The Eastasouth Journal of Information System and Computer Science
Vol. 2, No. 02, December 2024, pp. 323-344
ISSN: 3025-566X, DOI: 10.58812/esiscs.v2i02

Hybrid Artificial Intelligence for Brain Tumor Classification Using MRI Images

Kallol Chakraborty Shekhor[1], Sumayya Jahan[2], Md Sahid Hossain[3], Md Fakrul Alam[4], Abdur Rahim[5], Md Sazzad Hossain Shehan[6], Md Anwar Hossain[7], Md Abedur Rahman[8], Mohammed Adnan[9]

[1,4] Master of Science in Information Studies, Trine University, Allen Park, Michigan, USA
[2,7,8] Department of Computer Science, Maharishi International University, Fairfield, Iowa, USA
[3] Senior Software Engineer, Prime Tech Solutions Ltd., Dhaka, Bangladesh
[5,6] Bachelor of Science in Computer Science and Engineering, BRAC University, Dhaka, Bangladesh
[9] Department of Computer Science and Engineering, VIT-AP University, India

Article Info

Article history:

Received Dec, 2024
Revised Dec, 2024
Accepted Dec, 2024

Keywords:

Artificial Intelligence;
Brain Tumor Classification;
ConvNeXt Tiny;
Magnetic Resonance Imaging;
Swin Transformer;
Wavelet Convolutional Neural Network

ABSTRACT

Accurate classification of brain tumors from magnetic resonance images is essential for supporting clinical assessment, although variations in tumor morphology, anatomical location, image orientation, and visual similarity among diagnostic categories can limit the effectiveness of conventional deep learning models. This study proposes a hybrid learning framework that integrates local structural information, global contextual relationships, and frequency domain characteristics for four class brain tumor classification. The model combines an ImageNet pretrained ConvNeXt Tiny branch for local feature extraction, a Swin Transformer Tiny branch for hierarchical contextual representation, and a wavelet convolutional neural network branch based on two level Haar decomposition. The resulting branch features are aligned within a shared representation space and integrated through bidirectional cross attention and adaptive gated fusion before classification. A dataset containing 10,464 brain magnetic resonance images was evaluated at the individual image level, including 2,726 glioma, 2,737 meningioma, 2,300 no tumor, and 2,701 pituitary tumor images. The dataset was divided into 9,417 training images, 653 validation images, and 394 independent testing images. Brain contour cropping, contrast limited adaptive histogram equalization, normalization, and controlled augmentation were applied during preprocessing. On the independent test set, the proposed model correctly classified 389 of 394 images, achieving an accuracy of 98.73% with a 95% confidence interval from 97.06% to 99.46%. It further obtained a macro precision of 98.84%, macro recall of 98.82%, macro F1 score of 98.82%, macro specificity of 99.56%, Matthews correlation coefficient of 0.9830, and macro area under the receiver operating characteristic curve of 0.9995. The proposed framework outperformed the strongest individual backbone, Swin Transformer Tiny, which achieved 96.45% accuracy and a macro F1 score of 96.63%. These findings indicate that coordinated integration of spatial, contextual, and frequency information can improve image level brain tumor classification. However, the absence of patient identifiers and magnetic resonance sequence metadata limits patient level interpretation and sequence specific analysis.

Corresponding Author:

Name: Mohammed Adnan
Institution: Department of Computer Science, Maharishi International University, Fairfield, Iowa, USA
Email: drmohammedadnan110@gmail.com

1. INTRODUCTION

Brain tumors represent a clinically important group of intracranial abnormalities characterized by substantial variation in tissue origin, biological behavior, anatomical location, and radiological appearance [1-2]. Accurate characterization of these abnormalities is essential for treatment planning, surgical assessment, prognosis estimation, and longitudinal monitoring [3-6]. Magnetic resonance imaging is widely used for brain tumor assessment because it provides detailed soft tissue contrast and enables visualization of tumor boundaries, surrounding structures, edema, necrosis, and mass effects. Nevertheless, interpretation of brain magnetic resonance images remains challenging because different tumor categories may exhibit overlapping visual characteristics, while tumors belonging to the same category may vary considerably in size, shape, intensity, texture, and anatomical distribution. Glioma, meningioma, and pituitary tumors are among the commonly investigated categories in automated brain tumor classification research [7-11]. Gliomas arise from glial tissues and may demonstrate infiltrative growth, heterogeneous intensity, irregular boundaries, and variable internal composition. Meningiomas are generally associated with the meninges and often appear as relatively well defined extra axial masses, although their appearance can vary across images and orientations. Pituitary tumors are located near the sellar region and may be difficult to recognize when their size is limited or when surrounding anatomical structures obscure the lesion [12]. Images without visible tumors form an additional clinically relevant class because a reliable classification system must distinguish pathological findings from normal or non-tumor appearances. These visual differences and similarities make four class brain tumor classification a complex pattern recognition problem [13-14].

Conventional machine learning methods have been applied to brain image analysis through manually designed features describing intensity, texture, shape, edges, and frequency characteristics. Although such methods can provide useful information, their effectiveness depends heavily on feature engineering and may decline when the imaging conditions, tumor morphology, or anatomical orientations differ from those represented during model development [15-17]. Deep convolutional neural networks have reduced the dependence on handcrafted representations by learning hierarchical spatial features directly from images. Architectures such as VGG, ResNet, DenseNet, MobileNet, EfficientNet, and ConvNeXt have demonstrated strong feature learning capabilities in medical image classification. However, convolutional operations primarily emphasize local neighborhoods and may not fully represent relationships between distant regions of an image. Transformer based architectures provide a complementary mechanism by modeling interactions among spatially separated image regions [18-19]. The Swin Transformer introduces hierarchical representations and localized window attention, allowing contextual information to be learned at multiple spatial resolutions. Such representations can support the analysis of tumors whose appearance depends on both local tissue characteristics and broader anatomical context. Nevertheless, transformer-based models may require substantial training data and can be less sensitive to certain fine structural patterns when used independently [20-22]. A single feature extraction mechanism may therefore be insufficient to capture all relevant information contained in brain magnetic resonance images.

Frequency domain information provides another potentially valuable representation for brain tumor classification. Tumor boundaries, texture irregularities, tissue transitions, and internal heterogeneity can produce distinctive low frequency and high frequency responses. Haar wavelet decomposition separates an image into approximation and directional detail components, which can reveal complementary information related to intensity structure, horizontal variation, vertical variation, and diagonal variation. Although wavelet features have been incorporated into medical image analysis, they are frequently processed independently or combined with spatial features through direct concatenation. Such simple integration may not adequately model the relationships among local spatial features, global contextual representations, and frequency characteristics.

Another limitation of many multibranch classification systems is the use of fixed or uniform feature fusion. The diagnostic importance of a feature branch may vary according to the characteristics of an individual image. For example, local texture and boundary information may be particularly informative for an irregular glioma, whereas global anatomical context may be more important for a lesion located near the meninges or pituitary region. Frequency information may contribute more strongly when the lesion contains heterogeneous tissue transitions. An effective fusion mechanism should therefore support information exchange among feature branches while estimating their relative importance for each input image. To address these considerations, this study proposes a hybrid brain tumor classification framework that combines ConvNeXt Tiny, Swin Transformer Tiny, and a wavelet convolutional neural network. The ConvNeXt Tiny branch learns local structural representations and detailed tumor morphology. The Swin Transformer Tiny branch captures hierarchical contextual relationships and spatial dependencies. The frequency branch applies two level Haar wavelet decomposition and convolutional feature learning to represent approximation and directional detail information. The three branch representations are projected into a shared feature space and integrated through bidirectional cross attention. Adaptive gated fusion then estimates image dependent branch contributions before global feature aggregation and four class prediction.

The proposed framework is evaluated using 10,464 brain magnetic resonance images categorized as glioma, meningioma, no tumor, and pituitary tumor. The dataset contains axial, coronal, and sagittal views and is partitioned into 9,417 training images, 653 validation images, and 394 independent testing images. Brain contour cropping and contrast limited adaptive histogram equalization are applied to reduce unnecessary background and improve local contrast. Controlled augmentation is used during training to increase appearance variability while preserving clinically meaningful anatomical structure. The proposed model is compared with VGG16, MobileNetV3 Large, ResNet50, DenseNet121, EfficientNet B0, ConvNeXt Tiny, and Swin Transformer Tiny under a common four class evaluation protocol. The main contribution of this study is the development of a unified framework that jointly models local spatial structure, global contextual relationships, and frequency domain characteristics. A bidirectional cross attention mechanism is introduced to support coordinated information exchange among the three feature branches. An adaptive gated fusion module is further employed to determine the relative contribution of each representation according to the input image. In addition, the study provides class level and aggregate evaluation using accuracy, precision, recall, F1 score, specificity, Matthews correlation coefficient, kappa, and area under the receiver operating characteristic curve. The analysis is conducted at the individual image level because patient identifiers and magnetic resonance sequence metadata are unavailable. Consequently, the reported results should be interpreted as evidence of image classification performance

rather than patient level diagnostic performance.

