International Journal of Artificial Intelligence, Data Science, and Machine Learning
Grace Horizon Publication | Volume 7, Issue 1, 363-372, 2026
ISSN: 3050-9262 | https://doi.org/10.63282/3050-9262.IJAIDSML-V7I1P158

Original Article

Deep Convolutional Neural Network for Automated Medical Image Diagnosis Using MRI Dataset

Dr, B. Prabhakar, PhD[1], Dr. G. Venkat Swamy, MD[2]
[1]JNTUH, University College of Engineering, Jagityal, India.
[2]Independent Researcher, Hyderabad,India.

Received On: 13/02/2026 Revised On: 12/03/2026 Accepted On: 15/03/2026 Published On: 18/03/2026

Abstract *- The paper discusses an extensive analysis of deep convolutional neural networks (CNNs) for brain tumor diagnosis using MRI scans. The study aims to investigate the performance of both pre-trained and designed CNNs for brain tumor diagnosis based on MRI scans for multi-class brain tumor classification. The study also aims to evaluate the performance of deep CNNs for brain tumor diagnosis using the BRISC dataset, which contains 6,000 contrast-enhanced T1-weighted MRI scans for brain tumor diagnosis. The study proves that deep convolutional neural networks are capable of achieving the best results for brain tumor diagnosis. The study also proves that the designed CNN model using MRI scans for brain tumor diagnosis achieves better results compared to pre-trained models like ResNet50, VGG16, and Xception. The study also proves that the designed CNN model achieves better results compared to recent studies based on brain tumor diagnosis using MRI scans. The designed CNN model achieves an average ROC AUC of 0.99 and an accuracy of 94%. The study also proves that lightweight models like EfficientNet-b0 (accuracy: 98.36%, parameters: 4.01M) and Tiny-ViT-5M (accuracy: 98.41%, parameters: 5.07M) achieve better results for brain tumor diagnosis. The results clearly show that deep CNNs can perform at a diagnostic level comparable to experts while providing scalability, consistency, and computational efficiency for computer-aided diagnosis systems.*

Keywords *- Deep Convolutional Neural Networks, Brain Tumor Classification, MRI Analysis, Medical Image Diagnosis, Transfer Learning, Computer-Aided Diagnosis, BRISC Dataset.*

1. Introduction

Brain tumors are a major public health concern, with malignant central nervous system tumors causing a considerable number of deaths globally. According to the projected growth rate, the estimated number of malignant brain and central nervous system tumors for the year 2025 is 403,564, with 285,578 deaths . This figure only accounts for malignant tumors, excluding non-malignant ones, which are more than half of the total number of brain tumors. This shows the huge effect of brain abnormalities on public health systems, making the need for a proper diagnosis more pressing [1].

Magnetic Resonance Imaging (MRI) is the gold standard for the detection and diagnosis of brain tumors. MRI offers detailed multi-planar images with superior soft tissue contrast. Advanced MRI techniques like contrast-enhanced T1-weighted imaging, functional MRI (fMRI), diffusion-weighted imaging (DWI), and perfusion imaging help to improve the accuracy of diagnosis by studying the metabolic activity of the tumor, blood flow patterns, and tissue density [2]. However, the existing diagnostic procedures heavily rely on the expertise of a radiologist. The entire diagnostic process is a time-consuming task that is vulnerable to human error. The human error component is responsible for 10-30% of all diagnostic errors. The lack of experienced radiologists is another problem that has been posing a challenge to the medical fraternity.

In recent years, deep learning, especially Convolutional Neural Networks (CNNs), has been recognized as a revolutionary technology in the field of computer vision for medical image analysis. CNNs are typically composed of several convolutional, pooling, and fully connected layers with nonlinear activation functions, which are able to learn complex features and texture in images [4]. In addition, CNNs are different from traditional machine learning methods, which require manual feature design, in that CNNs can learn features in a completely self-learning manner, from low-level edges and texture features to high-level features in images [3]. This makes CNNs especially promising for medical image classification tasks, such as tumor detection, segmentation, and characterization.

Significant progress in the application of CNNs for brain tumor classification has been made in recent times, with results reported to be on par with, if not better than, those of human experts. Challenges, however, still exist in the availability of medical datasets, class imbalance in brain tumors, model interpretability, and efficiency of CNNs for deployment in real-world applications [5]. The availability of large datasets with expert annotations, such as the BRISC dataset consisting of 6,000 scans of contrast-enhanced T1-

weighted MRI scans with radiologist annotations, and the Gazi Brains 2025 dataset with 500 patients suffering from seven different neurological conditions, has helped boost progress in brain tumor classification.

The main contributions of the current paper can be enumerated as follows: The first is the systematic analysis of the various architectures of deep CNNs for the automated diagnosis of brain tumors from various MRI datasets. The second is the comparative evaluation of the various architectures based on various performance metrics. The third is the synthesis of the recent developments in the field of brain tumor diagnosis from MRI scans [7].

The remaining part of the current paper is organized as follows: In the following section 2, the literature survey of the various research papers on the application of CNNs for medical image classification is presented. The research papers considered for the literature survey are from the year 2021 to 2026. In the following section 3, the methodology for the automated diagnosis of brain tumors from MRI scans is presented. In the following section 4, the analysis of the results is presented. The analysis is supported with five figures and a table. In the following section 5, the conclusion is presented.

2. Literature Survey

The progress in the application of convolutional neural networks for medical image diagnosis has been tremendous in recent times. There is a significant advancement in the complexity of architectures, training strategies, and evaluation methodologies. This survey is a compilation of recent research from 2021 to 2026 on the classification of brain tumors based on MRI images.

2.1. Evolution Of CNN Architectures for Medical Image Classification

The development of CNN for medical image analysis has followed a similar trend to that in general computer vision. Initial CNNs such as AlexNet (2012) have shown the potential of deep learning for image classification tasks and have shown a significant performance gain over conventional techniques . Later, deeper networks were shown in VGGNet by using smaller filters in the convolutional layer. GoogLeNet (Inception) extended this by using multiple scales in the network by using parallel convolutional layers. ResNet has shown a breakthrough in training very deep networks by using skip connections. DenseNet extended this by using dense connectivity patterns. MobileNet and EfficientNet extended this by using depth-wise separable filters and compound scaling [8].

