Information Sciences 746 (2026) 123423

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

TumorNet: A hybrid lightweight framework for brain tumor classification and reasoning

Hanxiang Wang [a,1], Muhammad Zaqeem [b,1], Muhammad Fayaz [c,*], Defu Qiu [a], Sajjad Ahadzadeh [d], Tan N. Nguyen [e,2], L. Minh Dang [f,g,2]

[a] School of Computer Science, Qufu Normal University, Rizhao, 276826, China
[b] Department of Computer Science, Brains Institute Peshawar, Peshawar, Pakistan
[c] Department of Computer Science and Engineering, Sejong University, Seoul 05006, Republic of Korea
[d] GeoSpatial Sensors and Unmanned Platforms (GeoSUP) College of Engineering, University of Tehran, Tehran, Islamic Republic of Iran
[e] Department of Architectural Engineering, Sejong University, Seoul, 05006, Republic of Korea
[f] The Institute of Research and Development, Duy Tan University, Da Nang, 550000, Viet Nam
[g] Faculty of Information Technology, Duy Tan University, Da Nang, 550000, Viet Nam

ARTICLE INFO

Keywords:
Brain tumor classification
Attention mechanism
Magnetic resonance imaging (MRI)
Hybrid models
Transformer
CNN

ABSTRACT

Brain tumors are reported more frequently among children of older age, with cases constantly increasing among older adults. Although automated systems for MRI analysis have advanced, their routine clinical adoption remains limited in many settings. The current methods are mostly based on deep and complex networks, which require high computational power and specialized equipment that may not be readily available in hospitals or clinics with limited resources. We devised TumorNet, a minimal hybrid network that performs classification of brain tumors in an efficient manner without a decrease in accuracy. The model begins with a lightweight convolutional front end to capture detailed structural information in MRI scans. Its application requires multi-scale convolutional filters to identify tumor characteristics at different scales, and a Squeeze-and-Excitation module focuses on areas of importance and downplays those that are not informative. To be able to identify long-range dependencies and global-level patterns, a MobileViT-XXS backbone uses transformer-style representations that are compatible with very lightweight models. The training process takes two phases: first, the transformer backbone is frozen to stabilize learning, then it is fine-tuned to the task. TumorNet achieves a test accuracy of 99.85% on the primary MRI dataset and generalizes with outstanding accuracy of 99.91% and 100% on two public datasets. The model has only 1.26 million trainable parameters, making it lightweight and computationally inexpensive. An interactive module coupled with the Gemini API is used to produce clinically informed explanatory reports that can help users interpret the predictions and visual evidence of the model.

1. Introduction

One of the most severe and life-threatening neurological disorders that cause a significant impact on morbidity, mortality, and quality of life on a global scale is brain tumors. The Central Brain Tumor Registry of the United States estimates that there are

* Corresponding author.
Email addresses: hanxiang@qfnu.edu.cn (H. Wang), muhammadzaqeem@gmail.com (M. Zaqeem), muhammadfayaz@sju.ac.kr (M. Fayaz), defu.qiu@qfnu.edu.cn (D. Qiu), sajjad.ahadzadeh@gmail.com (S. Ahadzadeh), tnnguyen@sejong.ac.kr (T.N. Nguyen), danglienminh@duytan.edu.vn (L.M. Dang).
[1] Hanxiang Wang and Muhammad Zaqeem are co-first authors and contributed equally to the work.
[2] These authors are joint corresponding authors for this work.

https://doi.org/10.1016/j.ins.2026.123423
Received 31 December 2025; Received in revised form 24 March 2026; Accepted 24 March 2026
Available online 27 March 2026

The backbone is used as a feature extractor with its classification head removed; its output F_{vit} is the global embedding. The backbone is initially frozen and later fine-tuned. Mathematically:

$$F_{vit} = \text{MobileViT\_XXS}(I') \approx \text{MHSA}(W_q X, W_k X, W_v X), \tag{4}$$

where W_q, W_k, W_v are learnable projection matrices inside the transformer's attention blocks.

3.3. Attention-based fusion and classification

Fusion of local multi-scale features F_{ms} and global transformer features F_{vits} occurs in a shared latent space in the Attention Fusion Module (AFM), a second transformer. This module supports dynamic interactions among both the local and global features, whereby the network can modify the significance of local representations based on the global context. This cross-scale analysis is necessary to identify any fine-tuning of tumor differences, as well as long-range relationships in MRI images. Both descriptors are projected to a common hidden dimension (128 in the implementation):

$$C = W_c F_{ms}, \qquad V = W_v F_{vit}. \tag{5}$$

The two vectors are stacked as a two-token sequence $S = [C; V]$ and serve as input to scaled dot-product attention:

$$Q = W_Q S, \quad K = W_K S, \quad V' = W_V S, \tag{6}$$

$$A = \text{Softmax}\left(\frac{QK^\top}{\sqrt{d}}\right). \tag{7}$$

Attention mechanism in this case acts as a second separate transformer layer. It finds an input in the multi-scale local and global feature tokens, it successively performs cross-feature self attendance and an output is provided in the contextual fusion representation, F_{fused}.

The attention-weighted values are aggregated by averaging across tokens and projected to yield the fused embedding:

$$F_{fused} = \text{Proj}\big(\text{Mean}(A\,V')\big). \tag{8}$$

The fused vector F_{fused} is fed to a compact classification head: a three-layer MLP with dimensions $128 \rightarrow 96 \rightarrow 64$, ReLU activations, dropout (0.3 after the first dense layer, 0.2 after the second), and a final linear map to the number of classes. Probabilities are obtained via softmax:

$$P(y \mid x) = \text{Softmax}(W_{cls} F_{fused} + b_{cls}). \tag{9}$$

All hyperparameters (SE reduction $r = 8$, branch channels, fusion hidden dimension, classifier size, MobileViT usage) are implemented as in the codebase. Training specifics (loss, optimizer, scheduler, early stopping) are detailed in the Experiments Section 4.

