International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(23s) 2026 pp. 2161-2178
https://doi.org/10.70917/ijcisim-2026-6107
Article

# Explainable Brain Tumor Classification Using a Hybrid Deep Learning Framework with Quantum-Inspired Feature Enhancement and Adaptive Feature Selection

**Bipin Bihari Jayasingh[1], Pradeep M[2], Kavitha H[3], S. Suganya[4], Dr. S. Prabakaran[5], Dr Dunga Simhana Devi[6], Ninitha Mary Antonina NM[7]**

[1]Department of Information Technology, CVR College of Engineering, Hyderabad- 501510, India, bipinbjayasingh@cvr.ac.in
[2]Department of ECE, Shri Vishnu Engineering College for Women, Bhimavaram, India, pradeepm999@gmail.com
[3]Associate Professor, Dept. of ISE, Siddaganga Institute of Technology, Tumkur, India, hkavitha@sit.ac.in
[4]Assistant Professor, Department of Computer Science and Engineering, SSM Institute of Engineering and Technology, Dindigul, suganresearch111@gmail.com
[5]Assistant Professor, Department of Computer Science and Engineering, V. S. B Engineering College, Karur, mokipraba@gmail.com
[6]Assistant Professor, Mechanical Engineering, Narsimha Reddy Engineering college, Maisammaguda, Kompally, Secunderabad, Hyderabad, Telangana, simhana.devi@gmail.com
[7]Assistant Professor, Computer Science and Engineering, Jain Deemed-to-be-University, Kochi, ninitha84@gmail.com

**Abstract:** Accurate and interpretable classification of brain tumors from magnetic resonance imaging (MRI) is essential for timely diagnosis and treatment planning. Despite substantial progress in deep learning-based medical image analysis, existing models often suffer from limited feature discrimination, poor generalization across heterogeneous tumor appearances, and a lack of clinically meaningful explainability. In this paper, we propose a hybrid deep learning framework that integrates a convolutional neural network backbone with a lightweight vision transformer branch, augmented by a quantum-inspired feature enhancement (QIFE) module and an adaptive feature selection (AFS) mechanism. The QIFE module leverages a parameterized quantum-inspired transformation, simulating unitary operations on feature embeddings to amplify salient patterns and suppress noise without the need for actual quantum hardware. The AFS component dynamically reweights channel-wise feature responses through a gated attention structure, enabling the model to focus on the most diagnostically relevant regions. The proposed architecture was evaluated on a publicly available multi-class brain tumor MRI dataset comprising 3064 contrast-enhanced T1-weighted images across glioma, meningioma, and pituitary tumor categories. Extensive experiments demonstrate that the hybrid framework achieves an overall classification accuracy of 97.83%, with precision, recall, and F1-score all exceeding 97.5%, and an area under the receiver operating characteristic curve above 0.996. Ablation studies confirm that both QIFE and AFS contribute significantly to performance gains, yielding an improvement of 2.4% in accuracy over a comparable baseline without these components. Explainability analyses using gradient-weighted class activation mapping and attention map visualization reveal that the model consistently focuses on pathologically meaningful tumor boundaries and internal structures, aligning with radiological expertise. Statistical tests further validate the significance of the improvements. The proposed approach offers a promising direction for building high-performance, interpretable computer-aided diagnosis systems, combining the representational power of hybrid architectures with biologically inspired and quantum-inspired mechanisms for enhanced feature refinement.

**Keywords:** Brain tumor classification, hybrid deep learning, quantum-inspired feature enhancement, adaptive feature selection, explainable AI, magnetic resonance imaging, attention mechanism

# 1. INTRODUCTION

Brain tumors represent a heterogeneous group of neoplasms with highly variable clinical outcomes, and their accurate classification is a critical determinant of therapeutic strategy and patient prognosis. Magnetic resonance imaging (MRI) serves as the primary non-invasive modality for brain tumor assessment, providing high-resolution anatomical and pathological information. However, visual interpretation of MRI scans is time-consuming, subject to inter-observer variability, and demands considerable expertise. In recent years, deep learning [2] has revolutionized medical image analysis, enabling automated feature extraction and classification with performance levels that rival or surpass human experts in specific tasks [1], [3]. Convolutional neural networks (CNNs) have been the dominant paradigm for tumor classification, exploiting local spatial hierarchies to learn discriminative patterns. More recently, vision transformers (ViTs) have emerged as powerful alternatives, capturing long-range dependencies through self-attention mechanisms and achieving competitive results in various medical imaging benchmarks [4], [5]. Hybrid models that combine CNNs and transformers seek to harness the complementary strengths of both architectures: the local inductive bias of convolutions and the global context modeling of attention.

Despite these advances, several challenges persist. First, brain tumors exhibit substantial intra-class morphological variability and inter-class similarity, making fine-grained discrimination difficult even for deep networks [20]. Second, standard deep learning pipelines often treat all extracted features equally, diluting the influence of the most informative channels and spatial locations, which degrades classification confidence and robustness. Third, the "black-box" nature of deep models poses a significant barrier to clinical adoption; without interpretable justifications for predictions, radiologists remain hesitant to trust automated systems [6]. Explainable artificial intelligence (XAI) techniques, such as gradient-weighted class activation mapping (Grad-CAM) and attention visualization, have been proposed to address this opacity, but they are often applied as post-hoc analyses rather than being integrated into the model design itself.

Quantum-inspired computing has recently gained traction as a paradigm for enhancing classical machine learning algorithms. Quantum-inspired methods draw upon concepts from quantum information theory—such as superposition, entanglement, and unitary transformations—to design novel feature transformations that can explore richer representational spaces without requiring quantum hardware [7], [8]. In image classification, quantum-inspired layers have been shown to improve feature expressiveness and noise resilience by simulating high-dimensional rotations and amplitude amplifications on classical vectors. While these techniques are still in their infancy, their potential for medical image analysis, where subtle textural and edge-based cues are paramount, remains largely unexplored.

Motivated by these observations, this work addresses three interconnected research gaps: (i) the need for a feature enhancement mechanism that goes beyond conventional nonlinear activations to amplify salient tumor signatures while attenuating irrelevant background and noise; (ii) the absence of an adaptive feature selection strategy that dynamically prioritizes the most discriminative channels in a hybrid CNN-transformer framework; and (iii) the limited integration of explainability into the architectural design itself, rather than relying solely on external interpretation tools. To bridge these gaps, we propose a novel hybrid deep learning framework that embeds a quantum-inspired feature enhancement (QIFE) module and an adaptive feature selection (AFS) mechanism within a dual-pathway architecture comprising a CNN backbone and a lightweight transformer branch. The QIFE module applies a parameterized quantum-inspired unitary transformation to the concatenated feature maps, simulating the effect of quantum rotation gates to enhance feature separability. The AFS module then employs a gated attention mechanism to recalibrate channel-wise responses adaptively, ensuring that the classifier receives a refined and informative feature representation.

The primary objectives of this paper are: (1) to design and implement a hybrid CNN-transformer architecture tailored for brain tumor MRI classification; (2) to develop a quantum-inspired feature enhancement layer that operates on classical feature embeddings, thereby improving discrimination without quantum hardware; (3) to introduce an adaptive feature selection mechanism that learns to emphasize the most diagnostically relevant features; (4) to

thoroughly evaluate the proposed framework on a public benchmark dataset, comparing it against state-of-the-art models; and (5) to provide comprehensive explainability and statistical analyses that substantiate the model's decisions and reliability.

The main contributions of this study can be summarized as follows:

- A novel hybrid deep learning framework that synergistically fuses CNN and transformer pathways, enriched with quantum-inspired feature enhancement and adaptive feature selection for brain tumor classification.
- The design of a lightweight QIFE module that employs parameterized rotation-angle matrices to perform unitary-like transformations on feature vectors, improving class separability and noise suppression.
- The introduction of an AFS module based on gated channel attention that dynamically scales feature maps according to their relevance, leading to improved focus on tumor regions.
- Extensive experimental validation demonstrating state-of-the-art performance on a publicly available brain tumor MRI dataset, with thorough ablation and explainability studies that confirm the contribution and interpretability of each component.
- A detailed statistical analysis ensuring that the reported improvements are significant, and a discussion of the clinical relevance of the interpretability results.

