# Advancing Brain Tumor MRI Classification with Interpretable Attention-Enhanced Deep Learning

***Abstract*—Neuro-oncology relies on magnetic resonance imaging (MRI) brain tumor classification since automated methods can greatly improve diagnosis. Despite advances in deep learning, long-range spatial interactions, model interpretability, and computational efficiency for real-time deployment remain obstacles. This paper introduces an attention-enhanced DenseNet201 network with Multi-Head Self-Attention (MHSA) to solve these constraints. DenseNet201 emphasizes dense feature reuse, and MHSA focuses on concentrated areas surrounding the image's spatial frame and improves discrimination. MRI images were prepared by rescaling, intensity normalization, and contrast-limited adaptive histogram equalization. Data augmentation via rotations, translations, horizontal and vertical flips, and zoom operations strengthened model generalization. Gradient-Weighted Class Activation Mapping (Grad-CAM), an explainable artificial intelligence method, produced class-specific saliency maps that visualized tumoral locations and gave clinical interpretability. Test accuracy of $0.9966$, macro-average F1-score of $0.9966$, Cohen $\kappa$ value of $1.0000$, and consistent five-fold cross-validation accuracy of $0.9734$ demonstrate superior performance. Individual MRI scan inference takes less than 24 milliseconds per image, making it suitable for real-time use. Comparative assessments of contemporary convolutional neural networks and transformer-based architectures indicate that the DenseNet201 + MHSA framework effectively classifies brain tumors with commendable accuracy and interpretability, establishing it as a viable therapeutic tool.**

***Index Terms*—Brain tumor classification, MRI, DenseNet201, Multi-Head Self-Attention, Explainable AI, Grad-CAM visualization.**

## I. Introduction

Brain tumors represent a significant global health concern, with the World Health Organization reporting over 300,000 cases of incidence annually [1], [2]. Prompt and precise diagnostic assessment is essential for devising successful treatment strategies; diagnostic inconsistency can markedly diminish survival prospects and adversely affect quality of life [3], [4]. MRI is extensively employed as a non-invasive technique for detecting cerebral neoplasms, owing to its superior soft tissue contrast capabilities [5]. Nonetheless, conventional manual analysis of MRI scans incurs significant labor costs and is susceptible to human error, necessitating the involvement of highly trained radiologists. This can lead to misdiagnosis, delays in therapeutic intervention, and increased healthcare expenses [6]. Recent advancements in deep learning methodologies have yielded encouraging outcomes in the automated diagnosis and categorization of brain tumors. Transfer learning methods, convolutional neural networks, and attention-based mechanisms have significantly enhanced classification performance [7].

Despite these advancements, several obstacles remain. Numerous existing models depend on limited data samples or imbalanced data sets, hence constraining their application to heterogeneous populations [8]. Moreover, although achieving great predictive accuracy, the explainability of these models is frequently constrained, posing a barrier to the therapeutic use of algorithmic predictions. The computational efficiency is a barrier, as extensive neural networks may be impractical for real-time medical applications [9].

This research proposes an attention-enhanced DenseNet201 framework that incorporates MHSA to diagnose brain tumors, addressing the aforementioned deficiencies. The DenseNet201 component facilitates effective feature reuse by its dense connectedness, while the MHSA component ensures the network captures long-range spatial characteristics, enabling selective attention to clinically significant tumor locations in the MRI. Contrast Limited Adaptive Histogram Equalization (CLAHE) and intensity normalization are incorporated in image preprocessing to improve image quality. Data augmentation methods, including rotation, flipping, and scaling manipulations, are employed to mitigate overfitting and enhance the model's generalization with a relatively limited dataset. The implementation of dropout regularization and early stopping conditions in training optimization fosters model convergence and stability.

Grad CAM is employed to generate visual explanations of network predictions, thereby pinpointing tumor areas for subsequent clinical validation, enhancing interpretability. This approach provides a robust solution for brain tumor identification by computer vision, characterized by its superior predictive accuracy, interpretability, and computational economy.

The subsequent sections of this work are structured as follows: Section II delineates pertinent literature and highlights research deficiencies; Section III delineates the dataset, preprocessing, augmentation, and recommended approach; Section IV presents experimental findings and assessments, addresses model interpretability, efficiency, and comparison analysis; additionally, Section V summarizes and outlines future research directions.

## II. Related Works

Numerous deep learning and machine learning approaches have been proposed in recent years for the automated diagnosis and classification of brain tumors using MRI scans. This section reviews representative studies, emphasizing their methodological contributions, performance, and remaining limitations.

One study introduced the *IC-SNN-SVM* model for brain tumor detection using the CE-MRI database [10]. Noise reduction was performed with an Adaptive Wiener filter, followed by classification via a Radial Basis Function (RBF) neural network. Independent Component Analysis (ICA) enhanced image contrast, and the integration of ICA with a Self-Normalizing Neural Network (SNN) and a Support Vector Machine (SVM) yielded strong results: sensitivity $0.991$, specificity $0.989$, accuracy $0.989$, and Dice score $0.981$, with per-image processing time of $0.43$ s.