3.4. Conversational diagnostic reasoning

This framework also integrates a Conversational Diagnostic Reasoning Module powered by a large language model (Gemini LLM). Predicted labels, probabilities, and attention maps are converted into structured prompts and processed to produce concise, human-readable explanations of model decisions. This module does not alter classifier outputs but allows interactive interrogation of the model logic, improving transparency, interpretability, and user confidence.

4. Experimental results and discussion

We provide a comprehensive exposition on the issue of assessing the efficiency and effectiveness of the TumorNet. It gives a description of the datasets that were used with the framework proposed and the experimental setup that was used to conduct the training and testing process in detail. Then, the metrics of the evaluation to be picked to consider the performance in comparison with other methods. Additionally, the analysis includes the quantitative and qualitative results for comparing the proposed framework with state-of-the-art methods and the visualized results of the proposed framework during testing time. Finally, the section also presents the ablation study, which comprehensively evaluates and validates the proposed framework insights, including the real-time performance, diagnostic reliability, and the computational efficiency.

4.1. Datasets

The proposed **TumorNet** framework was evaluated and cross-validated on multiple publicly available MRI-based brain tumor datasets to ensure robust performance and strong generalization across varied clinical conditions. For the experimental validation of TumorNet, we utilized three different Brain MRI Tumor dataset Brain Tumor MRI Dataset [32], MRI for Brain Tumor with Bounding Boxes [33], and Labeled MRI Brain Tumor Dataset [34], which are publicly available.

4.2. Primary dataset

The *Brain Tumor MRI Dataset* consists of 7023 Brain-MRI images, which were developed by Masoud Nickparvar [32]. The dataset consists of high-resolution images categorized into four clinically significant classes: Glioma, Meningioma, Pituitary Tumor, and

Table 2
Class-wise distribution of the primary dataset.

Class	Total	Train	Test
Glioma	1621	1321	300
Meningioma	1645	1339	306
Pituitary Tumor	1757	1457	300
No Tumor (Healthy)	2000	1595	405
Total	7023	5712	1311

Table 3
Comprehensive quantitative comparison of recent state of the art models on the primary brain tumor MRI dataset. Best results are highlighted in bold.

	Author [Year]	Backbone	Params (M)	Accuracy	Macro Precision	Macro Recall	Macro F1	Weighted Precision	Weighted F1
CNN	Hossain et al. [35]	DenseNet121	22.9	0.9695	0.9679	0.9669	0.9671	0.9699	0.9694
	Kumar et al. [36]	ResNet50	25.6	0.9542	0.9543	0.9504	0.9505	0.9572	0.9540
	Swati et al. [37]	EfficientNet-B0	4.3	0.9718	0.9699	0.9694	0.9695	0.9721	0.9717
	Ghassemi et al. [38]	InceptionV3	23.9	0.9329	0.9294	0.9277	0.9282	0.9332	0.9327
	Alanazi et al. [39]	ResNeXt101	20.2	0.9588	0.9564	0.9555	0.9558	0.9584	0.9585
	Patel et al. [40]	DenseNet169	15.7	0.9789	0.9776	0.9768	0.9770	0.9788	0.9787
	Chen et al. [20]	CNN–CapsuleNet	50.3	0.9820	0.9810	0.9790	0.9800	0.9830	0.9818
	Tariq et al. [41]	RegNetY-8GF	39.5	0.9780	0.9771	0.9758	0.9762	0.9779	0.9779
	Singh et al. [42]	ResNet152	60.2	0.9875	0.9868	0.9860	0.9863	0.9875	0.9873
	Kar et al. [43]	EfficientNetV2	25.3	0.9902	0.9888	0.9879	0.9882	0.9903	0.9901
	Khan et al. [44]	ConvNeXt-T	27.9	0.9898	0.9880	0.9875	0.9877	0.9899	0.9897
	Pacal et al. [45]	RegNetX-6GF	32.8	0.9890	0.9875	0.9869	0.9872	0.9891	0.9889
	Afroj et al. [46]	DenseNet201	20.6	0.9885	0.9869	0.9861	0.9865	0.9887	0.9884
Transformer	Liu et al. [47]	ViT-B/32	44.7	0.9741	0.9725	0.9719	0.9722	0.9740	0.9740
	Zhang et al. [48]	Vision Transformer	32.1	0.9790	0.9785	0.9770	0.9777	0.9795	0.9788
	Hanif et al. [49]	ViT-Small	9.4	0.9862	0.9854	0.9848	0.9849	0.9863	0.9861
	Ahmed et al. [50]	ViT-GRU	22.4	0.9897	0.9700	0.9700	0.9700	0.9710	0.9708
	Proposed (TumorNet)	**CNN + MobileViT**	**1.26**	**0.9985**	**0.9950**	**0.9950**	**0.9950**	**0.9955**	**0.9954**

No Tumor (healthy). During development, images were collected from various medical centers. There are numerous kinds of MRI modalities in the dataset including T1-weighted, T2-weighted, FLAIR, and DWI sequences, which provide structural and pathological information. No CT images are included. To maintain reproducibility, the data were split into a training set of 5712 images and a testing set of 1311 images. The class-wise distribution is provided in Table 2.

Even though MRI is in principle a volumetric modality, the data only contains pre-extracted two dimensional slices and not three dimensional volumes. The training and testing were performed on all the available slices of the dataset, and no manual or automatic slice selection was undertaken. TumorNet was also trained and tested at the slice level, without volumetric reconstruction or inter-slice modeling.

The 80%20% train-test split was chosen according to previous research and general practice and offers an adequate amount of training data and independent cases to train and test generalization. The same division was always used for the comparison of TumorNet with the current methods Table 3 and guaranteed fair and reproducible evaluation.

Additionally, the dataset samples for each class are given in Fig. 2. While the images are visually consistent, subtle intra-class variations exist, reflecting realistic clinical conditions and posing challenges for classification. This may affect the model's performance in real-world scenarios.

