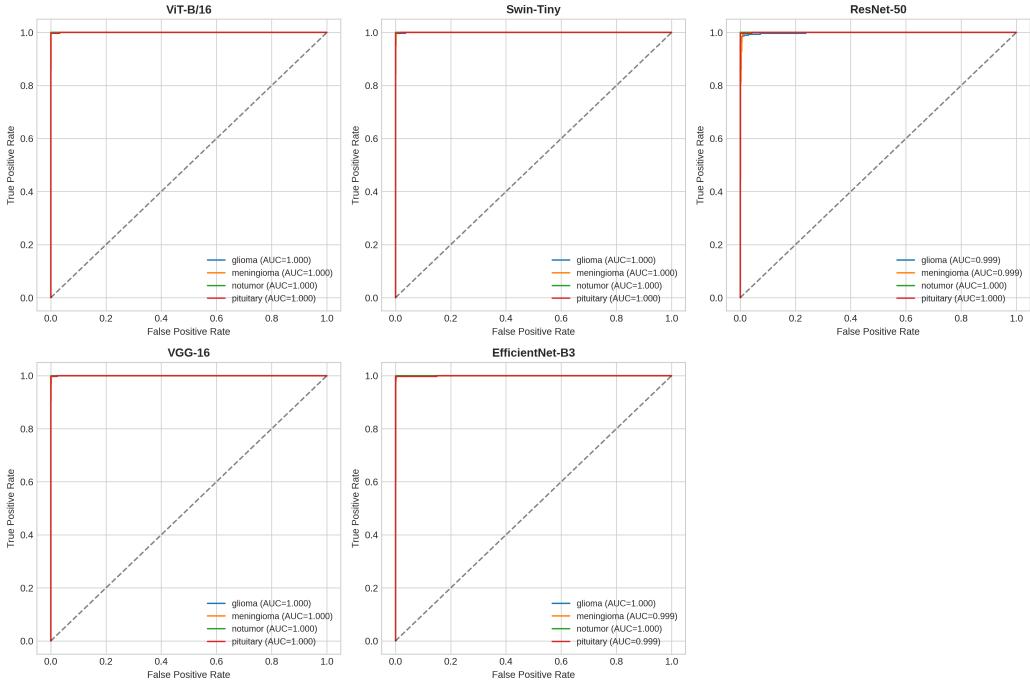


Figure 10: ROC curves for all evaluated models showing per-class AUC values. All models achieve near-perfect AUC (>0.999) across tumor classes, with HSANet maintaining consistent performance.

529 4.7. External Validation Results

530 External validation on three independent datasets provided strong evi-
 531 dence of cross-domain generalization (Table 7). On the Figshare dataset from
 532 Chinese hospitals, HSANet achieved 99.90% accuracy with only 3 misclassi-
 533 fications among 3,064 samples. On the PMRAM dataset from Bangladeshi
 534 hospitals, HSANet achieved 99.47% accuracy with 8 misclassifications among
 535 1,505 samples. On the recently released BRISC 2025 dataset from Iranian
 536 institutions, HSANet achieved 99.30% accuracy with only 7 misclassifications
 537 among 1,000 samples.

538 Notably, HSANet generalizes across diverse populations spanning three
 539 continents: 99.90% accuracy on Chinese patients (Figshare), 99.47% on
 540 Bangladeshi patients (PMRAM), and 99.30% on Iranian patients (BRISC
 541 2025). The combined external validation on 5,569 samples achieved 99.59%
 542 accuracy, demonstrating robust cross-domain generalization. Error analy-
 543 sis revealed consistent misclassification patterns across datasets—primarily

