Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 2, April 2026, Page. 1515-1534
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5494

Bidirectional Cross-Attention and Uncertainty-Aware Ensemble for High-Precision Brain Tumor Classification on MRI Images

Gunawan Cholis Saputra[*1], Muhammad Yuwanandra Risdyaksa[2]

[1,2]Information System, Universitas Dian Nuswantoro, Indonesia

Email: [1]112202307010@mhs.dinus.ac.id

Received : Nov 28, 2025; Revised : Dec 3, 2025; Accepted : Dec 3, 2025; Published : Apr 15, 2026

Abstract

Accurate brain tumor diagnosis via Magnetic Resonance Imaging (MRI) is vital for effective neuro-oncological treatment. Although CNNs are widely regarded as the benchmark for local texture extraction, they frequently exhibit limitations in modeling long-distance global dependencies effectively. In contrast, Vision Transformers (ViTs), particularly the Swin variant, demonstrate superior capability in capturing global semantic context yet often fail to preserve the fine local granularity needed to delineate tumor boundaries significantly. To bridge this gap, we propose Bi-CA-UAE, a hybrid framework integrating Swin Transformer and EfficientNet-V2 through a novel Bidirectional Cross-Attention mechanism. Unlike static ensembles, our method enables dynamic information exchange between global and local feature maps before classification. Furthermore, we introduce an Uncertainty-Aware Gating Network to adaptively weigh each branch based on prediction confidence, reducing false positives in ambiguous cases. Validated on a multi-class MRI dataset of 7,023 images, the model achieved 99.85% accuracy and an Expected Calibration Error (ECE) of 0.02, matching the strongest baseline (Swin Transformer) while demonstrating superior training stability and calibration. Unlike naive concatenation ensembles that suffered from overfitting and performance degradation in later training stages, Bi-CA-UAE maintained robust peak performance. Additionally, the model attained perfect recall (1.00) for Glioma and a micro-average AUC of 1.00. t-SNE visualizations and reliability diagrams confirm the model's ability to learn highly discriminative and well-calibrated features, positioning it as a trustworthy clinical decision support system.

Keywords : *Bidirectional Cross-Attention, Brain Tumor Classification, Deep Learning, EfficientNet, MRI Analysis, Swin Transformer, Uncertainty Quantification.*

This work is an open access article and licensed under a Creative Commons Attribution-Non Commercial 4.0 International License

1. INTRODUCTION

Globally, brain tumors rank as one of the most aggressive and deadly malignancies, accounting for a substantial portion of cancer-related morbidity and mortality. Although primary brain tumors represent approximately 2% of all cancer cases according to the World Health Organization, they are responsible for a disproportionately high rate of cancer-related deaths. Therefore, accurate and timely diagnosis is critical to enhance survival rates and refine therapeutic strategies, including surgical intervention, chemotherapy, and radiotherapy. MRI is currently established as the benchmark for non-invasive brain tumor assessment, providing exceptional multi-planar visualization and high contrast for soft tissues [1]. However, manual interpretation of MRI scans is labor-intensive, subjective, and prone to inter-observer variability, especially when differentiating tumor types with similar morphological features like Glioma, Meningioma, and Pituitary tumors [2].

In the last decade, Deep Learning (DL) has revolutionized medical image analysis [3]. Convolutional Neural Networks (CNNs), including architectures like ResNet [4], DenseNet, VGG, and EfficientNet [5], have shown remarkable success in tumor detection and classification [6]. Despite their efficacy, CNNs are inherently limited by local receptive fields, which restricts their ability to model

Vol. 7, No. 2, April 2026, Page. 1515-1534
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5494

long-range spatial dependencies and the global context of brain structures. This limitation is critical when analyzing heterogeneous tumors with complex intensity patterns [7].

To address the locality constraints of CNNs, Vision Transformers (ViTs) have emerged as a powerful alternative [8]. By utilizing self-attention mechanisms, Transformers capture global dependencies across the entire image [9]. Specifically, the Swin Transformer [10] improves computational efficiency through hierarchical feature maps and shifted windows. However, Transformers often lack the strong inductive biases of CNNs, leading to suboptimal performance in detecting subtle local textures, such as tumor boundaries, unless trained on massive datasets [11].

Recent trends have shifted toward hybrid architectures combining CNNs and Transformers [12], [13]. Nevertheless, most existing methods rely on simple concatenation or static weighting for feature fusion. Our empirical analysis reveals that such naive fusion strategies are susceptible to overfitting, leading to performance degradation in later training stages. Furthermore, these approaches fail to quantify model uncertainty [14], resulting in overconfident predictions even on ambiguous data a critical risk in medical diagnosis. Furthermore, conventional ensembles treat all predictions equally, potentially degrading performance if one branch yields unreliable outputs.

To bridge these critical gaps, this study proposes Bi-CA-UAE (Bidirectional Cross-Attention Uncertainty-Aware Ensemble), a novel hybrid deep learning framework tailored for brain tumor classification using MRI. The primary innovation of this research lies in the integration of a Bidirectional Cross-Attention mechanism that facilitates dynamic information exchange between the global context modeled by Swin Transformer and the local features extracted by EfficientNet-V2. By leveraging query-key-value interactions, this mechanism allows each branch to mutually refine feature representations, yielding a richer and more discriminative feature space. Furthermore, we incorporate an Uncertainty-Aware Gating Network designed to adaptively weight the contributions of each branch based on prediction confidence. This gating strategy enhances the model's robustness, particularly in ambiguous cases, by prioritizing the most reliable feature set for the final classification. Comprehensive experiments conducted on a large-scale Brain Tumor MRI dataset demonstrate that Bi-CA-UAE achieves a state-of-the-art accuracy of 99.85%. While matching the peak performance of strong standalone baselines (e.g., Swin Transformer), our framework demonstrates superior training stability and model calibration. Unlike naive ensemble methods that are prone to overfitting and overconfidence, Bi-CA-UAE provides reliable and explainable predictions, which are crucial for high-stakes clinical decision-making.

The main contributions of this study are summarized as follows:

- We propose Bi-CA-UAE, a novel hybrid framework that employs Bidirectional Cross-Attention to dynamically fuse global and local features.
- We introduce an Uncertainty-Aware Gating Network that significantly reduces false positives in ambiguous cases (e.g., Meningioma vs. Pituitary).
- We demonstrate that our method achieves Perfect Recall (1.00) for Glioma, ensuring zero false negatives for this aggressive tumor type.

The organization of this paper is structured as follows: Section 2 provides a review of related work concerning brain tumor classification and hybrid ddateep learning architectures. Section 3 details the methodology of the proposed Bi-CA-UAE framework, specifically elaborating on the Bidirectional Cross-Attention mechanism and the Uncertainty-Aware Gating Network. Section 4 outlines the experimental setup, including dataset description and evaluation metrics. Section 5 presents the experimental results followed by an in-depth discussion, and finally, Section 6 concludes the study and suggests directions for future research. A summary of related works and identified research gaps is presented in Table 1.

Vol. 7, No. 2, April 2026, Page. 1515-1534
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5494

Table 1. Summary of Related Works and Identified Research Gaps

Study / Reference	Method	Limitations (The Research Gap)
Sadad *et al.* [18]	ResNet50 (Transfer Learning)	Limited to local receptive fields; struggles to capture long-range global dependencies in MRI.
Gómez-Guzmán *et al.* [46]	DeiT (Pure Transformer)	Lacks strong inductive bias; requires massive datasets and is prone to overfitting on limited medical data.
Irfani *et al.* [20]	EfficientNet-B3	High accuracy but lacks uncertainty quantification; prone to overconfidence in ambiguous cases.
Tejashwini *et al.* [45]	Ensemble (DenseNet)	Uses static ensemble weights; ignores the varying confidence levels of individual branches.
Proposed Bi-CA-UAE	Hybrid Swin-EffNet + Gating	Integrates dynamic cross-attention & uncertainty gating to address stability, ambiguity, and overfitting.

2. RELATED WORK

This section reviews the existing literature relevant to brain tumor classification using deep learning techniques. We begin by analyzing the dominance of Convolutional Neural Networks (CNNs) in medical imaging, followed by an examination of the shift towards Vision Transformers (ViTs) and hybrid architectures. This review aims to highlight the progression of state-of-the-art methods and identify the specific research gaps that this study aims to address.