4.3. Cross-dataset 1

The labeled-mri-brain-tumor-dataset consists of 2175 MRI scan images that fall under four categories: glioma, meningioma, pituitary tumors and normal brain scans. There were several sources of data with one of them being Hayatabad Medical Complex (Peshawar) which brings variance in the acquisition procedures and imaging conditions. All the images in this dataset were only externally tested but not trained, validated or internally split to have a strict cross-dataset testing configuration.

4.4. Cross-dataset 2

The MRI for Brain Tumor with Bounding Boxes dataset was developed and curated by Ahmed Sorour [33] and is publicly available. It consists of 5249 MRI scan images covering diverse information. The dataset is annotated with bounding boxes and provides the ground truth for classification and localization tasks covering four different categories, including Glioma, Meningioma, Pituitary, and Normal (No Tumor). During the annotation, the annotations were handled by experienced annotators, ensuring the accuracy and reliability with correct label localization. Similarly, the dataset comprises MRI slices from axial, coronal, and sagittal planes, providing broad anatomical coverage. The detailed distribution of the dataset is provided below. The dataset was split into training and validation sets, consisting of 4737 and 512, respectively. The number of images in each class comprises 1153, 1449, 711, and

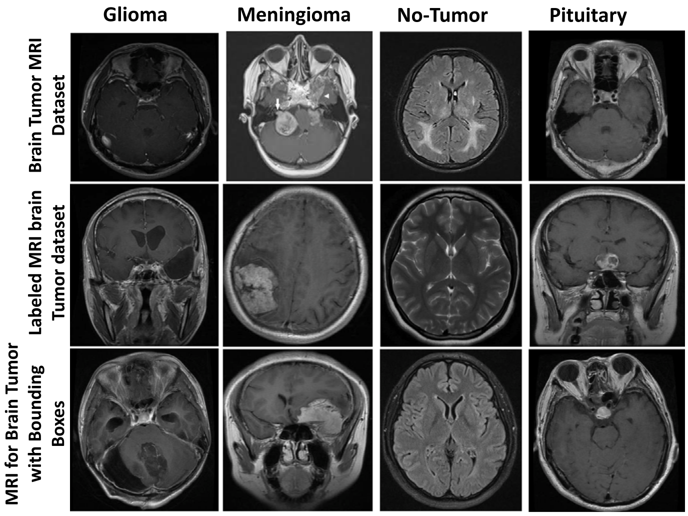

Fig. 2. Sample MRI images from three brain tumor datasets primary Dataset, Cross dataset 1, and Cross dataset 2 classes: (a) Glioma, (b) Meningioma, (c) No Tumor, and (d) Pituitary.

1424 in Glioma, Meningioma, No Tumor, and Pituitary, respectively, in the training set. While in the validation set, which we used for cross-validation as testing, it consists of 136, 140, 100, and 136 in Glioma, Meningioma, No Tumor, and Pituitary.

The primary dataset served as the training data for the TumorNet framework development during the experimental phase. While the remaining two datasets were utilized for cross-validation, the proposed framework performed well on the cross-validated dataset, particularly in terms of real-time predictions across various dataset distributions.

4.5. Experiment setup

All experiments were conducted on the Kaggle cloud platform equipped with dual NVIDIA Tesla T4 GPUs and an Intel Xeon processor using the PyTorch framework with CUDA support. The proposed **TumorNet** model was trained and evaluated on the Brain Tumor MRI dataset comprising four categories: Glioma, Meningioma, Pituitary Tumor, and No Tumor. All of the MRI images were comprehensively re-sized to 224×224 pixels with pre-normalization from ImageNet statistics. Also, the network was trained for a total of 50 epochs with a batch size of 16 using the AdamW optimizer with a learning rate of 2×10^{-4} along with a weight decay of 1×10^{-5} with a cosine annealing learning rate scheduler. Moreover, label smoothing with $\epsilon = 0.1$ was implemented to combat over-confidence, along with eight epoch patience for early stopping. Furthermore, the MobileViT network was frozen for the initial three training epochs followed by fine-tuning. Lastly, with respect to robustness issues, random resizing, random flipping, along with the implementation of color jittering and random erasing were employed as augmentation techniques. There was smooth convergence with approximately 1.26 million trainable parameters, sufficient for achieving an effective trade-off among convergence, efficiency, and network performance.

4.6. Evaluation metrics

The outcome of the suggested model has been gauged with standard classification metrics comprising accuracy, precision, recall, and F1-score. In order to deal with class imbalance, macro as well as weighted averages were provided. To make it easier, the following terms are used: **True Positive (TP)** is the identification of positive cases that are confirmed by the model, **False Positive (FP)** refers to those negative cases, which are falsely classified as positive, **True Negative (TN)** stands for the model correctly identifying the negative cases, and **False Negative (FN)** points to positive samples being mislabeled as negative.

1) Accuracy: Measures the overall correctness of the model:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

2) Precision: Indicates the proportion of correctly predicted positive samples among all predicted positives:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{11}$$

3) Recall (Sensitivity): Represents the proportion of actual positives correctly identified by the model:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{12}$$

4) F1-Score: Provides the harmonic mean of precision and recall, offering a balanced measure of both:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{13}$$

4.7. Performance evaluation of TumorNet

We conduct an extensive performance analysis of the novel **TumorNet** system against the latest advanced state-of-the-art (SOTA) deep learning techniques. The comparison includes not only quantitative analysis but also visual analysis to demonstrate the effectiveness, stability, and efficiency of the TumorNet method for brain tumor classification tasks.

4.7.1. Quantitative comparison

Table 3 presents a quantitative comparison of TumorNet with recent CNN- and Transformer-based architectures on the Brain Tumor MRI dataset. In order to compare the models fairly, the same experimental conditions were employed for all the models, which included identical dataset splits and preprocessing pipelines. In particular, a constant 80% 20% train-test split was always used on TumorNet and on all the compared state-of-the-art models, to ensure a fair and reproducible evaluation protocol.