Building on advanced image processing, another study applied the EfficientNet-B0 architecture with additional fine-tuned layers [11]. The incorporation of image enhancement and extensive data augmentation improved training quality, enabling the model to achieve $98.87\%$ accuracy (reported with accuracy, precision, recall, and AUC).

A transfer-learning approach based on the ResNet-50 architecture employed fine-tuning and data augmentation (rotation, shifting, zooming) to increase dataset diversity [12]. The model reached $99\%$ accuracy, with precision $98.35\%$, recall $98.32\%$, and F1-score $98.31\%$. A notable attribute was its exceptionally high AUC ($0.999$),

underscoring strong discriminative power between tumor and non-tumor cases.

In a comparative study, multiple transfer-learning models were evaluated against classical classifiers [13]. Seven deep models were tested in combination with five conventional algorithms (SVM, Random Forest, Decision Tree, AdaBoost, Gradient Boosting). The VGG-19 model paired with a classifier delivered the best performance, achieving $99.39\%$ accuracy under 10-fold cross-validation.

Beyond conventional CNNs, another framework integrated the Contourlet Transform, a time-adaptive self-organizing map (TASOM), and the Whale Optimization Algorithm for feature extraction and optimization [14]. This pipeline achieved $> 98.5\%$ accuracy; adaptive learning rates and Otsu thresholding [15] improved both accuracy and real-time processing speed.

Transfer learning with the InceptionV4 architecture was also investigated [16]. The model achieved $98.7\%$ accuracy, outperforming traditional methods in diagnostic precision and computational efficiency, indicating potential for real-time clinical deployment.

The *VS-BEAM* model combined squeeze-excitation spatial attention with ensemble learning via a voting mechanism and incorporated Bayesian learning to improve classification and segmentation [17]. The approach yielded $98.91\%$ accuracy with sensitivity and specificity scores of $0.981$ and $0.987$ respectively.

Furthermore, a recent study used a finely calibrated EfficientNet-B4 model and achieved state-of-the-art results with $99.33\%$ accuracy, $100\%$ sensitivity, $98.65\%$ precision, and $98.67\%$ specificity [18].

Despite the promising findings shown in these investigations, numerous research deficiencies persist. Numerous methodologies remain restricted by limited or class-imbalanced datasets, hindering models' capacity to generalize across varied patient groups. Interpretability and the transparent explanation of model decisions are generally not prioritized, despite the fact that high classification accuracy is frequently attained. This aspect is essential for the development of clinical trust and adoption. Furthermore, while real-time inference is emphasized as a primary goal, several models still encounter difficulties with computing efficiency, limiting their scalability and practicality in extensive clinical applications.

## III. Methodology

This section outlines the methodology that will be suggested to classify brain tumors through deep learning methods. The framework is designed in a way that the data is processed systematically to enhance its quality and train a robust model on which data is precisely classified. Figure 1 outlines the workflow of the proposed method in detail, including the steps taken between the acquisition of data and the evaluation.

### A. Data Collection

In this current study, a binary classification dataset of brain tumors was used, which was obtained on Kaggle [19]. The dataset includes MRI images that are divided into the two different classes namely: tumor and healthy. The tumor class includes 392 images and the healthy class includes 408 images. Images are all in the JPG format and in single channel (grayscale).

Figure 2 below shows representative samples of the dataset hence demonstrating imagery of the tumor and healthy classes.

### B. Preprocessing and Feature Enhancement

In order to achieve the best performance of the model, preprocessing and feature enhancement methods were applied to the MRI images. First of all, the images were downsampled to a uniform

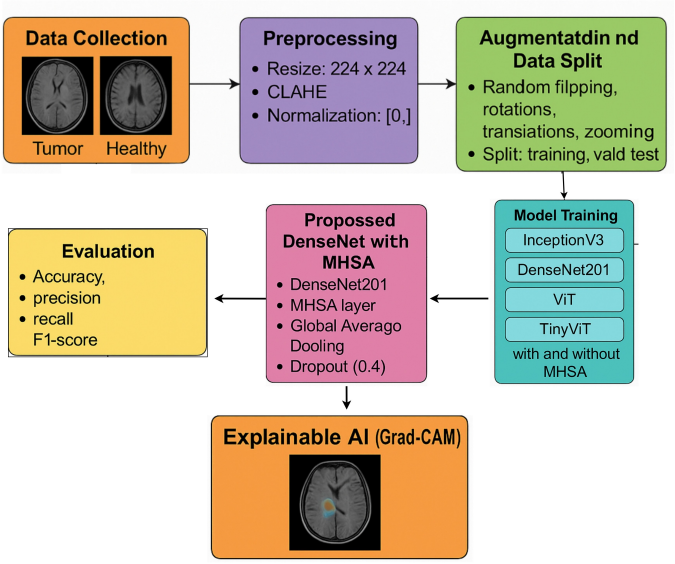


Fig. 1. The proposed brain tumor classification framework's overall workflow.