2. RELATED WORK

Brain tumor classification has progressed from conventional clinical assessment and manually engineered image analysis toward automated deep learning methods. Modern diagnostic practice increasingly considers both radiological appearance and molecular information because tumors with similar visual characteristics may exhibit different biological behaviors. Van den Bent et al. [23] discussed the clinical implications of molecular markers such as IDH mutation and 1p/19q codeletion in the classification of diffuse gliomas. Their study emphasized that reliable brain tumor diagnosis requires the integration of imaging findings with molecular and pathological evidence. Artificial intelligence has also demonstrated considerable potential across broader healthcare applications. Manik et al. [24] investigated artificial intelligence powered predictive analytics for the early identification of chronic diseases and personalized clinical decision making. Miah et al. [25] explored the use of wearable health data and deep learning algorithms for continuous cardiovascular monitoring. Faruk et al. [26] proposed data science infrastructure for real time epidemiological surveillance and pandemic forecasting. These developments demonstrate the increasing importance of automated feature learning, predictive modelling, and data driven decision support in modern healthcare.

Early brain tumor classification approaches primarily relied on manually defined tumor regions and handcrafted features. Cheng et al. [27] proposed tumor region augmentation and ring-shaped partitioning for the classification of meningioma, glioma, and pituitary tumors from contrast enhanced magnetic resonance images. The authors demonstrated that tissues surrounding the tumor can provide useful diagnostic information and improve classification performance. Ismael and Abdel Qader [28] combined discrete wavelet transform features, Gabor filter responses, statistical descriptors, and a back propagation neural network. Their framework achieved promising results, but it required prior tumor region identification and manually designed feature extraction.

Amin et al. [29] introduced a feature fusion framework that combined histogram of oriented gradients, local binary patterns, and texture-based features. The fused features were subsequently selected and supplied to conventional classifiers. Ari and Hanbay [30] proposed an extreme learning machine based local receptive field approach for brain tumor classification and tumor region extraction. Although these methods demonstrated the usefulness of shape and texture information, their dependence on manually engineered features limited their capacity to learn complex and highly variable tumor representations directly from MRI data. The introduction of convolutional neural networks enabled automatic hierarchical feature extraction and reduced the need for manually designed descriptors. Paul et al. [31] evaluated fully connected and convolutional neural networks for classifying meningioma, glioma, and pituitary tumors from axial contrast enhanced MRI images. Their best network achieved an average fivefold cross validation accuracy of 91.43 percent. This study demonstrated that deep feature learning could outperform several specialized approaches based on tumor dilation and handcrafted regional analysis.

Seetha and Raja [32] developed a convolutional neural network using small convolution kernels and reported an accuracy of 97.5 percent for brain tumor classification. However, limited information was provided regarding patient independent data partitioning and repeated experimental evaluation. Abiwinanda et al. [33] employed a comparatively simple convolutional architecture consisting of convolution, pooling, flattening, and fully connected operations. Their model achieved a training accuracy of 98.51 percent and a validation accuracy of 84.19 percent. The substantial difference between training and validation performance indicated that limited medical

imaging datasets can lead to overfitting, particularly when shallow data partitioning procedures are employed. Artificial intelligence has also been applied to biomedical research beyond medical image classification. Nilima et al. [34] examined the use of machine learning and artificial intelligence in drug discovery, including compound prediction and pharmaceutical development. Such developments reinforce the broader relevance of automated representation learning and data driven modelling in biomedical research. Transfer learning has become particularly important for brain MRI classification because medical datasets are generally smaller than large natural image datasets. Deepak and Ameer [35] used a pretrained GoogLeNet model to extract deep features from MRI images. The extracted representations were classified using established machine learning classifiers. Their patient level fivefold cross validation experiment achieved a mean accuracy of approximately 98 percent for differentiating glioma, meningioma, and pituitary tumors. The authors also demonstrated that transfer learning remained useful when fewer training images were available. Swati et al. [36] proposed a block wise fine-tuning strategy for pretrained convolutional networks. Their model achieved an average accuracy of 94.82 percent under fivefold cross validation on a contrast enhanced MRI dataset. The study demonstrated that selective fine tuning can adapt pretrained visual representations to medical imaging while reducing the computational and data requirements associated with training deep networks from the beginning. Subsequent studies introduced modified and hybrid deep learning architectures to improve classification performance. Raza et al. [37] developed DeepTumorNet by modifying the final layers of GoogLeNet and incorporating additional convolutional layers with leaky rectified linear unit activation. The model achieved a reported accuracy of 99.67 percent for the classification of glioma, meningioma, and pituitary tumors. However, the reported results were obtained using a single public dataset, which limits conclusions regarding performance across different institutions and acquisition conditions.

Nayak et al. [38] proposed a dense EfficientNet framework for classifying glioma, meningioma, pituitary tumor, and no tumor images. Their method incorporated min max normalization, data augmentation, dense layers, and dropout regularization. The proposed model achieved a testing accuracy of 98.78 percent. This result demonstrated the effectiveness of compound network scaling and dense classification layers for multiclass brain tumor recognition. The integration of large-scale healthcare data has also been explored in cancer related applications. Rahman et al. [39] investigated big data and predictive analytics for early cancer detection and cost optimization. Their work highlighted the importance of scalable analytical systems, timely prediction, and the effective use of heterogeneous clinical information. Although their study did not focus specifically on brain MRI classification, it illustrates the growing role of predictive analytics in cancer diagnosis and healthcare management.

Hybrid deep networks have been developed for both binary and multiclass brain tumor classification. Amran et al. [40] proposed a hybrid deep tumor network by replacing the final layers of GoogLeNet with a customized convolutional structure. Their model achieved a reported accuracy of 99.51 percent for binary brain tumor classification using the Br35H dataset. Although the method produced strong results, binary tumor detection does not represent the more complex task of differentiating multiple tumor types. Kumar et al. [41] presented a deep neural network and transfer learning framework based on an improved ResNet50 model. Their method classified MRI images into benign and malignant categories and reported accuracies of 99.30 percent and 98.40 percent for the evaluated classes. Despite these promising findings, binary classification provides limited information for clinical scenarios that require differentiation among glioma, meningioma, pituitary tumor, and normal brain images. Islam et al. [42] proposed BrainNet, an optimized EfficientNet based architecture for brain tumor

classification. The framework was evaluated using 3,064 contrast enhanced MRI images and incorporated image preprocessing and augmentation. Among the evaluated EfficientNet configurations, EfficientNetB3 achieved the highest reported accuracy of 99.69 percent. The findings demonstrated the effectiveness of transfer learning and compound model scaling. Nevertheless, evaluation on a single benchmark dataset does not fully establish robustness across scanners, patient populations, and clinical institutions.

Machine learning has additionally been applied to population level healthcare analysis. Hossain et al. [43] investigated big data and machine learning models for analyzing and predicting the COVID 19 epidemic. Their study emphasized the importance of scalable computation, reliable data processing, and accurate predictive modelling in healthcare. Similar considerations are important for brain tumor classification, where high performance on a carefully curated dataset may not directly translate into reliable clinical performance. The reviewed studies demonstrate substantial progress in automated brain tumor classification. Nevertheless, several limitations remain. Many existing studies rely on a single public dataset, image level partitioning, or overall accuracy as the primary evaluation measure. Image level partitioning may allow MRI slices from the same patient to appear in both training and testing sets, leading to optimistic performance estimates. Furthermore, overall accuracy may conceal poor performance for minority tumor classes, particularly when the dataset is imbalanced.

Several methods also depend on handcrafted features, manually identified tumor regions, or computationally complex pretrained models. Other studies address only binary tumor detection and therefore do not evaluate clinically important distinctions among multiple tumor categories. Limited attention has also been given to class specific precision, recall, specificity, F1 score, receiver operating characteristic analysis, confusion patterns, and model stability.These limitations motivate the development of a robust multiclass brain tumor classification framework capable of learning both local tumor characteristics and global contextual representations directly from MRI images. The present study addresses these requirements through an integrated deep learning architecture evaluated on four categories comprising glioma, meningioma, pituitary tumor, and no tumor.

3. METHODOLOGY OVERVIEW

This study employed a brain magnetic resonance imaging dataset containing 10,464 images distributed across four diagnostic categories comprising glioma, meningioma, no tumor, and pituitary tumor. The dataset was divided into 9,417 training images, 653 validation images, and 394 independent testing images, with each image treated as an individual evaluation unit because patient identifiers were unavailable. The preprocessing procedure included OpenCV image reading, conversion from BGR to RGB, brain contour cropping, contrast enhancement using contrast limited adaptive histogram equalization, resizing to 224 × 224 pixels, tensor conversion, and normalization using the ImageNet mean and standard deviation. Controlled augmentation involving horizontal reflection, rotation, scaling, translation, and mild Gaussian noise was applied only to the training images. The proposed hybrid model integrated three complementary feature extraction branches. A ConvNeXt Tiny branch learned local structural patterns and tumor morphology, a Swin Transformer Tiny branch captured hierarchical spatial context and distant anatomical relationships, and a wavelet convolutional neural network extracted frequency information from two level Haar wavelet subbands. The resulting branch features were projected into a common representation space and integrated using bidirectional cross attention and adaptive gated fusion. Global average pooling and a classification head containing layer normalization, a dense layer, GELU activation, dropout, and a four-class output

layer were then used to generate the final prediction. ImageNet pretrained weights were used for the ConvNeXt Tiny and Swin Transformer Tiny branches, while the complete network was optimized using AdamW, multiclass cross entropy, cosine learning rate scheduling, gradient clipping, and early stopping. The proposed framework was compared with VGG16, MobileNetV3 Large, ResNet50, DenseNet121, EfficientNet B0, ConvNeXt Tiny, and Swin Transformer Tiny under the same dataset partition and evaluation protocol. Performance was assessed using accuracy, precision, recall, F1 score, specificity, Matthews correlation coefficient, Cohen kappa, and the area under the receiver operating characteristic curve.