The proposed TumorNet is leading in performance on all the evaluation metrics, such as accuracy, precision, recall, and F1-score (both macro and weighted averages). TumorNet beats larger models like ViT-B/32, EfficientNetV2, ConvNeXt-T, and RegNetX that have more than 25M parameters, even though it uses only 1.26M parameters. This is a proof that TumorNet is an effective tool in balancing the requirements of computational efficiency and discriminative power. TumorNet is a classification system which can classify cancer types with higher accuracy and better generalization than those based on CNN and include DenseNet201, ResNet152, and InceptionV3. Their classification accuracy and generalization capability are overpowered chiefly by TumorNet. Transformer deep variants (ViT-Small, ViT-GRU, and Vision Transformer) are strong in performance, but they are less efficient in terms of training cost and inference time. By integrating the local and global features through a hybrid CNN-MobileViT backbone, TumorNet can retrieve more detailed spatial and contextual representations that normally can explain its quite consistent better precision and F1 score for all tumor classes everywhere.

The first point to notice from the results displayed in Table 3 is that TumorNet achieves the **highest accuracy of 99.85%**, thus leading the way beyond all the other methods. The close match between the macro and weighted averages is another indication of the model's balanced performance across classes, thereby not showing any bias towards a particular tumor type. In addition, due to its compact architecture and low parameter count, TumorNet can be easily integrated into the real-time clinical diagnostic systems, where both accuracy and efficiency are of paramount importance.

Comparison with Transformer-Based Models: Transformer based backbones such as ViT-B/32 and EfficientNetV2 have been able to extract features very well in medical imaging. Nevertheless, due to their high model complexity and computational cost, they are not very suitable for lightweight medical applications. On the other hand, TumorNet is able to deliver better results with only a small fraction of the parameters. Without a doubt, the model outperforms ViT-B/32 (97.41% accuracy) and EfficientNetV2 (99.02% accuracy) while the computational overhead is still at a minimum.

Along with the comparative outcomes, the training behavior of TumorNet was investigated to check for stability and convergence. Fig. 3 presents the accuracy of training and validation over epochs. The two curves demonstrate a steady increase and merge smoothly without any fluctuations, thus pointing to a stable optimization process and good generalization. Moreover, the tight fit between the curves serves as a further indication that the model is not overfitted during the training phase.

Comparison with Deep CNN Models: Traditional CNN architectures such as DenseNet, ResNet, and InceptionV3 are capable of producing fair results but they encounter the problem of vanishing gradients and require a lot of memory when they are scaled deeper. TumorNet resolves these problems by efficient feature fusion in its hybrid CNN-MobileViT architecture, thus, it leads to higher accuracy (99.85%) and a more balanced F1-score (0.9950) besides parameter count is reduced drastically. Quantitative analysis, in general, is a proof that TumorNet is at the forefront of technology in terms of performance with great model efficiency and capability for generalization.

4.7.2. Qualitative results

To validate the performance of TumorNet, we conducted a qualitative analysis, which provides insights into the real-time results of the proposed framework. These results highlight the generalization in real-time, emphasizing the applicability to real-time devices. Followed by the interpretability analysis, which explains the visualization of how the network focuses and filters the important regions, enabling the framework to accurately predict and classify. Finally, the chatbot integration further enhances the clinical interpretability and makes the predictions more transparent.

The Tumor Net is evaluated with the primary dataset and two other datasets during testing time. The performance on these datasets is provided by the Table 4. The results present the class-wise metrics for the datasets. It can be seen that the Tumor Net predicts with higher precision, recall, and f1-score on the primary dataset on the testing split, highlighting its generalization. Specifically, the recall in the No Tumor class achieved 100%, which clearly distinguishes it from the tumor classes, enabling the system to avoid false negative predictions. Similarly, the precision and F1-score also achieved 99.94% and 99.97% respectively for the No Tumor class. Moreover, during cross-validation, the Tumor Net showed outstanding performance as it was able to achieve 100% precision, recall, and f1-score for all tumor classes, thereby substantiating the correctness of the predictions in real-time on the new data. Finally, the framework on the third cross-validation dataset was able to achieve a precision, recall, and f1-score average of 99.91%, 99.92%, and 99.91% respectively for all the classes, thus, confirming the excellent performance of the proposed framework on the new data.

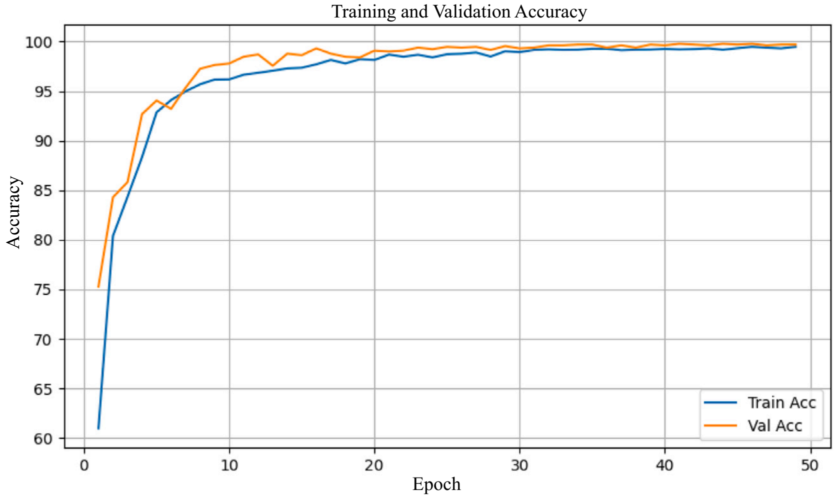

Fig. 3. Training and validation accuracy curves of TumorNet across epochs.

Table 4
Unified classification report on the primary test split and two cross datasets.