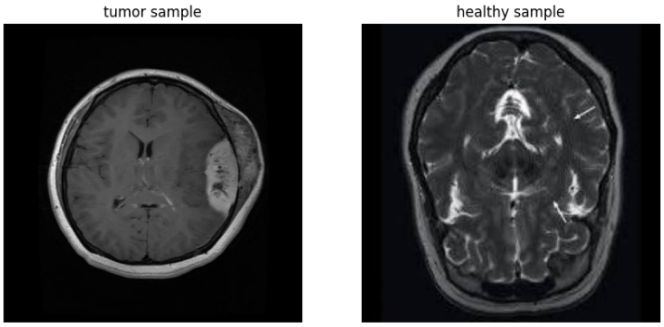


Fig. 2. Sample data from the brain tumor binary dataset: Healthy (right) and Tumor (left).

resolution of $224 \times 224$ pixels. The operation is important in insuring consistency in all input samples, which is vital in neural-network training. In order to increase contrast and visibility of salient features, CLAHE was used. The CLAHE is also especially beneficial to highlighting low contrast areas that are common in medical images and therefore avoid missing important information in the model. The images were also normalized by scaling pixel values to the range between 0 and 1 to make the process of training faster. This normalization is important since it makes the model stable with no high gradients in back-propagation. The transformed images were then archived in a new directory based on their classes by the preprocessing pipeline so that they could later be used in training the model.

The images were then processed then stored to a new folder by the class to ease the training of the model.

### C. Augmentation and Splitting Data

The data augmentation represents a significant initiative to enable more diversity of the training set, especially in the case of relatively small datasets. In this study, MRI images were augmented in many ways to facilitate model generalization. The augmentations that were

applied were horizontal flipping, random translations (shifting), random zoom, and random rotations. Namely, the probability of flipping of images was 50, the probability of rotation was uniformly sampled $(-20, +20)$, the probability of translation by pixel displacements $(-10, +10)$ was uniform, and the probability of zoom was uniform $(0.9, 1.1)$. Each image was run through these transformations to produce a total of $3,200$ augmented instances.

After augmentation, the data was divided into training, validation and test subsets in an $80 : 10 : 10$ ratio to evaluate the model performance on unseen data. The partitioning process is aimed at maintaining the balance in the classes, hence making sure that every class (healthy and tumor) is properly represented in all subsets. The dataset distribution is presented in Figure 3, which shows the amount of images per class in each split.

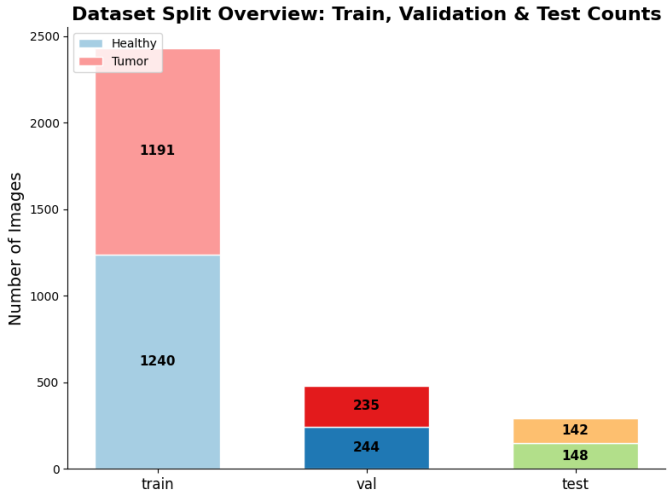


Fig. 3. Dataset split showing the number of images for each class in the training, validation, and test sets.

### D. Model Training

The first model considered is InceptionV3, a powerful convolutional neural network that processes visual data through parallel receptive fields of varying spatial sizes, making it effective for feature extraction in medical imaging. To enhance its ability to capture complex spatial and contextual relationships, InceptionV3 was augmented with Multi-Head Self-Attention (MHSA), allowing the model to focus on important regions within the brain tumor images. In parallel, MHSA was integrated with DenseNet201, a network renowned for its highly interconnected layers that facilitate efficient feature reuse and mitigate gradient vanishing issues. This combination allowed the model to better discriminate tumor regions. Additionally, transformer-based architectures, such as the Vision Transformer (ViT), were explored. ViT processes images as sequences of non-overlapping patches, enabling it to learn global relationships between image areas using self-attention. To optimize for computational efficiency without sacrificing performance, a lightweight version called TinyViT was trained, designed for environments with limited computational resources. This hybrid approach combines the strengths of convolutional networks and transformers, making it a robust model for brain tumor classification in resource-constrained settings.