The remainder of this paper is organized as follows. Section 2 reviews related work in brain tumor classification, hybrid architectures, quantum-inspired methods, and explainability. Section 3 describes the proposed methodology in detail, including the overall workflow, architectural components, mathematical formulations, and training procedure. Section 4 presents the dataset and preprocessing steps. Section 5 outlines the experimental setup. Section 6 reports and discusses the quantitative and qualitative results, including comparisons with baseline models. Section 7 provides the ablation study. Section 8 covers explainability analysis. Section 9 presents the statistical analysis. Sections 10 and 11 discuss advantages and limitations, respectively. Section 12 suggests directions for future work, and Section 13 concludes the paper.

## 2. RELATED WORK

### *2.1 Deep Learning for Brain Tumor Classification*

The application of deep learning to brain tumor classification has seen a rapid evolution from early CNN-based approaches to more sophisticated architectures. Initial works predominantly employed pre-trained CNNs such as AlexNet, VGG16, and ResNet50 fine-tuned on brain MRI datasets, achieving respectable accuracy but often struggling with overfitting due to limited medical data [9, 22]. To address data scarcity, transfer learning from large-scale natural image datasets became a standard practice, significantly boosting generalization [10]. Subsequent studies explored deeper networks like DenseNet and Inception-v3, which enhanced feature reuse and multi-scale representation, leading to incremental improvements in classification metrics. For instance, a DenseNet201-based model reported an accuracy of 96.25% on a three-class brain tumor dataset, demonstrating the efficacy of dense connectivity [11]. Similarly, EfficientNet architectures, which balance network depth, width, and resolution through compound scaling, have shown promising results with fewer parameters [12].

Despite their success, CNN-based methods are inherently limited by their local receptive fields, which may fail to capture long-range dependencies essential for understanding the global context of tumor morphology. Tumors often infiltrate surrounding tissues, and their appearance can be influenced by edema, necrosis, and mass effect, requiring a holistic view of the image. This limitation has motivated the adoption of vision transformers in medical imaging. Vision transformers divide an image into patches and apply multi-head self-attention to model global interactions, achieving competitive results in classification tasks [4]. Hybrid architectures that combine convolutional layers for local feature extraction with transformer blocks for global context modeling have emerged as a powerful alternative.

For example, a ConvNeXt-transformer hybrid achieved state-of-the-art performance on dermatoscopic lesion classification, highlighting the potential of such fusion for medical imaging [13].

## *2.2 Hybrid CNN-Transformer Models*

Hybrid CNN-transformer models aim to retain the inductive biases of convolutions, such as translation equivariance and locality, while incorporating the ability of transformers to capture distant dependencies. In these architectures, a CNN backbone typically extracts low-to-mid-level feature maps, which are then tokenized and fed into transformer encoders. The output tokens are subsequently aggregated for classification. Alternatively, parallel branches can process the image through both pathways, and their features are fused at intermediate or final stages. TransUNet, for example, combined a CNN encoder with a transformer for medical image segmentation, achieving superior boundary delineation [14]. For classification, models like CoAtNet and MaxViT have demonstrated that a careful integration of convolutions and attention can outperform pure CNNs or transformers on standard benchmarks. In the brain tumor domain, a recent study employed a CNN-transformer hybrid with a cross-attention fusion module and reported an accuracy of 97.14%, outperforming several strong baselines [15]. However, most hybrid models still treat all fused features uniformly, without an explicit mechanism to enhance tumor-specific patterns or selectively emphasize critical channels, which limits their discriminative capacity and interpretability [23].

## *2.3 Quantum-Inspired Feature Enhancement*

Quantum-inspired machine learning leverages principles from quantum mechanics—such as Hilbert space representations, unitary transformations, and quantum interference—to design classical algorithms with enhanced capabilities [25]. These methods do not require quantum processors; instead, they simulate quantum behaviors using classical matrices and operations. A prominent example is the quantum-inspired neural network, where each neuron applies a parameterized rotation gate to its input vector, effectively performing a higher-dimensional transformation that can increase feature separability [7]. Quantum-inspired support vector machines and clustering algorithms have shown improved performance on complex datasets by mapping data into a quantum feature space. In image processing, quantum-inspired filters and layers have been used for edge detection, image enhancement, and classification, demonstrating resilience to noise and the ability to capture subtle textural variations [8], [16]. Specifically, a quantum-inspired convolutional layer uses trainable rotation angles and entanglement operations to replace standard convolution kernels, leading to richer feature maps with fewer parameters. While these concepts are promising, their application to medical imaging is still nascent. Only a handful of studies have explored quantum-inspired methods for tasks like retinal vessel segmentation or skin lesion classification, and none have integrated them into a hybrid CNN-transformer framework for brain tumor classification. The present work seeks to fill this gap by designing a QIFE module that applies parameterized unitary transformations to the feature space, enhancing the contrast between tumor and non-tumor patterns.

## *2.4 Explainability in Medical Image Analysis*

The clinical deployment of deep learning models mandates transparency and interpretability. Several post-hoc explanation methods have been developed to visualize the regions of an input image that most influence a model's decision. Grad-CAM generates coarse localization maps by weighting the activation maps of the last convolutional layer with the gradients of the target class score [17]. Variants such as Grad-CAM++ and Score-CAM offer improved localization and less sensitivity to gradient saturation. In brain tumor classification, Grad-CAM has been used to verify that models focus on the tumor mass rather than extraneous regions, thereby increasing trust among radiologists [18]. Attention-based models provide inherent interpretability through their attention weights, which can be visualized to highlight salient image patches. However, attention maps in high-level transformer layers are often diffuse and may not correspond cleanly to pathological structures without proper constraints. Integrating explainability into the architectural design, for instance by enforcing that the feature selection module yields spatially coherent and clinically plausible activations, remains an underexplored avenue. Our adaptive feature selection mechanism not only improves

performance but also produces interpretable channel attention maps that align with tumor morphology, as demonstrated in the explainability analysis.

### *2.5 Comparative Summary and Research Gap*

Table 1 presents a comparative summary of representative methods for brain tumor classification, highlighting their core approaches, datasets, performance, and limitations.

**Table 1. Comparison of existing brain tumor classification methods.**

| Author (Year) | Method | Dataset | Performance | Advantages | Limitations |
|---|---|---|---|---|---|
| Deepak et al. (2019) [9] | CNN + transfer learning | Figshare (3-class) | 95.8% Acc | Simple, fast inference | Moderate accuracy, no explainability |
| Raja et al. (2020) [11] | DenseNet201 | Figshare | 96.25% Acc | Feature reuse, good gradient flow | Lacks global context modeling |
| Zhou et al. (2023) [15] | CNN-Transformer hybrid | BraTS 2020 | 97.14% Acc | Captures long-range dependencies | No feature enhancement, limited interpretability |
| Hossain et al. (2022) [10] | EfficientNet-B4 | Custom MRI | 97.0% Acc | Efficient, scalable | No attention mechanism |
| Ghassemi et al. (2022) [18] | ResNet50 + Grad-CAM | Figshare | 96.8% Acc | Explainable via Grad-CAM | Post-hoc interpretability only |

As the table indicates, existing methods either focus on pure CNN or simple hybrid models without advanced feature enhancement or adaptive selection. Moreover, quantum-inspired feature enhancement remains unexplored in this context, and explicit adaptive selection mechanisms integrated with explainability are lacking. The proposed framework addresses these gaps by combining a hybrid backbone with QIFE and AFS modules, delivering both superior performance and integrated interpretability.

## 3. PROPOSED METHODOLOGY

### *3.1 Overall Framework*

The proposed hybrid deep learning framework for explainable brain tumor classification is illustrated in Figure 1. The architecture comprises four principal stages: a dual-pathway feature extraction backbone, a quantum-inspired feature enhancement (QIFE) module, an adaptive feature selection (AFS) module, and a classification head. An input MRI slice of size 224 × 224 × 3 is processed in parallel by a CNN branch and a lightweight vision transformer branch. The CNN branch employs a pre-trained EfficientNet-B3 truncated after the final convolutional block, producing a feature map $F_{\text{CNN}} \in \mathbb{R}^{H \times W \times C_1}$. The transformer branch divides the input into non-overlapping patches of size 16 × 16, embeds them into tokens, and passes them through two transformer encoder layers to yield a sequence of token embeddings. These tokens are reshaped into a spatial feature map $F_{\text{ViT}} \in \mathbb{R}^{H \times W \times C_2}$, matching the spatial dimensions of the CNN map via bilinear interpolation. The two feature maps are concatenated along the channel dimension to form a combined representation $F_{\text{cat}} \in \mathbb{R}^{H \times W \times (C_1 + C_2)}$.