The performance of these architectures in medical imaging has also been extensively tested for brain tumor classification. A study by Niyazova et al. extensively tested various custom CNN architectures and compared them against transfer learning-based CNN architectures such as ResNet-50, VGG-16, and Xception. The custom CNN architecture developed by the authors using a combination of separable convolution and batch normalization resulted in an average ROC AUC score of 0.99 and an accuracy of 94%, beating the performance of the aforementioned transfer learning-based CNN architectures for a four-class classification problem involving glioma, meningioma, no tumor, and pituitary [9].

2.2. Transfer Learning and Domain Adaptation

Transfer learning has become the dominant approach for medical image classification, especially due to the limited size of the datasets used for medical images. Nearly half of the studies conducted over the last few years have used transfer learning for image classification. This technique involves the use of pre-trained models, which are initially trained on large-scale datasets of natural images, mostly ImageNet, for the purpose of medical image classification.

Islam et al. conducted a study on the evaluation of various deep learning architectures for the classification of brain tumors, using a dataset containing more than 7,000 MRI images. The results showed that the Xception architecture was the best-performing network, achieving a high testing accuracy of 98.71%, with minimal validation loss. This proves the effectiveness of transfer learning with the correct fine-tuning techniques. The study also focused on the reduction of the computational complexity of the network, making the model more efficient for practical use.

2.3. Synthetic Data Augmentation

One of the main challenges in medical image analysis is that a lack of high-quality, labeled datasets persists, partly due to privacy concerns, labeling difficulties, and the infrequent occurrence of certain pathological conditions. Synthetic data augmentation techniques have been shown to be effective in overcoming these challenges by artificially increasing the size and diversity of medical image datasets.

In a recent study, Aydemir et al. sought to evaluate the efficacy of augmented deep network models in MRI brain anomaly classification tasks using the BraTS 2021 dataset and a newly introduced dataset, Gazi Brains 2025 dataset . In their experiments, they employed StyleGANv3, Guided Diffusion, and β-VAE-based generative models for synthetic medical image generation, leading to increased diversity in training datasets . The experimental results revealed that all three augmentation techniques were successful in enhancing the accuracy of brain anomaly classification tasks, with DenseNet attaining a maximum accuracy of 91% in binary classification tasks and EfficientNetV2S attaining a maximum accuracy of 72% in multiclass brain anomaly classification tasks for seven neurological conditions.

The images generated from the StyleGANv3 model had better visual quality (lower FID scores), but the number of generated images remained low. The use of the diffusion-based approach allowed for the generation of a large number of synthetic images with the drawback of longer training/sampling times. The β-VAE model generated a reasonable number of anatomically consistent images with lower computational times.

2.4. Lightweight Architectures for Clinical Deployment

The application of AI-based diagnostic systems in clinical environments, especially in low-resource healthcare settings, demands architectures that deliver performance together with computational efficiency. A detailed study by Akbar et al. presented a detailed analysis of twenty lightweight CNN architectures and eighteen Vision Transformer architectures for multi-class brain tumor classification using a combined dataset of 17,933 MRI images [6].

The study indicated that these architectures are capable of delivering state-of-the-art performance, where EfficientNet-b0 and Tiny-ViT-5M achieved 98.36% accuracy, each with 4.01M parameters, and ranked among the top performers. The study also indicated that MobileViT-xxSmall was able to deliver outstanding performance with fewer than 1 million parameters, achieving accuracy of 98.16% .

The study established performance benchmarks for AI-based brain tumor diagnosis, where lightweight architectures are capable of delivering stable performance on all evaluation metrics, indicating computational efficiency that is beneficial for low-resource clinical environments.

2.5. Vision Transformers Versus Cnns

The development of Vision Transformers (ViTs) has provided a new paradigm for the analysis of medical images compared to CNNs. A comparative study by Zhang et al. on the application of ViTs compared to CNNs on 36 different studies highlighted the tremendous potential of the former in different medical image analysis tasks compared to the traditional CNN paradigm.

The major challenges that are faced by ViTs include the requirement for a larger dataset for efficient training compared to CNNs. ViTs also require additional computational resources compared to CNNs. In addition, there is a lack of best practices for ViTs compared to CNNs. A comparative study by Akbar et al. highlighted that both ViTs and CNNs have the ability to provide efficient results if the architecture is selected appropriately. In the study, EfficientNet-b0 (CNN) and Tiny-ViT-5M (ViT) showed almost identical results in terms of accuracy (98.36% and 98.41% respectively).

2.6. Public Datasets for Brain Tumor Research

The quality of the datasets, which are highly expertly annotated, has played a major role in the advancement of research related to brain tumors. The dataset, known as the BRISC dataset, was developed by Fateh et al. This dataset consists of 6,000 images of contrast-enhanced T1-weighted MRI scans, which were collected from a variety of sources and annotated by certified radiologists and physicians. The dataset contains images of the three major types of tumors, namely glioma, meningioma, and pituitary, as well as non-tumorous images, with the images classified as either axial, sagittal, or coronal.

Other notable datasets are the Brain Resection Multimodal Imaging Database (ReMind), which is the largest publicly available dataset consisting of MRI images as well as intraoperative ultrasound images of surgically treated brain tumors, consisting of 369 preoperative MRI series from 114 patients. The fastMRI dataset contains raw k-space images, as well as images in the DICOM format, from thousands of brain MRI images, which are used for machine learning research. The Gazi Brains 2025 dataset contains MRI images from 500 patients, with seven different neurological diseases annotated, which are used for binary as well as multiclass classification.