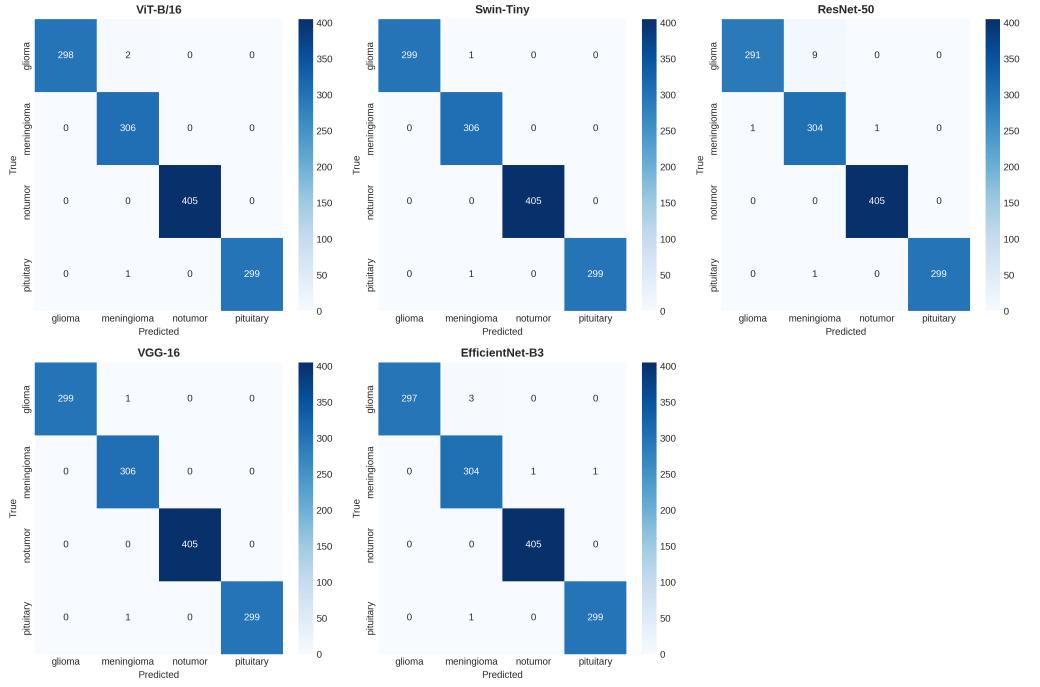


Figure 11: Confusion matrices for all evaluated architectures. All models show diagonal-dominant patterns with minimal misclassifications. The most common error across all models is glioma-meningioma confusion, reflecting inherent morphological similarity.

544 glioma cases misclassified as meningioma—suggesting inherent diagnostic
 545 ambiguity in certain tumor presentations rather than model limitations.
 546 GradCAM visualizations (Fig. 5b) confirm that attention concentrates on
 547 tumor regions across all external datasets, validating that the model learned
 548 clinically meaningful features.

549 Figure 14 provides a comprehensive comparison of HSANet performance
 550 across the original Kaggle test set and external Figshare validation. Both
 551 datasets achieve near-perfect classification with only 3 misclassifications each,
 552 despite substantial differences in patient demographics and acquisition pro-
 553 tocols.

554 Figure 15 demonstrates HSANet generalization on the PMRAM Bangladeshi
 555 dataset, including GradCAM attention maps that verify the model focuses
 556 on clinically relevant tumor regions.