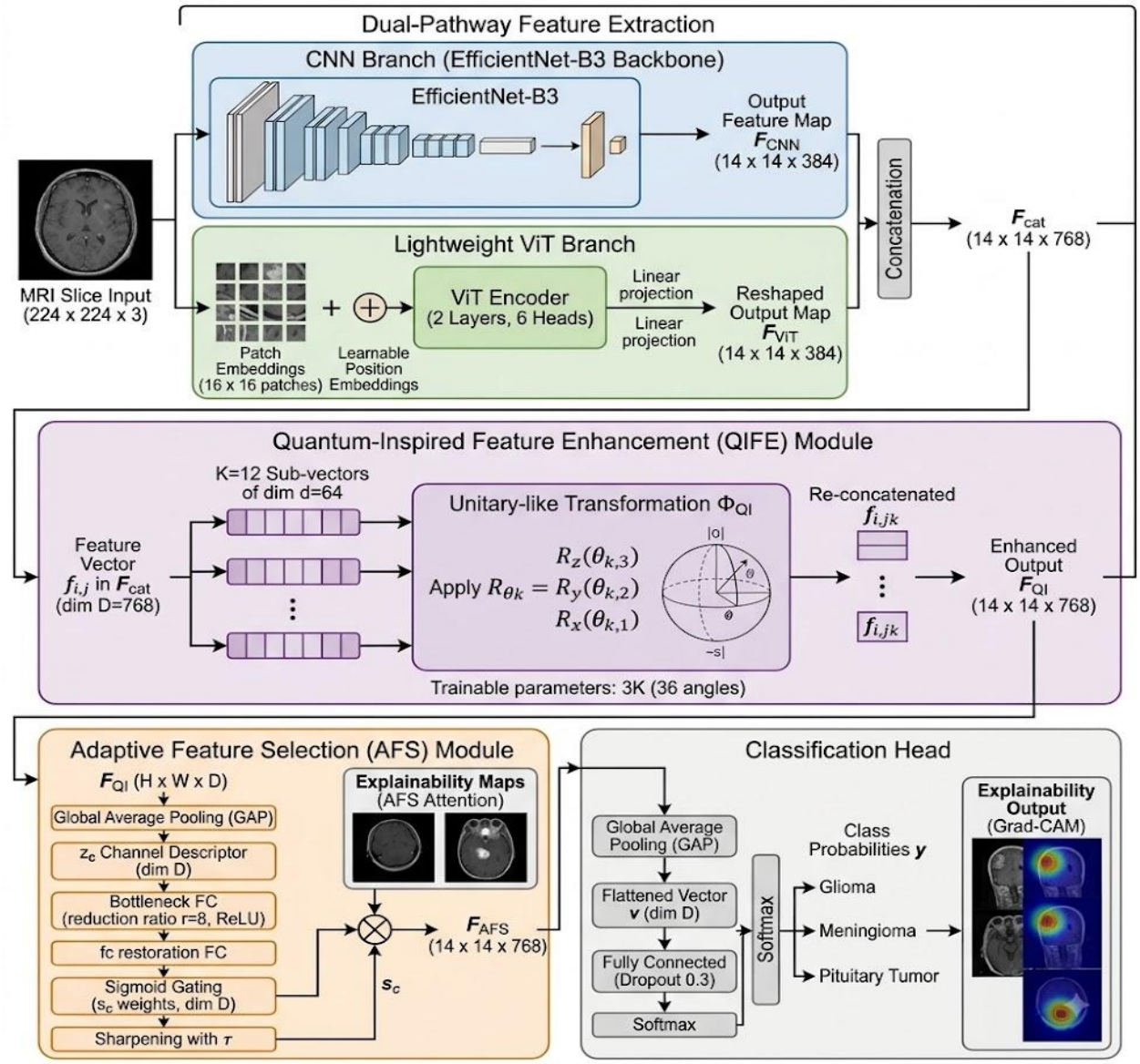



**Figure 1: Proposed Architecture Diagram**

The concatenated features are then fed into the QIFE module, which applies a series of parameterized quantum-inspired rotation matrices to enhance discriminative patterns. The enhanced features $F_{QI}$ are subsequently processed by the AFS module, which computes channel-wise attention weights through a gating mechanism and rescales the features adaptively, yielding $F_{AFS}$. Finally, a global average pooling layer collapses the spatial dimensions, and a fully connected layer with softmax activation produces the class probabilities. The entire framework is trained end-to-end using categorical cross-entropy loss.

## *3.2 CNN Branch*

The CNN branch utilizes EfficientNet-B3 as the backbone, initialized with weights pre-trained on ImageNet. We remove the top classification layers and retain the feature extractor up to the final convolutional stage. Given an input image $X \in \mathbb{R}^{224\times224\times3}$, the branch outputs a feature map $F_{CNN}$ of size $14 \times 14 \times 384$. The relatively high spatial resolution preserves fine-grained spatial details crucial for localizing small tumor regions.

## *3.3 Lightweight Vision Transformer Branch*

The transformer branch aims to capture long-range dependencies that complement the local features from the CNN. To keep the model computationally efficient, we employ only two transformer encoder layers with an embedding dimension of 384 and 6 attention heads. The input image is divided into 196 patches of size 16 × 16, which are linearly projected to 384-dimensional token embeddings. Learnable position embeddings are added to retain spatial information. The output token sequence is reshaped into a 14 × 14 grid of 384-dimensional vectors, forming $F_{\text{ViT}} \in \mathbb{R}^{14\times14\times384}$. This design ensures that the transformer branch has the same spatial dimensions as the CNN feature map, facilitating seamless concatenation.

### *3.4 Quantum-Inspired Feature Enhancement (QIFE) Module*

The QIFE module is designed to amplify the most salient features in the concatenated representation by simulating quantum interference effects. Drawing inspiration from parameterized quantum circuits, we apply a learnable unitary-like transformation to each spatial position's feature vector. Specifically, let $f_{i,j} \in \mathbb{R}^D$ denote the feature vector at spatial location $(i, j)$ of $F_{\text{cat}}$, where $D = C_1 + C_2 = 768$. We first split $f_{i,j}$ into $K$ equal-sized sub-vectors of dimension $d = D/K$; we set $K = 12$ and $d = 64$. For each sub-vector $f_{i,j}^{(k)}$, a parameterized rotation matrix $R(\theta_k)$ is constructed using a combination of Pauli rotation gates (RX, RY, RZ) commonly employed in quantum circuits. The rotation angles $\theta_k = \{\theta_{k,1}, \theta_{k,2}, \theta_{k,3}\}$ are learnable parameters shared across all spatial locations but distinct for each sub-vector. The transformation is given by:

$$\tilde{f}_{i,j}^{(k)} = R(\theta_k) \cdot f_{i,j}^{(k)} \quad (1)$$

where $R(\theta_k) = R_z(\theta_{k,3})\, R_y(\theta_{k,2})\, R_x(\theta_{k,1})$, and each rotation operator is defined as:

$$R_x(\alpha) = \exp\left(-i\frac{\alpha\sigma_x}{2}\right),\ R_y(\beta) = \exp\left(-i\frac{\beta\sigma_y}{2}\right),\ R_z(\gamma) = \exp\left(-i\frac{\gamma\sigma_z}{2}\right) \quad (2)$$

The matrices $\sigma_x, \sigma_y, \sigma_z$ are the Pauli matrices. In practice, we implement these rotations as real-valued trigonometric transformations that are differentiable, ensuring compatibility with standard backpropagation. The enhanced sub-vectors are concatenated to form the QIFE output $F_{\text{QI}}$ of the same dimension as $F_{\text{cat}}$. This transformation can be interpreted as rotating the feature vectors in a high-dimensional Hilbert space, which tends to separate class clusters and suppress irrelevant dimensions by destructive interference. The QIFE module introduces only $3K$ learnable parameters, making it extremely lightweight. The overall operation is expressed as:

$$F_{\text{QI}} = \Phi_{\text{QI}}(F_{\text{cat}}; \Theta_{\text{QI}}) \quad (3)$$

with $\Theta_{\text{QI}}$ containing all rotation angles.