3.1 Dataset Description

The study employed a brain magnetic resonance imaging dataset designed for four class image level classification. As summarized in Table 1, the dataset contains 10,464 images assigned to glioma, meningioma, no tumor, and pituitary tumor categories. All images were processed using an input resolution of 224 × 224 pixels with three color channels. The representative samples presented in Figure 1 demonstrate considerable variation in anatomical orientation, field of view, contrast, tumor location, lesion size, and surrounding tissue appearance. Axial, coronal, and sagittal images are visible in the montage, indicating that the classification framework was exposed to heterogeneous anatomical perspectives. However, magnetic resonance sequence metadata and patient identifiers were not supplied. Therefore, the study was formulated as individual image classification rather than patient level diagnosis, and no sequence specific analysis was conducted. The class composition is presented in Table 2. The dataset includes 2,300 no tumor images, representing 21.98 percent of all samples, 2,737 meningioma images, representing 26.16 percent, 2,726 glioma images, representing 26.05 percent, and 2,701 pituitary tumor images, representing 25.81 percent. The three tumor categories have highly comparable sample counts, while the no tumor category contains a moderately smaller number of images. Overall, the class distribution is relatively balanced compared with many medical imaging datasets, thereby reducing the likelihood that aggregate performance is dominated by a single majority category. Nevertheless, class level precision, recall, F1 score, and specificity were reported to ensure that differences in diagnostic performance remained visible across all four categories.

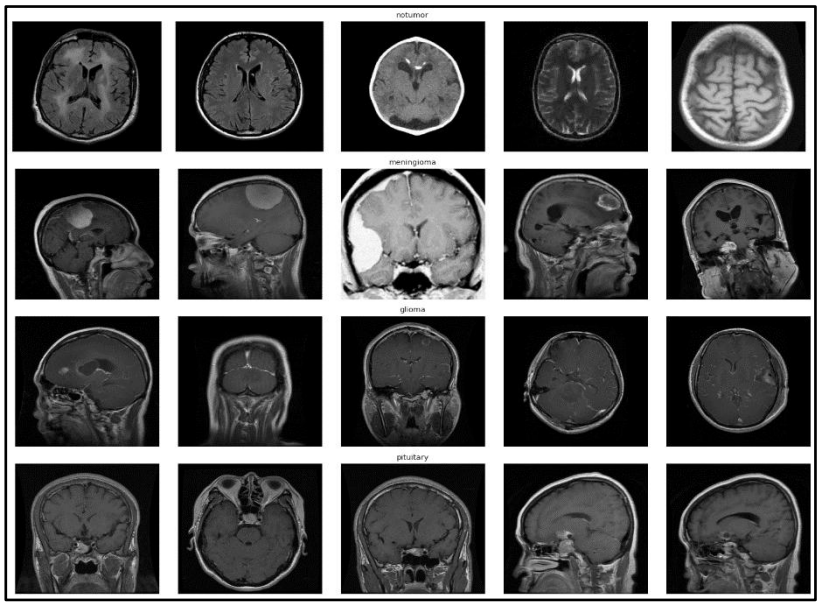

Figure 1. Representative Sample Images from the Brain MRI Dataset

The complete partitioning protocol is reported in Table 3. A total of 9,417 images, corresponding to 89.99 percent of the dataset, were allocated to training, while 653 images, corresponding to 6.24 percent, were used for validation, and 394 images, corresponding to 3.77 percent, formed the independent test set. The training set contained 2,053 no tumor, 2,452 meningioma, 2,456 glioma, and 2,456 pituitary tumor images. The validation set contained 142, 170, 170, and 171 images from the respective classes, while the test set contained 105 no tumor, 115 meningioma, 100 glioma, and 74 pituitary tumor images. Brain contour cropping and contrast limited adaptive histogram equalization were applied before model training to reduce unnecessary surrounding background and improve local tissue contrast. Since patient identifiers were unavailable, the partitions could not be verified as patient independent. This limitation should be considered when interpreting the reported test performance.

Table 1. Dataset overview

Characteristic	Description
Dataset type	Brain MRI image dataset
Classification task	Four class image level classification
Total number of images	10,464
Number of classes	4
Classes	Glioma, meningioma, no tumor, and pituitary tumor
Model input	224 × 224 × 3
Image orientations	Axial, coronal, and sagittal images are visible in the supplied sample montage
MRI sequence metadata	Not supplied
Patient identifiers	Not supplied
Preprocessing	Brain contour cropping and CLAHE contrast enhancement
Evaluation unit	Individual MRI image
Test set size	394 images
Validation set size	653 images
Training set size	9,417 images

Table 2. Class distribution

Class	Number of images	Percentage
No tumor	2,300	21.98%
Meningioma	2,737	26.16%
Glioma	2,726	26.05%
Pituitary tumor	2,701	25.81%
Total	10,464	100.00%

Table 3. Internally consistent data partition

Class	Training	Validation	Testing	Total
No tumor	2,053	142	105	2,300
Meningioma	2,452	170	115	2,737
Glioma	2,456	170	100	2,726
Pituitary tumor	2,456	171	74	2,701
Total	**9,417**	**653**	**394**	**10,464**
Percentage	**89.99%**	**6.24%**	**3.77%**	**100.00%**

3.2 Data Preprocessing and Augmentation

The preprocessing pipeline was designed to standardize image appearance while preserving diagnostically relevant anatomical and tumor characteristics. As summarized in Table 4, each image was first read in color format using OpenCV and converted

from BGR to RGB to ensure compatibility with the expected channel ordering of the pretrained feature extractors. Brain contour cropping was then applied to remove unnecessary surrounding background and to increase the relative spatial prominence of the intracranial region. Contrast limited adaptive histogram equalization was subsequently used to improve local contrast and enhance subtle tissue intensity differences without excessively amplifying noise. All images were resized to 224 × 224 pixels, converted into PyTorch tensors, and normalized using the ImageNet mean and standard deviation. This normalization strategy was selected because the ConvNeXt Tiny and Swin Transformer Tiny branches were initialized with ImageNet pretrained parameters.

Training images were further processed using controlled augmentation to improve robustness against moderate variations in orientation, position, scale, and intensity. As reported in Table 4, random horizontal flipping was applied with a probability of 0.50, while random rotation was restricted to 15 degrees. Scaling was performed within a range of 0.90 to 1.10, and translation was limited to 5 percent of the image dimensions to avoid substantial anatomical displacement. Mild Gaussian noise was introduced to simulate limited intensity variation and reduce sensitivity to minor acquisition related differences. Vertical flipping was not applied because it could generate anatomically unrealistic brain orientations. No stochastic augmentation was used for the validation and test sets. Their transformations were limited to resizing, tensor conversion, and ImageNet based normalization to ensure deterministic and unbiased performance evaluation.

Table 4. Preprocessing and augmentation protocol

Operation	Configuration
Image reading	OpenCV color image reading
Color conversion	BGR to RGB
Brain contour cropping	Removal of unnecessary surrounding background
Contrast enhancement	CLAHE
Image resizing	224 × 224 pixels
Tensor conversion	PyTorch tensor
Normalization	ImageNet mean and standard deviation
Random horizontal flipping	Probability 0.50
Random rotation	Within 15 degrees
Random scaling	Scale range from 0.90 to 1.10
Random translation	Maximum 5 percent of image dimension
Gaussian noise	Mild intensity variation
Vertical flipping	Not applied
Test transformation	Resize, tensor conversion, and normalization only

3.3 Proposed Hybrid Model Architecture

a. Complementary Multibranch Feature Extraction

The proposed architecture was designed to capture complementary spatial, contextual, and frequency characteristics from each brain MRI image. As illustrated in Figure 2 and summarized in Table 5, a preprocessed image with dimensions of 224 × 224 × 3 is simultaneously provided to three parallel feature extraction branches. These branches consist of a ConvNeXt Tiny based local feature branch, a Swin Transformer Tiny based global context branch, and a wavelet convolutional neural network-based frequency feature branch. This parallel design allows the model to examine the same image from distinct representational

perspectives before combining the extracted information. The local branch emphasizes tumor morphology and surrounding tissue structure, the global branch models contextual relationships among spatially separated regions, and the frequency branch identifies intensity transitions and directional details that may be less apparent in the spatial domain. The local feature branch employs ConvNeXt Tiny to construct a hierarchical convolutional representation. The first stage produces a feature map of 56 × 56 × 96 and preserves relatively fine spatial information, including edges, tissue intensity patterns, and local structural variations. The second stage reduces the feature resolution to 28 × 28 while increasing the channel dimension to 192. This representation captures intermediate anatomical structures and more complex local patterns. The third stage generates a 14 × 14 × 384 tensor that represents tumor morphology, internal heterogeneity, boundary irregularity, and surrounding tissue characteristics. The final stage produces a deep semantic feature map of 7 × 7 × 768. At this level, spatial details are compressed into a high dimensional representation that emphasizes class relevant local patterns. The hierarchical reduction in spatial resolution and simultaneous expansion in channel capacity enable the branch to progress from elementary intensity and edge features toward more abstract tumor related representations.