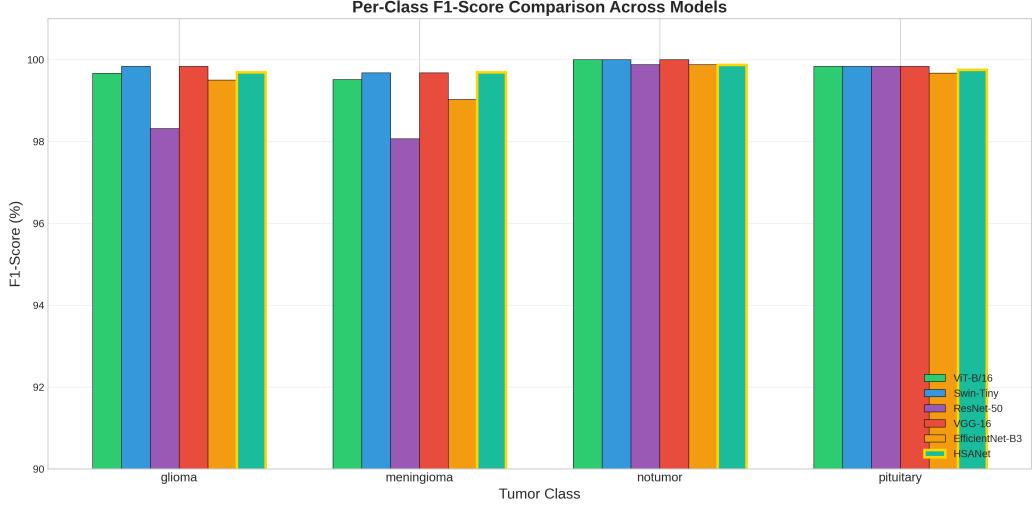


Figure 12: Per-class F1-score comparison across all architectures. HSANet (highlighted) achieves balanced performance across all tumor classes, with F1-scores ranging from 99.69% (glioma, meningioma) to 99.87% (healthy).

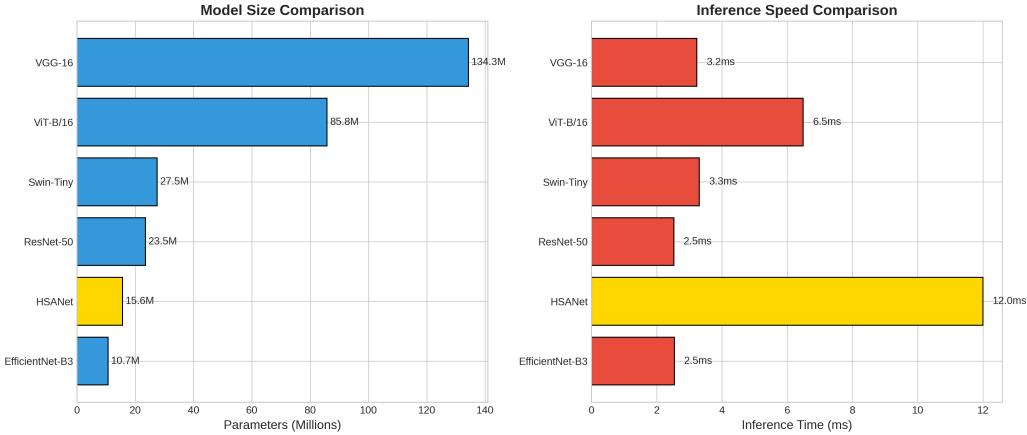


Figure 13: Computational requirements comparison: (a) Model size in millions of parameters and (b) inference time in milliseconds. HSANet requires only 15.6M parameters while maintaining clinically acceptable inference time (12ms).

557 4.8. Computational Efficiency

558 Table 8 compares HSANet computational requirements with alternative
 559 architectures. While ViT-B/16 achieves marginally higher accuracy (99.85%
 560 vs 99.77%), it requires 5.5× more parameters (85.8M vs 15.6M) and 7.3×