### E. Proposed DenseNet201–MHSA Model

This model is based on the combination of DenseNet201 and MHSA to enhance feature extraction and capture long-range dependencies. DenseNet201 is used as the feature extractor; the results are then passed through a reshaping layer before being fed to the attention mechanism. The MHSA is subsequently applied, preceded by a GlobalAveragePooling1D layer to reduce dimensionality. To mitigate overfitting, a Dropout layer with a rate of 0.4 is employed, and the final classification is obtained through a Dense layer with softmax activation.

Therefore, the overall architecture consists of DenseNet base layers, reshaping, MHSA, pooling, dropout, and output layers. The architecture of the proposed DenseNet201 model with MHSA is illustrated in Figure 4.

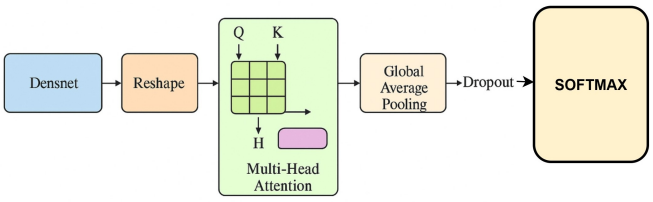


Fig. 4. Proposed hybrid architecture: DenseNet201 integrated with MHSA.

### F. Model Training Settings

To evaluate the performance of various base models (InceptionV3, DenseNet201, ViT, and TinyViT), a consistent training setup was used. This included the selection of a suitable optimizer, loss function, and hyperparameters such as batch size, dropout rate, and training epochs. Additionally, the use of Mixed-Precision Training on a Kaggle P100 GPU with $32GB$ of RAM was leveraged for efficient model training and inference. The following Table I outlines the specific training configurations applied across all models:

TABLE I
MODEL TRAINING SETTINGS FOR VARIOUS BASE MODELS

| Setting | Value |
|---|---|
| Base Model | InceptionV3 / DenseNet201/ ViT / TinyViT |
| Optimizer | Adam |
| Loss | Binary Cross-Entropy |
| Batch Size | 16 |
| Patience (Early Stopping) | 5 |
| Epochs | 40 |
| Dropout | 0.4 |
| MHSA (for CNNs) | 4 heads |
| GPU | Kaggle P100 |
| RAM | 32GB |

These settings were chosen to balance model performance and computational efficiency while ensuring generalization across different architectures.

### G. Explainable AI-GRAD CAM

Grad-CAM is utilized in the proposed DenseNet201 + MHSA model to identify the regions in MRI images that have the most significant effect on the classification performance of the model. By generating class-specific heatmaps, Grad-CAM facilitates the identification of salient focal areas, thereby augmenting the interpretability of the system. This methodological addition enhances the clarity of

the model so that attentional processes are sensitive to diagnostically relevant areas relevant to brain tumor classification.

## IV. Results and Discussion

The DenseNet201 integrated with MHSA improves brain tumor classification by capturing long-range dependencies in MRI images. Model performance was evaluated against baselines (InceptionV3, DenseNet201, ViT, TinyViT) using accuracy, precision, recall, F1-score, Cohen's $\kappa$, and test accuracy with 95% confidence intervals. McNemar's test confirmed statistical significance, while computational cost analysis ensured feasibility for real-world deployment.

### A. Validation and Testing Performances

A comparative analysis of the macro-average performance metrics of the InceptionV3, ViT, TinyViT, and the proposed DenseNet201 architecture with MHSA in terms of brain tumor classification is introduced in Table IV. InceptionV3 achieved balanced precision, recall, and F1-scores of 0.9875 and 0.9862, respectively, with validation and test scores of 0.985 and 0.9862 , respectively, which demonstrate that it is a better-performing model overall and able to extract features with high precision. ViT had a validation score of 0.8789 and a test score of 0.8690, which is moderate and slightly lower accuracy and recall in comparison with InceptionV3. TinyViT has validation and test scores of 0.8559 and 0.8552, respectively, indicating that TinyViT has lower discriminative capacity against brain tumor MRI. The proposed DenseNet201 with MHSA performed better than any competing models with the validation score of 0.9979 with macro-average precision, recall, and F1-scores of 0.9979 and 0.9966, respectively. These findings show that it has better generalization and strength. Taken together, the results prove that DenseNet201 combined with MHSA is effective in extracting salient features out of MRI images, thus making brain tumor-related classification highly accurate.

Figure 5 illustrates the loss curve of the validation with DenseNet201 with MHSA, which is important to highlight the stability of the training process.