Class	Dataset 1: Brain Tumor MRI Test Split (1,311 images)			Dataset 2: Cross-Dataset 2 (2,175 images)			Dataset 3: Cross-Dataset 1 (512 images)		
Metric	Precision (%)	Recall (%)	F1-score (%)	Precision (%)	Recall (%)	F1-score (%)	Precision (%)	Recall (%)	F1-score (%)
Glioma	99.78	99.67	99.72	100.00	100.00	100.00	100.00	100.00	100.00
Meningioma	99.85	99.42	99.63	99.64	100.00	99.82	100.00	100.00	100.00
Pituitary	99.56	99.67	99.61	100.00	99.68	99.84	100.00	100.00	100.00
No Tumor	99.94	100.00	99.97	100.00	100.00	100.00	100.00	100.00	100.00

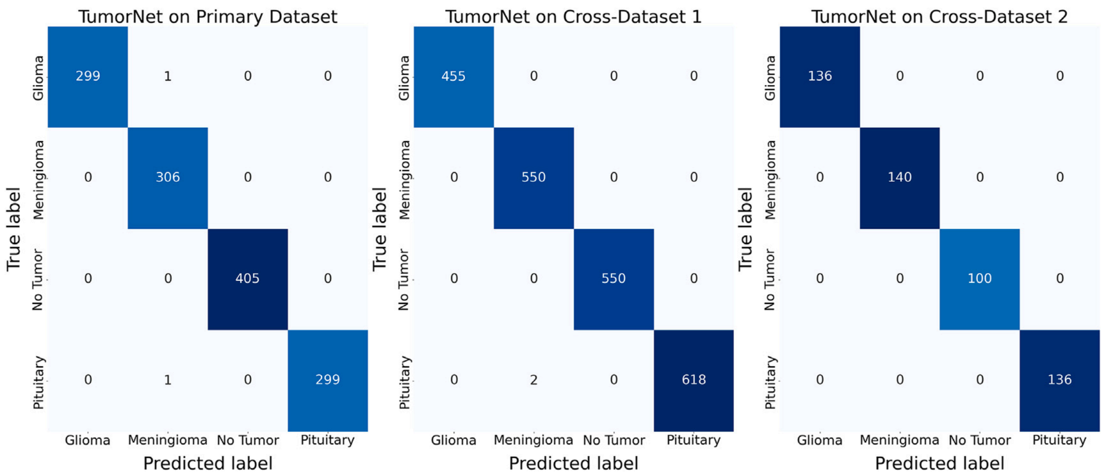

Fig. 4. TumorNet has a stable classification rate and good generalization across many different data distributions as shown by the combined confusion matrices of the primary test split(1,310) and two cross-dataset experiments (Cross-Dataset 1 = 2175 and Cross-Dataset 2 = 512).

To illustrate further the model prediction consistency, Fig. 4 shows the confusion matrices for the main dataset and both cross-dataset evaluations in one plot. This exhaustive overview indicates the class-wise classification accuracy for each tumor category and, thus, confirms the network stability as well as strong generalization capability when it is applied to unseen data from different distributions. The model's capability for making predictions in real-time is substantiated by Fig. 5 which displays the sample input images along with their corresponding predictions and ground truth labels. This representation is a vivid demonstration of the model's classification performance on the individual instances, thus, proving that the predictions and actual labels are in accordance with each other across different tumor categories. These prediction examples are supported by the quantitative results from the confusion matrices and hence, they provide tangible proof of the model's consistent performance in clinical simulation scenarios, thus, increasing the trustworthiness of the model for practical diagnostic applications.

Interpretability Analysis: Gradient-weighted Class Activation Mapping (Grad-CAM) was utilized to make figures understandable without the help of the text and to reduce the black-box nature of the proposed **TumorNet** framework. Grad-CAM images showed the

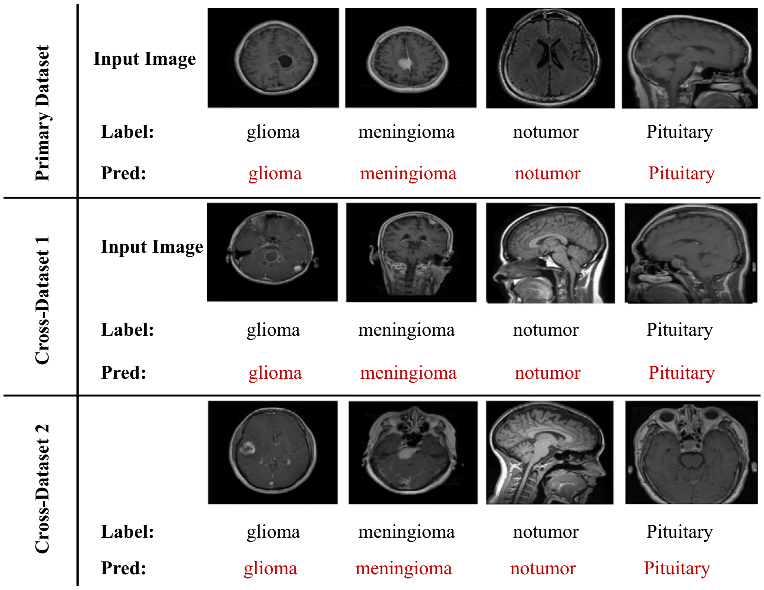

Fig. 5. Example predictions from TumorNet showing input images with corresponding model predictions and ground truth labels. The results demonstrate correct classification across different tumor types including glioma, meningioma, pituitary, and no tumor cases, illustrating the model's practical performance on individual samples.

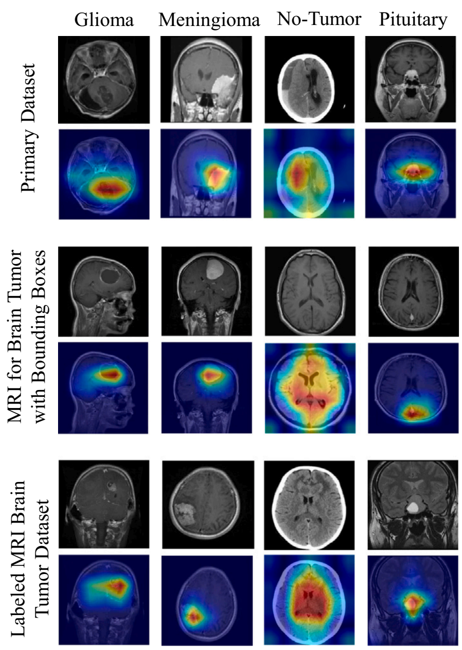

Fig. 6. Qualitative interpretability comparison for four different tumor categories by means of MobileViT encoder-based Grad-CAM visualizations. In each pair, the first picture shows the original MRI (Top), and the second one is its respective Grad-CAM heatmap (Bottom). The highlighted areas represent the parts of the image that contributed the most to TumorNet's decision. In the case of glioma and meningioma, the model focuses directly on the lesion areas; for pituitary, the activations are around the sella turcica; and for non-tumor scans, the activations are scattered and of low intensity, which is typical of normal tissues.

particular parts of the input that the model's decision was most heavily based on. The Grad-CAM images refer to the last convolutional layer of the MobileViT encoder, which combines the most abstract semantic representations needed for the classification task. The most advanced attention maps of these states were later connected to the corresponding MRI slices via image registration, thus showing the spatial areas that were the main contributors to the final prediction.