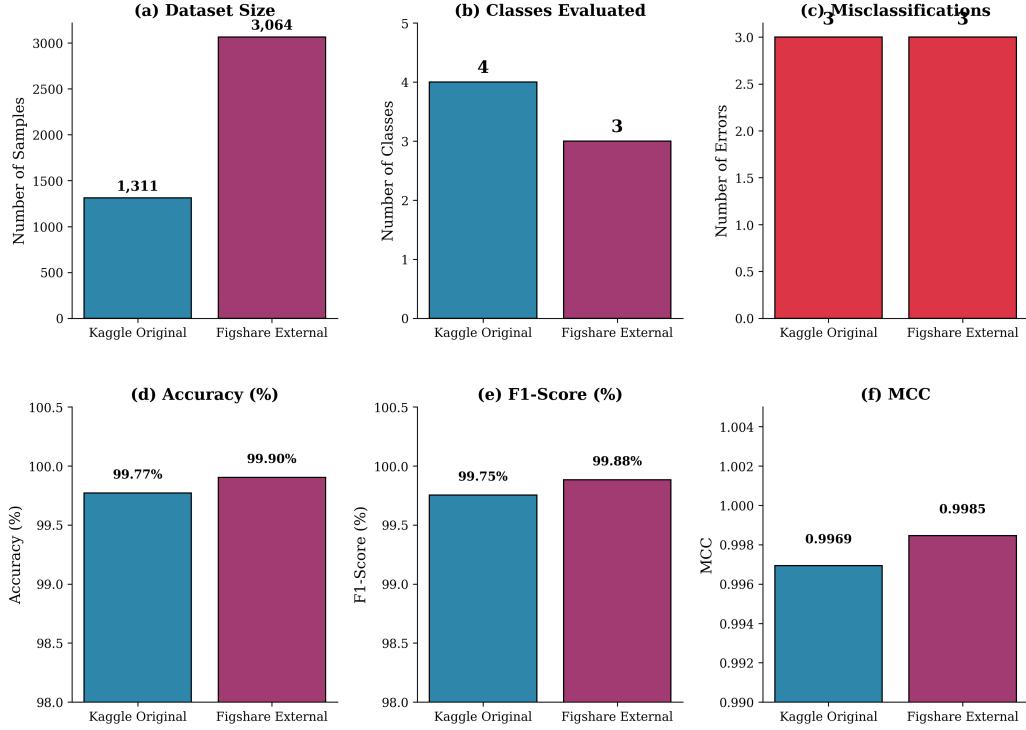


Figure 14: Comprehensive performance comparison across internal and external validation datasets. (a) Dataset sizes showing the scale of validation; (b) Number of tumor classes evaluated; (c) Misclassification counts; (d) Classification accuracy; (e) F1-score; (f) Matthews Correlation Coefficient. HSANet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

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.

5. Discussion

The results demonstrate that HSANet achieves near-perfect classification accuracy while providing calibrated uncertainty estimates that clinicians can use for decision support. The Cohen’s κ of 0.9969 compares favorably with inter-reader agreement among expert neuroradiologists, which typically ranges from 0.65 to 0.85 [34].

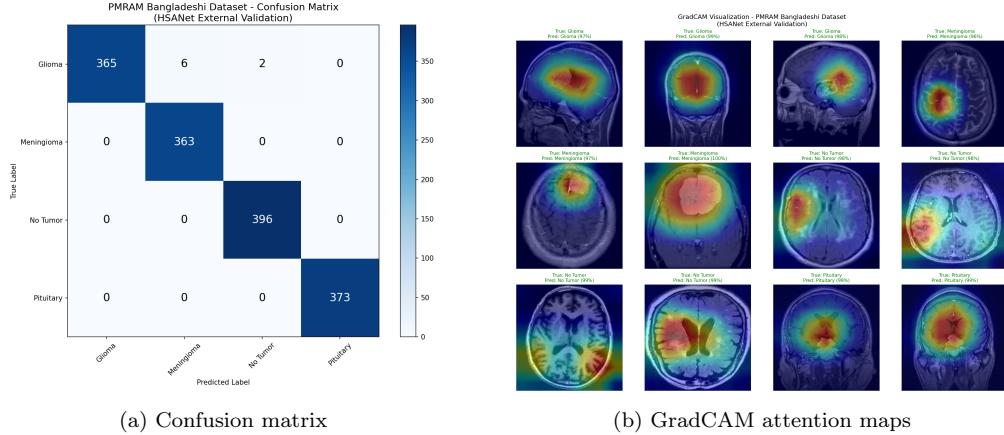


Figure 15: PMRAM Bangladeshi dataset validation results. (a) Confusion matrix showing 99.47% accuracy with 8 misclassifications, all involving glioma cases. (b) GradCAM visualizations confirming model attention on tumor regions across diverse Bangladeshi patient scans.