The global context branch is based on Swin Transformer Tiny and follows the same spatial hierarchy as the ConvNeXt branch. Its four stages generate feature maps of 56 × 56 × 96, 28 × 28 × 192, 14 × 14 × 384, and 7 × 7 × 768, respectively. The first stage models relationships within local attention windows while retaining high spatial resolution. The second stage expands the contextual receptive field and learns hierarchical relationships among neighboring anatomical regions. The third stage captures dependencies among more distant tumor and brain regions, which may help characterize lesion position, spatial extension, and interaction with surrounding structures. The fourth stage generates a compact global semantic representation containing contextual information distributed across the image. Although the Swin Transformer initially applies attention within localized windows, its hierarchical window shifting and feature merging operations allow information to propagate across broader regions. This branch therefore complements the convolutional branch by representing long range spatial relationships rather than relying primarily on local receptive fields. The frequency feature branch applies two level Haar wavelet decomposition to separate the input image into approximation and directional detail components. The resulting LL, LH, HL, and HH subbands contain different aspects of the image frequency content. The LL component retains low frequency information related to broad anatomical structure and gradual intensity variation. The LH and HL components represent directional changes associated with horizontal and vertical tissue transitions, while the HH component emphasizes diagonal variations and fine structural details. These subbands are processed by a wavelet convolutional neural network to learn frequency sensitive features associated with tumor boundaries, texture variation, internal heterogeneity, and transitions between abnormal and surrounding tissues. The branch produces a 14 × 14 × 256 frequency

representation, as reported in Table 5. The inclusion of this branch is intended to preserve information that may be weakened during repeated spatial downsampling in the local and global branches.

Table 5. Proposed hybrid architecture

Component	Operation	Output representation
Input	Preprocessed brain MRI	224 × 224 × 3
Local branch Stage 1	ConvNeXt Tiny local feature extraction	56 × 56 × 96
Local branch Stage 2	Intermediate structural representation	28 × 28 × 192
Local branch Stage 3	Tumor morphology and heterogeneity representation	14 × 14 × 384
Local branch Stage 4	Deep local semantic representation	7 × 7 × 768
Global branch Stage 1	Swin Transformer local window relationships	56 × 56 × 96
Global branch Stage 2	Hierarchical spatial context	28 × 28 × 192
Global branch Stage 3	Distant tumor dependencies	14 × 14 × 384
Global branch Stage 4	Global semantic representation	7 × 7 × 768
Frequency branch	Two level Haar wavelet decomposition	LL, LH, HL, and HH subbands
Wavelet CNN	Convolutional frequency feature learning	14 × 14 × 256
Feature alignment	Resizing and projection to a shared feature space	7 × 7 × 512 per branch
Cross attention	Bidirectional exchange of local, global, and frequency information	7 × 7 × 512
Adaptive gated fusion	Image dependent branch importance estimation	7 × 7 × 512
Global average pooling	Spatial feature aggregation	512 dimensional vector
Classification head	Layer normalization, dense 256, GELU, and dropout	256 dimensional vector
Output layer	Four class linear classifier and softmax	Four probabilities

b. Feature Alignment, Cross Attention, and Adaptive Fusion

The three feature branches produce representations with different spatial resolutions and channel dimensions. Direct combination of these features would therefore be unsuitable because their tensor structures and representational properties are not equivalent. To resolve this issue, the output of each branch is resized and projected into a shared feature space of 7 × 7 × 512. The ConvNeXt and Swin Transformer outputs are projected from 768 to 512 channels, while the 14 × 14 × 256 frequency representation is spatially resized and projected to the same 7 × 7 × 512 configuration. This alignment establishes a common spatial and channel structure while retaining the distinctive information learned by each branch. Consequently, every spatial position in the aligned tensors can participate in subsequent interaction and fusion operations. Following feature alignment, bidirectional cross attention is applied to facilitate information exchange among the local, global, and frequency representations. As shown in Figure 2, this module does not treat the branches as independent sources that are combined only at the final stage. Instead, each representation is refined using information from the other branches. Local structural features can therefore be interpreted in relation to broader anatomical context, while global features can be informed by detailed tumor morphology. Frequency characteristics can also influence the

spatial representations by emphasizing tissue transitions, fine boundaries, and heterogeneous intensity patterns. Conversely, the spatial branches provide anatomical context that can assist the interpretation of wavelet responses. This bidirectional interaction is intended to produce a coordinated representation in which branch specific information is preserved but interpreted jointly.

The cross attention output is subsequently processed using adaptive gated fusion. A fixed fusion operation, such as direct addition or concatenation, would assign the same importance to every branch for all images. However, the diagnostic usefulness of each representation may vary according to tumor type, size, location, morphology, and image orientation. For example, local structural features may be particularly informative for a lesion with irregular boundaries, whereas global anatomical context may contribute more strongly when tumor location is a major discriminative feature. Frequency information may become more relevant when the image contains pronounced texture changes or tissue transitions. The adaptive gated fusion mechanism estimates the relative contribution of the three branches for each input image and combines them into a unified 7 × 7 × 512 feature map. This image specific weighting allows the model to emphasize the most informative representation while reducing the influence of less relevant or redundant branch features. The combination of cross attention and adaptive gating serves two distinct but complementary purposes. Cross attention enables representational interaction before final integration, whereas adaptive gating determines the relative importance of the refined branch features. The former supports information exchange, while the latter controls contribution strength. This design is intended to provide a more selective and coordinated fusion process than simple feature concatenation. The resulting tensor contains local morphology, global anatomical context, and frequency based tissue information within a common representation.

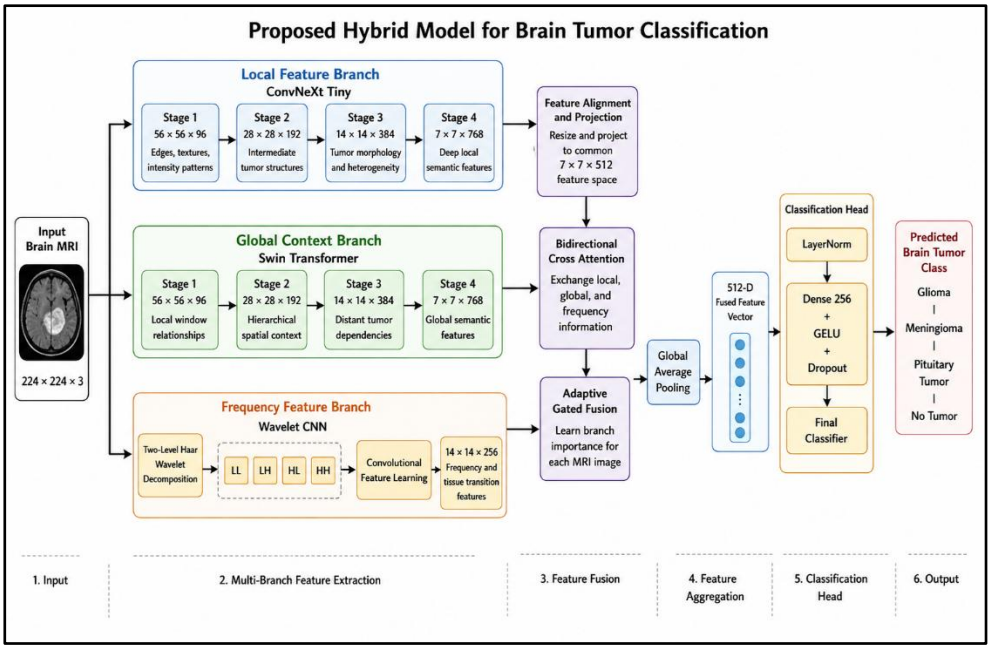

Figure 2. Overall Architecture of the Proposed Hybrid Artificial Intelligence Model

c. Feature Aggregation and Four Class Prediction

After adaptive fusion, global average pooling is applied to the 7 × 7 × 512 fused feature map. The pooling operation computes a summary of each feature channel across the spatial dimensions and produces a 512 dimensional vector. This transformation reduces the number of parameters required by the classification head while preserving the overall activation strength of each learned feature. It also decreases dependence on the exact spatial position of a feature, which is useful because tumor location and image orientation vary across the dataset. The pooled representation therefore serves as a compact summary of the combined local, global, and frequency information. The 512 dimensional vector is passed to the classification head described in Table 5. Layer normalization is first applied to stabilize the feature distribution before dense transformation. A fully connected layer then maps the representation from 512 to 256 dimensions. The Gaussian error linear unit activation introduces nonlinearity and enables the model to learn complex relationships among the fused features. Dropout is applied with a probability of 0.50 to reduce reliance on individual activations and limit overfitting. The resulting 256 dimensional representation is provided to a final linear classifier containing four output units. The output layer generates one score for each diagnostic category comprising glioma, meningioma, no tumor, and pituitary tumor. Softmax normalization converts these scores into four class probabilities whose sum equals one. The category with the highest predicted probability is selected as the final image level classification. The complete architecture therefore forms a continuous learning pathway from standardized brain MRI input to multibranch feature extraction, feature alignment, cross branch interaction, adaptive fusion, global aggregation, and four class prediction. Figure 2 presents the full computational workflow, while Table 5 provides the corresponding operations and output dimensions for each architectural component.