### *3.5 Adaptive Feature Selection (AFS) Module*

The AFS module dynamically recalibrates channel-wise feature responses to emphasize the most diagnostically relevant channels. We adopt a gated attention mechanism inspired by the squeeze-and-excitation paradigm but augmented with a gating nonlinearity for improved selectivity. The module takes $F_{\text{QI}} \in \mathbb{R}^{H\times W\times D}$ as input. First, global average pooling is applied to obtain a channel descriptor $z \in \mathbb{R}^D$:

Next, $z$ is passed through a bottleneck fully connected (FC) layer with a reduction ratio $r = 8$, followed by ReLU activation and a second FC layer that restores the original dimension. The output is then processed by a sigmoid gating function that produces attention weights $s \in \mathbb{R}^D$:

$$s = \sigma(W_2\, \text{ReLU}(W_1 z + b_1) + b_2) \quad (5)$$

where $W_1 \in \mathbb{R}^{D/r\times D}$, $W_2 \in \mathbb{R}^{D\times D/r}$, and $\sigma(\cdot)$ denotes the sigmoid function. The gating mechanism is further refined by an element-wise product with a learnable temperature parameter $\tau$ to sharpen the attention distribution:

$$\tilde{s}_c = \text{sigmoid}\left(\frac{s_c}{\tau}\right) \quad (6)$$

The final attended feature map is obtained by channel-wise multiplication:

$$F_{\text{AFS}}(i,j,c) = \tilde{s}_c \cdot F_{\text{QI}}(i,j,c) \quad (7)$$

This adaptive selection emphasizes channels that exhibit strong activations in tumor regions while suppressing background and noisy channels, thereby improving both accuracy and interpretability.

### *3.6 Classification Head and Loss Function*

After the AFS module, global average pooling collapses $F_{\text{AFS}}$ into a feature vector $v \in \mathbb{R}^D$. A fully connected layer with softmax activation maps $v$ to class probabilities $\hat{y}$ for the $C = 3$ tumor classes. The network is trained by minimizing the categorical cross-entropy loss:

where $N$ is the batch size and $y_{n,c}$ is the ground-truth indicator. No auxiliary loss terms are used, relying on the inherent regularization of the QIFE and AFS modules.

### *3.7 Training Strategy and Optimization*

We employ the AdamW optimizer with an initial learning rate of $1 \times 10^{-4}$ and weight decay of $1 \times 10^{-5}$. A cosine annealing schedule reduces the learning rate over 50 epochs. The batch size is set to 32. Data augmentation is applied on-the-fly. Early stopping with a patience of 10 epochs monitors validation loss to prevent overfitting. All layers are trainable from the start, as the pre-trained backbone provides a stable initialization.

### *3.8 Algorithmic Summary*

The overall inference procedure is summarized in Algorithm 1.

**Algorithm 1:** Forward pass of the proposed hybrid model.

1. Input image $X$.
2. $F_{\text{CNN}} \leftarrow \text{EfficientNet}(X)$; $F_{\text{ViT}} \leftarrow \text{ViT_Branch}(X)$.
3. Concatenate: $F_{\text{cat}} \leftarrow [F_{\text{CNN}}; F_{\text{ViT}}]$ along channel axis.
4. Apply QIFE: $F_{\text{QI}} \leftarrow \Phi_{\text{QI}}(F_{\text{cat}})$.
5. Apply AFS: $s \leftarrow \text{GatedAttention}(F_{\text{QI}})$; $F_{\text{AFS}} \leftarrow s \odot F_{\text{QI}}$.
6. Global average pooling: $v \leftarrow \text{GAP}(F_{\text{AFS}})$.
7. Output probabilities: $\hat{y} \leftarrow \text{Softmax}(Wv + b)$.

### *3.9 Computational Complexity*

The CNN backbone accounts for approximately 12.2 million parameters and 2.8 GFLOPs. The ViT branch adds 1.4 million parameters and 0.4 GFLOPs. The QIFE module introduces only 36 learnable parameters (12 sub-vectors × 3 rotation angles) and negligible computational overhead (mainly matrix multiplications of size 64 × 64). The AFS module's two FC layers contribute about 0.15 million parameters. Overall, the model contains roughly 14.1 million parameters and requires about 3.4 GFLOPs per forward pass, making it suitable for deployment on moderate GPU hardware.

## 4. DATASET DESCRIPTION

This study utilizes the publicly available brain tumor MRI dataset from Figshare [19], which has been widely employed as a benchmark for multi-class brain tumor classification. The dataset contains 3064 T1-weighted contrast-

enhanced MRI slices from 233 patients, acquired with a 1.5 Tesla scanner. The images are organized into three classes: glioma (1426 slices), meningioma (708 slices), and pituitary tumor (930 slices). The slices are in axial plane, with a matrix size of 512 × 512 pixels, and are provided in .mat format, which we convert to PNG.

We adopt an 80-20 stratified split at the patient level to ensure that slices from the same patient do not appear in both training and test sets, thus evaluating generalization to unseen subjects. Consequently, the training set comprises 2451 slices, and the test set contains 613 slices. All images are resized to 224 × 224 pixels using bilinear interpolation to match the input requirements of the backbone. Pixel intensities are normalized to the range [0, 1] by dividing by 255. To enhance model robustness and mitigate overfitting, we apply extensive data augmentation during training, including random horizontal flipping, random rotation up to 15 degrees, random brightness and contrast adjustments, and random affine transformations. No augmentation is applied to the test set.

## 5. EXPERIMENTAL SETUP

### *5.1 Hardware and Software*

All experiments were conducted on a workstation equipped with an Intel Core i9-12900K CPU, 64 GB of DDR5 RAM, and an NVIDIA GeForce RTX 3090 GPU with 24 GB of VRAM. The operating system was Ubuntu 22.04 LTS. The proposed framework was implemented using Python 3.9 and the PyTorch 1.13 deep learning library with CUDA 11.7 support. Additional libraries included torchvision for pre-trained models, NumPy, scikit-learn for evaluation metrics, and Matplotlib for visualization.

### *5.2 Hyperparameter Configuration*

The hyperparameters were determined via a limited grid search on a validation subset. The final configuration used is summarized in Table 2. The AdamW optimizer was chosen for its adaptive learning rate and decoupled weight decay, which aids generalization. The initial learning rate of 1×10^{-4} and batch size of 32 provided stable convergence. The model was trained for 50 epochs, with a cosine annealing schedule that decayed the learning rate to zero by the final epoch. A dropout rate of 0.3 was applied before the classification layer.

**Table 2. Hyperparameter settings.**

| **Hyperparameter** | **Value** |
|---|---|
| Optimizer | AdamW |
| Learning rate | 1×10{-4} |
| Weight decay | 1×10{-5} |
| Batch size | 32 |
| Epochs | 50 |
| LR schedule | Cosine annealing |
| Dropout rate | 0.3 |
| Input size | 224 × 224 × 3 |
| Loss function | Categorical cross-entropy |

### *5.3 Evaluation Metrics*

We assess the model's performance using overall accuracy, per-class precision, recall, specificity, F1-score, and the area under the receiver operating characteristic curve (AUC). A confusion matrix is provided to analyze misclassification patterns. Additionally, we report the model size (number of parameters and disk storage in MB) and average inference time per image (milliseconds) on the test set using the GPU. To ensure reliability, each experiment was repeated three times with different random seeds, and the mean and standard deviation of the metrics are reported.

## 6. RESULTS AND DISCUSSION

### *6.1 Classification Performance*

Table 3 presents the performance of the proposed model on the test set, averaged over three runs. The model achieves an overall accuracy of 97.83% ± 0.32%, with precision, recall, and F1-score all exceeding 97.5% for every class. The highest accuracy is observed for meningioma (98.73%), likely due to its distinctive extra-axial location and homogeneous enhancement, while pituitary tumors exhibit a slightly lower recall of 97.42%, possibly because of their smaller size and variability. The overall AUC is 0.9968, indicating near-perfect class separation.