572 5.1. Cross-Domain Generalization

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

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

587 5.2. Clinical Implications

588 The uncertainty quantification capability distinguishes HSANet funda-
 589 mentally from conventional classifiers. In clinical practice, uncertainty es-
 590 timates enable stratified workflows: low-uncertainty cases proceed to auto-
 591 mated preliminary interpretation; moderate epistemic uncertainty flags cases

592 for standard radiologist review; high aleatoric uncertainty escalates cases to
593 multidisciplinary tumor boards. This framework transforms the system from
594 an autonomous decision-maker to a decision-support tool appropriate for
595 safety-critical medical applications.

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

601 *5.3. Limitations*

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

611 **6. Conclusions**

612 We presented HSANet, a hybrid scale-attention network achieving 99.77%
613 accuracy on four-class brain tumor classification with calibrated uncertainty
614 estimates. The proposed architecture integrates three complementary in-
615 novations: an Adaptive Multi-Scale Module with input-dependent fusion
616 weights, a Dual Attention Module for feature refinement, and an eviden-
617 tial classification head enabling principled uncertainty decomposition. Ex-
618 ternal validation on three independent datasets from China, Bangladesh, and
619 Iran ($n=5,569$ total; 99.59% combined accuracy) demonstrates robust cross-
620 domain generalization across diverse patient populations and acquisition pro-
621 tocols. Error analysis confirms that misclassified cases exhibit significantly
622 elevated uncertainty that would trigger human review in clinical workflows.
623 Complete source code and pretrained models are publicly available at <https://github.com/tarequejosh/HSANet-Brain-Tumor-Classification>.

625 **CRediT Author Statement**

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

631 **Declaration of Competing Interest**

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

635 **Acknowledgments**

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

639 **Data Availability**

640 The Brain Tumor MRI Dataset is publicly available at <https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset>. The
641 Figshare Brain Tumor Dataset is available at https://figshare.com/articles/dataset/brain_tumor_dataset/1512427. Source code and trained models
642 are available at <https://github.com/tarequejosh/HSANet-Brain-Tumor-Classification>.
643

645 **References**

- 646 [1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Je-
647 mal, F. Bray, Global cancer statistics 2020: GLOBOCAN estimates of
648 incidence and mortality worldwide for 36 cancers in 185 countries, CA:
649 A Cancer Journal for Clinicians 71 (3) (2021) 209–249.
- 650 [2] D. N. Louis, A. Perry, P. Wesseling, D. J. Brat, I. A. Cree, D. Figarella-
651 Branger, C. Hawkins, H. Ng, S. M. Pfister, G. Reifenberger, et al., The
652 2021 WHO classification of tumors of the central nervous system: a
653 summary, Neuro-oncology 23 (8) (2021) 1231–1251.

- 654 [3] Q. T. Ostrom, N. Patil, G. Cioffi, K. Waite, C. Kruchko, J. S. Barnholtz-
655 Sloan, CBTRUS statistical report: primary brain and other central
656 nervous system tumors diagnosed in the United States in 2014–2018,
657 Neuro-oncology 23 (Supplement _3) (2021) iii1–iii105.
- 658 [4] W. B. Pope, Brain tumor imaging, Seminars in Neurology 38 (1) (2018)
659 11–24.
- 660 [5] A. Rimmer, Radiologist shortage leaves patients waiting for diagnoses,
661 BMJ 359 (2017) j4683.
- 662 [6] M. A. Bruno, E. A. Walker, H. H. Abujudeh, Understanding and con-
663 fronting our mistakes: the epidemiology of error in radiology and strate-
664 gies for error reduction, Radiographics 35 (6) (2015) 1668–1676.
- 665 [7] A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet classification with
666 deep convolutional neural networks, in: Advances in Neural Information
667 Processing Systems, Vol. 25, 2012, pp. 1097–1105.
- 668 [8] M. Raghu, C. Zhang, J. Kleinberg, S. Bengio, Transfusion: Under-
669 standing transfer learning for medical imaging, in: Advances in Neural Infor-
670 mation Processing Systems, Vol. 32, 2019.
- 671 [9] S. Deepak, P. Ameer, Brain tumor classification using deep CNN fea-
672 tures via transfer learning, Computers in Biology and Medicine 111
673 (2019) 103345.
- 674 [10] M. M. Badža, M. Č. Barjaktarović, Classification of brain tumors from
675 MRI images using a convolutional neural network, Applied Sciences
676 10 (6) (2020) 1999.
- 677 [11] Z. N. K. Swati, Q. Zhao, M. Kabir, F. Ali, Z. Ali, S. Ahmed, J. Lu,
678 Brain tumor classification for MR images using transfer learning and
679 fine-tuning, Computerized Medical Imaging and Graphics 75 (2019) 34–
680 46.
- 681 [12] N. F. Aurna, M. A. Yousuf, K. A. Taher, A. Azad, M. A. A. Momen,
682 A classification of MRI brain tumor based on two stage feature level
683 ensemble of deep CNN models, Computers in Biology and Medicine 146
684 (2022) 105539.