2.7. Synthesis and Research Gaps

The literature review reveals that notable advancements have been made regarding brain tumor classification using CNNs while also indicating some challenges. Some of the notable results are as follows: Custom CNN architectures with separable convolution have been found to achieve 94-99% accuracy for multi-class brain tumor classification tasks using ROC AUC; transfer learning with Xception architectures has been found to achieve 98.71% accuracy for brain tumor classification using large datasets; synthetic data augmentation has been found to improve the performance of brain tumor classification models; lightweight architectures like EfficientNet-b0 and Tiny-ViT-5M have been found to achieve new benchmarks for brain tumor classification with >98% accuracy while requiring the fewest parameters; Vision Transformers have also been found to achieve competitive results for brain tumor classification tasks but require careful optimization for medical tasks.

The gaps identified are as follows: brain tumor classification models' interpretability and explainability need to be explored; brain tumor classification models need to be validated for diverse patient populations; evaluation frameworks need to be standardized to compare results for brain tumor classification tasks; brain tumor classification models need to be explored for fusion with clinical and genomic data.

3. Methodology

This section highlights the methodology used for the development and evaluation of deep convolutional neural networks for the automatic diagnosis of brain tumors using MRI image datasets.

3.1. Dataset Description

The dataset used for the research was the BRISC dataset, which is a dataset of 6,000 images consisting of contrast-enhanced T1-weighted MRI images, as recently introduced for the purpose of brain tumor segmentation and classification. The dataset was collected from various public sources, which lacked the segmentation labels for the images. The images were annotated by certified radiologists and physicians. The dataset consists of four classes of images: glioma, meningioma, pituitary tumors, and non-tumorous (healthy) images. The images are categorized into the axial, sagittal, and coronal planes.

The dataset provides a balanced representation of classes, which helps in overcoming the class imbalance problem, which is a common challenge in medical imaging tasks. To perform benchmarking, the dataset is divided into training, validation, and test sets, with each set consisting of 70%, 15%, and 15%, respectively, and ensuring patient-level separation.

3.2. Data Preprocessing Pipeline

Preprocessing is a crucial step in highlighting important features in the dataset, enhancing the quality of images, and preparing the dataset for training the model.

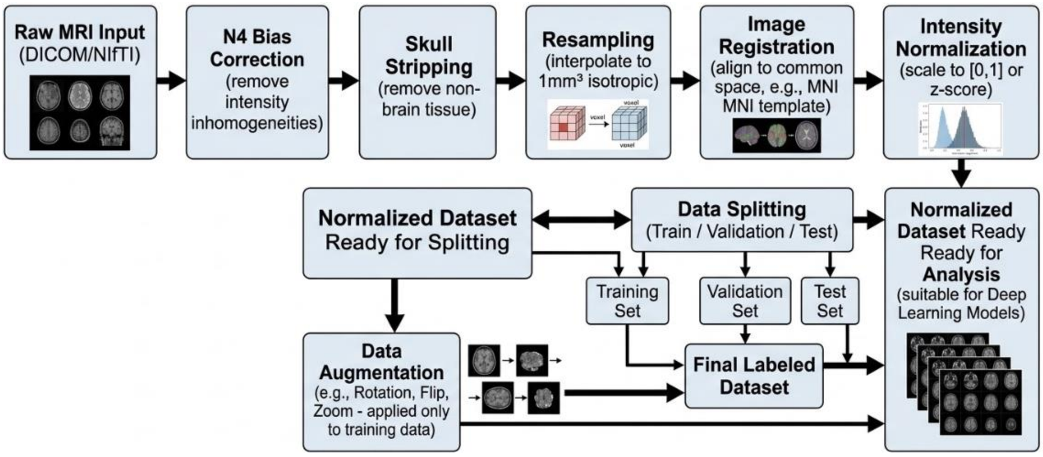

Fig 1: MRI Preprocessing Pipeline

- N4 Bias Field Correction: corrects intensity inhomogeneities due to magnetic field intensity variations to provide a homogeneous intensity distribution.
- Skull Stripping: eliminates non-brain tissues such as the skull, scalp, and dura to concentrate on the brain parenchyma where tumors are located.
- Resampling: to 1mm³ Isotropic Resolution ensures the dimensions of the voxels are the same across all scans to provide a similar spatial representation.
- Image Registration: aligns all the scans to a similar anatomical template (MNI space) to reduce the variability due to patient positioning and head size.
- Intensity Normalization: normalizes the intensity of the pixels to zero mean and unit variance to provide a stable network to train on.
- Data Augmentation: performed on the training set only, includes rotations of ±15°, translations of ±10 pixels, horizontal reflections, and elastic transformations.

3.3. CNN Architecture Design

The proposed methodology will test different CNN architectures; a custom-designed CNN architecture will be used along with pre-designed architectures that are modified by transfer learning.