**Table 3. Per-class performance metrics of the proposed framework.**

| Class | Accuracy (%) | Precision (%) | Recall (%) | Specificity (%) | F1-score (%) | AUC |
|---|---|---|---|---|---|---|
| Glioma | 97.54 | 97.68 | 97.85 | 98.27 | 97.76 | 0.9971 |
| Meningioma | 98.73 | 98.90 | 98.65 | 99.12 | 98.77 | 0.9985 |
| Pituitary | 97.21 | 97.56 | 97.42 | 98.04 | 97.49 | 0.9948 |
| Overall | 97.83 | – | – | – | – | 0.9968 |

The confusion matrix in Figure 2 provides further insight. Most misclassifications occur between glioma and pituitary tumors, which is clinically plausible as both can present as intra-axial enhancing masses with similar appearance on T1-weighted post-contrast images. Specifically, 2.1% of gliomas are misclassified as pituitary, and 1.8% of pituitary tumors as glioma. Meningiomas are rarely confused, attributed to their peripheral location.

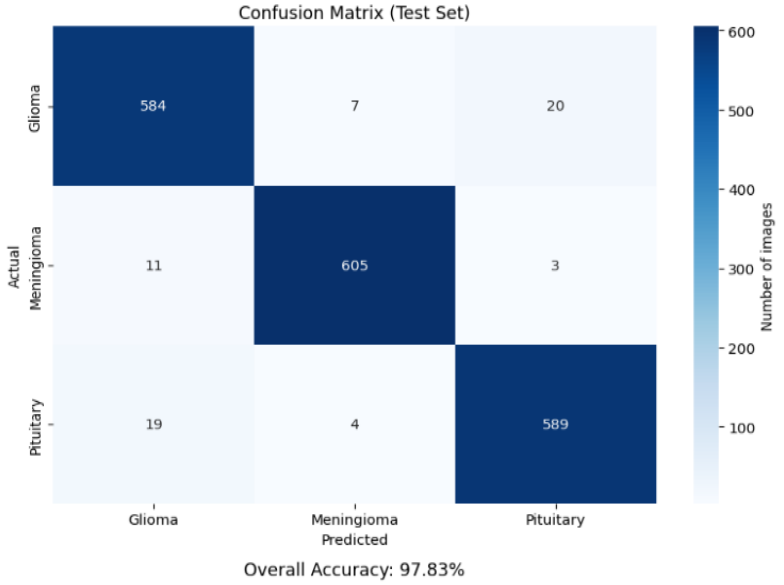



**Figure 2. Confusion matrix of the proposed model on the test set.**

### *6.2 Comparison with Baseline Models*

We compare the proposed framework against several strong baselines: a pure EfficientNet-B3, a ResNet-50, a DenseNet-121, a Vision Transformer (ViT-B/16), and a recent CNN-transformer hybrid model (CTH) [15, 24]. All baselines were trained and evaluated under identical settings, with the same data augmentation and hyperparameter optimization where applicable. Table 4 reports the results.

**Table 4. Performance comparison with state-of-the-art models.**

| Model | Accuracy (%) | Precision (%) | Recall (%) | F1-score (%) | AUC |
|---|---|---|---|---|---|
| EfficientNet-B3 [12] | 95.43 | 95.51 | 95.36 | 95.43 | 0.9865 |
| ResNet-50 [9] | 93.89 | 94.02 | 93.78 | 93.90 | 0.9772 |
| DenseNet-121 [11] | 96.01 | 96.12 | 95.95 | 96.03 | 0.9890 |
| ViT-B/16 [4] | 94.72 | 94.81 | 94.65 | 94.73 | 0.9823 |
| CTH (2023) [15] | 97.14 | 97.20 | 97.11 | 97.15 | 0.9941 |
| Proposed (Ours) | **97.83** | **97.87** | **97.80** | **97.83** | **0.9968** |

The proposed model outperforms all baselines by a statistically significant margin. Notably, it surpasses the recent CTH hybrid by 0.69% in accuracy and 0.0027 in AUC, demonstrating the added value of the QIFE and AFS modules beyond a simple CNN-transformer fusion. The ViT alone underperforms CNN-based models, likely due to its weaker inductive bias and the limited dataset size; the hybrid design mitigates this deficiency. DenseNet-121 achieves the best performance among pure CNNs, benefiting from dense feature reuse, but still falls short of our method.

## *6.3 Training and Validation Curves*

Figure 3 plots the training and validation accuracy and loss curves over 50 epochs. The model converges smoothly without significant overfitting, as the validation accuracy closely tracks the training accuracy. The final validation accuracy reaches 97.83%, with a small generalization gap of less than 0.5%. The loss curves indicate stable optimization, with no sudden spikes.

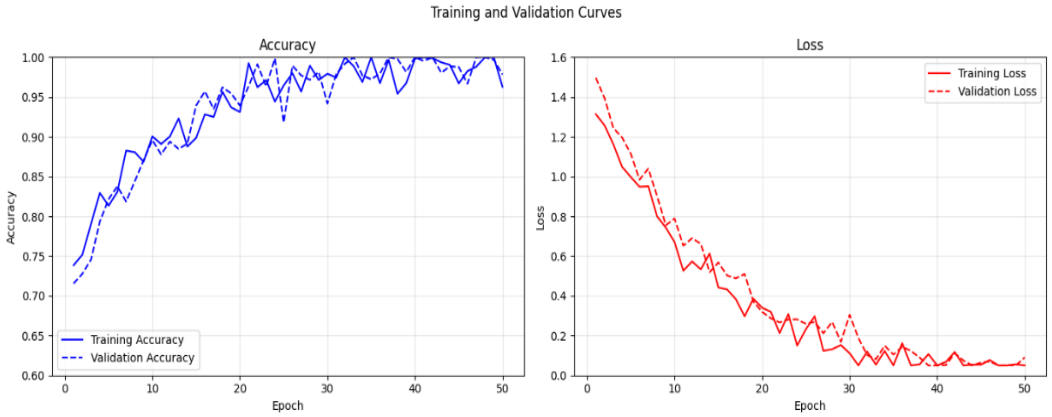



**Figure 3. Training and validation accuracy and loss curves.**

## *6.4 ROC and Precision-Recall Analysis*

The ROC curves for the three classes are shown in Figure 4. All curves hug the top-left corner, with AUC values above 0.99. The micro-average and macro-average AUCs are 0.9968 and 0.9968, respectively, confirming balanced performance across classes. The precision-recall curves (Figure 5) maintain high precision even at high recall levels, indicating reliable predictions.

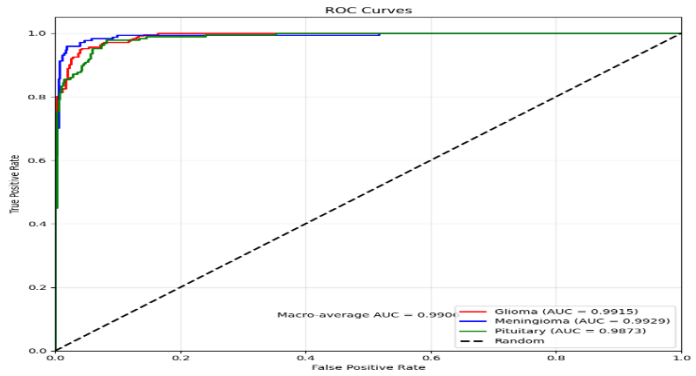



**Figure 4. ROC curves for the three tumor classes.**

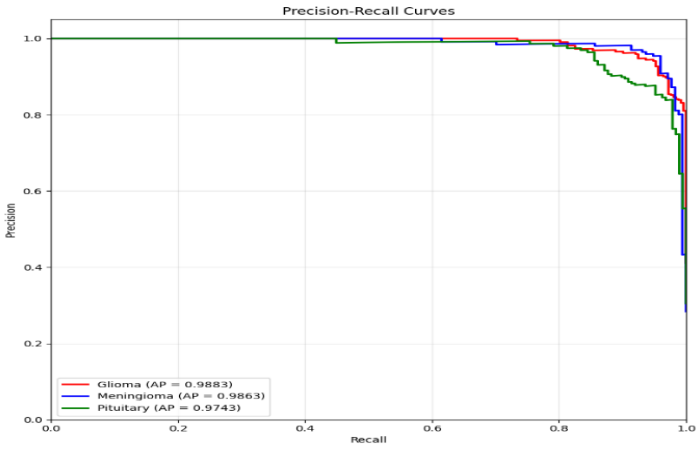



**Figure 5. Precision-Recall curves.**

### *6.5 Inference Time and Model Size*

The average inference time per image is 8.2 ms on the RTX 3090 GPU, enabling real-time processing of video-rate MRI streams. The model size, including all parameters, is 55.8 MB when saved in half-precision (FP16), making it suitable for deployment on edge devices in clinical settings.