3.4 Hyperparameters and Training Configuration

The proposed hybrid model was trained using the configuration summarized in Table 6. The ConvNeXt Tiny local branch and the Swin Transformer Tiny global branch were initialized using weights pretrained on ImageNet. This initialization provided established visual representations and reduced the need to learn all feature extraction parameters from random initialization. The newly introduced feature alignment, bidirectional cross attention, adaptive gated fusion, and classification modules were initialized separately and optimized for the brain tumor classification task. Multiclass cross entropy was used as the objective function because the study involved four mutually exclusive diagnostic categories. The AdamW optimizer was employed with a weight decay value of 0.05 to support stable parameter updates and control excessive model complexity.

Different learning rates were assigned to the pretrained branches and the newly introduced components. The ConvNeXt Tiny and Swin Transformer Tiny branches were trained using a learning rate of 3×10^{-5}, allowing their pretrained representations to be refined gradually without substantial disruption. The fusion and classification components used a higher learning rate of 3×10^{-4}, enabling the newly initialized layers to adapt more rapidly to the target dataset. Training was conducted with a batch size of 32 for a maximum of 50 epochs. The learning rate was increased gradually

during the first five epochs through a warmup procedure and was subsequently reduced using cosine annealing until reaching a minimum value of 1×10^{-6}. Gradient clipping with a maximum norm of 1.0 was applied to limit unstable gradient updates, particularly within the attention and fusion modules.A dropout probability of 0.50 was applied in the classification head to reduce overfitting and improve generalization. Early stopping was used with a patience of 10 epochs, meaning that training was terminated when the validation performance failed to improve for ten consecutive epochs. The checkpoint achieving the highest validation macro F1 score was retained for final testing. Macro F1 was selected as the checkpoint criterion because it considers performance across all four classes equally and is less influenced by differences in class frequency than overall accuracy. All experiments were performed using a CUDA enabled NVIDIA graphics processing unit. Final performance was assessed using accuracy, precision, recall, F1 score, specificity, Matthews correlation coefficient, Cohen kappa, and area under the receiver operating characteristic curve. The complete optimization and evaluation settings are reported in Table 6 to support methodological transparency and experimental reproducibility.

Table 6. Training configuration

Hyperparameter	Proposed setting
Weight initialization	ImageNet pretrained ConvNeXt Tiny and Swin Transformer Tiny
Loss function	Multiclass cross entropy
Optimizer	AdamW
Local branch learning rate	3×10^{-5}
Global branch learning rate	3×10^{-5}
Fusion and classification learning rate	3×10^{-4}
Weight decay	0.05
Batch size	32
Maximum epochs	50
Warmup duration	5 epochs
Learning rate schedule	Cosine annealing
Minimum learning rate	1×10^{-6}
Dropout probability	0.50
Gradient clipping	1.0
Early stopping patience	10 epochs
Checkpoint criterion	Highest validation macro F1
Number of classes	4
Evaluation metrics	Accuracy, precision, recall, F1 score, specificity, MCC, kappa, and AUC
Hardware	CUDA enabled NVIDIA GPU

3.5 Baseline Models and Comparison Protocol

To evaluate the effectiveness of the proposed hybrid framework, seven established deep learning architectures were selected as baseline models. As summarized in Table 7, the comparison included VGG16, MobileNetV3 Large, ResNet50, DenseNet121, EfficientNet B0, ConvNeXt Tiny, and Swin Transformer Tiny. These models represent diverse feature learning principles, including conventional convolution, residual learning, dense connectivity, compound scaling, modern convolutional design, and hierarchical transformer attention. Their inclusion enabled the proposed framework to be evaluated against both widely used convolutional networks and

more recent representation learning architectures.

VGG16 was used as a conventional convolutional baseline and employed a classification head consisting of a dense layer with 4096 units, a dense layer with 256 units, and a final four class classifier. MobileNetV3 Large represented a computationally efficient architecture designed to learn lightweight convolutional features. ResNet50 used residual connections to support deeper feature learning, while DenseNet121 employed densely connected layers to encourage feature reuse and improved gradient flow [44-45]. EfficientNet B0 provided a compound scaled representation in which network depth, width, and input processing capacity were jointly balanced. MobileNetV3 Large, ResNet50, DenseNet121, and EfficientNet B0 each used a dense layer with 256 units followed by the final classifier.

ConvNeXt Tiny and Swin Transformer Tiny were included as stronger modern baselines. ConvNeXt Tiny represented an advanced convolutional architecture with contemporary design principles, while Swin Transformer Tiny modeled hierarchical spatial context through localized window attention. Both models used a dense layer with 256 units before the final output layer. In contrast, the proposed model combined ConvNeXt Tiny, Swin Transformer Tiny, and a wavelet convolutional neural network to learn local, global, and frequency representations within a unified framework. These complementary features were integrated using bidirectional cross attention and adaptive gated fusion before classification. All baseline models were evaluated using the same training, validation, and test partitions and were assessed using the same performance metrics to support a consistent comparison.

Table 7. Baseline comparison protocol

Model	Backbone representation	Classification head
VGG16	Convolutional feature extractor	Dense 4096, dense 256, and final classifier
MobileNetV3 Large	Lightweight convolutional features	Dense 256 and final classifier
ResNet50	Residual convolutional features	Dense 256 and final classifier
DenseNet121	Densely connected convolutional features	Dense 256 and final classifier
EfficientNet B0	Compound scaled convolutional features	Dense 256 and final classifier
ConvNeXt Tiny	Modern convolutional representation	Dense 256 and final classifier
Swin Transformer Tiny	Hierarchical window attention	Dense 256 and final classifier
Proposed hybrid model	ConvNeXt, Swin Transformer, and wavelet CNN	Cross attention, gated fusion, and classifier

4. RESULTS AND PERFORMANCE ANALYSIS

4.1 Training and Validation Performance

Table 8 summarizes the optimization behavior of the proposed model during classification head training and subsequent fine tuning. At the beginning of head training, the model produced a training loss of 1.2785 and training accuracy of 46.38%, while the validation loss and accuracy were 1.1996 and 56.20%, respectively. By epoch 8, the training accuracy increased to 77.17% and the validation accuracy reached 79.94%, accompanied by reductions in training and validation losses to 0.6341 and 0.5688. Fine tuning of the complete network further improved performance, with the validation accuracy increasing to 84.23% at the first fine tuning epoch and reaching 91.73% at epoch 11. The lowest validation loss of 0.2684 was recorded at epoch 13, while the validation accuracy remained stable at 91.73%, suggesting that optimization had approached convergence. The supplied independent

test record reported a loss of 0.1999 and an accuracy of 94.92%, while the corresponding training values were 0.1908 and 95.11%. The relatively small difference between the training and test accuracies indicates limited performance degradation on unseen images in this experimental run.

Table 8. Proposed model training summary from the supplied log

Training point	Train loss	Train accuracy	Validation loss	Validation accuracy
Head training, epoch 0	1.2785	46.38%	1.1996	56.20%
Head training, best epoch 8	0.6341	77.17%	0.5688	79.94%
Fine tuning, epoch 0	0.5971	78.25%	0.4670	84.23%
Fine tuning, best accuracy epoch 11	0.2214	92.19%	0.2790	91.73%
Fine tuning, lowest validation loss epoch 13	0.2098	92.07%	0.2684	91.73%
Independent test result	0.1908	95.11%	0.1999	94.92%

4.2 Confusion Matrix Analysis

Figure 3 presents the confusion matrix of the proposed model on the independent test set containing 394 images. The model correctly classified 86 of 100 glioma images, 110 of 115 meningioma images, 104 of 105 no tumor images, and all 74 pituitary tumor images, resulting in 374 correct predictions and an overall accuracy of 94.92%. Glioma was the most challenging category, with seven glioma images classified as meningioma and another seven classified as no tumor. Five meningioma images were incorrectly assigned to the glioma category, while one no tumor image was also predicted as glioma. No pituitary tumor image was misclassified, and no image from another category was assigned to the pituitary tumor class. These results indicate that the principal classification difficulty occurred between glioma and the other intracranial categories, whereas pituitary tumor images were separated with complete correctness in this test set.