Fig. 6 also shows a clear before-after comparison in four different classes, namely, Glioma, Meningioma, No Tumor, and Pituitary as illustrated in the visualization. In the case of glioma, the network is centered on the irregular and heterogeneous masses, which are the tumor, successfully capturing the outline of the abnormal mass with high-intensity activations. In the case of meningioma, the Grad-CAM identifies the clear and localized tumor edges, which establishes that the model is focusing on the focal area of the lesion while dimming other brain tissues. The emphasized spot in the case of a pituitary tumor is located around the middle and lower part of the brain- which anatomically points towards the sella turcica, which shows that the model has identified the pituitary growths in the right place. On the other hand, the activations are small and diffused in the case of a non-tumor MRI scan, indicating that the model identifies healthy brain structures and suppresses non-diagnostic regions. All in all, the above interpretability findings indicate that TumorNet is not only capable of achieving high classification accuracy but also possesses clinically consistent spatial reasoning. The visual confirmation of the heatmaps shows that the focus of the model corresponds to radiologically meaningful areas, reinforcing its transparency and potential utility in real world medical diagnostic workflows.

Chatbot Integration for Clinical Assistance: To simplify TumorNet and make it easy to comprehend and apply in practice, we have created a chatbot built on Streamlit that connects the classifier of TumorNet with a language model based on Gemini. The chatbot is a system that is activated at the end of the classification phase and is not a decision-making one but an explanation interface. TumorNet classifies based on learning discriminative local and global visual patterns of brain MRI scans, such as lesion boundaries, texture heterogeneity, and anatomical localization, which are the foundations of tumor class predictions. The chatbot uses the output predicted class, confidence level, and probability scores that TumorNet generates together with these learned visual cues and translates them into summaries in natural language.

Upon uploading a brain MRI photo, the system shows the chosen type of tumor (glioma, meningioma, pituitary, or non-tumor) and the corresponding confidence score. After this, the chatbot develops an explanatory summary of the visual evidence behind the prediction directly by linking the predicted classification with the morphological and textual features as determined by the model. The system also allows users to ask questions to discover the rationale behind the decision, differences between tumor classes, or implications in a medical setting in general. As an example, the chatbot could describe how, in the case of glioma, the prediction is linked to irregular lesion boundaries and heterogeneous textures, whereas meningioma or pituitary cases are related to more regular contours and anatomically localized regions.

Such an interaction shows that the acquired representations are meaningful visual patterns. The chatbot enhances transparency by translating numerical model results into explanations that can be understood and serves as a research, educational, and interpretability resource for students, trainees, and researchers. It is not designed to be used in clinical diagnosis or patient-facing decision support.

Example Chatbot Interaction To demonstrate the practical usage of the Gemini-based chatbot, consider the following interaction after TumorNet predicts a glioma case with 98% confidence:

- **User:** “Why did this scan get the gioma classification?”
- **Chatbot:** “The model has found irregular lesion margins and irregular texture patterns in the temporal lobe, which are typical of glioma as compared to meningioma or pituitary tumors that have characteristically regular margins.”
- **User:** “To what extent is this prediction certain?”
- **Chatbot:** “The classification confidence is 98% which indicates a high level of certainty based on the extracted spatial and contextual features.”
- **User:** “What differentiates it from a meningioma?”
- **Chatbot:** “he cases of meningioma are usually characterized by well-circscribed and homogenous lesions located in the proximity of the meninges. No visual indicators are present in this scan nowadays.”

The examples demonstrate the benefits of the chatbot: the model outputs are translated into clear explanations, the thought process behind the predictions is explained, and the user is able to notice the differences between the types of tumors in an interactive format.

4.8. Ablation study

An extensive ablation study was conducted within the unified TumorNet framework to assess the individual contributions of architectural components and backbone variations. The summarized Table 5 shows the experimental results. The first part examines the impact of the removal of particular modules. The exclusion of either multi-scale CNN or attention fusion blocks leads to a notable decrease in accuracy thereby confirming that both local and contextual representations are necessary for robust tumor discrimination. The SE block, however, provides a smaller yet consistent advantage.

Whereas the contribution of the Attention Fusion Module (the second transformer) to increasing the accuracy is relatively small (99.77 to 99.85 Prestige), the module remains central to the dynamic combination of local multi-scale features and global transformer representations. The cross-scale interaction aided by the module allows the neural network to better localize tumors by incorporating global contextual information, especially when it comes to distinguishing subtle differences in tumors and long-scale dependencies at a distance that are particularly characteristic of MRI data. By extension, the AFM generates qualitative improvement in the feature representation despite the comparatively minor improvement in the total accuracy.

The second part of the work studies different CNN and Transformer backbones that are combined under the same hybrid configuration with the training kept constant. DenseNet121 and EfficientNet-B0 offer the most powerful performance-to-parameter ratio while Swin-Tiny and ConvNeXt-Tiny can achieve similar accuracy but at a significantly higher computational cost. Thus, the findings are consistent with the idea that TumorNet delivers the best trade-off between accuracy and parameter efficiency. The use of multi-scale CNN feature extraction in combination with a lightweight attention fusion module represents an ideal trade-off between discriminative performance and computational cost.