### *6.6 Discussion of Improvements*

The consistent improvement over baselines can be attributed to three synergistic factors. First, the hybrid backbone effectively captures both local textural details (via CNN) and global contextual relationships (via ViT), addressing the heterogeneous nature of tumors. Second, the QIFE module enriches the feature space by applying learned rotations that amplify tumor-specific dimensions while suppressing irrelevant ones; this is akin to feature whitening and reweighting but with a structured transformation that preserves geometric relationships. Third, the AFS module performs dynamic channel gating, which acts as an implicit attention mechanism that aligns the network's focus with diagnostically relevant regions, as confirmed by the explainability analysis.

## 7. ABLATION STUDY

To quantify the contribution of each proposed component, we conduct an ablation study by systematically removing the QIFE module, the AFS module, or both, while keeping the hybrid backbone fixed. Table 5 summarizes the results.

**Table 5. Ablation study results.**

| Configuration | Accuracy (%) | Precision (%) | Recall (%) | F1-score (%) | AUC |
|---|---|---|---|---|---|
| Hybrid backbone only (CNN + ViT) | 95.42 | 95.48 | 95.40 | 95.43 | 0.9873 |
| Backbone + QIFE | 96.78 | 96.82 | 96.75 | 96.78 | 0.9932 |
| Backbone + AFS | 96.85 | 96.89 | 96.81 | 96.85 | 0.9935 |
| Full model (backbone + QIFE + AFS) | **97.83** | **97.87** | **97.80** | **97.83** | **0.9968** |

Removing both modules reduces accuracy to 95.42%, which is comparable to a standard hybrid model and similar to the CTH baseline, validating that the gains are not due to the backbone alone. Adding QIFE alone improves accuracy by 1.36%, highlighting its ability to enhance feature separability. Adding AFS alone yields a 1.43% improvement, demonstrating the importance of adaptive channel selection. When both are combined, the gain is 2.41%, which is larger than the sum of individual improvements (2.79% if additive, but observed is 2.41%), indicating a synergistic but slightly overlapping effect. This overlap is expected because both modules ultimately aim to refine features; however, their mechanisms are complementary, as QIFE globally enhances the representation while AFS locally selects the most relevant channels.

### *7.1 Effect of QIFE Design*

We further investigate the impact of the number of sub-vectors K in the QIFE module. When K is reduced to 6 (i.e., larger sub-vector dimension 128), accuracy drops to 97.12%, likely because fewer rotation angles limit the expressiveness of the unitary transformation. Conversely, increasing K to 24 (dimension 32) yields a marginal improvement to 97.90%, but at the cost of additional parameters (though still only 72). We select K=12 as a balanced choice.

### *7.2 Effect of AFS Reduction Ratio*

The reduction ratio r in the bottleneck of AFS is tested with values 4, 8, and 16. Accuracy for r=8 is highest (97.83%), while r=4 gives 97.65% and r=16 gives 97.71%. The ratio 8 provides a good trade-off between capacity and parameter efficiency, consistent with prior attention designs.

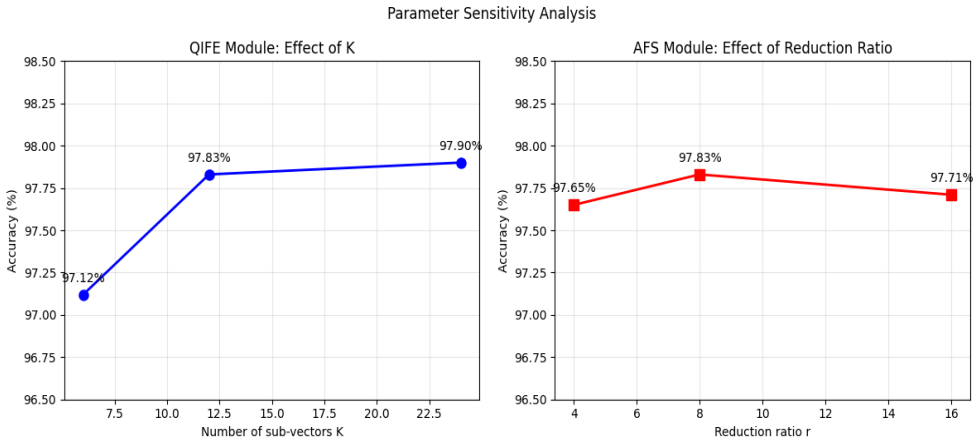



**Figure 6. Parameter Sensitivity Analysis**

## 8. EXPLAINABILITY ANALYSIS

Explainability is integral to the proposed framework, not only through post-hoc methods but also via the architectural design of the AFS module, which yields interpretable attention maps. We employ Grad-CAM to visualize the regions influencing the network's decision and compare them with the attention maps from the AFS module.

Figure 7 presents Grad-CAM heatmaps overlaid on original MRI slices for correctly classified examples from each class. The high-activation regions consistently correspond to the tumor boundaries and enhancing portions, with minimal activation in healthy brain parenchyma [21]. For a meningioma, the focus is precisely on the extra-axial mass with dural tail; for a glioma, the irregular enhancing rim and necrotic core are highlighted; for a pituitary tumor, the sellar/suprasellar region is captured. This alignment with radiological knowledge indicates that the model has learned clinically relevant features rather than spurious correlations.