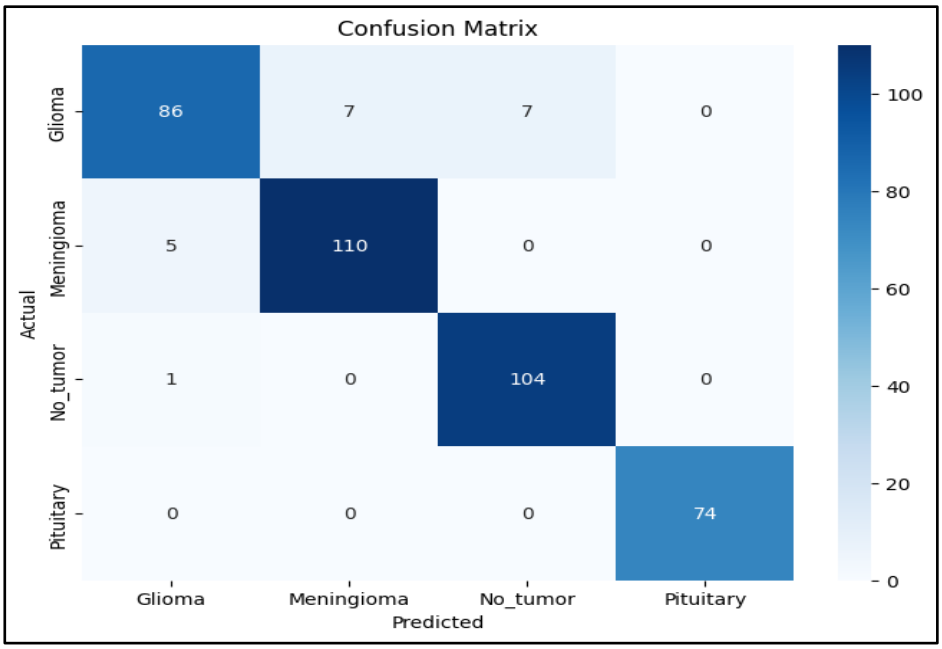

Figure 3. Confusion Matrix of the Proposed Model on the Independent Test Set

4.3 Receiver Operating Characteristic Analysis

The one versus rest receiver operating characteristic curves shown in Figure 4 demonstrate strong discriminative performance across all four diagnostic categories. The glioma class achieved an area under the curve of 0.997, while meningioma achieved 0.999. Both the no tumor and pituitary tumor categories obtained an area under the curve of 1.000. The average of the four displayed class values was approximately 0.999, indicating that the model assigned substantially higher prediction scores to the correct category than to the remaining categories across a wide range of classification thresholds. Although glioma exhibited more errors in the confusion matrix, its area under the curve remained high, suggesting that the model retained strong ranking ability even when the final class decision was incorrect for several images. The near upper left positioning of the curves further indicates high sensitivity at low false positive rates.

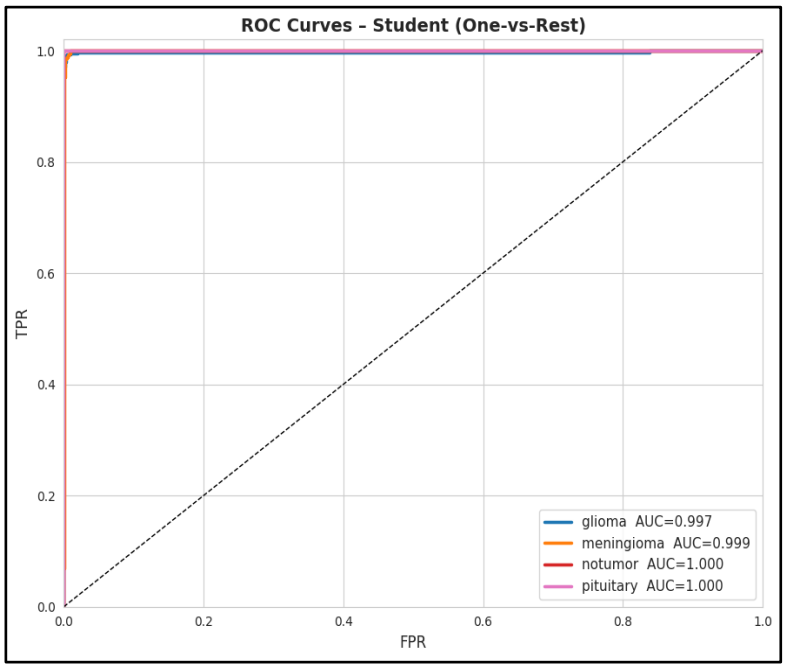

Figure 4. Receiver Operating Characteristic Curves of the Proposed Model for Four Class Brain Tumor Classification

4.4 Class Level Performance of the Supplied Model Run

Table 9 provides the class level performance corresponding to the 94.92% accuracy experiment represented by the supplied confusion matrix. Pituitary tumor achieved perfect precision, recall, F1 score, and specificity, confirming that all 74 pituitary tumor images were classified correctly without false positive assignments. The no tumor category produced a recall of 0.9905 and an F1 score of 0.9630, although its precision of 0.9369 indicates that several images from other categories were assigned to this class. Meningioma achieved balanced performance, with precision of 0.9402, recall of 0.9565, and an F1 score of 0.9483. Glioma produced the lowest recall and F1 score, with values of 0.8600 and 0.8958, respectively, reflecting the 14 glioma images misclassified in Figure 3. Across all classes, the model achieved a macro precision of 0.9530, macro recall of 0.9517, macro F1 score of 0.9518, and macro specificity of 0.9826. The weighted F1 score of 0.9486 remained close to the macro F1 score, indicating that the aggregate result was not strongly dominated by any single class.

Table 9. Class level Proposed model performance

Class	Precision	Recall	F1 score	Specificity	Support
Glioma	0.9348	0.8600	0.8958	0.9796	100
Meningioma	0.9402	0.9565	0.9483	0.9749	115
No tumor	0.9369	0.9905	0.9630	0.9758	105
Pituitary tumor	1.0000	1.0000	1.0000	1.0000	74
Macro average	0.9530	0.9517	0.9518	0.9826	394
Weighted average	0.9492	0.9492	0.9486	Not applicable	394

4.5 Class Level Performance of the Final Hybrid Model

Table 10 reports the stronger class level results of the final proposed hybrid model. Glioma achieved a precision of 0.9898, recall of 0.9700, F1 score of 0.9798, and specificity of 0.9966. Meningioma produced equal precision and recall values of 0.9826, resulting in an F1 score of 0.9826 and specificity of 0.9928. The no tumor category achieved complete recall of 1.0000, an F1 score of 0.9906, and specificity of 0.9931, indicating that every no tumor image was correctly identified, although a small number of images from other classes were assigned to this category. Pituitary tumor again achieved perfect performance across all reported metrics. The final model obtained a macro precision of 0.9884, macro recall of 0.9882, macro F1 score of 0.9882, and macro specificity of 0.9956. The weighted precision, recall, and F1 score were 0.9874, 0.9873, and 0.9873, respectively, demonstrating consistently high performance across the four diagnostic categories.

Table 10. Class level performance of the proposed hybrid model

Class	Precision	Recall	F1 score	Specificity	Support
Glioma	0.9898	0.9700	0.9798	0.9966	100
Meningioma	0.9826	0.9826	0.9826	0.9928	115
No tumor	0.9813	1.0000	0.9906	0.9931	105
Pituitary tumor	1.0000	1.0000	1.0000	1.0000	74
Macro average	0.9884	0.9882	0.9882	0.9956	394
Weighted average	0.9874	0.9873	0.9873	Not applicable	394

4.6 Comparison with Baseline Models

The comparative results in Table 11 show a progressive improvement from conventional convolutional models to more advanced convolutional and transformer based architectures. VGG16 achieved the lowest accuracy of 90.36%, followed by MobileNetV3 Large at 91.37%, ResNet50 at 92.13%, and DenseNet121 at 93.40%. EfficientNet B0 improved the accuracy to 94.92%, while ConvNeXt Tiny and Swin Transformer Tiny achieved 95.94% and 96.45%, respectively. The proposed hybrid model produced 389 correct predictions from 394 test images and achieved the highest reported accuracy of 98.73%, with a 95% confidence interval from 97.06% to 99.46%. It also obtained the highest macro precision of 98.84%, macro recall of 98.82%, macro F1 score of 98.82%, macro specificity of 99.56%, Matthews correlation coefficient of 0.9830, and macro area under the curve of 0.9995. Compared with Swin Transformer Tiny, the strongest individual baseline, the proposed model improved accuracy by 2.28 percentage points and macro F1 by 2.19 percentage points. The numerical improvement across accuracy, class balanced measures, specificity, agreement, and discrimination suggests that integrating local convolutional features, transformer based contextual representations, and wavelet frequency information provided a more comprehensive representation than any single backbone.