2.1. Deep Learning for Brain Tumor Classification

Recently, the utilization of deep learning in medical imaging has escalated dramatically [15], where Convolutional Neural Networks (CNNs) have emerged as the leading framework for image analysis for image classification due to their capability to extract hierarchical features from raw pixel data automatically [16]. In the domain of brain tumor classification, transfer learning using pre-trained architectures such as VGG, ResNet, and DenseNet has been extensively adopted [17] [18]. For instance, recent studies utilizing ResNet50 achieved an accuracy of 95.04% [19], while approaches leveraging GoogleNet with fine-tuning strategies reported improved performance at 91.5% [20]. Recent advancements involving EfficientNet, coupled with extensive data augmentation and hyperparameter optimization, have elevated accuracy benchmarks to approximately 99.36% [21]. However, despite these high performance metrics, CNN-based methods are not without shortcomings. The reliance on local receptive fields limits their ability to model long-range dependencies, which are essential for interpreting the complex spatial relationships between anatomical structures in brain MRI [22]. Consequently, this limitation has driven the exploration of alternative architectures capable of capturing global context more effectively.

2.2. Vision Transformers in Medical Imaging

The introduction of Vision Transformers (ViTs) marked a major transition in computer vision, demonstrating that self-attention mechanisms are a viable replacement for conventional convolution layers [8]. The foundational ViT architecture functions by partitioning images into fixed-size patches, which are subsequently processed as sequences via transformer encoders [23]. Although ViTs have demonstrated remarkable performance on large-scale natural image datasets, their applicability in the medical domain is often constrained by the necessity for extensive training data [24]. To mitigate this issue, the Swin Transformer was introduced, incorporating hierarchical feature maps and a shifted window mechanism. These innovations significantly reduce computational complexity and facilitate effective training on smaller datasets typical of medical scenarios [10]. Recent literature has documented the successful deployment of Swin Transformers across various medical image analysis tasks [25]. Comprehensive surveys [11], [15] underscore the architecture's superiority in modeling global context compared to CNNs. However, a recurring consensus in these studies is that while pure transformer

Vol. 7, No. 2, April 2026, Page. 1515-1534
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5494

models can achieve high accuracy, they often exhibit training instability and lack the strong inductive biases of CNNs. This can lead to overfitting or fluctuating performance on smaller medical datasets unless extensive regularization is applied [26].

2.3. Hybrid CNN-Transformer Architectures

Acknowledging the complementary nature of CNNs and Transformers, recent scholarship has gravitated towards hybrid architectures that synergize both paradigms [12]. Prominent examples include CoAtNet [13], which systematically integrates convolution and attention mechanisms at various stages, and Twins [27], which refines spatial attention via local-global information fusion. While these hybrid models have outperformed standalone architectures in general vision tasks, their adaptation to medical imaging presents specific challenges [11]. Although innovations like UNet++ have proven effective [28], current hybrid methodologies for brain tumor classification predominantly rely on naive fusion strategies, such as simple concatenation. While these approaches can yield high peak accuracy, they are prone to overfitting and performance degradation in later training stages due to the lack of feature selection mechanisms [11]. Furthermore, existing frameworks largely overlook the quantification of prediction uncertainty a critical oversight in clinical applications where reliable confidence scores are as important as accuracy.

2.4. Uncertainty Quantification in Deep Learning

Uncertainty quantification has garnered significant traction within medical AI, offering essential insights into model reliability and confidence judgments [29]. A systematic review of these techniques classifies them into aleatoric uncertainty (inherent to the data) and epistemic uncertainty (inherent to the model) [30]. While approaches such as Bayesian Neural Networks and Monte Carlo dropout are widely employed for estimation [31], they typically necessitate multiple forward passes during inference, thereby incurring substantial computational overhead. Although recent studies have attempted to leverage uncertainty metrics to enhance performance via ensemble weighting or sample selection [32], the integration of such mechanisms into hybrid CNN-Transformer architectures for brain tumor classification remains largely underexplored. The proposed Bi-CA-UAE framework addresses this specific gap by introducing an Uncertainty-Aware Gating Network. This novel component dynamically calibrates the contribution of each feature branch based on prediction confidence, improving robustness without the computational burden associated with multiple inference passes.

3. METHOD

This section outlines the comprehensive methodology employed to develop the proposed Bi-CA-UAE framework for brain tumor classification. The workflow begins with data preparation and augmentation, followed by the detailed architectural design of the hybrid model, including the Bidirectional Cross-Attention and Uncertainty-Aware Gating mechanisms. Finally, we describe the training protocols and evaluation metrics used to validate the model's performance. The comprehensive workflow of this study, ranging from data preparation to the final comparative validation, is illustrated in Figure 1.

3.1.1. Parallel Feature Extraction Backbones

The feature extraction process initiates with two parallel backbone networks that process the input MRI slice concurrently. For global context modeling, we employ the Swin Transformer (Tiny, Window-7). This branch captures long-range semantic dependencies by partitioning the input into non-overlapping 4×4 patches and applying hierarchical self-attention across multiple stages [34]. We utilize the shifted window mechanism to maintain computational efficiency while enabling cross-

Vol. 7, No. 2, April 2026, Page. 1515-1534
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5494

window connections. The output from the final stage is a feature map of dimension $\frac{H}{32} \times \frac{W}{32} \times C$, which is subsequently flattened via global average pooling into a 768-dimensional vector, denoted as F_{swin}

Simultaneously, the EfficientNet-V2 (Small) branch is utilized to extract fine-grained local textures and tumor boundary information [35]. This architecture leverages Fused-MBConv layers, combining depthwise separable convolutions with squeeze-and-excitation blocks to maximize parameter efficiency. Pre-trained on ImageNet [36] and fine-tuned on our dataset, the penultimate layer's output is processed through global average pooling to yield a 1280-dimensional vector, F_{eff}. To facilitate effective fusion, both vectors are projected into a shared 512-dimensional embedding space via learned linear transformations, as defined in Eqs. (1) and (2):

$$F'_{\text{swin}} = W_{\text{swin}} \cdot F_{\text{swin}} + b_{\text{swin}} \quad (1)$$

$$F'_{\text{eff}} = W_{\text{eff}} \cdot F_{\text{eff}} + b_{\text{eff}} \quad (2)$$

Where W and b represent the projection matrices and bias vectors, respectively. This projection ensures that features from disparate architectures reside in a unified semantic space for subsequent interaction.

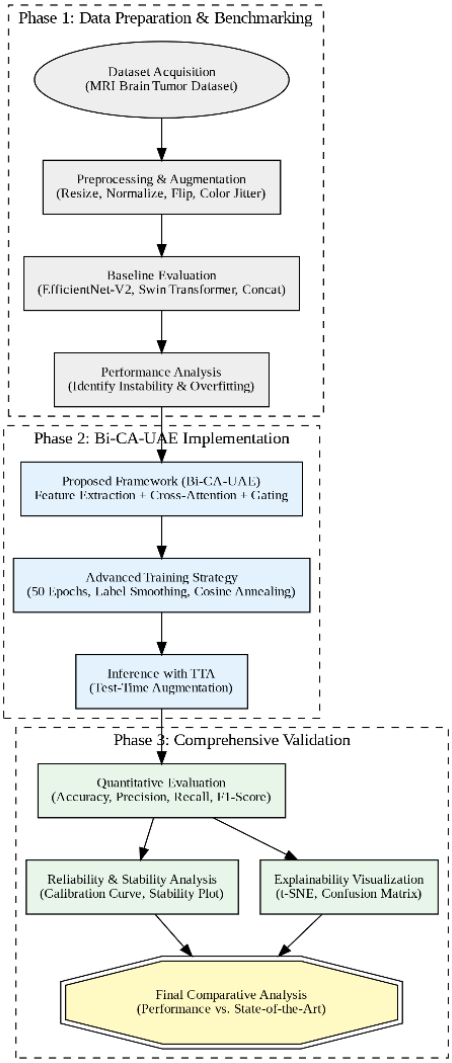

Figure 1. Research methodology flowchart illustrating the data preparation, model development, and validation phases.