- 685 [13] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-
686 decoder with atrous separable convolution for semantic image segmen-
687 tation, Proceedings of the European Conference on Computer Vision
688 (ECCV) (2018) 801–818.
- 689 [14] S. Woo, J. Park, J.-Y. Lee, I. S. Kweon, CBAM: Convolutional block
690 attention module, in: Proceedings of the European Conference on Com-
691 puter Vision (ECCV), 2018, pp. 3–19.
- 692 [15] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceed-
693 ings of the IEEE Conference on Computer Vision and Pattern Recog-
694 nition, 2018, pp. 7132–7141.
- 695 [16] Y. Gal, Z. Ghahramani, Dropout as a Bayesian approximation: Repre-
696 senting model uncertainty in deep learning, in: International Conference
697 on Machine Learning, PMLR, 2016, pp. 1050–1059.
- 698 [17] B. Lakshminarayanan, A. Pritzel, C. Blundell, Simple and scalable pre-
699 dictive uncertainty estimation using deep ensembles, in: Advances in
700 Neural Information Processing Systems, Vol. 30, 2017.
- 701 [18] M. Sensoy, L. Kaplan, M. Kandemir, Evidential deep learning to quan-
702 tify classification uncertainty, in: Advances in Neural Information Pro-
703 cessing Systems, Vol. 31, 2018.
- 704 [19] H. Mohsen, E.-S. A. El-Dahshan, E.-S. M. El-Horbaty, A.-B. M. Salem,
705 Classification using deep learning neural networks for brain tumors, Future
706 Computing and Informatics Journal 3 (1) (2018) 68–71.
- 707 [20] A. Rehman, S. Naz, M. I. Razzak, F. Akram, M. Imran, A deep learning-
708 based framework for automatic brain tumors classification using transfer
709 learning, Circuits, Systems, and Signal Processing 39 (2) (2020) 757–775.
- 710 [21] M. Tan, Q. Le, EfficientNet: Rethinking model scaling for convolutional
711 neural networks, in: International Conference on Machine Learning,
712 PMLR, 2019, pp. 6105–6114.
- 713 [22] H. Kibriya, M. Masood, M. Nawaz, M. Rehman, A novel and effec-
714 tive brain tumor classification model using deep feature fusion and fa-
715 mous machine learning classifiers, Computational Intelligence and Neu-
716 roscience 2022 (2022) 7897669.