TABLE II
Performance Comparison of Different Models on Brain Tumor Classification (Macro Average Metrics)

| Validation Performance | | | | |
|---|---|---|---|---|
| **Model** | **Accuracy (%)** | **Precision (%)** | **Recall (%)** | **F1-Score (%)** |
| InceptionV3 | 98.75 | 98.74 | 98.75 | 98.75 |
| ViT | 87.89 | 88.95 | 87.74 | 87.77 |
| TinyViT | 85.59 | 86.53 | 85.44 | 85.46 |
| **Proposed Hybrid Model** | **99.79** | **99.80** | **99.79** | **99.79** |

| Test Performance | | | | |
|---|---|---|---|---|
| **Model** | **Accuracy (%)** | **Precision (%)** | **Recall (%)** | **F1-Score (%)** |
| InceptionV3 | 98.62 | 98.63 | 98.65 | 98.62 |
| ViT | 86.90 | 88.02 | 86.72 | 86.75 |
| TinyViT | 85.52 | 86.42 | 85.35 | 85.38 |
| **Proposed Hybrid Model** | **99.66** | **100.00** | **99.32** | **99.66** |

The learning curve of the proposed DenseNet201 with MHSA is depicted with a 9-epoch training and validation of the model. The accuracy of the training steadily rises from 88.34% at the start of the first epoch to over 99%, and the validation accuracy reaches a maximum of 99.79%, thus indicating good learning and good generalization. The loss in training is decreasing at a high rate: the value falls from below 0.3313 to 0.05 in epoch 6, which means that the model learns the discriminative features of the training data rapidly. The loss on the validation is at least 0.0032, which reflects very accurate predictions made on unseen data. Small oscillations in later epochs are noticed in both the accuracy and loss curves; however, the general patterns of both confirm constant convergence and good performance of the model during the training period. These curves facilitate the stability and predictability of MHSA with DenseNet201 in the classification of brain tumors in MRI. In order to further support this strength, 5-fold cross-validation of the proposed model was also conducted.

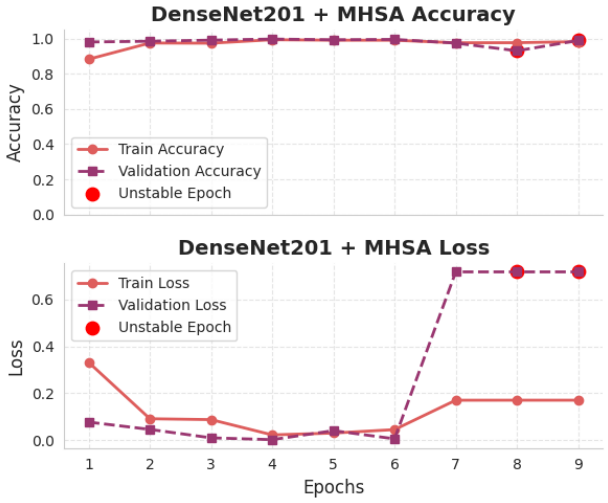


Fig. 5. The proposed DenseNet201 with Multi-Head Self-Attention model's 9-epoch training and validation learning curves show accuracy improvement and loss reduction that indicate stable convergence.

Additional evidence of the strength of the suggested DenseNet201 architecture with MHSA module can be found in the 5-fold results of cross-validation provided in Table III. Training accuracy is always high across all folds, with four folds recording accuracy above 98, and Fold 4 recording a relatively lower accuracy of 89.39%. The measured mean accuracy of 97.34% suggests that the model has high generalization properties even when the dataset is split into far off sections, thus limiting the chances of the model overfitting to a specific subset.

TABLE III
5-Fold Validation Accuracy of DenseNet201 + MHSA

| **Fold** | **Training Accuracy** |
|---|---|
| Fold 1 | 0.9966 |
| Fold 2 | 0.9987 |
| Fold 3 | 0.9923 |
| Fold 4 | 0.8939 |
| Fold 5 | 0.9854 |
| **Mean** | 0.9734 |

### B. Comparative Model Evaluation and Statistical Analysis

The relative analysis as shown in Table IV reveals that there are significant performance variations between the models. Inception V3 and DenseNet201 + MHSA showed a perfect score of class-wise AUC and PR-AUC (1.0000), which indicates excellent class-wise discriminative ability of both healthy and tumorous classes. However, all other models had lower overall reliability and statistical consistency as shown by the 95% confidence interval of [99.65, 99.96] and the Cohen $\kappa$ of 1.0000, which supports the perfect

agreement and strong generalization across the models compared to the proposed DenseNet201 + MHSA. The McNemar test value of 0.0 (p = 0.000000) has additionally shown that the predictions of the model are much the same and are of great superiority to chance. In contrast, ViT and TinyViT exhibited lower AUC, PR-AUC, and Cohen's $\kappa$ values, along with significant McNemar test results ($p < 0.001$), indicating less stable predictions and reduced reliability. All in all, these results indicate that the proposed DenseNet201 + MHSA does not only deliver high classification accuracy, but also guarantees strong, stable, and statistically confident results in the classification of MRI brain tumors, making it the best option in the group of the tested models.