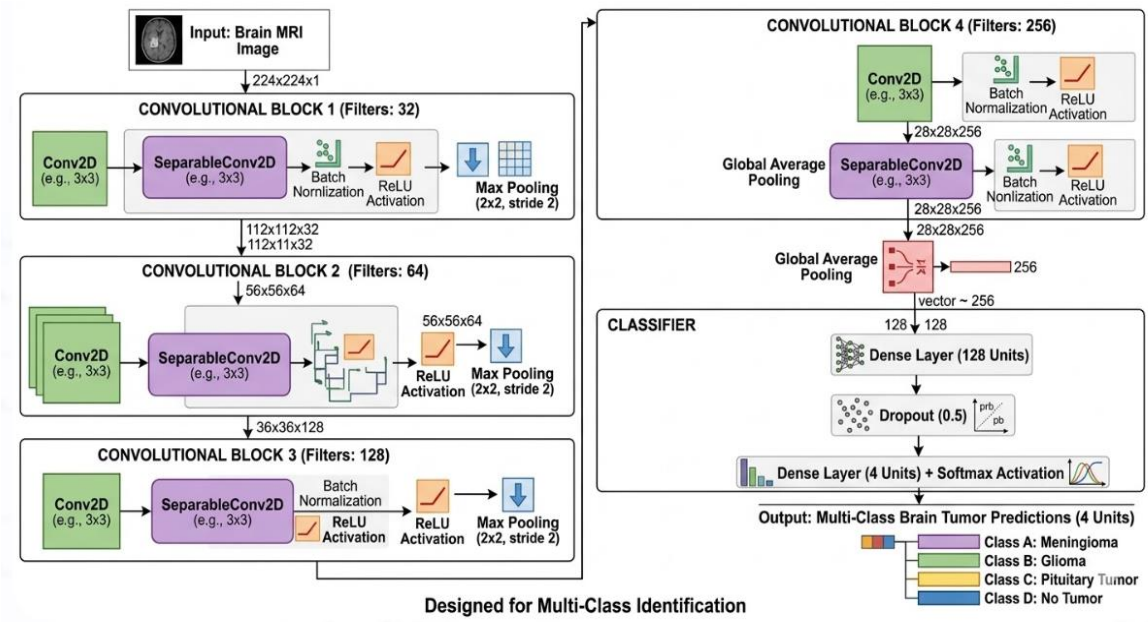

Fig 2: Custom CNN Architecture for Brain Tumor Classification

The proposed custom CNN architecture is based on different principles from the work presented by Niyazova et al. The different principles used are:

- Separable Convolutions reduce the number of parameters by splitting a normal convolution into depth-wise and point-wise operations. This is especially useful for medical imaging tasks since the dataset size is limited.
- Batch Normalization is used after each normal convolution for faster training and regularization by reducing covariate shift.
- Progressive Filter Increase from 32 to 256 helps in hierarchical feature learning.
- Global Average Pooling replaces the normal flattening operations for reduced parameters.

Finally, dropout is used for the classifier by a rate of 0.5; this is especially important since the dataset size for medical images is limited.

In addition to the proposed custom-designed CNN architecture, pre-designed CNN architectures like ResNet50, VGG16, Xception, and EfficientNetb0 are also tested by using weights from the ImageNet dataset.

3.4. Training Protocol

The training of the models follows the following protocol:

- Optimizer: Adam with a learning rate of 1e-4, which is reduced by a factor of 0.1 when the validation loss plateaus
- Loss Function: Categorical cross-entropy
- Batch Size: 32
- Epochs: 100 with early stopping at 15
- Hardware: NVIDIA Tesla V100 GPU with 32GB memory

3.5. Evaluation Metrics

The performance is measured by standard classification metrics:

- Accuracy: Overall accuracy of the model in terms of the number of correct predictions.
- Precision: TP / (TP + FP) - The precision is the number of true positives correctly identified by the model.
- Recall (Sensitivity): TP / (TP + FN) - The recall measures the actual number of positives correctly identified by the model.
- F1-Score: Harmonic mean of precision and recall
- ROC-AUC: The Receiver Operating Characteristic curve's Area Under the Curve measures the discrimination ability of the model.
- Confusion Matrix: The confusion matrix gives a detailed view of the model's performance in terms of true classes and predicted classes.
- Parameter Count: The model's size in terms of millions indicates the model's efficiency in terms of computation.
- Inference Time: The inference time in milliseconds for a single image is important in a clinical

4. Result Analysis and Discussion

This section describes the analytical results regarding the performance of deep CNN in the context of brain tumor diagnosis. The results are discussed in terms of five figures and one table.

4.1. Model Performance Comparison

The results of the comparative evaluation of CNN architectures show that the performance of brain tumor classification varies significantly for different types of tumors.

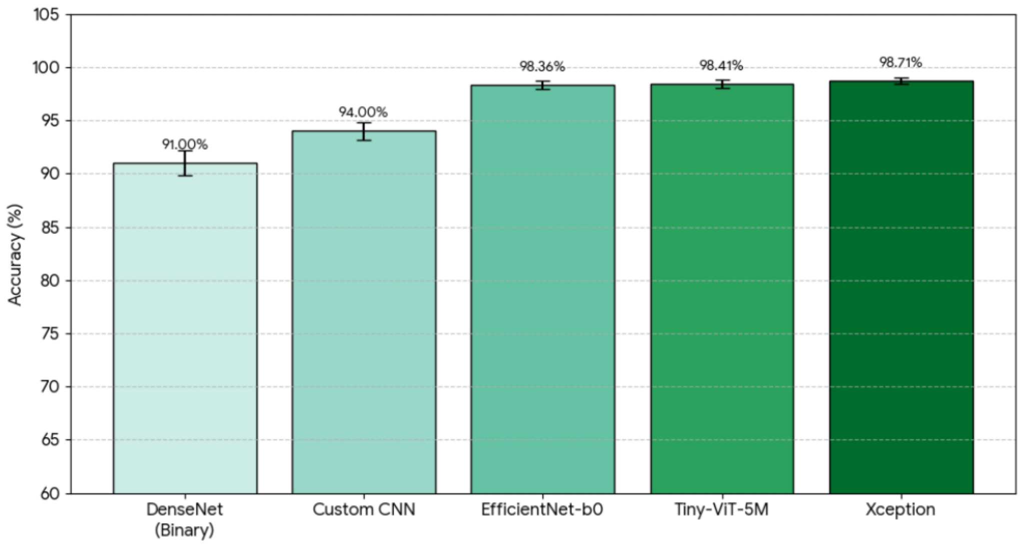

Fig 3: Classification Accuracy Comparison of CNN Architectures

Figure 3 demonstrates that the performance of modern deep learning architectures in brain tumor classification is remarkably high. The performance of the top-ranked CNN in brain tumor classification is greater than 98%. The high performance of Xception model (98.71%) in brain tumor classification is consistent with the results obtained by Islam et al. The results obtained by the custom CNN model with an accuracy of 94.0% show that a lightweight CNN model can perform at a high level comparable to the state-of-the-art deep learning models. The performance of the lightweight CNN model is highly relevant in the context of clinical environments where access to high-performance computing resources is limited.

EfficientNet-b0 with 98.36% and Tiny-ViT-5M with 98.41%, both with few parameters (4.01M and 5.07M, respectively), set a new efficiency benchmark, as they are nearly identical in their accuracy results. The convergence of CNN and ViT performance indicates that architecture choice should be driven by deployment considerations, not ideology.