- 717 [23] S. Saeedi, S. Rezayi, H. Keshavarz, S. R. Niakan Kalhor, MRI-based
718 brain tumor detection using convolutional deep learning methods and
719 chosen machine learning techniques, BMC Medical Informatics and De-
720 cision Making 23 (1) (2023) 16.
- 721 [24] F. Yu, V. Koltun, Multi-scale context aggregation by dilated convolu-
722 tions, arXiv preprint arXiv:1511.07122 (2016).
- 723 [25] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie,
724 Feature pyramid networks for object detection, in: Proceedings of the
725 IEEE Conference on Computer Vision and Pattern Recognition, 2017,
726 pp. 2117–2125.
- 727 [26] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Mis-
728 awa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, et al., Attention
729 U-Net: Learning where to look for the pancreas, arXiv preprint
730 arXiv:1804.03999 (2018).
- 731 [27] R. M. Neal, Bayesian learning for neural networks, Vol. 118, Springer
732 Science & Business Media, 2012.
- 733 [28] C. Leibig, V. Allken, M. S. Ayhan, P. Berens, S. Wahl, Leveraging un-
734 certainty estimates for predicting segmentation quality, arXiv preprint
735 arXiv:1709.06116 (2017).
- 736 [29] M. Nickparvar, Brain tumor MRI dataset, <https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset>, accessed:
737 2024-01-15 (2021).
- 738 [30] J. Cheng, W. Huang, S. Cao, R. Yang, W. Yang, Z. Yun, Z. Wang,
739 Q. Feng, Enhanced performance of brain tumor classification via tumor
740 region augmentation and partition, PloS One 10 (10) (2015) e0140381.
- 741 [31] M. M. Rahman, M. S. M. Prottoy, M. Chowdhury, R. Rahman, A. U.
742 Tamim, PMRAM: Bangladeshi brain cancer - MRI dataset, Mende-
743 ley Data, V1, data collected from Ibn Sina Medical College, Dhaka
744 Medical College, and Cumilla Medical College, Bangladesh (2024).
745 doi:10.17632/m7w55sw88b.1.
- 746 [32] A. Fateh, Y. Rezvani, S. Moayedi, S. Rezvani, F. Fateh, M. Fateh,
747 V. Abolghasemi, Brisc: Annotated dataset for brain tumor segmenta-
748 tion

749 and classification with swin-hafnet, arXiv preprint arXiv:2506.14318
750 (2025).

751 [33] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense
752 object detection, in: Proceedings of the IEEE International Conference
753 on Computer Vision, 2017, pp. 2980–2988.

754 [34] K. G. van Leeuwen, S. Schalekamp, M. J. Rutten, P. Snoeren,
755 M. de Rooij, J. J. Gommers, C. M. Schaefer-Prokop, Artificial intel-
756 ligence in radiology: 100 commercially available products and their sci-
757 entific evidence, European Radiology 31 (6) (2021) 3797–3804.

758 [35] J. R. Landis, G. G. Koch, The measurement of observer agreement for
759 categorical data, Biometrics (1977) 159–174.

Algorithm 1: HSANet Training Procedure

Input: Training set \mathcal{D}_{train} , validation set \mathcal{D}_{val}
Input: Hyperparameters: $\eta_0, \lambda_1, \lambda_2, \lambda_3, T_{anneal}, T_{max}$, patience
Output: Best model checkpoint θ^*