TABLE IV
COMPARATIVE EVALUATION OF DIFFERENT MODELS ON MRI BRAIN TUMOR CLASSIFICATION

| Model | Class Healthy (AUC / PR-AUC) | Class Tumor (AUC / PR-AUC) | 95% CI | Cohen's $\kappa$ | McNemar | p-value |
|---|---|---|---|---|---|---|
| Inception V3 | 1.0000 / 1.0000 | 1.0000 / 1.0000 | [0.9690, 0.9966] | 0.9724 | 0.0 | 0.1250 |
| ViT | 0.9353 / 0.9185 | 0.9353 / 0.9470 | [0.8310, 0.9069] | 0.7369 | 7.0 | 0.0001 |
| TinyViT | 0.9248 / 0.9131 | 0.9248 / 0.9399 | [0.8172, 0.8931] | 0.7093 | 10.0 | 0.0009 |
| **Proposed hybrid model** | **1.0000 / 1.0000** | **1.0000 / 1.0000** | **[99.65, 99.96]** | **1.0000** | **0.0** | **1.0000** |

### C. Misclassification Analysis Utilizing Confusion Matrix

The distribution of misclassifications is analyzed in Figure 6 using a confusion matrix of the proposed DenseNet201 architecture that is integrated with MHSA. A comprehensive analysis of the healthy cohort reveals that there are no erroneous classifications of any healthy specimen as tumorous, with all 148 accurately recognized as healthy. Within the tumor classification, 141 instances are accurately identified; however, one tumor case is erroneously classified as healthy. The confusion matrix indicates that the misclassification rate is negligible, confined to a solitary instance of tumor, hence affirming the model's robust discriminate capability between healthy and tumorous conditions.

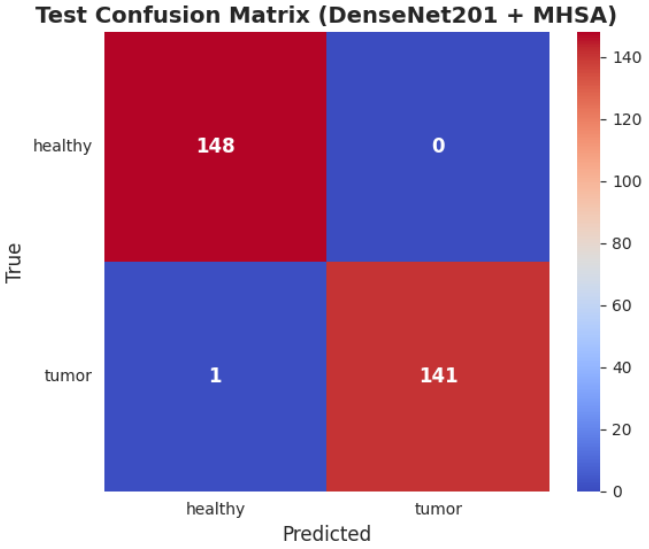


Fig. 6. Confusion matrix of the proposed DenseNet201 with MHSA model for MRI brain tumor classification, showing true and predicted class distributions for healthy and tumor samples.

### D. Training and Inference Efficiency

Figure 7 illustrates the training and inference efficiency of Inception V3, DenseNet201 MHSA, ViT and TinyViT. In comparison to Inception V3 (288.33 seconds to train), DenseNet201 with MHSA takes 233.36 s to train, and only uses 874.75 MB of memory (as compared to 2830.73 MB of Inception V3). MHSA + DenseNet201 takes 0.0237 second to process per image, consuming 13.99 GB of memory which is less than that of transformer based models. ViT and TinyViT will have slightly higher per-image throughput at the expense of higher memory consumption, in comparison. Inception V3 requires less inference memory but requires longer training time and has a slightly lower accuracy.

Together, DenseNet201 with MHSA provides an effective training and inference trade off between speed and memory consumption and model classification.

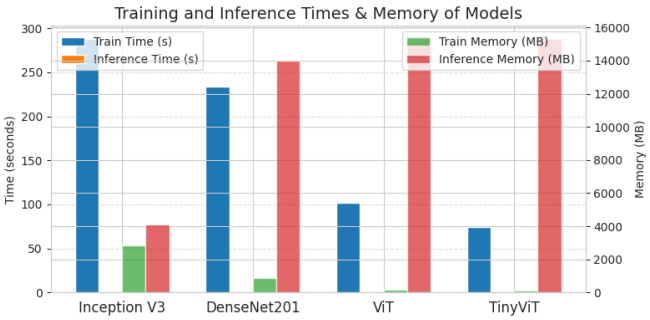


Fig. 7. Comparison of training and inference efficiency for different models, including training time, training memory, inference time per image, and inference memory usage.

