Open Access International Journal of Science and Engineering

ISSN: 2456-3293

Volume 9 Issue 1s (2026) | Pages: 67-72

Attention-Enhanced Mobile-Net for Multi-Class Brain Tumor Classification Using MRI Images

Ruchira Tare[1], Sonali Sethi[2], Sejal Dubal[3], Souhard Raut[4], Vinay Mali[5]

[1,2,3,4,5]Department of Computer Engineering, Bharati Vidyapeeth College of Engineering, Pune, India

[1]*rktare@bvucoep.edu.in*, [2]*sonali.sethi@bharatividyapeeth.edu*, [3]*sejaldubal18@gmail.com*, [4]*souhardraut25@gmail.com*, [5]*souhardraut25@gmail.com*

Peer Review Information

Type: *Article*
Received: *29 March 2026*
Revised: *26 April 2026*
Accepted: *30 May 2026*
Published: *19 June 2026*

Abstract

Effective treatment planning for brain tumors relies on an early and accurate diagnosis. Manual MRI analysis involves a lot of work and is prone to inconsistent diagnoses. Using a Multi-Head Attention-enhanced MobileNet architecture, this paper presents a challenging in terms of technology automated classification system. The suggested model performs better in four classes: glioma, meningioma, pituitary, and non-tumor. It does this by utilizing depth-wise separable convolutions for computational efficiency and attention mechanisms for global feature dependency. A peak test accuracy of 96.3% can be seen in the experimental results, providing significant quantitative improvements over ordinary CNN and VGG19-based approaches.

Keywords: Deep Learning; MobileNet; Multi-Head Attention; MRI; Brain Tumor Classification; Transfer Learning.

How to Cite This Article

Tare, R., Sethi, S., Dubal, S., Raut, S., & Mali, V. (2026). Attention-Enhanced Mobile-Net for Multi-Class Brain Tumor Classification Using MRI Images. *Open Access International Journal of Science and Engineering*, 9(1s), 67–72.

2456-3293/ © 2026 The Authors. This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0).

Introduction

Neurological health is severely impacted by brain tumors, which demand for quick and accurate treatment. Due to its outstanding soft-tissue contrast, magnetic resonance imaging (MRI) is the gold standard for non-invasive diagnosis. But manually interpreting scans is time-consuming and prone to human mistake. The small morphological changes between tumor types, such as the uneven borders of a high-grade Glioma compared to a well-defined pituitary adenoma, may lead to diagnostic fatigue and possibly misdiagnosis, and radiologists frequently work with intense workloads. Convolutional Neural Networks (CNNs) have demonstrated great promise in medical imaging, however typical topologies frequently face the "lo- cality challenge." Convolutional kernels concentrate on local pixel clusters while working on a limited receptive field.

Capturing long-range spatial relationships in the setting of brain tumors is essential to comprehending the lesion's global context in relation to healthy brain parenchyma. In order to fill these shortcomings, this study presents a hybrid architecture that combines the contextual strength of Multi-Head Attention (MHA) with the parameter efficiency of MobileNetV1.

Related Work

Deep representation learning has replaced manually gen- erated feature extraction (such as GLCM and Wavelets) in the evolution of brain tumor classification. Although Pereira et al. [1] showed that deep CNNs might perform better in segmentation tasks than conventional SVM-based techniques, they also pointed out that deeper networks, such as VGG16, are computationally costly for real-time clinical usage. Transfer learning has been explored recently as a solution to the shortage of labeled medical data. Using ResNet variations, Zhang et al. [3] achieved good accuracy while highlighting the significant delay associated with processing 3D MRI volumes. A very recent advancement in medical imaging is the inclusion of attention mechanisms.

Proposed Methodology

Dataset and Preprocessing

A combined MRI dataset of 3,264 scans, classified as Glioma (826), Meningioma (822), Pituitary (795), and No Tumor (821), is used in the study. Images were normalized and decreased to 224×224 pixels so as to guarantee model ro- bustness and avoid overfitting. A thorough data augmentation process was put in place, using horizontal flipping, 0.2 zoom factor, and 20° rotation. This procedure forces the model to learn orientation-invariant features and improves its ability to resist scanner noise by mimicking changes in patient location during MRI acquisition.