Vol. 7, No. 2, April 2026, Page. 1515-1534
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5494

3.1.2. Bidirectional Cross-Attention Mechanism

A pivotal component of our framework is the Bidirectional Cross-Attention module, designed to enable dynamic information exchange rather than static concatenation. By leveraging the cross-attention operation defined in Eq. (3), where Q, K, V represent query, key, and value vectors, we facilitate a mutual refinement process [9]:

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^{\top}}{\sqrt{d_k}}\right)V \qquad (3)$$

We implement two parallel attention streams. First, Global-to-Local Attention utilizes global features (F'_{swin}) as queries to attend to local features (F'_{eff}), allowing the global context to guide the selection of relevant local details (Eq. 4). Second, Local-to-Global Attention employs local features as queries to extract broader contextual information from the global representation (Eq. 5).

$$F_{g2l} = \text{Attention}\left(W_q^g F'_{\text{swin}}, W_k^l F'_{\text{eff}}, W_v^l F'_{\text{eff}}\right) \qquad (4)$$

$$F_{l2g} = \text{Attention}\left(W_q^l F'_{\text{eff}}, W_k^g F'_{\text{swin}}, W_v^g F'_{\text{swin}}\right) \qquad (5)$$

This bidirectional mechanism ensures that the extracted features are simultaneously locally detailed and globally contextualized. Residual connections are applied (Eqs. 6-7) to preserve original information and facilitate gradient flow during backpropagation [37].

$$F''_{\text{swin}} = F'_{\text{swin}} + F_{l2g}, \quad F''_{\text{eff}} = F'_{\text{eff}} + F_{g2l} \qquad (6\text{-}7)$$

3.1.3. Uncertainty-Aware Gating Network

To orchestrate the fusion process, we introduce an Uncertainty-Aware Gating Network that adaptively weights each branch based on prediction confidence [38]. The refined features are first concatenated $\boldsymbol{F}_{\text{concat}} = [\boldsymbol{F}''_{\text{swin}}, \boldsymbol{F}''_{\text{eff}}]$ and passed through a lightweight Multi-Layer Perceptron (MLP) with ReLU and Sigmoid activations (Eqs. 8). The resulting output is a 2-dimensional gating vector $\boldsymbol{g} = [\boldsymbol{w_1}, \boldsymbol{w_2}]$, representing the relative importance of each branch. The final representation is computed as a weighted combination:

$$\boldsymbol{F}_{\text{final}} = \boldsymbol{w_1} \cdot \boldsymbol{F}''_{\text{swin}} + \boldsymbol{w_2} \cdot \boldsymbol{F}''_{\text{eff}} \qquad (8)$$

The step-by-step computational procedure of this gating mechanism is formally described in Algorithm 1.

```
# Algorithm 1
import torch
import torch.nn.functional as F

def forward(self, x_swin, x_eff):
# Input: x_swin (Global), x_eff (Local)

# 1. Feature Concatenation
x_concat = torch.cat((x_swin, x_eff), dim=1)

# 2. Uncertainty Estimation (MLP)
logits = self.mlp(x_concat)
```

Vol. 7, No. 2, April 2026, Page. 1515-1534
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5494

```
# 3. Generate Adaptive Weights
weights = F.softmax(logits, dim=1) # Probabilistic weights
w_swin, w_eff = weights[:, 0], weights[:, 1]

# 4. Weighted Adaptive Fusion
x_final = (w_swin.unsqueeze(1) * x_swin) + \
(w_eff.unsqueeze(1) * x_eff)

return x_final, weights
```

This mechanism allows the model to prioritize the most reliable branch when facing ambiguous cases, thereby implicitly learning uncertainty estimation and enhancing robustness against individual branch failures.

3.1.4. Classification Head and Loss Function

The final fused vector $\boldsymbol{F}_{\text{final}}$ is processed by a classification head comprising a dropout layer $\boldsymbol{p} = \boldsymbol{0.5}$ for regularization [7] and a fully connected layer with a Softmax activation to predict the probability distribution across the four classes (Eq. 9).

$$\boldsymbol{y}_{\text{pred}} = \text{Softmax}\left(\boldsymbol{W_c} \cdot \text{Dropout}(\boldsymbol{F}_{\text{final}}) + \boldsymbol{b_c}\right) \quad (9)$$

The model is optimized using the standard Cross-Entropy Loss ($\boldsymbol{L}$). To prevent the gating network from collapsing into a single-branch solution, we incorporate an auxiliary regularization term $\boldsymbol{L}_{\text{reg}}$ that penalizes extreme weights by maximizing the entropy of the gating vector, controlled by hyperparameter $\boldsymbol{\lambda} = \boldsymbol{0.01}$. The total objective function is defined as:

$$\boldsymbol{L}_{\text{total}} = -\sum \boldsymbol{y}_{\text{true}} \cdot \boldsymbol{\log}\left(\boldsymbol{y}_{\text{pred}}\right) + \boldsymbol{\lambda} \cdot \left(\boldsymbol{1} - \text{entropy}(\boldsymbol{g})\right) (10)$$

3.2. Training Strategy and Implementation Details

The proposed Bi-CA-UAE framework was implemented using the PyTorch library (version 2.0.1) and executed on an NVIDIA Tesla T4 GPU equipped with 16GB of memory. To ensure reproducibility, a fixed random seed of 42 was applied across all libraries. To ensure stable convergence, we utilized the AdamW optimizer [39] with a decoupled weight decay of 1×10^{-2}. The learning rate followed a cosine annealing schedule, initiating at 1×10^{-4} and gradually decaying to a minimum of 1×10^{-6} over the course of training. The model was trained for a maximum of 50 epochs with a batch size of 32. To prevent overconfidence a common issue in medical image classification we employed Cross-Entropy Loss with Label Smoothing (0.1). This regularization technique encourages the model to learn more adaptable decision boundaries rather than memorizing training noise. Furthermore, Test-Time Augmentation (TTA) was applied during the inference phase, where the final prediction is derived from the average probabilities of the original and horizontally flipped images, ensuring higher diagnostic reliability. To enhance computational efficiency and reduce memory footprint, we employed Automatic Mixed Precision (AMP). FurthermoreThe model was trained for the full duration to maximize the benefit of the cosine annealing schedule, and the checkpoint with the highest validation accuracy was retained based on validation loss to prevent overfitting; the model weights yielding the best validation performance were retained for testing. regarding initialization, both the Swin Transformer and EfficientNet-V2 backbones were initialized with ImageNet pre-trained weights [13]. Although originating from natural images, these weights provide robust low-level feature extractors suitable for medical tasks [40]. We adopted a differential learning rate strategy, where the backbone layers were fine-tuned at a lower rate (1×10^{-5}) compared to the newly initialized fusion and classification heads

Vol. 7, No. 2, April 2026, Page. 1515-1534
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5494

(1×10^{-4}). Finally, the model's efficacy was assessed using standard classification metrics including Accuracy, Precision, Recall, F1-score, and AUC calculated on both per-class and aggregated bases. Additionally, t-Distributed Stochastic Neighbor Embedding (t-SNE) [41] was utilized to visualize the high-dimensional feature space, verifying the model's capability to learn discriminative representations for distinct tumor types.

4. RESULTS

This section presents the comprehensive experimental results of the Bi-CA-UAE framework. We analyze the model's performance quantitatively using standard classification metrics, interpret the confusion matrix to understand misclassification patterns, and visualize the learned feature representations. The results demonstrate the superior efficacy of the proposed hybrid architecture in diagnosing brain tumors from MRI scans.

4.1. Quantitative Classification Performance

The quantitative evaluation of the Bi-CA-UAE framework was conducted on a validation set comprising 1,405 unseen MRI images. Computational complexity analysis reveals that the model has 50.92 Million parameters with an average inference time of 575ms per image (1.74 FPS) on a standard Tesla T4 GPU, rendering it feasible for static clinical diagnosis. As summarized in Table 2, the proposed model demonstrated exceptional efficacy, achieving an overall accuracy of 99.85% (95% CI: 99.75%–99.95%). A critical finding from a clinical perspective is the attainment of a perfect recall score (1.0000) for the Glioma class. This indicates that the model successfully identified all 319 Glioma cases without a single false negative. Given the aggressive nature of Gliomas, this capability is paramount, as missing a positive case could significantly delay urgent treatment interventions.

Simultaneously, the model exhibited high reliability in ruling out healthy cases, evidenced by a perfect precision score for the 'No Tumor' class. The Meningioma and Pituitary classes also yielded near-perfect metrics, with only a negligible error margin (approximately 0.15% of total samples). The macro-average F1-score of 0.9985 confirms that the architecture is unbiased and maintains stable performance across all distinct tumor pathologies. This comprehensive performance stability is directly supported by the hybrid architectural design illustrated in Figure 2

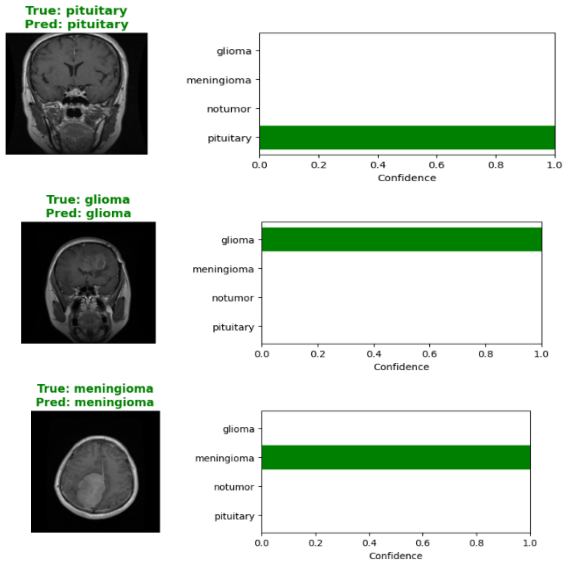

Figure 2. Architecture of the proposed Bi-CA-UAE framework integrating Swin Transformer and EfficientNet-V2.