Table 5
Unified ablation study analyzing the influence of architectural modules and backbone substitutions within the TumorNet framework.Where MR, Acc, and F1 represent Module Removal, Accuracy (%), and F1-Score (%).

Variant / Backbone	Category	Acc	F1
Backbone Substitution Study			
ViT-Tiny (Patch16)	Transformer	95.27	95.10
Swin-Tiny (Patch4)	–	98.63	98.55
EfficientNet-B0	CNN	**99.77**	**99.75**
DenseNet121	–	**99.77**	**99.75**
GoogLeNet	–	99.62	99.59
ConvNeXt-Tiny	Hybrid	99.69	99.67
Component-Level Analysis			
Without SE Blocks	MR	99.82	99.80
Without Multi-Scale CNN	MR	99.69	99.67
Without Attention Fusion	MR	99.77	99.76
MobileViT Only	Reduced Variant	99.54	99.50
Full Model (TumorNet)	Baseline	**99.85**	**99.84**

Table 6
Model complexity and inference performance of the proposed TumorNet.

Metric	Value	Unit
Trainable Parameters	1.26	Million
Model Size	5	MB
Computational Cost	0.39	FLOPs
Average Inference Time	10.79	ms

4.9. *Model complexity and inference efficiency*

Table 6 gives a summary of the computational complexity and the inference performance of the *TumorNet* model, which is our proposal. TumorNet is essentially a very efficient tool for real-time clinical use as it has only 1.26M parameters and a small memory size of 5 MB. The minimal computational cost of 0.39 FLOPs and the average inference time of 10.79 ms per image are great examples of how a balance between accuracy and runtime efficiency can be achieved on both GPU and CPU platforms.

5. Conclusion and future direction

We present the TumorNet Framework, which uses MRI images to classify brain tumors. This architecture is able to leverage the ability of convolutional neural networks for spatial feature extraction while capturing the global understanding of vision transformers with a high degree of accuracy of 99.85% for the initial database, as well as excellent generalization skills for two public databases of MRI images. Additionally, for improving the clinical credibility and explainability of the proposed solution, a conversational diagnostic reasoning component enabled with a large language model is incorporated into the system that would enable doctors to engage with the machine learning solution, examine the rationale behind the image classification, and precisely provide medically valid interpretations for the classifications carried out by the solution with their clarity of explanation.

CRediT authorship contribution statement

Hanxiang Wang: Writing – original draft, Methodology, Data curation, Conceptualization. **Muhammad Zaqeem:** Writing – original draft, Methodology, Data curation, Conceptualization. **Muhammad Fayaz:** Writing – review & editing, Validation, Formal analysis. **Defu Qiu:** Writing – review & editing, Investigation. **Sajjad Ahadzadeh:** Writing – review & editing, Visualization, Formal analysis. **Tan N. Nguyen:** Supervision, Resources, Project administration, Funding acquisition. **L. Minh Dang:** Supervision, Software, Resources, Project administration, Investigation, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the Young Scientists Fund of the National Natural Science Foundation of China under Grant 62502271 and Grant 62503276, in part by the Young Scientists Fund of the Natural Science Foundation of Shandong Province under Grant ZR2025QC630 and Grant ZR2025QC670, and in part by the Natural Science Foundation of Rizhao City under Grant RZ2024ZR33. We also want to express our gratitude for the AI tools that ensured content clarity throughout.

Data availability

Data will be made available on request.