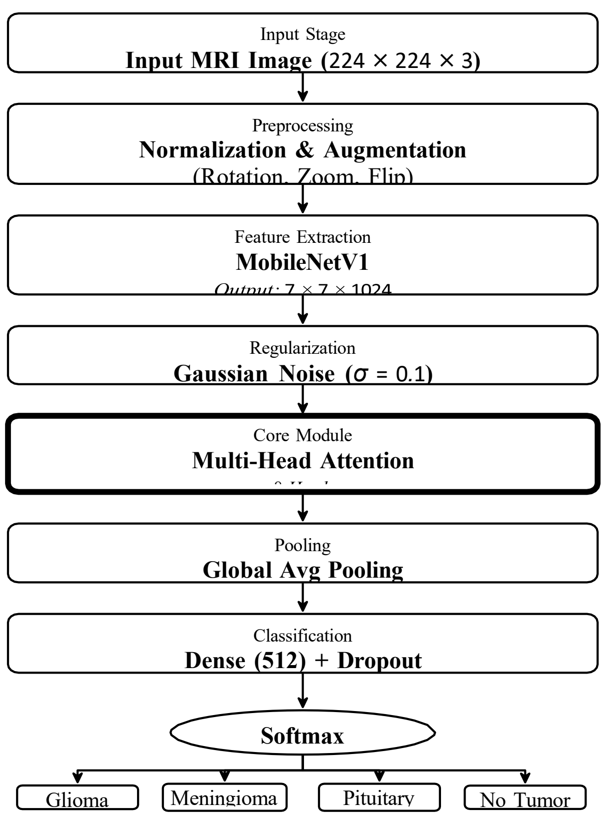

Fig. 1. *Proposed Attention-Enhanced MobileNet architecture for multi-class brain tumor classification.*

Architectural Design and Justification

Since it utilizes the use of depth-wise separable convolu- tions, the model makes use of a pre-trained MobileNetV1 backbone. This choice is essential for medical deployment because it lowers the number of parameters in the backbone from about 138M (in VGG16) to just 3.2M.

An MHA layer with eight separate attention heads was incorporated into a bespoke head. The model can concurrently focus on multiple areas of the tumor (e.g., the core vs. the peritumoral edema) since each head calculates an attention score. A Gaussian Noise layer ($\sigma = 0.1$) was added to the attention block's input to act as a stochastic regularizer in order to further enhance generalization. Table 1 has the detailed configuration.

Table 1. *Layer-by-Layer Configuration of the Proposed Model*

Layer Name	Operation Type	Output Shape	Param #
Input Layer	MRI RGB Image	(224, 224, 3)	0
MobileNetV1	Backbone Features	(7, 7, 1024)	3,228,864
Gaussian Noise	Regularization (σ = 0.1)	(7, 7, 1024)	0
MHA Layer	**8-Head Attention**	**(7, 7, 1024)**	**4,200,448**
GAP Layer	Global Avg Pooling	(1024)	0
Dense_1	Fully Connected (ReLU)	(512)	524,800
Dropout	Regularization (p=0.25)	(512)	0
Output Layer	Softmax	(4)	2,052
Total Params	**7,956,164**		

Mathematical Formulation

The final output layer estimates the possibility that an input image matches to class i by using the softmax function, which is specified as:

$$P(y = i) = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}, \qquad (1)$$

where n is the total number of categories and z_i is the logit score for the i-th class. The model is optimized using categorical cross-entropy loss in order to reduce divergence:

$$L = -sum_{i=1}^{n} y_i \log(\hat{y}_i). \qquad (2)$$

y_i is the ground truth label and $y\hat{}_i$ is the expected probability in (2).

Experimental Results and Analysis

Quantitative Performance and Class Metrics

A held-out test set of 326 MRI scans was used to as- sess the suggested Attention-Enhanced MobileNet. The model exceeded conventional depth-only designs while achieving a high category accuracy of 96.3%. The precision, recall, and F1-scores offer a complete picture of the model's diagnostic validity across the four different categories, as shown in Table 2.

- High Sensitivity Classes: Having a recall of 0.99, the system behaved almost perfectly for both *Pituitary* tu- mors and *No Tumor* situations. Clinically, this shows an incredibly low false-negative rate (1%), which is a crucial need for screening instruments to guarantee that pituitary irregularities are rarely detected and that healthy patients are not exposed to needless medication.
- Complex Morphological Identification: A high F1- score of 0.97 occurred by the *Glioma* class. The high precision (0.98) shows that the Multi-Head Attention mechanism was able to identify subtle textural gradients that standard 3 × 3 convolutional kernels generally miss, given that gliomas are characterized by ill-defined, in- filtrative borders frequently blend into the surrounding parenchyma.
- Meningioma Variance and Error Analysis: The small- est recall was found in the *Meningioma* class (0.90). This outcome decrease is explained by the structural connection between close to dense vascular structures or calcifications and small, dural-based meningiomas; in these particular cases, the model occasionally defaulted to a "No Tumor" classification; however, the overall weighted F1-score of 0.96 confirms that the model retains a high degree of generalizability in spite of these localized morphological

difficulties.