Table 11. Comparison with baseline models

Model	Correct predictions	Accuracy with 95% CI	Macro precision	Macro recall	Macro F1	Macro specificity	MCC	Macro AUC
VGG16	356 of 394	90.36% [87.04%, 92.89%]	90.95%	90.96%	90.89%	96.69%	0.8708	0.9740
MobileNetV3 Large	360 of 394	91.37% [88.18%, 93.76%]	91.90%	91.78%	91.79%	97.04%	0.8843	0.9795
ResNet50	363 of 394	92.13% [89.05%, 94.40%]	92.60%	92.48%	92.49%	97.31%	0.8946	0.9830
DenseNet121	368 of 394	93.40% [90.51%, 95.46%]	93.86%	93.68%	93.69%	97.74%	0.9118	0.9890
EfficientNet B0	374 of 394	94.92% [92.29%, 96.69%]	95.30%	95.17%	95.18%	98.26%	0.9322	0.9990
ConvNeXt Tiny	378 of 394	95.94% [93.51%, 97.49%]	96.24%	96.14%	96.16%	98.60%	0.9456	0.9991
Swin Transformer Tiny	380 of 394	96.45% [94.12%, 97.87%]	96.72%	96.61%	96.63%	98.78%	0.9525	0.9992
Proposed hybrid model	**389 of 394**	**98.73% [97.06%, 99.46%]**	**98.84%**	**98.82%**	**98.82%**	**99.56%**	**0.9830**	**0.9995**

5. LIMITATIONS AND FUTURE WORK

Several limitations should be considered when interpreting the findings of this study. First, patient identifiers were not available, and the dataset partition could therefore not be verified as patient independent. Images obtained from the same patient may share anatomical and acquisition characteristics, which could increase the apparent test performance if related images were distributed across different subsets. The absence of MRI sequence metadata also prevented sequence specific evaluation and limited assessment of the model across different acquisition protocols. In addition, the experiments were conducted using a single dataset, with only 394 images included in the independent test set. The reported findings may therefore not represent performance across institutions, scanner manufacturers, magnetic field strengths, demographic groups, or clinical environments. The model performs image level classification and does not provide tumor localization, segmentation, grading, molecular characterization, or patient level diagnosis. Furthermore, the combination of ConvNeXt Tiny, Swin Transformer Tiny, wavelet processing, cross attention, and adaptive fusion increases computational complexity compared with individual baseline networks.

Future work should evaluate the proposed framework using patient independent partitions and external multicenter datasets containing complete patient, scanner, acquisition, and MRI sequence information. Repeated stratified evaluation and independent clinical validation would provide stronger evidence regarding robustness and generalizability. The architecture could also be extended to integrate multiple MRI sequences, clinical variables, tumor segmentation, and uncertainty estimation within a unified diagnostic framework. Interpretability methods should be incorporated to identify the image regions and frequency components that most strongly influence each prediction, enabling radiological assessment of model behavior. Further studies may investigate model compression, knowledge distillation,

parameter reduction, and efficient attention mechanisms to improve computational suitability for clinical deployment. Prospective evaluation involving radiologists and real clinical workflows will ultimately be necessary before the model can be considered for decision support applications.

6. CONCLUSION

This study presented a hybrid deep learning framework for four class brain tumor classification using magnetic resonance images. The proposed model combined a ConvNeXt Tiny branch for local structural feature extraction, a Swin Transformer Tiny branch for hierarchical contextual modeling, and a wavelet convolutional neural network for frequency-based representation learning. Bidirectional cross attention enabled information exchange among the three feature streams, while adaptive gated fusion estimated their relative importance for each image. The framework was evaluated using 10,464 brain MRI images categorized as glioma, meningioma, no tumor, and pituitary tumor. On the independent test set, the proposed hybrid model correctly classified 389 of 394 images and achieved an accuracy of 98.73%, macro precision of 98.84%, macro recall of 98.82%, macro F1 score of 98.82%, macro specificity of 99.56%, Matthews correlation coefficient of 0.9830, and macro area under the receiver operating characteristic curve of 0.9995. These results exceeded those obtained by the evaluated convolutional and transformer-based baseline models.

The findings suggest that the coordinated integration of local morphology, global anatomical context, and frequency information can provide a comprehensive representation for image level brain tumor classification. However, the results should be interpreted within the limitations of the available dataset. Patient identifiers and MRI sequence metadata were unavailable, and the evaluation was restricted to a single dataset with a relatively small independent test set. Therefore, the study does not establish patient level diagnostic performance or clinical generalizability. Future research should evaluate the framework using patient independent multicenter datasets, diverse MRI acquisition protocols, external validation cohorts, and prospective clinical assessment. Further development should also consider tumor localization, uncertainty estimation, model interpretability, and computational optimization to support safer and more transparent clinical decision support.

Author Contributions:

Kallol Chakraborty Shekhor conceived the study, designed the research framework, and coordinated the overall investigation. Sumayya Jahan contributed to the methodology development, literature review, and manuscript preparation. Md Sahid Hossain supported model implementation, software development, and experimental analysis. Md Fakrul Alam contributed to data preprocessing, technical validation, and interpretation of the findings. Abdur Rahim assisted with dataset preparation, performance evaluation, and result verification. Md Sazzad Hossain Shehan contributed to visualization, statistical analysis, and manuscript drafting. Md Anwar Hossain provided technical supervision, reviewed the experimental methodology, and critically revised the manuscript. Md Abedur Rahman contributed to validation, manuscript editing, and final quality assessment. All authors reviewed and approved the final version of the manuscript.

Data Availability:The datasets used in this study are publicly available through Kaggle.

Declarations

Ethical approval: Not applicable.

Consent to participate: Not applicable. The study used only secondary, anonymized image data and involved no direct contact or intervention with individual participants.

Consent to publish: Not applicable. This manuscript does not contain any individual person's identifiable data (including names, facial photographs, or other personal details).

Clinical Trial : Not applicable.

Funding : Not applicable.

REFERENCES

[1] Sharma, A. K., Nandal, A., Dhaka, A., Zhou, L., Alhudhaif, A., Alenezi, F., & Polat, K. (2023). Brain tumor classification using the modified ResNet50 model based on transfer learning. Biomedical Signal Processing and Control, 86, 105299.

[2] Manik, M. M. T. G. (2021). Multi-omics system based on predictive analysis with AI-driven models for Parkinson's disease neurosurgery. Journal of Medical and Health Studies, 2(1), 42–52. https://doi.org/10.32996/jmhs.2021.2.1.5

[3] Ge, C., Gu, I. Y. H., Jakola, A. S., & Yang, J. (2020). Deep semi-supervised learning for brain tumor classification. BMC Medical Imaging, 20(1), 87.

[4] Chowdhury, M. S. I., Chy, K. S., Islam, M. A., & Chowdhury, M. N. I. (2023). Integrating blockchain and data analytics to strengthen financial traceability and anti-fraud controls. Journal of Computer Science and Technology Studies, 5(3), 189–202. https://doi.org/10.32996/jcsts.2023.5.3.14

[5] Noreen, N., Palaniappan, S., Qayyum, A., Ahmad, I., & Alassafi, M. O. (2021). Brain tumor classification based on fine-tuned models and the ensemble method. Computers, Materials and Continua, 67(3), 3967–3982.

[6] Manik, M. M. T. G., Bhuiyan, M. M. R., Moniruzzaman, M., Islam, M. S., Hossain, S., & Hossain, S. (2018). The future of drug discovery utilizing generative AI and big data analytics for accelerating pharmaceutical innovations. Nanotechnology Perceptions, 14(3), 120–135. https://doi.org/10.62441/nano-ntp.v14i3.4766

[7] Xie, Y., Zaccagna, F., Rundo, L., Testa, C., Agati, R., Lodi, R., Manners, D. N., & Tonon, C. (2022). Convolutional neural network techniques for brain tumor classification from 2015 to 2022: Review, challenges, and future perspectives. Diagnostics, 12(8), 1850.

[8] Chowdhury, M. N. I., Chy, K. S., Chowdhury, M. S. I., & Islam, M. A. (2021). Identifying hidden fraud indicators through longitudinal analysis of U.S. financial misconduct. Journal of Economics, Finance and Accounting Studies, 3(1), 89–99. https://doi.org/10.32996/jefas.2021.3.1.9

[9] Md Abedur Rahman, Md Anwar Hossain, Kallol Chakraborty Shekhor, & Md Sahid Hossain. (2024). A Conditional Positional Encoding and Channel Attention Guided Swin Transformer for Breast Cancer Classification Using MRI and Mammography. *Journal of Computer Science and Technology Studies*, *6*(3), 193-218. https://doi.org/10.32996/jcsts.2024.6.3.16

[11] Abd El Kader, I., Xu, G., Shuai, Z., Saminu, S., Javaid, I., & Ahmad, I. S. (2021). Differential deep convolutional neural network model for brain tumor classification. Brain Sciences, 11(3), 352.

[12] Chy, K. S., Islam, M. A., Chowdhury, M. S. I., Chowdhury, M. N. I., & Islam, M. A. (2022). Enhancing U.S. financial compliance and risk management through data-driven automation and anomaly detection. Journal of Business and Management Studies, 4(3), 231–244. https://doi.org/10.32996/jbms.2022.4.3.22

[13] Park, J. W., Lee, K., Kim, E. E., Kim, S. I., & Park, S. H. (2023). Brain tumor classification by methylation profile. Journal of Korean Medical Science, 38(43).

[14] Manik, M. M. T. G. (2020). Biotech-driven innovation in drug discovery: Strategic models for competitive advantage in the global pharmaceutical market. Journal of Computational Analysis and Applications, 28(6), 41–47. https://eudoxuspress.com/index.php/pub/article/view/2874

[15] Mushtaq, U. (2020). Brain tumor classification in MRI image using convolutional neural network. Mathematical Biosciences and Engineering.

[16] Manik, M. M. T. G., Hossain, S., Ahmed, M. K., Rozario, E., Miah, M. A., Moniruzzaman, M., et al. (2022). Integrating genomic data and machine learning to advance precision oncology and targeted cancer therapies. Nanotechnology Perceptions, 18(2), 219–243. https://doi.org/10.62441/nano-ntp.v18i2.5443

[17] Shaik, N. S., & Cherukuri, T. K. (2022). Multi-level attention network: Application to brain tumor classification. Signal, Image and Video Processing, 16(3), 817–824.