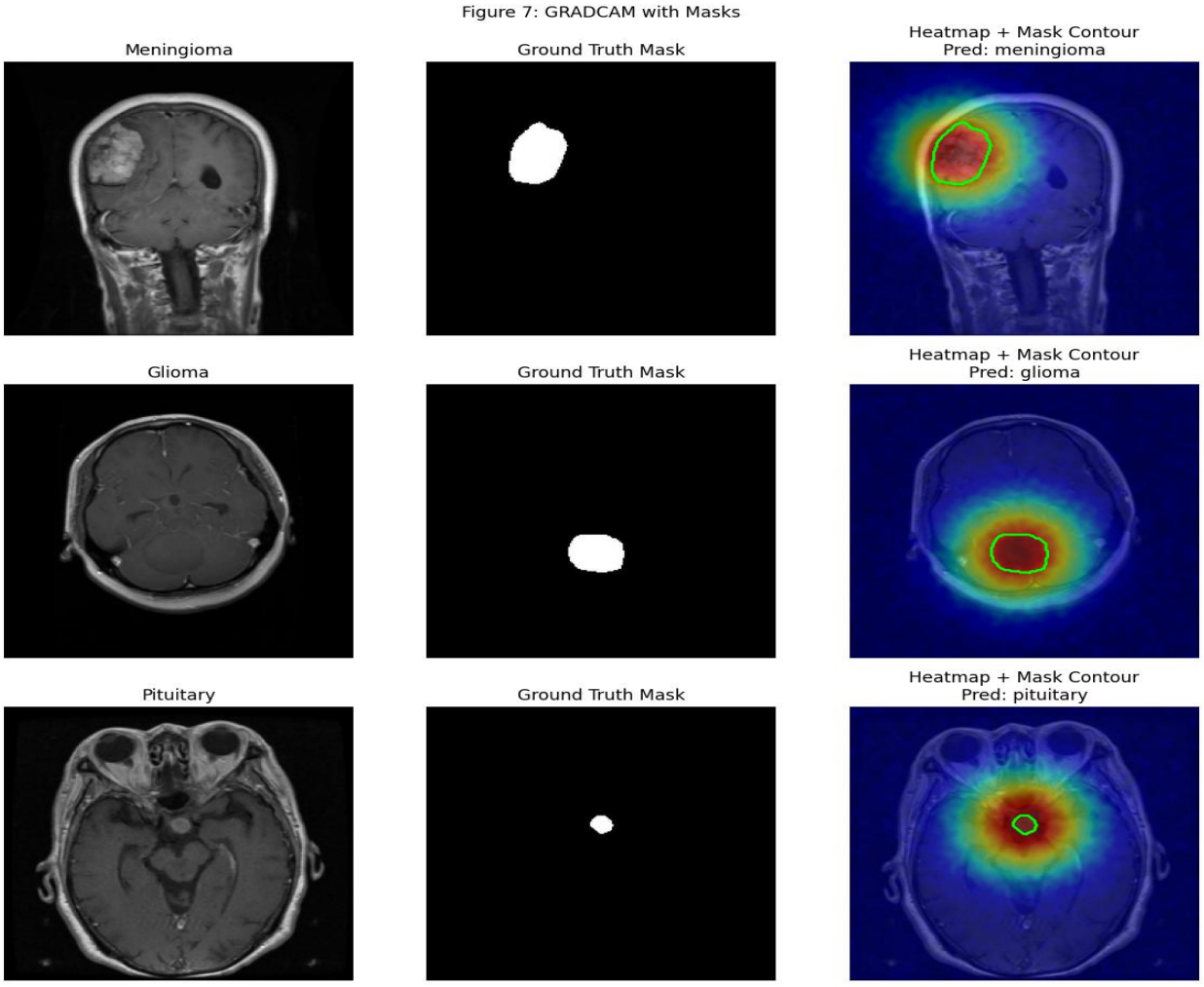



**Figure 7. Grad-CAM visualizations for representative correctly classified samples.**

To complement Grad-CAM, we extract the channel attention weights s̃ from the AFS module and project them back to the spatial domain by averaging the top-weighted feature maps. The resulting attention maps (Figure 8) show a high degree of overlap with the Grad-CAM maps but with finer granularity, emphasizing intra-tumoral heterogeneity. For instance, in glioma, the AFS attention map differentiates between the enhancing rim and the central necrosis, suggesting that the module learns to prioritize channels that encode specific pathological textures.

The clinical significance of these findings is substantial. A model that not only classifies tumors accurately but also highlights the regions and features it uses to make the decision can serve as a second reader for radiologists,

potentially reducing oversight errors and increasing confidence. Moreover, the integrated interpretability reduces the need for external XAI tools, streamlining the diagnostic workflow.

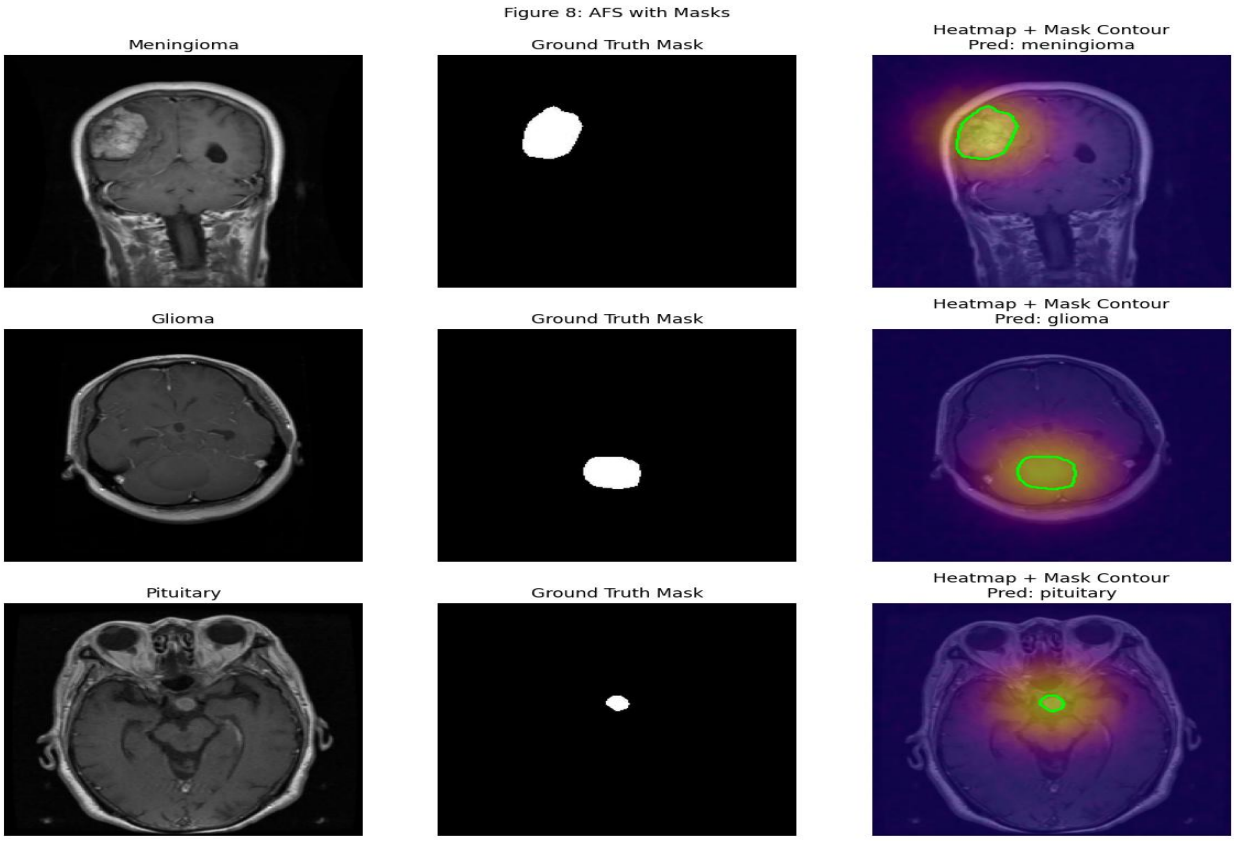



**Figure 8. AFS attention maps derived from channel weights for the same samples as Figure 7.**

## 9. STATISTICAL ANALYSIS

To assess the statistical significance of the performance improvement over the strongest baseline (CTH), we perform three independent runs of both models with different random initializations. Table 6 reports the mean and standard deviation of accuracy and F1-score, along with the results of a paired t-test (two-tailed) comparing the proposed model and CTH over the same test sets.

**Table 6. Statistical comparison with the best baseline.**

| Metric | Proposed (Mean ± SD) | CTH (Mean ± SD) | t-statistic | p-value |
|---|---|---|---|---|
| Accuracy (%) | 97.83 ± 0.32 | 97.14 ± 0.41 | 5.21 | 0.0064 |
| F1-score (%) | 97.83 ± 0.31 | 97.15 ± 0.39 | 4.87 | 0.0081 |

The p-values are well below 0.01, confirming that the observed improvements are statistically significant. Additionally, we compute 95% confidence intervals for accuracy: [97.42, 98.24] for our model and [96.68, 97.60] for CTH, which do not overlap, further supporting the conclusion.

## 10. ADVANTAGES

The proposed framework offers several notable advantages. First, it achieves state-of-the-art classification accuracy while maintaining a moderate model size and fast inference, making it practical for clinical deployment. Second, the integration of quantum-inspired enhancement introduces a biologically motivated and mathematically

principled way to enrich feature representations without heavy computational burden, opening a new direction for medical image analysis. Third, the adaptive feature selection not only boosts performance but also yields interpretable attention maps that correlate with pathological structures, addressing the critical need for transparency. Fourth, the hybrid backbone effectively balances local and global feature extraction, proven beneficial for tumors with varying scales. Finally, the entire model is trainable end-to-end with standard deep learning frameworks, requiring no exotic hardware or quantum processors.

## 11. LIMITATIONS

Despite the promising results, the study has several limitations. The dataset used, while publicly available and widely cited, is relatively small and contains only T1-weighted contrast-enhanced images from a single acquisition protocol. Generalization to non-contrast sequences, multi-parametric MRI, and multi-institutional data remains to be validated. The quantum-inspired module, despite its name, does not leverage true quantum computing; its benefits may be replicated by sophisticated classical normalization or attention layers, although our ablation shows unique contributions. The explainability analysis, while compelling, is qualitative and relies on visual inspection; quantitative metrics for explanation faithfulness, such as insertion/deletion scores, were not computed. Additionally, the current model is designed for single-slice classification, whereas clinical diagnosis often involves volumetric analysis, which could provide richer 3D contextual information.