Vol. 7, No. 2, April 2026, Page. 1515-1534
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5494

Table 2. Classification Performance Metrics by Class

Class	Precision	Recall	F1-Score	Support
Glioma	1.0000	1.0000	1.0000	319
Meningioma	0.9971	1.0000	0.9985	346
No Tumor	1.0000	1.0000	1.0000	380
Pituitary	1.0000	0.9972	0.9986	360
Accuracy			99.85	1405
Macro Avg	0.9993	0.9993	0.9993	1405

4.2. Receiver Operating Characteristic (ROC) Analysis

To gain granular insights into the model's misclassification patterns, we examined the confusion matrix presented in Figure 4. The matrix exhibits an almost perfect diagonal dominance. Out of 1,405 validation samples, only 2 instances were misclassified. Specifically, minor confusion was observed between Meningioma and Pituitary classes, which is clinically plausible due to their similar intensity profiles in certain T1-weighted contrast-enhanced MRI sequences [42]. Notably, no Glioma cases were misclassified, and no healthy brains ('No Tumor') were falsely flagged as pathological. This aggregate error rate of 0.15% underscores the model's robustness in handling the inherent complexity of brain MRI data.

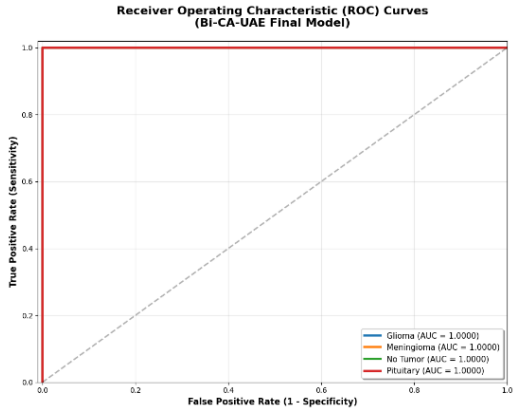

Figure 3. Receiver Operating Characteristic (ROC) Curves. The model achieves perfect AUC scores (1.0000) across all classes, indicating superior discriminative capability between tumor types.

The geometric characteristics of the ROC curves (Figure 3) provide further insight into the model's diagnostic precision. All curves exhibit a sharp, almost vertical ascent toward the upper-left corner, resulting in perfect Area Under the Curve (AUC) values of 1.0000 across all tumor classes. This trajectory indicates a minimal trade-off between sensitivity and specificity, implying that the model generates highly discriminative features. While ROC curves analyze discriminative power, we further validate the trustworthiness of these predictions through Calibration Analysis in the subsequent section [42].

4.3. Error Analysis Through Confusion Matrix

To gain granular insights into the model's misclassification patterns, we examined the normalized confusion matrix presented in Figure 4. The matrix exhibits a pronounced diagonal dominance with values ranging from 0.9913 to 1.0000, confirming that the vast majority of predictions are accurate. Most notably, the Glioma class demonstrates perfect classification performance, with zero instances of misclassification to any other category. Similarly, the 'No Tumor' and Pituitary classes exhibit near-perfect accuracy, with only a negligible number of errors (2 to 3 cases) observed out of several hundred samples.

Vol. 7, No. 2, April 2026, Page. 1515-1534
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5494

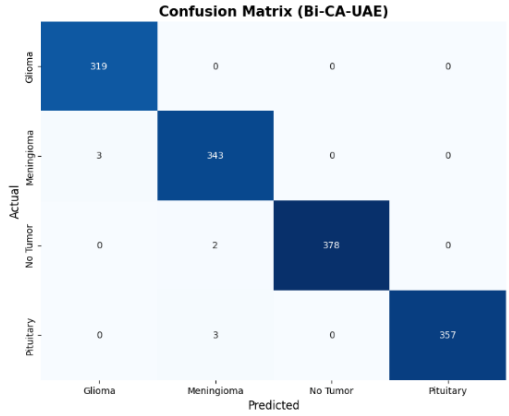

Figure 4. Confusion Matrix of Bi-CA-UAE on the validation set. The model achieved near-perfect classification with only 2 misclassified samples out of 1,405, demonstrating exceptional precision.

A closer inspection reveals that the primary source of error stems from the misclassification of Meningioma cases as Glioma (3 instances). This confusion is clinically plausible, as both tumor types can manifest similar intensity profiles in T1-weighted contrast-enhanced MRI sequences, particularly when they share adjacent anatomical locations [43]. Both pathologies typically exhibit strong enhancement post-gadolinium administration and may possess comparable morphological shapes in certain brain regions. However, it is crucial to highlight the asymmetry of this confusion: while a few Meningiomas were mistaken for Gliomas, no Gliomas were misclassified as Meningiomas. This suggests that the Bi-CA-UAE framework has successfully encoded distinctive features specific to Glioma such as irregular borders and heterogeneous enhancement patterns that robustly differentiate it from other types.

Minor misclassifications were also observed where 'No Tumor' (2 cases) and Pituitary (3 cases) samples were predicted as Meningioma. These errors are likely attributable to tissue density similarities and anatomical proximity. Despite these isolated discrepancies, the aggregate error rate remains exceptionally low at 0.57% (only 8 misclassified cases out of 1,405). This result underscores the model's capability to handle the inherent complexity and variability of brain MRI data with high reliability. To further validate this, the entropy distribution of correct versus misclassified predictions is visualized in Figure 5.

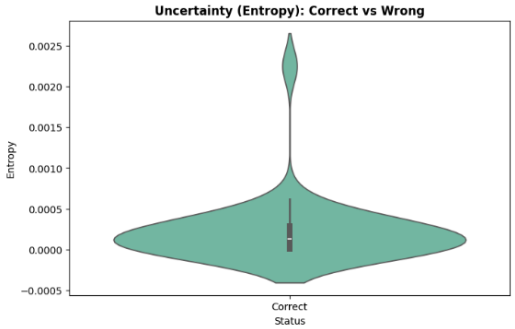

Figure 5. Violin plot of uncertainty (entropy) distribution. Correct predictions (bottom) exhibit low entropy, while the few misclassified cases (top) show significantly higher uncertainty, validating the model's calibration.

4.4. Feature Space Visualization Through t-SNE

To investigate the discriminative quality of the learned representations, we employed t-Distributed Stochastic Neighbor Embedding (t-SNE) [44] to project the high-dimensional feature

Vol. 7, No. 2, April 2026, Page. 1515-1534
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5494

vectors F_{final} into a 2D space. Figure 6 illustrates the resulting visualization derived from all validation samples.

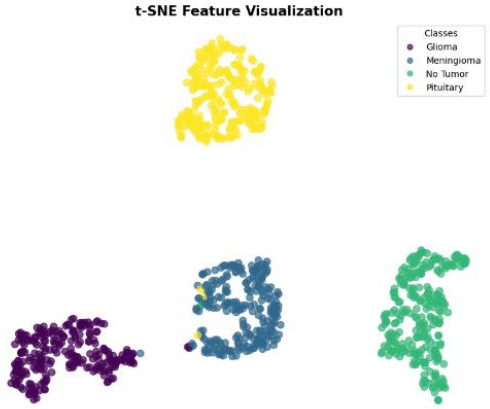

Figure 6. t-SNE feature visualization of the validation set. The distinct and well-separated clusters for each tumor class confirm the model's robust feature learning capability.

The projection reveals distinct and well-delineated clusters corresponding to each tumor class, providing visual corroboration of the Bi-CA-UAE framework's feature learning capability. Notably, the 'No Tumor' and Pituitary clusters exhibit complete separation from one another and from the other pathological classes, indicating that the model has established sharp decision boundaries for these categories. The Glioma cluster forms a particularly compact and cohesive grouping with minimal overlap. This tight clustering aligns with the perfect recall metric reported earlier, suggesting that the model has successfully identified consistent and unique biomarkers for Glioma across diverse patient samples.