References

[1] Q.T. Ostrom, J.S. Barnholtz-Sloan, et al., Primary brain and other central nervous system tumors diagnosed in the United States in 2014–2018, Neuro-Oncology 23 (Supplement 3) (2021) iii1–iii105.
[2] S. Bauer, M. Reyes, Magnetic resonance imaging of brain tumors: a review, Neuro-Oncology 24 (3) (2022) 345–357.
[3] B.H. Menze, et al., The multimodal brain tumor image segmentation benchmark, Med. Image Anal. 25 (2021) 135–145.
[4] A.K. Anaraki, M. Ayati, Magnetic resonance imaging-based brain tumor classification using machine learning, Artif. Intell. Med. 104 (2020) 101813.
[5] Z. Akkus, D. Rubin, Deep learning for brain MRI analysis, J. Digit. Imaging 35 (2022) 278–289.
[6] X. Wang, et al., Non-local neural networks, IEEE Trans. Pattern Anal. Mach. Intell. 44 (4) (2022) 1711–1724.
[7] A. Dosovitskiy, et al., An image is worth 16x16 words: transformers for image recognition at scale, in: International Conference on Learning Representation, 2021, https://openreview.net/forum?id=YicbFdNTTy.
[8] X. Chen, et al., Efficient deep learning models for medical image analysis, IEEE Rev. Biomed. Eng. 16 (2023) 45–60.
[9] A. Hatamizadeh, et al., Hybrid CNN–transformer architectures for medical image analysis, Med. Image Anal. 86 (2024) 102802.
[10] G. Litjens, et al., A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88.
[11] Z. Akkus, et al., Deep learning for brain MRI segmentation: state of the art and future directions, J. Digit. Imaging 30 (4) (2017) 449–459.
[12] A. Hatamizadeh, et al., Unetr: transformers for 3D medical image segmentation, Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (2022) 574–584.
[13] B. Urooj, M. Fayaz, S. Ali, L.M. Dang, K.W. Kim, Large language models in medical image analysis: a systematic survey and future directions, Bioengineering 12 (8) (2025) 818.
[14] K. Rao, H. Singh, Handcrafted features versus deep features in brain tumor classification, Pattern Recognit. Lett. 120 (2019) 45–52.
[15] F. Mehmood, R. Ali, A. Khan, Fine-tuned resnet50 for glioma grading using brats-2019 MRI dataset, Biomed. Signal Process. Control 93 (2024) 105626.
[16] S. Shargunam, R. Rajakumar, Hybrid CNN-SVM model for glioma classification using MRI, Comput. Biol. Med. 164 (2024) 107402.
[17] M. Abd-Ellah, A. Awad, A. Khalaf, Cascaded CNN with u-net symmetry for brain tumor MRI classification, Expert Syst. Appl. 235 (2024) 121245.
[18] Y. Li, F. Zhao, L. Wang, Attention-augmented resnet for brain tumor MRI classification, IEEE Access 12 (2024) 58522–58534.
[19] M. Sreedevi, R. Kumar, V. Sharma, Resnet50 with bigru and dual attention for multi-class brain tumor classification, J. King Saud Univ. Comput. Inf. Sci. 36 (7) (2024) 1245–1257.
[20] J. Chen, Y. Gao, T. Sun, Fusion of densenet and resnet features for brain tumor MRI classification, Pattern Recognit. 145 (2024) 109009.
[21] S. Zebari, A. Karim, D. Saeed, Belief network-based fusion of resnet50 and vgg16 features for brain tumor classification, Expert Syst. Appl. 233 (2024) 121121.
[22] M. Haque, A. Rahman, T. Ahmed, Modified VGG19 architecture for brain tumor MRI classification, Biomed. Signal Process. Control 94 (2024) 105679.
[23] L. Wang, P. Zhang, J. Liu, Transformer-based architecture for differentiating glioblastoma and CNS lymphoma from MRI, Magn. Reson. Imaging 105 (2024) 1–9.
[24] R. Patel, P. Singh, Hybrid CNN-transformer architectures for MRI classification: balancing accuracy and efficiency, IEEE J. Biomed. Health Inform. 28 (2024) 1–10.
[25] A.K. Anaraki, M. Ayati, F. Kazemi, Magnetic resonance imaging-based brain tumor classification using deep learning techniques, Biomed. Signal Process. Control 57 (2019) 101678.
[26] E.I. Zacharaki, et al., Classification of brain tumor type and grade using MRI texture and shape features, Magn. Reson. Med. 62 (6) (2009) 1609–1618.
[27] C.J. Rao, S.S. Singh, Brain tumor classification using handcrafted features and machine learning techniques, J. Med. Syst. 43 (7) (2019) 1–12.
[28] M.A. Alam, A. Mahmood, Performance analysis of SVM-based brain tumor classification, Comput. Biol. Med. 123 (2020) 103865.
[29] S. Sultana, et al., Traditional machine learning methods for brain tumor analysis: a review, Comput. Biol. Med. 155 (2023) 106624.
[30] T. Nguyen, J. Park, Limitations of convolutional networks in modelling global context for medical images, IEEE Trans. Med. Imaging 72 (2022) 102120.
[31] H. Cheng, L. Zhao, Convolutional neural networks in medical image analysis: a comprehensive review, Artif. Intell. Rev. 85 (2023) 102415.
[32] M. Nick Parvar, Brain tumor MRI dataset, Kaggle repository (2022.
[33] A. Sorour, MRI for brain tumor with bounding boxes, Kaggle repository (2023.
[34] A. Rostami, Labeled MRI brain tumor dataset, 2024. https://universe.roboflow.com/ali-rostami/labeled-mri-brain-tumor-dataset.
[35] M.A. Hossain, M.R. Islam, N. Kurata, Brain tumor classification from MRI images using optimized densenet121 and inceptionv3, Diagnostics 13 (15) (2023) 2533.
[36] A. Kumar, R. Singh, V. Sharma, A resnet-based convolutional neural network for efficient brain tumor detection from MRI scans, Comput. Methods Programs Biomed. 240 (2023) 107703.
[37] Z.N. Swati, R. Qureshi, S. Hussain, Efficientnet-based hybrid transfer learning model for brain tumor classification, Expert Syst. Appl. 228 (2023) 120462.
[38] N. Ghassemi, S. Shojaeilangari, L. Wang, Inceptionv3 feature fusion for MRI-based brain tumor recognition, Multim. Tools Appl. 82 (9) (2023) 14127–14146.
[39] S. Alanazi, A. Alqahtani, R. Khan, Deep resnext architecture for multiclass brain tumor MRI classification, J. Biomed. Inform. 164 (2023) 107302.
[40] H. Patel, S. Sharma, R. Patel, DenseNet169-based multi-class brain tumor classification with MRI imaging, Comput. Electr. Eng. 113 (2024) 109034.
[41] M. Tariq, Z. Iqbal, A. Hussain, Regnety-based deep neural network for high-performance brain tumor MRI classification, IEEE Trans. Instrum. Meas. 73 (2024) 2500813.
[42] R. Singh, K. Sharma, J. Patel, ResNet152 with CBAM attention for enhanced brain tumor classification, IEEE Access 12 (2024) 59983–59995.
[43] P. Kar, S. Banerjee, T. Ghosh, EfficientNetV2-driven deep hybrid network for high-precision brain tumor MRI classification, Artif. Intell. Med. 153 (2025) 102013.
[44] L. Wang, X. Li, J. Zhang, Convnext-based deep visual recognition for brain tumor MRI classification, Neural Comput. Appl. 35 (2023) 18971–18983.
[45] I. Pacal, A novel hybrid deep learning model based on local and global feature fusion for brain tumor classification using MRI images, Biomed. Eng. Lett. 99 (2025) 106880.
[46] M. Afroj, S. Khan, T. Ahmed, Mobdensenet: densenet201-based hybrid architecture for high-accuracy brain tumor MRI classification, Pattern Recognit. 136 (2025) 109123.
[47] Y. Luo, H. Zhang, P. Wang, Vision transformer for brain tumor MRI classification: a comparative analysis, Pattern Recognit. Lett. 172 (2023) 55–63.
[48] Y. Zhang, J. Li, F. Zhou, Transformer-based hybrid learning model for brain tumor classification using MRI, Neural Netw. 172 (2024) 106002.
[49] M. Hanif, S.A. Raza, M. Aslam, Hybrid CNN–vit framework for lightweight and accurate brain tumor detection, Comput. Biol. Med. 169 (2024) 107622.
[50] N. Ahmed, R. Ali, S.A. Khan, Attention-guided densenet-169 for robust brain tumor MRI classification, IEEE Access 12 (2024) 45421–45433.