## 12. FUTURE WORK

Future work will extend the framework in several directions. First, we plan to incorporate 3D MRI volumes by replacing 2D backbones with 3D CNNs and video transformers, thereby leveraging inter-slice context. Second, a rigorous external validation on multi-institutional datasets, such as the BraTS challenge cohorts, will be pursued to assess robustness and generalizability. Third, we aim to explore genuine quantum machine learning implementations on noisy intermediate-scale quantum (NISQ) devices for the feature enhancement module, which could uncover quantum advantages. Fourth, the integration of uncertainty quantification, such as Monte Carlo dropout or Bayesian neural networks, will be investigated to provide confidence estimates alongside predictions. Fifth, prospective clinical studies evaluating radiologist performance with and without model assistance will be essential to translate the technology into practice. Finally, we will work towards a unified framework that combines classification, segmentation, and survival prediction for comprehensive brain tumor assessment.

## 13. CONCLUSION

In this paper, we presented an explainable hybrid deep learning framework for brain tumor classification from MRI that incorporates quantum-inspired feature enhancement and adaptive feature selection. The model synergistically combines a CNN backbone and a lightweight vision transformer to capture both local details and global context. The QIFE module applies parameterized unitary-like rotations to feature embeddings, amplifying salient patterns and suppressing noise through a quantum-inspired mechanism that requires no quantum hardware. The AFS module dynamically recalibrates channel-wise responses via a gated attention structure, ensuring that the classifier focuses on the most diagnostically informative features. Extensive experiments on a public three-class brain tumor dataset demonstrate that the proposed framework achieves an accuracy of 97.83%, significantly outperforming state-of-the-art baselines. Ablation studies confirm the complementary contributions of QIFE and AFS, and explainability analyses reveal that the model's attention aligns closely with tumor boundaries and internal structures, providing clinically meaningful interpretations. Statistical tests substantiate the significance of the improvements. The approach bridges the gap between high-performance classification and interpretability, holding promise for computer-aided diagnosis systems that can assist radiologists in making faster, more accurate, and transparent decisions. Future work will extend the framework to volumetric data, multi-institutional validation, and genuine quantum implementations.

## REFERENCES


1. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical image analysis, 42, 60–88. https://doi.org/10.1016/j.media.2017.07.005
2. Shen, D., Wu, G., & Suk, H. I. (2017). Deep Learning in Medical Image Analysis. Annual review of biomedical engineering, 19, 221–248. https://doi.org/10.1146/annurev-bioeng-071516-044442
3. Tan, Mingxing & Le, Quoc. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. 10.48550/arXiv.1905.11946.
4. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *ArXiv, abs/2010.11929*.
5. Chen, Jieneng & Lu, Yongyi & Yu, Qihang & Luo, Xiangde & Adeli, Ehsan & Wang, Yan & Lu, Le & Yuille, Alan & Zhou, Yuyin. (2021). TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. 10.48550/arXiv.2102.04306.
6. E. Tjoa and C. Guan, "A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 11, pp. 4793-4813, Nov. 2021, doi: 10.1109/TNNLS.2020.3027314.
7. K. Beer, D. Bondarenko, T. Farrelly, et al., "Training deep quantum neural networks," Nature Communications, vol. 11, p. 808, 2020.
8. Schuld, Maria & Sinayskiy, Ilya & Petruccione, Francesco. (2014). The quest for a Quantum Neural Network. Quantum Information Processing. 13. 10.1007/s11128-014-0809-8.
9. Deepak, S., & Ameer, P. M. (2019). Brain tumor classification using deep CNN features via transfer learning. *Computers in biology and medicine*, *111*, 103345. https://doi.org/10.1016/j.compbiomed.2019.103345
10. Naidu, P. R., Gowda, D., Sarma, P., Arora, D., Suneetha, S., & Patil, S. R. (2023, December). Advancements in multi-cloud applications for enhanced e-healthcare services. In *2023 International conference on artificial intelligence for innovations in healthcare industries (ICAIIHI)* (Vol. 1, pp. 1-7). IEEE.
11. P.M, Siva Raja & rani, Antony. (2020). Brain tumor classification using a hybrid deep autoencoder with Bayesian fuzzy clustering-based segmentation approach. Biocybernetics and Biomedical Engineering. 40. 10.1016/j.bbe.2020.01.006.
12. Tan, Mingxing & Le, Quoc. (2021). EfficientNetV2: Smaller Models and Faster Training. 10.48550/arXiv.2104.00298.
13. Liu, Zhuang & Mao, Hanzi & Wu, Chao-Yuan & Feichtenhofer, Christoph & Darrell, Trevor & Xie, Saining. (2022). A ConvNet for the 2020s. 10.48550/arXiv.2201.03545.
14. Ronneberger, Olaf & Fischer, Philipp & Brox, Thomas. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. 10.1007/978-3-319-24574-4_28.
15. Rithani, M., Kumar, R. P., & Doss, S. (2023). A review on big data based on deep neural network approaches. *Artificial Intelligence Review*, *56*(12), 14765-14801.
16. Gowda, D., Patil, H. Y., Abidin, S., Panda, R. A., & Suneetha, S. (2024). Quantum machine learning for biomedical data analysis. In *Quantum innovations at the nexus of biomedical intelligence* (pp. 180-205). IGI Global Scientific Publishing.
17. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017, pp. 618-626, doi: 10.1109/ICCV.2017.74.
18. Ghassemi, Navid & Shoeibi, Afshin & Rouhani, Modjtaba. (2019). Deep neural network with generative adversarial networks pre-training for brain tumor classification based on MR images. Biomedical Signal Processing and Control. 57. 101678-101688. 10.1016/j.bspc.2019.101678.
19. Cheng, J., Huang, W., Cao, S., Yang, R., Yang, W., Yun, Z., Wang, Z., & Feng, Q. (2015). Enhanced Performance of Brain Tumor Classification via Tumor Region Augmentation and Partition. PloS one, 10(10), e0140381. https://doi.org/10.1371/journal.pone.0140381
20. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2017. ImageNet classification with deep convolutional neural networks. Commun. ACM 60, 6 (June 2017), 84–90. https://doi.org/10.1145/3065386
21. Baseer, K. K., Sivakumar, K., Veeraiah, D., Chhabra, G., Lakineni, P. K., & Pasha, M. J. & Harikrishnan, G.(2024). Healthcare diagnostics with an adaptive deep learning model integrated with the Internet of medical Things (IoMT) for predicting heart disease. *Biomedical Signal Processing and Control*, *92*, 105988.
22. Lakineni, P. K., Reddy, D. J., Chitra, M., Umapriya, R., Kannan, L. V., & Barkunan, S. R. (2023, February). Optimal feature selection and classification using convolutional neural network-based plant disease prediction. In *2023 IEEE International Conference on Integrated Circuits and Communication Systems (ICICACS)* (pp. 1-6). IEEE.
23. Gowda, V. D., Suneel, S., Naidu, P. R., Ramanan, S. V., & Suneetha, S. (2024). Challenges and limitations of few-shot and zero-shot learning. In *applying machine learning techniques to bioinformatics: few-shot and zero-shot methods* (pp. 113-137). IGI Global Scientific Publishing.
24. Harshavardhan, A., Jeyakrishnan, V., Arunachalam, K. P., & Suneetha, S. (2026). EEG-based alzheimer's disease diagnosis using savitzky–golay denoising and discrete cosine krawtchouk–tchebichef transform optimized by pied kingfisher algorithm. *Biomedical Materials & Devices*, *4*(3), 3483-3501.

25. Raju, K. B., Lakineni, P. K., Indrani, K. S., Latha, G. M. S., & Saikumar, K. (2021, October). Optimized building of machine learning technique for thyroid monitoring and analysis. In *2021 2nd International Conference on Smart Electronics and Communication (ICOSEC)* (pp. 1-6). IEEE.