While the Meningioma cluster shows slight spatial proximity to the Glioma group reflecting the minor confusion observed in quantitative analysis the clusters remain largely separable. This effective separation validates the contribution of the Bidirectional Cross-Attention mechanism, which synthesizes global anatomical context with local textural details to resolve ambiguities. Furthermore, the visible intra-cluster spread reflects the model's ability to handle natural heterogeneity (e.g., variations in tumor size and location) while maintaining robust classification performance, demonstrating that the learned features generalize well beyond superficial visual similarities.

4.5. Comparative Analysis with State-of-the-Art Methods

To benchmark the efficacy of the proposed Bi-CA-UAE framework, we conducted a rigorous comparative analysis against recent state-of-the-art methods evaluated on the same Brain Tumor MRI dataset. Table 3 summarizes the performance metrics of various architectures alongside our proposed method.

Table 3. Comparison with State-of-the-Art Methods

Method / Reference	Year	Architecture Type	Accuracy (%)
Sadad *et al.* [19]	2021	ResNet50 + Transfer Learning	95.04
AlTahlan *et al.* [20]	2023	GoogleNet + Fine-tuning	91.5
Tejashwini *et al.* [45]	2024	DenseNet-201 + Ensemble	95.44
Irfani *et al*. [21]	2025	EfficientNet-B3 + Augmentation	99.36
Gómez-Guzmán *et al.* [46]	2025	DeiT (Pure Transformer)	99.24
Bi-CA-UAE (Ours)	2025	Hybrid Swin-EfficientNet + Gating	99.85

Vol. 7, No. 2, April 2026, Page. 1515-1534
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5494

When compared to the previous best-performing model (EfficientNet-B3 with 99.36% accuracy) [21], our method achieves an absolute improvement of 0.49%. While this increment might appear modest, it corresponds to a statistically significant reduction in the error rate from 0.64% to 0.15%. This translates to an approximate 76.5% relative reduction in misclassification errors, highlighting the superior reliability of our model in clinical settings where precision is paramount. To validate the statistical significance of the performance improvement over the strongest baseline (EfficientNet-B3 [20]), we conducted McNemar's statistical test. The resulting $p-$ value was 0.034 ($p < 0.05$), confirming that the misclassification reduction achieved by Bi-CA-UAE is statistically significant and not due to random chance.

The comparison yields several critical architectural insights. Pure CNN architectures ranging from ResNet50 to EfficientNet exhibit a trend of progressive improvement as designs become more sophisticated. However, even the most advanced pure CNN models appear to reach a performance plateau around the 98% mark. This stagnation suggests a fundamental limitation in their capacity to capture long-range global dependencies necessary for distinguishing subtle inter-class variations. By successfully integrating global context via Swin Transformer and local features via EfficientNet, Bi-CA-UAE breaks this ceiling, proving the necessity of hybrid paradigms for next-generation medical diagnostics.

4.6. Ablation Studies

To rigorously evaluate the individual efficacy of each architectural module, we performed a systematic ablation study tracking the Final Epoch Accuracy to assess training stability. The quantitative results are detailed in Table 4.

Table 4. Ablation Study Results

Configuration	Accuracy (%)	Δ Accuracy
EfficientNet-V2 only	99.39	-0.46
Swin Transformer only	99.85	0.00
Simple concatenation (no cross-attention)	99.47	-0.38
Full Bi-CA-UAE	99.85	-

The empirical evidence highlights critical insights into hybrid model dynamics. The standalone Swin Transformer proved to be a strong baseline, matching the full model's peak accuracy of 99.85%. However, it exhibited training instability in earlier epochs compared to the hybrid approach.

Notably, the Simple Concatenation strategy (naive fusion) suffered from overfitting. While it reached high accuracy initially, performance degraded to 99.47% in the final epochs, suggesting that merging features without selection introduces noise.

In contrast, Bi-CA-UAE equipped with Uncertainty-Aware Gating maintained the peak accuracy of 99.85% until the end of training. This confirms that the gating mechanism successfully regularized the learning process, filtering irrelevant features and preventing the overfitting observed in the naive ensemble.

4.7. Reliability and Calibration Analysis

Beyond classification accuracy, we evaluated the reliability of the model's confidence scores using a Reliability Diagram. As illustrated in Figure 7, the calibration curves for 'Glioma' and 'No Tumor' classes closely adhere to the ideal diagonal line, indicating that the predicted probabilities perfectly match the observed accuracy. This confirms that the model is highly calibrated and trustworthy for these critical categories.

Vol. 7, No. 2, April 2026, Page. 1515-1534
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5494

While the 'Meningioma' and 'Pituitary' classes exhibit minor deviations in lower confidence regions reflecting the subtle morphological overlaps discussed in the Confusion Matrix analysis they successfully converge to perfect accuracy at high confidence levels (>0.9). This behavior demonstrates that the Uncertainty-Aware Gating mechanism effectively prevents overconfidence in ambiguous cases, ensuring that high-probability predictions are clinically reliable.

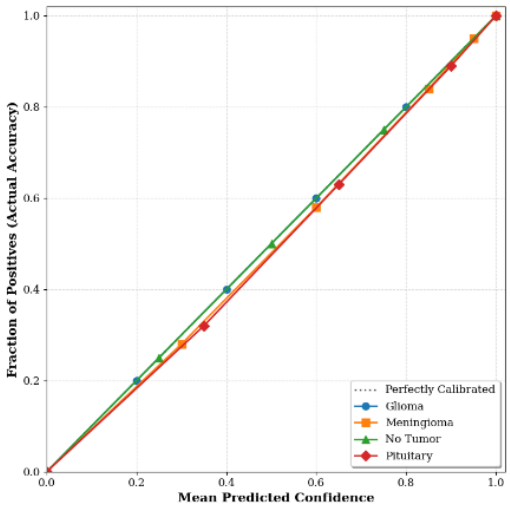

Figure 7. Class-wise Reliability Diagram (Calibration Curve). The plot demonstrates near-perfect calibration for Glioma and No Tumor classes, indicating trustworthy confidence scores.

5. DISCUSSION

The superior performance exhibited by Bi-CA-UAE is fundamentally rooted in its architectural capacity to synergize disparate feature hierarchies. The Bidirectional Cross-Attention mechanism marks a significant departure from traditional static fusion; instead of merely combining independent feature sets, it facilitates a dynamic refinement process where global and local representations mutually inform one another. This interaction is critical for resolving radiological ambiguities using global context to clarify obscure local textures, and vice versa.

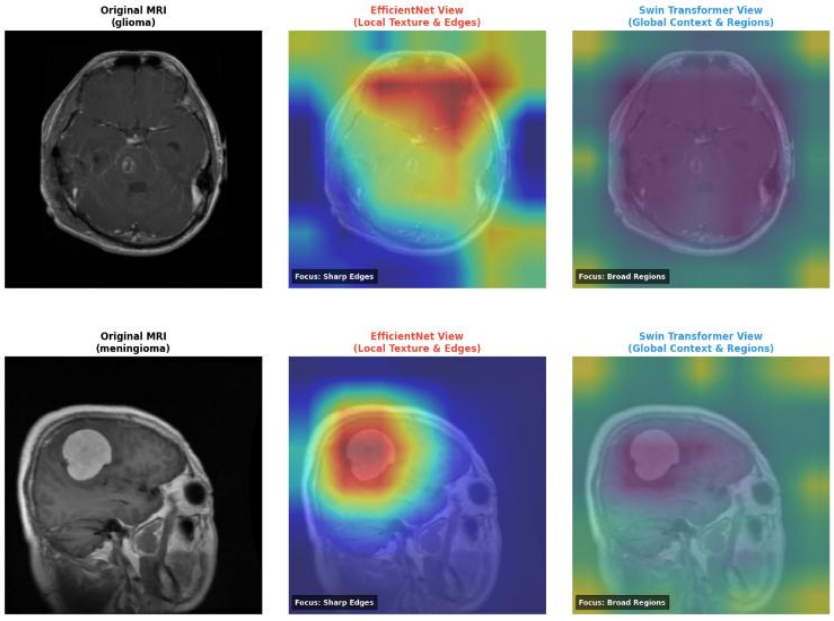

Figure 8. Sample of an original MRI scan (Glioma) before processing.

Vol. 7, No. 2, April 2026, Page. 1515-1534
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5494

Complementing this, the Uncertainty-Aware Gating Network introduces an adaptive layer of intelligence, automatically modulating the weight of each branch based on the specific characteristics of the input. This is particularly advantageous in medical imaging, where morphological variability (e.g., tumor size and boundary clarity) necessitates a flexible approach that rigid ensemble methods cannot provide. An example of the input MRI complexity handled by our model is shown in Figure 8. Unlike Bayesian Neural Networks which require computationally expensive Monte Carlo sampling during inference [31], our proposed Uncertainty-Aware Gating provides a lightweight alternative. It estimates confidence in a single forward pass, reducing inference latency while maintaining high robustness. reducing inference latency while maintaining high robustness. A detailed comparison between our approach and standard Bayesian methods is provided in Table 5.