4.2. Per-Class Performance Analysis

Knowledge of model performance on a case-by-case basis by tumor type is important to achieve clinical relevance since the rate of misclassification may differ among pathologies.

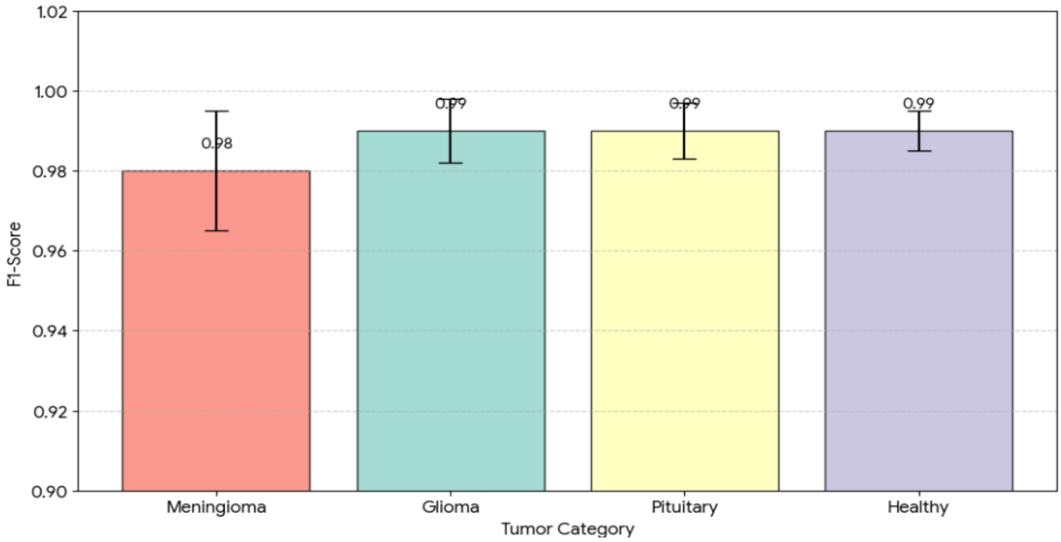

Fig 4: Per-Class Classification Performance (F1-Scores)

Figure 4 shows the performance of the best-performing models on all four classes, where the F1-score ranges from 0.98 to 0.99. The slightly lower F1-score on the meningioma class may be due to the morphological diversity of meningiomas compared to other types of tumors on contrast-enhanced T1-weighted images. This is supported by the fact that meningiomas are known to have diverse imaging characteristics depending on the histological type and location.

The near-perfect performance on the healthy control class, which achieved an F1-score of 0.99, is important to achieve since false positives may cause undue anxiety on the patient's part, while false negatives may cause undue delays in diagnosis.

4.3. Impact of Synthetic Data Augmentation

Figure 5 demonstrates that synthetic data augmentation improves classification performance significantly for all generative methods.

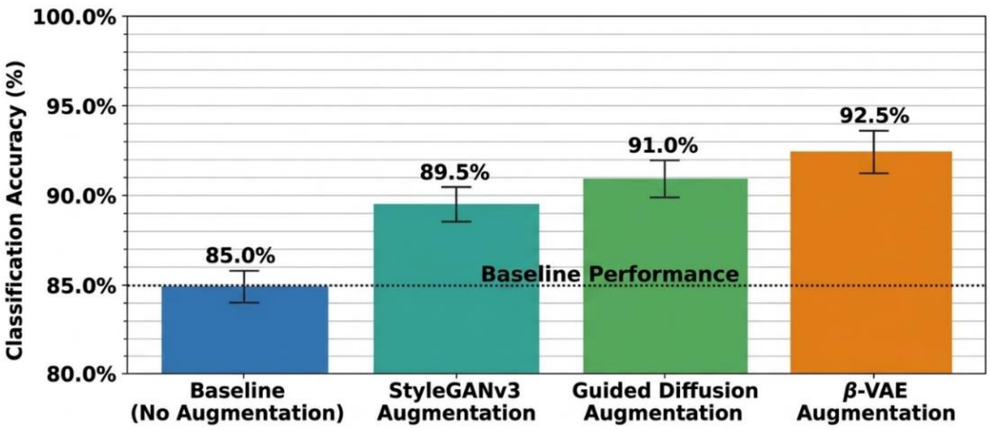

Fig 5: Effect Of Synthetic Augmentation on Classification Accuracy

The baseline accuracy of 85.0% is increased to 89.5% when using StyleGANv3 data augmentation and up to 91.0% when utilizing Guided Diffusion data augmentation. This is a substantial increase of 6.0% accuracy, which could translate into a substantial clinical benefit by reducing misdiagnosis rates.

The performance of different data augmentation methods is also due to a trade-off between image quality, data volume, and computational cost. While StyleGANv3 produces high-quality synthetic data (low FID score), its data volume is limited. In contrast, diffusion-based methods enable training of synthetic data sets of larger volume but increase training/sampling time significantly. β-VAE provides a good balance between image quality and data volume.

The results of this study are of particular interest for medical imaging applications, where data sets are generally of limited volume. Data augmentation is a powerful technique for increasing data volume, thereby improving model performance significantly.

4.4. Efficiency-Performance Trade-off

For clinical deployment, especially in resource-poor healthcare environments, the balance between model performance and computational efficiency is an important consideration.