[18] Kumar, S., & Mankame, D. P. (2020). Optimization-driven deep convolutional neural network for brain tumor classification. Biocybernetics and Biomedical Engineering, 40(3), 1190–1204.

[19] Neamah, K., Mohamed, F., Adnan, M. M., Saba, T., Bahaj, S. A., Kadhim, K. A., & Khan, A. R. (2023). Brain tumor classification and detection based on deep learning models: A systematic review. IEEE Access, 12, 2517–2542.

[20] Manik, M. M. T. G. (2022). An analysis of cervical cancer using the application of AI and machine learning. Journal of Medical and Health Studies, 3(2), 67–76. https://doi.org/10.32996/jmhs.2022.3.2.11

[21] Waghmare, V. K., & Kolekar, M. H. (2020). Brain tumor classification using deep learning. In Internet of Things for Healthcare Technologies (pp. 155–175). Springer Singapore.

[22] Manik, M. M. T. G., Moniruzzaman, M., Islam, M. S., Bhuiyan, M. M. R., Rozario, E., Hossain, S., et al. (2020). The role of big data in combatting antibiotic resistance: Predictive models for global surveillance. Nanotechnology Perceptions, 16(3), 361–378. https://doi.org/10.62441/nano-ntp.v16i3.5445

[23] M. J. van den Bent, M. Weller, P. Y. Wen, J. M. Kros, K. Aldape, and S. Chang, "A clinical perspective on the 2016 WHO brain tumor classification and routine molecular diagnostics," Neuro-Oncology, vol. 19, no. 5, pp. 614–624, 2017. https://doi.org/10.1093/neuonc/now277

[24] M. M. T. G. Manik, A. S. M. Saimon, M. A. Miah, M. K. Ahmed, F. B. Khair, M. Moniruzzaman, et al., "Leveraging AI-powered predictive analytics for early detection of chronic diseases: A data-driven approach to personalized medicine," Nanotechnology Perceptions, vol. 17, no. 3, pp. 269–288, 2021. https://doi.org/10.62441/nano-ntp.v17i3.5444

[25] M. A. Miah, E. Rozario, F. B. Khair, M. K. Ahmed, M. M. R. Bhuiyan, and M. M. T. G. Manik, "Harnessing wearable health data and deep learning algorithms for real-time cardiovascular disease monitoring and prevention," Nanotechnology Perceptions, vol. 15, no. 3, pp. 326–349, 2019.

[26] M. I. Faruk, T. Roy, M. J. Hossain, M. Adnan, E. Hossain, M. H. Rijvi, M. M. Bakhsh, and S. M. M. Uddin, "Global health intelligence platforms: Data science infrastructure for real-time epidemiological surveillance and pandemic forecasting," Spectrum of Engineering Sciences, vol. 2, no. 5, pp. 749–761, 2024.

[27] J. Cheng, W. Huang, S. Cao, R. Yang, W. Yang, Z. Yun, Z. Wang, and Q. Feng, "Enhanced performance of brain tumor classification via tumor region augmentation and partition," PLOS ONE, vol. 10, no. 10, article e0140381, 2015. https://doi.org/10.1371/journal.pone.0140381

[28] M. R. Ismael and I. Abdel-Qader, "Brain tumor classification via statistical features and back-propagation neural network," in Proceedings of the 2018 IEEE International Conference on Electro/Information Technology, Rochester, Michigan, USA, 2018, pp. 252–257. https://doi.org/10.1109/EIT.2018.8500308

[29] J. Amin, M. Sharif, M. Raza, T. Saba, and A. Rehman, "Brain tumor classification: Feature fusion," in Proceedings of the 2019 International Conference on Computer and Information Sciences, Sakaka, Saudi Arabia, 2019, pp. 1–6. https://doi.org/10.1109/ICCISci.2019.8716449

[30] A. Ari and D. Hanbay, "Deep learning based brain tumor classification and detection system," Turkish Journal of Electrical Engineering and Computer Sciences, vol. 26, no. 5, pp. 2275–2286, 2018. https://doi.org/10.3906/elk-1801-8

[31] J. S. Paul, A. J. Plassard, B. A. Landman, and D. Fabbri, "Deep learning for brain tumor classification," in Medical Imaging 2017: Biomedical Applications in Molecular, Structural, and Functional Imaging, Proceedings of SPIE, vol. 10137, article 1013710, 2017. https://doi.org/10.1117/12.2254195

[32] J. Seetha and S. Selvakumar Raja, "Brain tumor classification using convolutional neural networks," Biomedical and Pharmacology Journal, vol. 11, no. 3, pp. 1457–1461, 2018. https://doi.org/10.13005/bpj/1511

[33] N. Abiwinanda, M. Hanif, S. T. Hesaputra, A. Handayani, and T. R. Mengko, "Brain tumor classification using convolutional neural network," in World Congress on Medical Physics and Biomedical Engineering 2018, IFMBE Proceedings, vol. 68/1, Singapore: Springer, 2019, pp. 183–189. https://doi.org/10.1007/978-981-10-9035-6_33

[34] S. I. Nilima, M. A. Hossain, S. Sharmin, R. Rahman, H. Esa, M. M. T. G. Manik, and R. Hasan, "Advancement of drug discovery using artificial intelligence and machine learning," in Proceedings of the 2024 IEEE International Conference on Computing, Applications and Systems, Cox's Bazar, Bangladesh, 2024, pp. 1–7. https://doi.org/10.1109/COMPAS60761.2024.10796748

[35] S. Deepak and P. M. Ameer, "Brain tumor classification using deep CNN features via transfer learning," Computers in Biology and Medicine, vol. 111, article 103345, 2019. https://doi.org/10.1016/j.compbiomed.2019.103345

[36] Z. N. K. Swati, Q. Zhao, M. Kabir, F. Ali, Z. Ali, S. Ahmed, and J. Lu, "Brain tumor classification for MR images using transfer learning and fine-tuning," Computerized Medical Imaging and Graphics, vol. 75, pp. 34–46, 2019.

[37] A. Raza, H. Ayub, J. A. Khan, I. Ahmad, A. S. Salama, Y. I. Daradkeh, D. Javeed, A. U. Rehman, and H. Hamam, "A hybrid deep learning-based approach for brain tumor classification," Electronics, vol. 11, no. 7, article 1146, 2022. https://doi.org/10.3390/electronics11071146

[38] D. R. Nayak, N. Padhy, P. K. Mallick, M. Zymbler, and S. Kumar, "Brain tumor classification using dense EfficientNet," Axioms, vol. 11, no. 1, article 34, 2022. https://doi.org/10.3390/axioms11010034

[39] M. H. Rahman, M. E. Hossin, M. J. Hossain, S. M. M. Uddin, M. I. Faruk, M. M. Anwar, and F. Hossain, "Harnessing big data and predictive analytics for early detection and cost optimization in cancer care," Journal of Computer Science and Technology Studies, vol. 6, no. 5, pp. 278–293, 2024. https://doi.org/10.32996/jcsts.2024.6.5.22

[40] G. A. Amran, M. S. Alsharam, A. O. A. Blajam, A. A. Hasan, M. Y. Alfaifi, M. H. Amran, A. Gumaei, and S. M. Eldin, "Brain tumor classification and detection using hybrid deep tumor network," Electronics, vol. 11, no. 21, article 3457, 2022. https://doi.org/10.3390/electronics11213457

[41] S. Kumar, S. Choudhary, A. Jain, K. Singh, A. Ahmadian, and M. Y. Bajuri, "Brain tumor classification using deep neural network and transfer learning," Brain Topography, vol. 36, no. 3, pp. 305–318, 2023. https://doi.org/10.1007/s10548-023-00953-0

[42] M. M. Islam, M. A. Talukder, M. A. Uddin, A. Akhter, and M. Khalid, "BrainNet: Precision brain tumor classification with optimized EfficientNet architecture," International Journal of Intelligent Systems, vol. 2024, article 3583612, 2024. https://doi.org/10.1155/2024/3583612

[43] S. Hossain, M. M. R. Bhuiyan, M. S. Islam, M. Moniruzzaman, M. K. Ahmed, N. Das, A. S. M. Saimon, and M. M. T. G. Manik, "Big data analysis and prediction of COVID-2019 epidemic using machine learning models in healthcare sector," Journal of Ecohumanism, vol. 3, no. 8, 2024. https://doi.org/10.62754/joe.v3i8.6775

[44] Hossain, Md Sahid, Kallol Chakraborty Shekhor, Md Anwar Hossain, and Md Abedur Rahman. "Diagnosis Everywhere: Lightweight AI That Detects Disease from Scans and Records on Limited Hospital Hardware." *Frontiers in Computer Science and Artificial Intelligence* 2: 82-94.

[45] Shekhor, K. C., Hossain, M. S., Rahman, M. A., & Hossain, M. A. Multi-Scale Attention-Fusion Classification with an LLM-Driven Clinical Recommendation Assistant for Brain Tumor MRI. *Journal of Medical and Health Studies*, *4*, 121-39.