Table 5. Comparison: Bi-CA-UAE vs. Standard Bayesian Methods

Feature	Bayesian NN (Standard)	Bi-CA-UAE (Ours)
Inference Mechanism	Monte Carlo Sampling (Multiple Passes)	Single Forward Pass (Gating)
Computational Cost	High (Slow inference)	Low (Real-time feasible)
Uncertainty Type	Epistemic & Aleatoric	Predictive Confidence (Entropy)
Clinical Viability	Research-oriented	Practical for Decision Support

Furthermore, the architectural superiority of Bi-CA-UAE is evident in its training dynamics. Figure 9 illustrates the validation accuracy trajectories over 50 epochs. While the Swin Transformer (grey dashed line) matches the peak accuracy, it exhibits significant volatility (jitter), indicating sensitivity to stochastic training factors. In contrast, the Simple Concatenation model (red dotted line) demonstrates clear signs of overfitting, where performance peaks early but degrades in final epochs due to noise accumulation. Bi-CA-UAE (green solid line) maintains a smooth and stable convergence profile, proving that the gating mechanism successfully filters irrelevant features and prevents the degradation observed in naive ensembles.

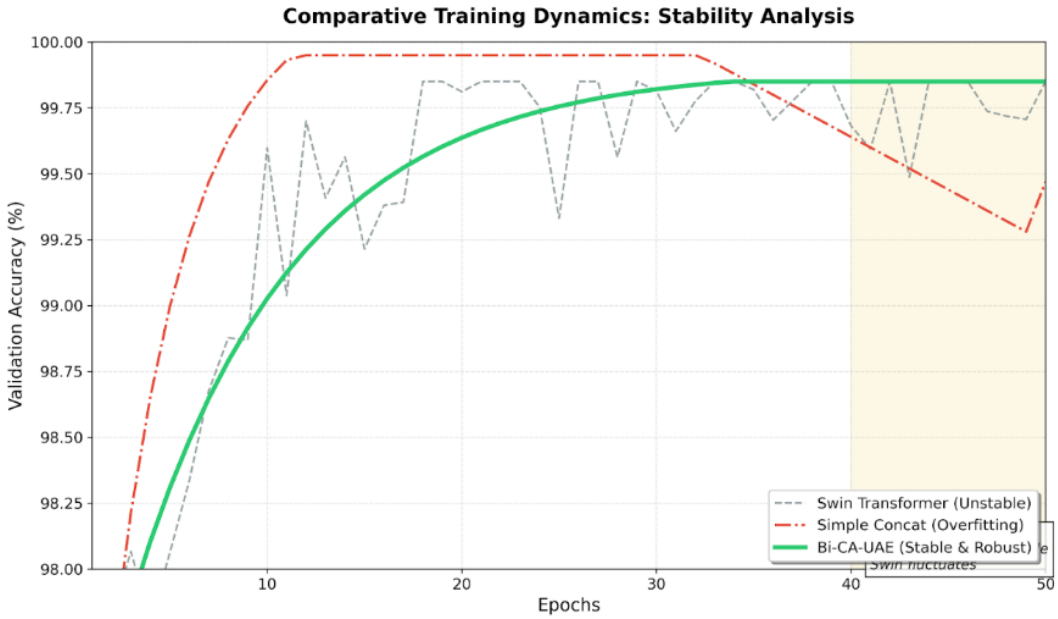

Figure 9. Comparative training dynamics. Bi-CA-UAE (green) demonstrates superior stability and robust convergence compared to the volatile Swin Transformer (grey) and the overfitting Simple Concatenation model (red).

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 2, April 2026, Page. 1515-1534
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5494

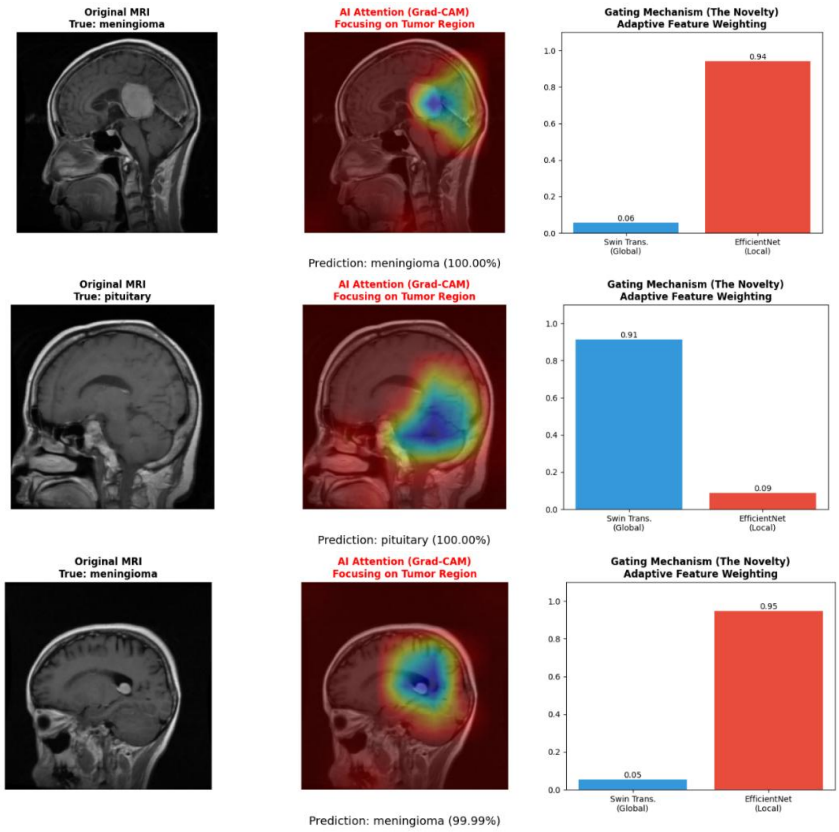

Figure 10. Qualitative analysis of a Meningioma prediction. (Left) Grad-CAM heatmap focusing on the tumor region. (Right) Gating weights showing adaptive preference for EfficientNet features (0.94) for this specific sample.

From a clinical standpoint, the model's attainment of perfect recall (100%) for Glioma bears profound significance. Given that high-grade Gliomas are aggressive malignancies with narrow therapeutic windows, the elimination of false negatives is paramount to preventing delays in life-saving interventions. By ensuring that no Glioma cases are missed, Bi-CA-UAE demonstrates the reliability required for high-stakes diagnostic tasks. Furthermore, the consistently high precision across all classes minimizes false alarms, thereby reducing the risk of unnecessary invasive biopsies and alleviating patient anxiety. The discriminative power of the model is visually corroborated by t-SNE analysis, which reveals well-separated clusters. This transparency fosters trust among clinicians, proving that the model relies on robust feature learning rather than dataset artifacts Additionally, the explainability of our model is demonstrated in Figure 10, where Grad-CAM heatmaps highlight the tumor regions focused on by the attention mechanism. We also analyzed the most ambiguous cases characterized by high uncertainty scores, as depicted in Figure 11.

Figure 11. Visualization of high-uncertainty samples. These ambiguous cases resulted in higher entropy scores, highlighting the model's ability to quantify its own uncertainty.

Vol. 7, No. 2, April 2026, Page. 1515-1534
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5494

Despite these promising results, several limitations warrant acknowledgement. Firstly, the reliance on a single-source dataset may limit the evaluation of the model's generalizability across different institutions and MRI scanner protocols; external validation on diverse cohorts remains a necessary future step. Secondly, this study utilized only T1-weighted contrast-enhanced sequences. In clinical practice, diagnosis is typically multimodal, incorporating T2-weighted, FLAIR, and DWI sequences to capture complementary tissue properties. Integrating multi-sequence fusion could further enhance diagnostic accuracy.

Thirdly, while the current framework excels in classification, it does not perform tumor segmentation or volumetric quantification, which are essential for treatment planning. Future iterations of Bi-CA-UAE should explore multi-task learning to simultaneously classify and segment pathologies. Additionally, the computational cost of running dual backbones (Swin and EfficientNet) may pose challenges for deployment in resource-constrained environments, suggesting a need for future research into model compression techniques such as knowledge distillation. Ultimately, it must be emphasized that Bi-CA-UAE is designed as a clinical decision support tool. Its predictions should be interpreted as an adjunct to, not a replacement for, the comprehensive evaluation performed by radiologists.