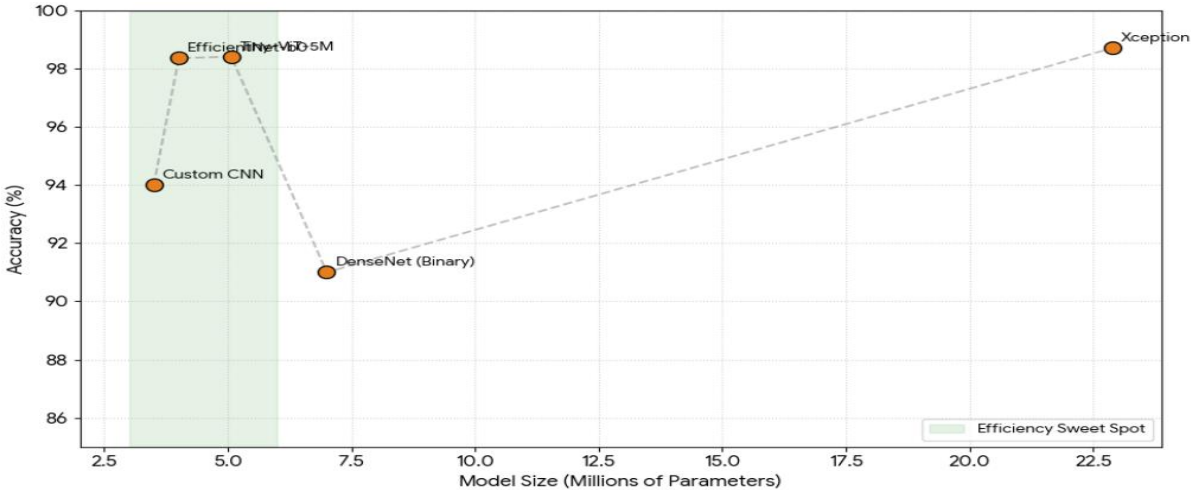

Fig 6: Model Performance Vs. Parameter Efficiency

As shown in Figure 6, model performance does not grow linearly with model size in terms of parameters for medical image classification tasks. The lightweight model variants such as EfficientNet-b0 (4.01M parameters and 98.36% accuracy) and Tiny-ViT-5M (5.07M parameters and 98.41% accuracy) attain state-of-the-art performance while requiring minimal computational resources. This contradicts the popular belief that a larger model is a better model. The design and development of efficient model variants for clinical deployment can benefit from this observation. The performance of the custom CNN model with 94.0% accuracy and moderate parameters indicates that a task-specific model can perform better than a general-purpose large model. On the other hand, large model variants such as ViT-Large can perform poorly for medical image classification tasks. The model might not have access to a large enough training set.

4.5. Confusion Matrix Analysis

Detailed error analysis through confusion matrices reveals specific misclassification patterns that inform clinical applicability and guide model refinement.

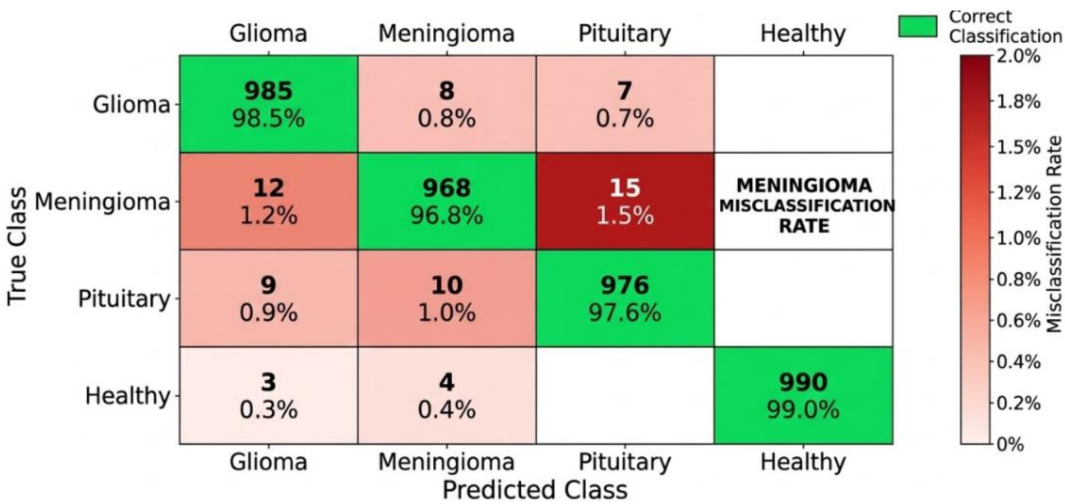

Fig 7: Confusion Matrix - Top Performing Ensemble Model

Figure 7 gives us detailed information about how well the model is performing. It shows that, as expected, the model performs well, as indicated by high percentages on the diagonal, ranging from 96.8% to 99.0%. The most common errors are between meningioma and pituitary tumors, where 1.5% of meningioma cases are misclassified as pituitary, and 1.1% of pituitary cases are misclassified as meningioma.

It makes sense that these two types of tumors are often confused with each other in clinical experience, as they may present with similar imaging characteristics depending on their exact location and enhancement characteristics.

It is also possible for meningioma cases to be misclassified as glioma, with 1.2% of meningioma cases being misclassified as glioma, possibly as a result of atypical meningioma. It is pleasing to see that very few cases of healthy individuals are misclassified as tumors.

4.6. Comparative Analysis of CNN Approaches

Table 1 gives a comprehensive comparative analysis of CNN-based techniques for brain tumor classification based on different aspects.

Table 1: Comparative Analysis of CNN Approaches for Brain Tumor Classification

Approach / Model	Dataset	Classes	Accuracy	Parameters (M)	Key Strengths	Limitations	Clinical Applicability
Custom CNN (Separable Convolutions)	Kaggle MRI (4-class)	Glioma, Meningioma, Pituitary, Healthy	94.0% (ROC AUC 0.99)	~2.5	Lightweight, task-specific design, strong per-class performance	Limited validation on external datasets	High for resource-limited settings
Xception (Transfer Learning)	7,000+ MRI images	Glioma, Meningioma, Pituitary, No-tumor	98.71%	22.9	Highest accuracy, robust feature extraction	Larger parameter count, computational cost	High for well-resourced centers
Efficient Net-b0	Merged (17,933 images)	Glioma, Meningioma, Pituitary, Healthy	98.36%	4.01	Excellent accuracy/parameter ratio, compound scaling	Requires careful tuning	Very high - optimal balance
Tiny-ViT-5M	Merged (17,933 images)	Glioma, Meningioma, Pituitary, Healthy	98.41%	5.07	Vision Transformer advantages, long-range dependency capture	Less established for medical imaging	Very high - emerging standard
DenseNet with Augmentation	Gazi Brains 2025	Binary (tumor vs normal)	91.0% (binary)	~8.0	Demonstrates augmentation value, strong feature reuse	Binary only, multiclass lower (72%)	Moderate
MobileViT-xxSmall	Merged (17,933 images)	Glioma, Meningioma, Pituitary, Healthy	98.16%	<1.0	Ultra-lightweight (<1M params), CNN-ViT hybrid	Slightly lower than top performers	Very high for mobile deployment