### E. Proposed DenseNet201-MHSA Interpretability through Grad-CAM Visualization

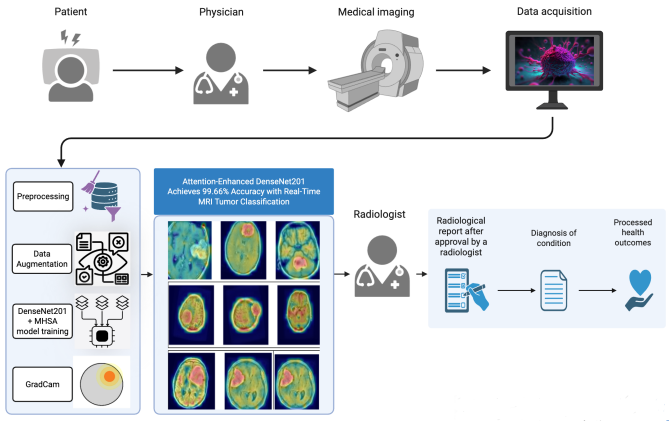


Fig. 8. Role of Grad-CAM (explainability and tumor localization)

Grad-CAM visualization was implemented to evaluate the efficacy of the proposed DenseNet201 with an MHSA module in accurately identifying clinically relevant regions in MRI scans. The model is able to generate heatmaps that clearly emphasize tumor regions while largely suppressing healthy tissue, as demonstrated in Figure 8. This confirms that the network is focused on meaningful pathological features rather than irrelevant patterns. The highlighted regions' close resemblance to the ground-truth tumor locations was further validated by manual inspection and expert analysis, indicating that the model's predictions are both medically informed and reliable. These visualizations are crucial for clinical deployment, as they increase transparency

and trust by offering a glimpse into the model's decision-making process. Clinicians can have greater confidence in the diagnostic utility of the network by guaranteeing that it consistently addresses the appropriate anatomical regions. In conclusion, the Grad-CAM analysis illustrates that DenseNet201 with MHSA not only obtains high predictive performance but also valid tumor localization, thereby effectively bridging the gap between automated classification and clinical interpretability.

### *F. Comparative Analysis of Brain Tumor Detection Models*

Table V shows a comparative study of brain tumor detection models based on classification accuracy. The presented model model using DenseNet201 and MHSA is able to achieve the best accuracy of $99.66\%$, which is significantly higher than the latest state of the art methods. In addition to its excellent results, the model incorporates Explainable AI (XAI) methods, which provide interpretability and inspire trust in clinical practice and, at the same time, maintain a computational cost comparable to that of real-time implementation.

TABLE V
COMPARATIVE ANALYSIS OF BRAIN TUMOR DETECTION MODELS

| Study | Model | Accuracy (%) |
|---|---|---|
| [10] | IC-SNN-SVM | 98.9 |
| [11] | EfficientNet-B0 | 98.87 |
| [12] | ResNet-50 (Transfer Learning) | 99 |
| [13] | VGG-19 + Classifier | 99.39 |
| [14] | Contourlet + TASOM + Whale Optimization | 98.5 |
| [16] | InceptionV4 | 98.7 |
| [17] | VS-BEAM | 98.91 |
| [18] | EfficientNet-B4 | 99.33 |
| **Proposed hybrid model** | DenseNet201 + MHSA | **99.66** |

## V. CONCLUSION AND FUTURE WORK

This study presents an attention enhanced architecture of DenseNet201 with MHSA as a way of obtaining brain tumor information in the form of highly accurate and interpretable classification of tumor types using scans of MRI. The dense connectivity of DenseNet201 and the spatial attention mechanism of MHSA are synergistically integrated in the proposed model, which allows capturing the local and long range contextual dependencies, outperforming the performance of the traditional convolutional neural network and transformer based models. The five fold cross-validation resulted in the test accuracy of $99.66\%$, the F1-score of $99.66\%$, the Cohen $\kappa$ of $1.0$, and the macro-average precision and recall of consistently high values. The inference has a low computational cost, and the time of processing each scan is about $24$ milliseconds. Grad-CAM results showed that the model is sensitive to clinically relevant areas of the tumor, thus improving interpretability and evidence of its possible clinical application. These findings support that the DenseNet201+MHSA architecture is a sensible, precise, and interpretable system of brain tumor detection in the fully automated mode. Despite its excellent performance, the study has been based on a region specific dataset in Bangladesh which may not be generalizable in other populations.

Further studies will aim at diversifying the data by introducing multi institutional and multi modal MRI acquisitions to enhance the strength of the model. Additionally, the use of more image modalities, production of lightweight model predictors capable of edge deployment, and the improvement of explainability are likely to help bring clinical applications to real time operation. It is also expected to facilitate the clinicians with daily work in the area of diagnostic activities through the design of user friendly software platform that will facilitate hospital integration.