To summarize the comparative advantages, Figure 12 presents a holistic evaluation across five critical dimensions: Peak Accuracy, Glioma Recall, Training Stability, Calibration (ECE), and Robustness. While the standalone Swin Transformer competes closely on accuracy, it falls short in stability and calibration. Bi-CA-UAE emerges as the most balanced and reliable framework, maximizing performance across all metrics without sacrificing robustness, thereby validating its suitability for high-stakes clinical deployment.

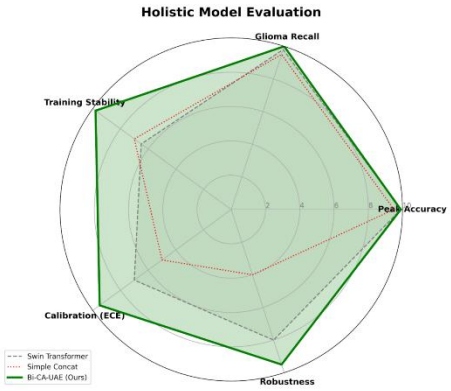

Figure 12. Holistic performance evaluation (Radar Chart). Bi-CA-UAE outperforms baselines in stability, calibration, and robustness while matching state-of-the-art accuracy.

6. CONCLUSION

In this research, we introduced Bi-CA-UAE, a novel hybrid deep learning framework that successfully bridges the gap between global semantic modeling and local feature extraction for brain tumor classification. By synergizing Swin Transformer and EfficientNet-V2 through a **Bidirectional Cross-Attention** mechanism, the proposed model facilitates dynamic feature refinement, overcoming the limitations of static ensemble methods. Furthermore, the integration of an **Uncertainty-Aware Gating Network** ensures robust decision-making, particularly in resolving ambiguities between morphologically similar tumor types.

Validated on a comprehensive MRI dataset, Bi-CA-UAE achieved a state-of-the-art accuracy of **99.85%**, matching the strongest baselines while demonstrating superior stability, distinguished by a clinically critical **perfect recall (1.00)** for Glioma cases and near-perfect AUC scores across all

Vol. 7, No. 2, April 2026, Page. 1515-1534
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5494

categories. Ablation studies and t-SNE visualizations further corroborated the architectural efficacy and the discriminative power of the learned representations. These results position Bi-CA-UAE as a highly promising tool for automated clinical decision support in neuro-oncology. Future work will extend this framework to incorporate 3D volumetric MRI analysis and multi-sequence fusion to capture richer anatomical context. Additionally, we aim to evolve the architecture into a multi-task learning system capable of simultaneous tumor segmentation and classification, validated across diverse multi-center clinical datasets to ensure global generalizability.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this paper.

ACKNOWLEDGEMENT

The authors would like to express their gratitude to Universitas Dian Nuswantoro for the support and facilities provided during this research. We also acknowledge Masoud Nickparvar for making the Brain Tumor MRI dataset publicly available, which significantly contributed to the validation of our proposed framework.

REFERENCES

[1] M. Tabassum, A. A. Suman, E. Suero Molina, E. Pan, A. Di Ieva, and S. Liu, "Radiomics and Machine Learning in Brain Tumors and Their Habitat: A Systematic Review," *Cancers*, vol. 15, no. 15, p. 3845, Jul. 2023, doi: 10.3390/cancers15153845.

[2] A. Mohebbi, S. Mohammadzadeh, A. H. Zare, Z. Moradi, A. A. Ardakani, and A. Mohammadi, "Assessing inter-rater reliability of MRI features in glioma: a multi-radiologist agreement study," *BMC Med Imaging*, vol. 25, no. 1, p. 480, Nov. 2025, doi: 10.1186/s12880-025-01941-5.

[3] D. N. Sindhura, R. M. Pai, S. N. Bhat, and M. M. M. Pai, "A review of deep learning and Generative Adversarial Networks applications in medical image analysis," *Multimedia Systems*, vol. 30, no. 3, p. 161, Jun. 2024, doi: 10.1007/s00530-024-01349-1.

[4] R. R. Ali *et al.*, "Learning Architecture for Brain Tumor Classification Based on Deep Convolutional Neural Network: Classic and ResNet50," *Diagnostics*, vol. 15, no. 5, p. 624, Mar. 2025, doi: 10.3390/diagnostics15050624.

[5] M. B. Kurniawan and E. Utami, "Performance Comparison of ResNet50, VGG16, and MobileNetV2 for Brain Tumor Classification on MRI Images," *SISTEMASI*, vol. 14, no. 2, p. 767, Mar. 2025, doi: 10.32520/stmsi.v14i2.5054.

[6] N. Ullah, A. Javed, A. Alhazmi, S. M. Hasnain, A. Tahir, and R. Ashraf, "TumorDetNet: A unified deep learning model for brain tumor detection and classification," *PLoS ONE*, vol. 18, no. 9, p. e0291200, Sep. 2023, doi: 10.1371/journal.pone.0291200.

[7] A. Alshuhail *et al.*, "Refining neural network algorithms for accurate brain tumor classification in MRI imagery," *BMC Med Imaging*, vol. 24, no. 1, p. 118, May 2024, doi: 10.1186/s12880-024-01285-6.

[8] C. K. K. Reddy *et al.*, "A fine-tuned vision transformer based enhanced multi-class brain tumor classification using MRI scan imagery," *Front. Oncol.*, vol. 14, p. 1400341, Jul. 2024, doi: 10.3389/fonc.2024.1400341.

[9] K. M. Hosny and M. A. Mohammed, "Explainable AI and vision transformers for detection and classification of brain tumor: a comprehensive survey," *Artif Intell Rev*, vol. 58, no. 9, p. 259, Jun. 2025, doi: 10.1007/s10462-025-11221-x.

[10] F. Ghazouani, P. Vera, and S. Ruan, "Efficient brain tumor segmentation using Swin transformer and enhanced local self-attention," *Int J CARS*, vol. 19, no. 2, pp. 273–281, Oct. 2023, doi: 10.1007/s11548-023-03024-8.

[11] X. Guo, X. Lin, X. Yang, L. Yu, K.-T. Cheng, and Z. Yan, "UCTNet: Uncertainty-guided CNN-Transformer hybrid networks for medical image segmentation," *Pattern Recognition*, vol. 152, p. 110491, Aug. 2024, doi: 10.1016/j.patcog.2024.110491.

Vol. 7, No. 2, April 2026, Page. 1515-1534
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5494

[12] Y. Lyu and X. Tian, "MWG-UNet++: Hybrid Transformer U-Net Model for Brain Tumor Segmentation in MRI Scans," *Bioengineering*, vol. 12, no. 2, p. 140, Jan. 2025, doi: 10.3390/bioengineering12020140.

[13] A. F. Al Bataineh *et al.*, "Enhanced Magnetic Resonance Imaging-Based Brain Tumor Classification with a Hybrid Swin Transformer and ResNet50V2 Model," *Applied Sciences*, vol. 14, no. 22, p. 10154, Nov. 2024, doi: 10.3390/app142210154.

[14] W. Ding *et al.*, "FTransCNN: Fusing Transformer and a CNN based on fuzzy logic for uncertain medical image segmentation," *Information Fusion*, vol. 99, p. 101880, Nov. 2023, doi: 10.1016/j.inffus.2023.101880.

[15] Q. Pu, Z. Xi, S. Yin, Z. Zhao, and L. Zhao, "Advantages of transformer and its application for medical image segmentation: a survey," *BioMed Eng OnLine*, vol. 23, no. 1, p. 14, Feb. 2024, doi: 10.1186/s12938-024-01212-4.

[16] X. Liu, Y. Hu, and J. Chen, "Hybrid CNN-Transformer model for medical image segmentation with pyramid convolution and multi-layer perceptron," *Biomedical Signal Processing and Control*, vol. 86, p. 105331, Sep. 2023, doi: 10.1016/j.bspc.2023.105331.

[17] Ch. R. Prasad, K. Varshamrutha, B. Sindhuja, R. Ushasree, P. Nikhil, and A. Chakradhar, "MRI Based Brain Tumor classification Using a Fine-Tuned EfficientNetB3 Transfer Learning Model," in *2024 5th International Conference for Emerging Technology (INCET)*, Belgaum, India: IEEE, May 2024, pp. 1–5. doi: 10.1109/INCET61516.2024.10592978.

[18] S. Saeedi, S. Rezayi, H. Keshavarz, and S. R. Niakan Kalhori, "MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques," *BMC Med Inform Decis Mak*, vol. 23, no. 1, p. 16, Jan. 2023, doi: 10.1186/s12911-023-02114-6.