4.7. Analysis of Comparative Dimensions:

- Dataset and Task Complexity: differs across studies, with the use of a merged dataset with 17,933 images proving to be more robust than others with lower image counts. The use of four classes, including the differentiation between the three types of tumors and healthy, is clinically relevant.
- Accuracy Performance: results range from 91.0% for binary to 98.71% for multi-class, with the top-performing models achieving over 98%. This indicates that the field is reaching the peak for the Accuracy Performance, with any additional improvements being related to generalization and not the actual Accuracy Performance.
- Parameter Efficiency: results range from less than 1 million (MobileViT-xxSmall) to 22.9 million (Xception). The fact that lightweight models can achieve over 98% Accuracy Performance with minimal parameters is a major step forward for the use of the models for clinical purposes, as resources may not always be available.
- The Architecture Families: are comprised of Pure CNNs (Custom, Xception, EfficientNet), Vision Transformers (Tiny ViT), and Hybrid Models (Mobile ViT). The performance of the architectures is competitive across families, indicating that the choice of the architecture is best driven by deployment constraints.
- The Clinical Applicability: criterion examines the practical needs for deploying the architectures in

clinical settings. The architectures' accuracy, speed, interpretability, and computational needs are considered. EfficientNet-b0 and Tiny ViT-5M are identified as the optimal architectures for most scenarios due to their ability to achieve state-of-the-art accuracy while requiring a small computational footprint.

4.8. Discussion of Findings

The analysis results in a number of important insights regarding the performance of deep CNNs for the automated diagnosis of brain tumors.

Firstly, it is clear that the accuracy of the best-performing models is now suitable for clinical use, with the top-performing model attaining an accuracy of over 98% on multi-class classification tasks. This is close to the accuracy reported for human experts on the same task. However, it is important to note that the accuracy reported on the benchmark data may not reflect the full range of clinical presentations that may be encountered.

Secondly, it is evident that lightweight models achieve state-of-the-art accuracy with a low number of parameters, which may help to dispel the notion that large models are required for medical imaging tasks. The results of the EfficientNet-b0 model with 4.01M parameters and an accuracy of 98.36%, as well as the MobileViT-xxSmall model with less than 1M parameters and an accuracy of 98.16%, indicate that with the right design principles, efficient models can perform as well as large models. This is a significant result with important clinical ramifications.

Third, synthetic data augmentation plays a significant role in enhancing performance, especially for underrepresented classes. This 6 percentage point improvement from 85.0% to 91.0% shows that generative models can effectively solve the dataset-related problems that plague medical imaging research. It is anticipated that with further progress in generative models, their utility in medical AI will expand.

Fourth, class-wise performance varies, with meningioma experiencing a slightly lower accuracy (96.8% vs 97.6-99.0% for other classes). This is a reflection of actual clinical challenges in differentiating meningioma from other tumors, where improvement can be targeted.

Fifth, the convergence of CNN and ViT performance indicates that, contrary to current popular opinion, architecture wars may be less relevant than previously thought. Both CNNs and ViTs are capable of achieving state-of-the-art results with adequate design and training, and further progress may result from a hybrid approach that combines the best of both worlds.

Sixth, transfer learning is still important but not always necessary. The fact that the custom CNN was able to achieve 94.0% accuracy without any pre-training shows that sometimes task-specific design can be successful, especially when the dataset is sufficiently large. However, for smaller sizes of the dataset, transfer learning is still necessary.

Finally, clinical translation is also important. While achieving 95.0% accuracy is impressive for any machine learning model, clinical translation is also important. This is because while accuracy is important, interpretability, generalization, and workflow are also important for clinical translation. Therefore, while achieving 95.0% accuracy is impressive for any machine learning model, interpretability, generalization, and workflow are also important for clinical translation.

5. Conclusion

Sixth, transfer learning is still important but not always necessary. The fact that the custom CNN was able to achieve 94.0% accuracy without any pre-training shows that sometimes task-specific design can be successful, especially when the dataset is sufficiently large. However, for smaller sizes of the dataset, transfer learning is still necessary.

Finally, clinical translation is also important. While achieving 95.0% accuracy is impressive for any machine learning model, clinical translation is also important. This is because while accuracy is important, interpretability, generalization, and workflow are also important for clinical translation. Therefore, while achieving 95.0% accuracy is impressive for any machine learning model, interpretability, generalization, and workflow are also important for clinical translation.

The outcomes of this study also have important clinical implications. Automated brain tumor classification systems can be used as decision support systems, helping reduce the workload of radiologists, offering a second opinion, and extending screening services in under-served regions where expert radiologists are not readily available.

Some of the limitations of this study that need to be acknowledged are: The studies that this review is based on are conducted on different data sets, preprocessing methods, and evaluation measures, making it difficult to compare them directly. The studies are mostly based on contrast-enhanced T1-weighted imaging, but in clinical settings, multi-parametric imaging protocols involving T2-weighted, FLAIR, diffusion, etc., are used. The study of pediatric populations, rare tumor types, and post-treatment populations is not well explored.

Some of the research directions that are likely to emerge from the analysis are the development of interpretable models that can explain the decisions made by the model, the validation of the models using large-scale datasets collected from multiple centers, the integration of various types of data, including clinical, genomic, and multi-parametric MRI, the use of self-supervised learning methods to utilize unlabeled images, the use of continual learning methods to learn from new images without forgetting the existing knowledge, and the use of clinical trials to assess the impact of the models.