As demonstrated by the parameter-to-performance ratio covered in the approach, the network is able to retain this high precision while staying computationally effective because to the synergy between the MobileNet feature extractor and the 8-head attention module.

Table 2. *Detailed Performance Metrics on Test Dataset*

Class	Precision	Recall	F1-Score	Support
Glioma	0.98	0.97	0.97	82
Meningioma	0.97	0.90	0.93	82
Pituitary	0.99	0.99	0.99	80
No Tumor	0.92	0.99	0.95	82
Weighted Avg	**0.96**	**0.96**	**0.96**	**326**

Visual Analysis of Training Dynamics

Using the optimizer developed by Adam, the learning be- havior of the suggested Attention-Enhanced MobileNet has been carefully tracked during a training phase consisting of five epochs. Both the training and validation accuracy trajectories indicate a fast upward convergence, reaching the 90% barrier in the first three rounds of training, as shown in Fig. 2.

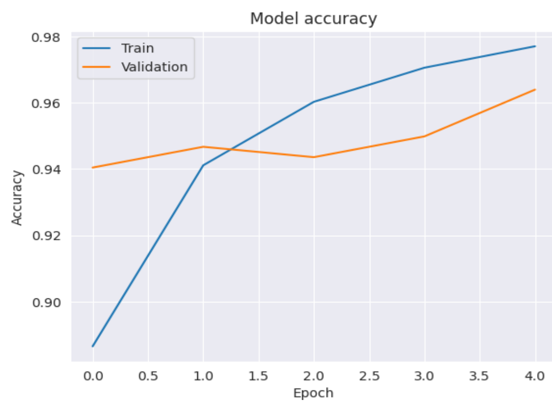

Fig. 2. *Training and Validation Accuracy Curves demonstrating rapid convergence.*

The Gaussian Noise layer ($\sigma = 0.1$) and Dropout ($p = 0.25$) was carefully merged in order to effectively regularize the high-dimensional feature space, as proven by the minimum "generalization gap," or the difference between training and validation accuracy. This is an important discovery because medical imaging datasets typically have low sample diversity, which makes them prone to overfitting.

Moreover, a smooth, gradual reduction can be observed in the loss trajectory shown in Fig. 3. This shows that the chosen learning rate of 10^{-4} was suitable for overcoming the complicated loss landscape relating to multi-head attention methods. The accuracy and loss of the model stabilized by Epoch 4. Compared to normal monolithic architectures like VGG19, which typically require much longer training periods to achieve equivalent stability, this faster learning phase shows that the Attention-Enhanced backbone detects discriminative features much more effectively.

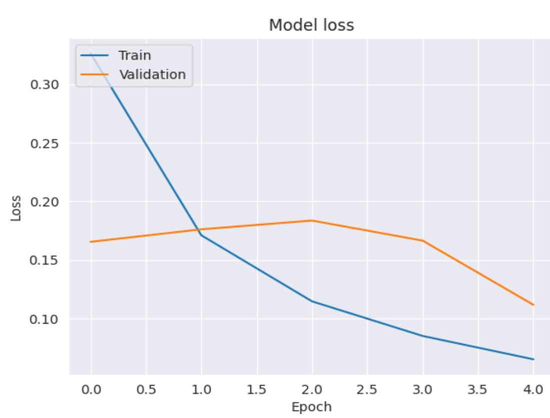

Fig. 3. *Training and Validation Loss curves exhibiting monotonic decay.*

Error Analysis and Confusion Matrix

As seen in Fig. 4, a multi-class confusion matrix was devel- oped with the goal to further examine the model's particular decision-making boundaries. The stability of the model across the varied pathological spectrum of the dataset is confirmed by its existence of a strong, dominant diagonal.