## REFERENCES


[1] Y. Fan, X. Zhang, C. Gao, S. Jiang, H. Wu, Z. Liu, and T. Dou, "Burden and trends of brain and central nervous system cancer from 1990 to 2019 at the global, regional, and country levels," *Archives of Public Health*, vol. 80, 2022.

[2] I. Ilic and M. Ilic, "International patterns and trends in the brain cancer incidence and mortality: An observational study based on the global burden of disease," *Heliyon*, vol. 9, no. 7, p. e18222, 2023.

[3] E. Hassan and H. Ghadiri, "Advancing brain tumor classification: A robust framework using efficientnetv2 transfer learning and statistical analysis," *Computers in Biology and Medicine*, vol. 185, 2025.

[4] S. Banik, M. Das, T. Banik, and M. E. Haque, "Mobiledenseattn:a dual-stream architecture for accurate and interpretable brain tumor detection," 2025. [Online]. Available: https://arxiv.org/abs/2508.18294

[5] K. Lamba, S. Rani, and M. Shabaz, "Synergizing advanced algorithm of explainable artificial intelligence with hybrid model for enhanced brain tumor detection in healthcare," *Scientific Reports*, vol. 15, 2025.

[6] M. E. Haque, A. Fahim, S. Dey, S. A. Jahan, S. M. J. Islam, S. Rokoni, and M. S. Morshed, "A modified vgg19-based framework for accurate and interpretable real-time bone fracture detection," 2025. [Online]. Available: https://arxiv.org/abs/2508.03739

[7] T. Berghout, "The neural frontier of future medical imaging: A review of deep learning for brain tumor detection," *J. Imaging*, vol. 11, no. 1, 2025.

[8] R. Kaifi, "A review of recent advances in brain tumor diagnosis based on ai-based classification," *Diagnostics (Basel)*, vol. 13, no. 18, 2023.

[9] T. Rahman, M. S. Islam, and J. Uddin, "Mri-based brain tumor classification using a dilated parallel deep convolutional neural network," *Digital*, vol. 4, no. 3, pp. 529–554, 2024.

[10] A. A. Asiri, T. A. Soomro, A. A. Shah, G. Pogrebna, M. Irfan, and S. Alqahtani, "Optimized brain tumor detection: A dual-module approach for mri image enhancement and tumor classification," *IEEE Access*, vol. 12, pp. 42 868–42 887, 2024.

[11] H. A. Shah, F. Saeed, S. Yun, J.-H. Park, A. Paul, and J.-M. Kang, "A robust approach for brain tumor detection in magnetic resonance images using finetuned efficientnet," *IEEE Access*, vol. 10, pp. 65 426–65 438, 2022.

[12] A. Younis, Q. Li, Z. Afzal, M. Jajere Adamu, H. Bello Kawuwa, F. Hussain, and H. Hussain, "Abnormal brain tumors classification using resnet50 and its comprehensive evaluation," *IEEE Access*, vol. 12, pp. 78 843–78 853, 2024.

[13] S. Ahmad and P. K. Choudhury, "On the performance of deep transfer learning networks for brain tumor detection using mr images," *IEEE Access*, vol. 10, pp. 59 099–59 114, 2022.

[14] A. Farzamnia, S. H. Hazaveh, S. S. Siadat, and E. G. Moung, "Mri brain tumor detection methods using contourlet transform based on time adaptive self-organizing map," *IEEE Access*, vol. 11, pp. 113 480–113 492, 2023.

[15] W. A. H. Jumiawi and A. El-Zaart, "Otsu thresholding model using heterogeneous mean filters for precise images segmentation," in *2022 international conference of advanced Technology in Electronic and Electrical Engineering (ICATEEE)*. IEEE, 2022, pp. 1–6.

[16] N. Bibi, F. Wahid, Y. Ma, S. Ali, I. A. Abbasi, A. Alkhayyat, and Khyber, "A transfer learning-based approach for brain tumor classification," *IEEE Access*, vol. 12, pp. 111 218–111 238, 2024.

[17] S. M. A. H. Shah, A. Ullah, J. Iqbal, S. Bourouis, S. S. Ullah, S. Hussain, M. Q. Khan, Y. A. Shah, and G. Mustafa, "Classifying and localizing abnormalities in brain mri using channel attention based semi-bayesian ensemble voting mechanism and convolutional auto-encoder," *IEEE Access*, vol. 11, pp. 75 528–75 545, 2023.

[18] R. Preetha, M. J. P. Priyadarsini, and J. S. Nisha, "Automated brain tumor detection from magnetic resonance images using fine-tuned efficientnet-b4 convolutional neural network," *IEEE Access*, vol. 12, pp. 112 181–112 195, 2024.

[19] H. Habib, "Brain cancer detection mri images," 2023, accessed: 2023-09-05. [Online]. Available: https://www.kaggle.com/datasets/hamzahabib47/brain-cancer-detection-mri-images