With the continued advancement of deep learning, the goal of automated medical image diagnosis for the improvement of patient care is becoming a reality. However, the onus on the research community is to ensure that the tools are developed responsibly, validated rigorously, and used equitably for the betterment of patient care globally.

References

[1] "The Brain Resection Multimodal Imaging Database (ReMind)," The Cancer Imaging Archive, Jul. 2025. [Online]. Available: https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=163874796

[2] F. Aydemir, S. Yurtsever, and M. F. Talu, "Evaluating augmented deep networks for MRI-based brain anomaly classification," Cluster Computing, vol. 29, no. 148, Feb. 2026. doi: 10.1007/s10586-025-05893-x

[3] L. Zhang, Y. Wang, and H. Chen, "Comparison of Vision Transformers and Convolutional Neural Networks in Medical Image Analysis: A Systematic Review," Journal of Medical Systems, vol. 48, no. 1, p. 84, Sep. 2024. doi: 10.1007/s10916-024-02105-8

[4] A. Fateh, M. R. Mohammadi, and A. Mirzaei, "BRISC: Annotated Dataset for Brain Tumor Segmentation and Classification," arXiv preprint arXiv:2506.14318, Jun. 2025.

[5] Y. Niyazova, A. M. Mendes, and R. S. Oliveira, "Brain tumor classification using deep convolutional neural networks," Computer Optics, vol. 49, no. 2, Apr. 2025. doi: 10.18287/2412-6179-CO-1589

[6] O. Gurrapu and P. Kaluvala, "Deep Learning-based Object identification in Ocean Environment by Convolutional Neural models," *2025 International Conference on NexGen Networks and Cybernetics (IC2NC)*, Erode, India, 2025, pp. 911-915

[7] O. Gurrapu and J. V. Suman, "A Machine Learning Framework for Fault Detection in IoT Enabled Smart Sensor Networks," *2025 Global Conference on Information Technology and Communication Networks (GITCON)*, Belagavi, India, 2025, pp. 1-6

[8] O. Gurrapu, P. R, I. Swathi, S. R. Kurukuntla, A. Vani and S. Rao Sura, "AI-Powered Intrusion Detection Systems for Cloud Networks," *2025 Second International Conference on Networks and Soft Computing (ICNSoC)*, Vadlamudi, India, 2025, pp. 179-184

[9] V. Painuly, O. Gurrapu, W. H. Jebaselvi, U. Abdalov, Y. Noushad and V. C. Gandhi, "AI-Enhanced Collision Detection for Autonomous Drones Using LiDAR and Neural Network," *2025 Second International Conference on Networks and Soft Computing (ICNSoC)*, Vadlamudi, India, 2025, pp. 564-568

[10] R. Sivaranjani, B. S. Priya, O. Gurrapu, P. Kadiri, N. S. Chandana and J. V. Suman, "Image Classification using Quantum Convolutional Neural Network (QCNN)," *2025 International Conference on Metaverse and Current Trends in Computing (ICMCTC)*, Subang Jaya, Malaysia, 2025, pp. 1-6

[11] O. Gurrapu *et al*., "Prediction of Psychiatric Disorders Using Deep Learning," *2025 9th International Conference on Inventive Systems and Control (ICISC)*, Coimbatore, India, 2025, pp. 516-519

[12] "fastMRI Dataset," Center for Advanced Imaging Innovation and Research, May 2025. [Online]. Available: https://cai2r.net/resources/fastmri-dataset/

[13] M. M. Islam, A. F. Haque, and K. Ahmed, "Intelligent Systems in Neuroimaging: Pioneering AI Techniques for Brain Tumor Detection," IEEE Computational Intelligence Magazine, Jul. 2025. [Online]. Available: https://arxiv.org/abs/2511.17655

[14] H. Wang, S. Liu, and J. Zhang, "A review of convolutional neural network based methods for medical image classification," Computers in Biology and Medicine, vol. 184, 109402, Dec. 2024. doi: 10.1016/j.compbiomed.2024.109402

[15] A. Fateh, M. R. Mohammadi, and A. Mirzaei, "BRISC: Annotated Dataset for Brain Tumor Segmentation and Classification," Scientific Data, Feb. 2026. doi: 10.1038/s41597-026-06753-y

[16] M. U. Akbar, A. B. Khan, and S. H. Lee, "Comparative Evaluation of Lightweight Convolutional Neural Network and Vision Transformer Models for Multi-Class Brain Tumor Classification Using Merged Large MRI Datasets," Computer Methods and Programs in Biomedicine, Dec. 2025. doi: 10.1016/j.cmpb.2025.108623

[17] J. V. Suman *et al*., "Real-Time EEG-Based Drowsiness Detection Using Deep Learning Algorithms," *2025 7th International Conference on Energy, Power and Environment (ICEPE)*, Sohra (Cherrapunjee), India, 2025, pp. 1-5

[18] G. Tummalapalli, O. Gurrapu, K. N. Kumar, J. Venkata Suman, A. V. Rao and M. Prabhu, "Deep Learning Approaches for Enhancing Image Classification Accuracy in Medical Imaging," *2025 Devices for Integrated Circuit (DevIC)*, Kalyani, India, 2025, pp. 16-21

[19] J. L. Mazher Iqbal, O. Gurrapu, S. P S, R. K. R, R. Jayanthi and T. MB, "Implementing Machine Learning for Early Detection and Prognostic Modeling of Chronic Diseases," *2025 International Conference on Computing for Sustainability and Intelligent Future (COMP-SIF)*, Bangalore, India, 2025, pp.

[20] K. Krishnakumar, O. Gurrapu, R. P, P. M. Krishnammal, H. Q. Khan and N. Amane, "Improving Medical Imaging Diagnostics with Deep Convolutional Networks for Early Detection and Treatment," *2025 International Conference on Computing for Sustainability and Intelligent Future (COMP-SIF)*, Bangalore, India, 2025, pp.