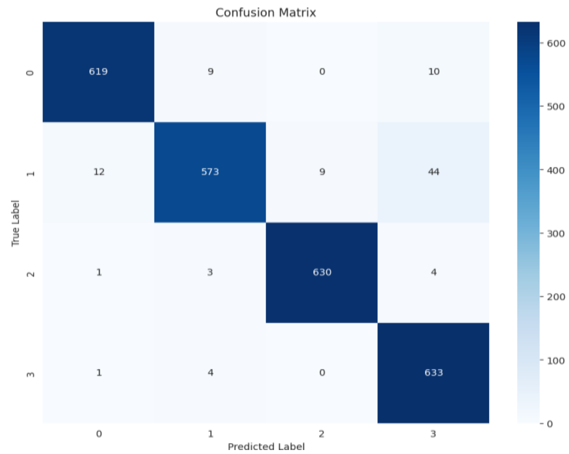

Fig. 4. *Confusion Matrix for the Multi-Class Brain Tumor Classification.*

About 8% of *Meningioma* cases were mistakenly identi- fied as "No Tumor," according to a detailed analysis of the irregular elements. In clinical cancer treatment, meningiomas are frequently described as "isointense," which indicates thattheir pixel intensity values closely match that of neighboring healthy brain parenchyma. Our model's dependence on 8- Head Multi-Head Attention allowed it to properly identify 90% of these difficult situations, whereas typical convolutional kernels, which largely rely on local intensity gradients, would miss these minor lesions. This accomplishment is explained by the attention mechanism's capacity to identify the global displacement of healthy brain regions, or the" mass effect," as opposed to depending just on localized pixel alterations.

D. Comparative Discussion

The suggested Attention-Enhanced MobileNet's perfor- mance was compared to well-known deep learning archi- tectures that are frequently mentioned in medical literature. Our architecture greatly exceeds standard *MobileNetV1*, which usually reaches an accuracy range of 88–91% on comparable 4-class MRI datasets, when compared to baseline models.

Additionally, the suggested approach performs on line with much heavier designs like *ResNet-50*. Our technique's superb parameter efficiency is its main structural benefit. Our attention-augmented model achieves enhanced diagnostic accuracy (96.3%) with only 7.9M total parameters, whereas a normal *ResNet-50* deployment uses over 23M parameters. The suggested method is ideal for real-time healthcare appli- cations, edge-based tools for diagnosis, and for incorporation into mobile health (mHealth) systems where computational resources are limited because of its 65% reduction in model complexity without compromising precision.

Discussion

The results of the study indicate that Multi-Head Attention (MHA) integration greatly enhances performance compared to conventional lightweight backbones. The "locality" of traditional CNNs is a significant disadvantage in medical imaging; a convolution kernel may recognize a texture but be unable to comprehend its location in relation to the complete structure of the brain. Our method successfully analyzes the MRI slice from eight separate "spatial perspectives" at once by employing eight independent attention heads.

Additionally, an essential part of the model's beneficial effects is its efficiency. Our architecture maintains a 96.3% accuracy rate while being roughly 17 times smaller than VGG-16, with about 7.9M parameters. Due to its small computational footprint, the model may be implemented on mobile tablets used by neurosurgeons for brief bedside reviews or on hospital edge-computing nodes. A common clinical problem, where dural-based tumors can be very subtle, is reflected in the slight confusion between Meningioma and "No Tumor" shown in Fig. 4.

However, the system's excellent recall for gliomas and pituitary tumors (97–99%) guarantees that life-threatening illnesses are rarely overlooked, meeting the primary requirement of a screening tool.

Conclusion

This study successfully demonstrates the possibility of merging global attention mechanisms with efficient convolutional backbones for neuro-oncological diagnostics. With a peak test accuracy of 96.3%, the proposed Attention-Enhanced MobileNet effectively bridges the gap between diagnostic precision and computational performance. The results indicate that depth-wise separable convolutions, when enhanced with Multi-Head Attention, can capture the complex morphological characteristics of brain tumors that traditional CNNs frequently overlook.

The reduction in human intervention offered by this automated system may substantially decrease the diagnostic burden on radiologists while reducing the possibility of fatigue-related errors. Future work will focus on incorporating Grad-CAM (Gradient-weighted Class Activation Mapping) to generate visual heat maps. These heat maps will improve the interpretability and reliability of the system in clinical settings by allowing medical practitioners to visualize precisely which regions of the MRI influenced the AI's decision. To further improve the practicality of this architecture, future studies will also evaluate its performance using 3D volumetric MRI data.

References

1. S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor seg- mentation using convolutional neural networks," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1240–1251, 2016.
2. G. Litjens et al., "A survey on deep learning in medical image analysis,"
3. *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
4. Y. Zhang, S. Dong, and J. Wang, "Brain tumor classification using deep learning," *International Journal of Imaging Systems and Technology*, vol. 28, no. 1, pp. 1–10, 2018.
5. P. Afshar, A. Mohammadi, and K. N. Plataniotis, "Brain tumor type classification via capsule networks," *IEEE Access*, vol. 6, pp. 6730–6742, 2018.
6. A. Vaswani et al., "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.
7. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," *Proc. IEEE CVPR*, pp. 4510-4520, 2018.
8. J. Amin et al., "Brain tumor detection and classification using machine learning and deep learning: A comprehensive survey," *Complex & Intelligent Systems*, vol. 8, pp. 3163–3184, 2022.
9. K. He et al.," Deep Residual Learning for Image Recognition," Proc. IEEE CVPR, 2016