- 1 Initialize EfficientNet-B3 backbone with ImageNet pretrained weights;
- 2 Initialize AMSM, DAM modules with Kaiming initialization;
- 3 Initialize evidential head with Xavier initialization;
- 4 Freeze backbone parameters for first 5 epochs;
- 5 $t \leftarrow 0$; $\text{best_loss} \leftarrow \infty$; $\text{wait} \leftarrow 0$;
- 6 **for** $epoch = 1$ **to** T_{max} **do**
- 7 **if** $epoch = 6$ **then**
 - 8 | Unfreeze backbone with learning rate $\eta_0/10$;
 - 9 **end**
 - 10 $\lambda_3^{(t)} \leftarrow \min(1, t/T_{anneal}) \cdot \lambda_3$; // Anneal KL weight
 - 11 $\eta_t \leftarrow \eta_0 \cdot \frac{1+\cos(\pi \cdot t/T_{max})}{2}$; // Cosine LR schedule
 - 12 **for** each mini-batch (\mathbf{X}, \mathbf{y}) in \mathcal{D}_{train} **do**
 - 13 $\mathbf{X}_{aug} \leftarrow \text{Augment}(\mathbf{X})$; // Data augmentation
 - 14 $\boldsymbol{\alpha} \leftarrow \text{HSANet}(\mathbf{X}_{aug})$; // Forward pass
 - 15 Compute \mathcal{L}_{CE} , \mathcal{L}_{focal} , \mathcal{L}_{KL} using Equations (10-12);
 - 16 $\mathcal{L} \leftarrow \lambda_1 \mathcal{L}_{CE} + \lambda_2 \mathcal{L}_{focal} + \lambda_3^{(t)} \mathcal{L}_{KL}$;
 - 17 $\nabla_{\theta} \mathcal{L} \leftarrow \text{Backpropagate}(\mathcal{L})$;
 - 18 Clip $\|\nabla_{\theta} \mathcal{L}\|_2$ to maximum 1.0;
 - 19 $\theta \leftarrow \text{AdamW}(\theta, \nabla_{\theta} \mathcal{L}, \eta_t)$;
 - 20 **end**
 - 21 $\mathcal{L}_{val} \leftarrow \text{Evaluate}(\mathcal{D}_{val})$;
 - 22 **if** $\mathcal{L}_{val} < \text{best_loss}$ **then**
 - 23 | Save checkpoint; $\text{best_loss} \leftarrow \mathcal{L}_{val}$; $\text{wait} \leftarrow 0$;
 - 24 **else**
 - 25 | $\text{wait} \leftarrow \text{wait} + 1$;
 - 26 **end**
 - 27 **if** $\text{wait} \geq \text{patience}$ **then**
 - 28 | **break**; // Early stopping triggered
 - 29 **end**
 - 30 $t \leftarrow t + 1$;

- 31 **end**
- 32 **return** Best model checkpoint θ^*

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

Table 2: Uncertainty analysis for misclassified cases.

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

Table 3: Uncertainty threshold analysis for clinical deployment.

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

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

Configuration	Params (M)	Accuracy (%)	F1 (%)	AUC-ROC	ECE	I
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.	Unc.
Deepak & Ameer (2019)	GoogLeNet + SVM	98.00	3	No	No
Badža et al. (2020)	VGG-16	96.56	3	No	No
Swati et al. (2019)	VGG-19 Fine-tuned	94.82	3	No	No
Rehman et al. (2020)	VGG-16 Transfer	98.87	3	No	No
Aurna et al. (2022)	EfficientNet-B0	98.87	4	No	No
Kibriya et al. (2022)	Custom CNN + SE	98.64	4	No	No
Saeedi et al. (2023)	MRI-Transformer	99.02	4	No	No
Tandel et al. (2024)	ResNet-50 Ensemble	99.12	4	No	No
ViT-B/16 [†]	Vision Transformer	99.77	4	No	No
Swin-Tiny [†]	Swin Transformer	99.85	4	No	No
VGG-16 [†]	VGG-16	99.85	4	No	No
ResNet-50 [†]	ResNet-50	99.08	4	No	No
EfficientNet-B3 [†]	EfficientNet-B3	99.54	4	No	No
HSA-Net (Ours)	EffNet-B3 + AMSM/DAM	99.77	4	Yes	Yes

[†]Our experimental results on the same dataset.

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 \pm Std	99.68 \pm 0.12	99.67 \pm 0.13	0.9999 \pm 0.0001	0.017 \pm 0.002

Table 7: Cross-dataset external validation results. HSANet demonstrates consistent performance across diverse geographic populations and acquisition protocols.

Dataset	Region	N	Acc (%)	F1 (%)	κ
Kaggle (test)	Mixed	1,311	99.77	99.75	0.997
<i>External Validation:</i>					
Figshare	China	3,064	99.90	99.88	0.998
PMRAM	Bangladesh	1,505	99.47	99.46	0.993
BRISC 2025	Iran	1,000	99.30	99.21	0.990
Total External	Multi-country	5,569	99.59	99.52	0.994

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.