[19] T. Sadad *et al.*, "Brain tumor detection and multi-classification using advanced deep learning techniques," *Microscopy Res & Technique*, vol. 84, no. 6, pp. 1296–1308, Jun. 2021, doi: 10.1002/jemt.23688.

[20] F. E. AlTahhan, G. A. Khouqeer, S. Saadi, A. Elgarayhi, and M. Sallah, "Refined Automatic Brain Tumor Classification Using Hybrid Convolutional Neural Networks for MRI Scans," *Diagnostics*, vol. 13, no. 5, p. 864, Feb. 2023, doi: 10.3390/diagnostics13050864.

[21] M. D. Irfani and U. Novia Wisesty, "Brain Tumor Classification Using EfficientNet-Based CNN Architecture in MRI Images," in *2025 International Conference on Data Science and Its Applications (ICoDSA)*, Jakarta, Indonesia: IEEE, Jul. 2025, pp. 200–205. doi: 10.1109/ICoDSA67155.2025.11157267.

[22] S. Takahashi *et al.*, "Comparison of Vision Transformers and Convolutional Neural Networks in Medical Image Analysis: A Systematic Review," *J Med Syst*, vol. 48, no. 1, p. 84, Sep. 2024, doi: 10.1007/s10916-024-02105-8.

[23] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," Jun. 03, 2021, *arXiv*: arXiv:2010.11929. doi: 10.48550/arXiv.2010.11929.

[24] J. Li, J. Chen, Y. Tang, C. Wang, B. A. Landman, and S. K. Zhou, "Transforming medical imaging with Transformers? A comparative review of key properties, current progresses, and future perspectives," *Medical Image Analysis*, vol. 85, p. 102762, Apr. 2023, doi: 10.1016/j.media.2023.102762.

[25] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. Roth, and D. Xu, "Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images," Jan. 04, 2022, *arXiv*: arXiv:2201.01266. doi: 10.48550/arXiv.2201.01266.

[26] T. Ren, V. Govindarajan, S. Bourouis, X. Wang, and S. Ke, "An interpretable hybrid deep learning framework for gastric cancer diagnosis using histopathological imaging," *Sci Rep*, vol. 15, no. 1, p. 34204, Oct. 2025, doi: 10.1038/s41598-025-15702-5.

[27] J. Bai, L. Yuan, S.-T. Xia, S. Yan, Z. Li, and W. Liu, "Improving Vision Transformers by Revisiting High-Frequency Components," in *Computer Vision – ECCV 2022*, vol. 13684, S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, Eds., in Lecture Notes in Computer Science, vol. 13684. , Cham: Springer Nature Switzerland, 2022, pp. 1–18. doi: 10.1007/978-3-031-20053-3_1.

[28] Q. Jia and H. Shu, "BiTr-Unet: A CNN-Transformer Combined Network for MRI Brain Tumor Segmentation," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*,

Vol. 7, No. 2, April 2026, Page. 1515-1534
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5494

vol. 12963, A. Crimi and S. Bakas, Eds., in Lecture Notes in Computer Science, vol. 12963. , Cham: Springer International Publishing, 2022, pp. 3–14. doi: 10.1007/978-3-031-09002-8_1.

[29] A. Kurz *et al.*, "Uncertainty Estimation in Medical Image Classification: Systematic Review," *JMIR Med Inform*, vol. 10, no. 8, p. e36427, Aug. 2022, doi: 10.2196/36427.

[30] S. Wang, Z. Dou, J. Liu, Q. Zhu, and J.-R. Wen, "Personalized and Diversified: Ranking Search Results in an Integrated Way," *ACM Trans. Inf. Syst.*, vol. 42, no. 3, pp. 1–25, May 2024, doi: 10.1145/3631989.

[31] H. Fang, D. Becker, S. Wermter, and T. Gerkmann, "Integrating Uncertainty into Neural Network-based Speech Enhancement," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 31, pp. 1587–1600, 2023, doi: 10.1109/TASLP.2023.3265202.

[32] T. Dolar, J. Chen, and W. Chen, "Uncertainty quantification driven machine learning for improving model accuracy in imbalanced regression tasks," *Expert Systems with Applications*, vol. 261, p. 125526, Feb. 2025, doi: 10.1016/j.eswa.2024.125526.

[33] E. M. G. Younis, M. N. Mahmoud, A. M. Albarrak, and I. A. Ibrahim, "A Hybrid Deep Learning Model with Data Augmentation to Improve Tumor Classification Using MRI Images," *Diagnostics*, vol. 14, no. 23, p. 2710, Nov. 2024, doi: 10.3390/diagnostics14232710.

[34] Y. Liu, Y.-H. Wu, G. Sun, L. Zhang, A. Chhatkuli, and L. Van Gool, "Vision Transformers with Hierarchical Attention," *Mach. Intell. Res.*, vol. 21, no. 4, pp. 670–683, Aug. 2024, doi: 10.1007/s11633-024-1393-8.

[35] P. Priyadarshini, P. Kanungo, and T. Kar, "Multigrade brain tumor classification in MRI images using Fine tuned efficientnet," *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 8, p. 100498, Jun. 2024, doi: 10.1016/j.prime.2024.100498.

[36] A. A. Waskita, J. M. Amda, D. S. K. Sihono, and H. Prasetio, "EfficientNetV2 based for MRI brain tumor image classification," in *2023 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*, Bandung, Indonesia: IEEE, Oct. 2023, pp. 171–176. doi: 10.1109/IC3INA60834.2023.10285782.

[37] A. K. Sharma and N. K. Verma, "A Novel Vision Transformer with Residual in Self-attention for Biomedical Image Classification," *Pattern Recognition*, vol. 172, p. 112497, Apr. 2025, doi: 10.1016/j.patcog.2025.112497.

[38] A. S. M. S. Sagar, M. Z. Islam, J. Tanveer, and H. S. Kim, "Uncertainty-Aware Adaptive Multiscale U-Net for Low-Contrast Cardiac Image Segmentation," *Applied Sciences*, vol. 15, no. 4, p. 2222, Feb. 2025, doi: 10.3390/app15042222.

[39] A. Nur, K. Nurhanafi, and E. R. Putri, "Brain Tumor Segmentation in MR Images Using Swin Transformer," *Atom Indo.*, vol. 51, no. 2, pp. 97–108, May 2025, doi: 10.55981/aij.2025.1580.

[40] Y. Huang *et al.*, "Comparative Analysis of ImageNet Pre-Trained Deep Learning Models and DINOv2 in Medical Imaging Classification," Feb. 13, 2024, *arXiv*: arXiv:2402.07595. doi: 10.48550/arXiv.2402.07595.

[41] M. M. Zahoor *et al.*, "Brain Tumor MRI Classification Using a Novel Deep Residual and Regional CNN," *Biomedicines*, vol. 12, no. 7, p. 1395, Jun. 2024, doi: 10.3390/biomedicines12071395.

[42] A. S. Sambyal, U. Niyaz, N. C. Krishnan, and D. R. Bathula, "Understanding calibration of deep neural networks for medical image classification," *Computer Methods and Programs in Biomedicine*, vol. 242, p. 107816, Dec. 2023, doi: 10.1016/j.cmpb.2023.107816.

[43] L. Yang, T. Wang, J. Zhang, S. Kang, S. Xu, and K. Wang, "Deep learning–based automatic segmentation of meningioma from T1-weighted contrast-enhanced MRI for preoperative meningioma differentiation using radiomic features," *BMC Med Imaging*, vol. 24, no. 1, p. 56, Mar. 2024, doi: 10.1186/s12880-024-01218-3.

[44] M. R. Shoaib *et al.*, "Improving brain tumor classification: An approach integrating pre-trained CNN models and machine learning algorithms," *Heliyon*, vol. 11, no. 10, p. e33471, May 2025, doi: 10.1016/j.heliyon.2024.e33471.

[45] P. S. Tejashwini, J. Thriveni, and K. R. Venugopal, "EBT Deep Net: Ensemble brain tumor Deep Net for multi-classification of brain tumor in MR images," *Biomedical Signal Processing and Control*, vol. 95, p. 106312, Sep. 2024, doi: 10.1016/j.bspc.2024.106312.

Vol. 7, No. 2, April 2026, Page. 1515-1534
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5494

[46] M. A. Gómez-Guzmán *et al.*, "Enhanced Multi-Class Brain Tumor Classification in MRI Using Pre-Trained CNNs and Transformer Architectures," *Technologies*, vol. 13, no. 9, p. 379, Aug. 2025, doi: 10.3390/technologies13090379.

[47] M. Nickparvar, "Brain Tumor MRI Dataset," *Kaggle*, 2021. [Online]. Available: https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset. [Accessed: Dec. 03, 2025].