Pakistan Journal of Medical & Cardiological Review
https://pakjmcr.com/index.php/1/about
Online ISSN 3007-2387
Print ISSN 3007-2379

Deep Learning-Based Comparative Framework for Automated Brain Tumor Diagnosis Using MRI

Habib Ullah
GDC Hayatabad, Peshawar, Higher Education Department, KP, Pakistan
Email: hkhabib105@gmail.com
Zafar Khan
GDC Hayatabad, Peshawar, Higher Education Department, KP, Pakistan
Email: zafar.khalil88@gmail.com
Abdullah Malik
GDC Hayatabad, Peshawar, Higher Education Department, KP, Pakistan
Email: abdullahmalik3483@gmail.com
Salman Khan
GDC Hayatabad, Peshawar, Higher Education Department, KP, Pakistan
Email: salmankhan270.ak@gmail.com
Irfan Ullah
GDC Hayatabad, Peshawar, Higher Education Department, KP, Pakistan
Email: irfanullahafridi046@gmail.com
Basirullah
GDC Hayatabad, Peshawar, Higher Education Department, KP, Pakistan
Email: basirullah972@gmail.com

Author Details

Keywords: Brain Tumor, MRI, Deep Learning, CNN, DenseNet201, EfficientNetB3, Transfer Learning, Medical Imaging, Tumor Detection.

Received on 25 Nov 2025
Accepted on 30 Dec 2025
Published on 16 Jan 2026

Corresponding E-mail & Author*:

Habib Ullah
GDC Hayatabad, Peshawar, Higher Education Department, KP, Pakistan
Email: hkhabib105@gmail.com

Abstract

Brain tumors are one of the most dangerous medical conditions because they are tough to detect at an early stage and even harder to treat successfully once they progress, which is why they are often linked with high death rates. MRI scans have become the standard method for brain tumor diagnosis since they produce detailed and high-resolution images of the brain in a safe and non-invasive way. However, despite their usefulness, MRI scans are not always easy to interpret because reviewing hundreds of scans can be time-consuming, repetitive, and stressful for radiologists, and the results may vary depending on the doctor's experience and judgment. With the growing use of Artificial Intelligence in healthcare, deep learning models, especially Convolutional Neural Networks (CNNs), have shown great promise in helping doctors quickly and accurately spot patterns in medical images that may be too subtle or time-intensive for humans to notice. In this research, we compared two robust CNN architectures, DenseNet201 and EfficientNetB3, both trained via transfer learning for brain MRI classification. Unlike earlier studies that often relied on smaller datasets with only around 3,000 images, we built our work on a much larger, balanced dataset of 8,000 MRI scans, evenly divided between tumor and non-tumor cases. This larger dataset helped the models learn more general and reliable features, which are crucial for medical applications where accuracy directly

impacts patient care. Both models were fine-tuned and carefully evaluated using multiple performance metrics, including precision, recall, F1-score, and ROC-AUC, to assess not only how well they detected tumors but also how consistent and reliable their predictions were. The results showed that DenseNet201 achieved a validation accuracy of 96.85%, while EfficientNetB3 reached 96.48%, with both models also performing strongly across precision, recall, and F1-score. These results highlight that dataset size and model architecture both play a massive role in how accurate a deep learning system can be. While each model had its strengths, the overall takeaway is that deep learning can genuinely support medical experts by reducing errors, saving time, and improving consistency in brain tumor diagnosis. This technology has the potential to reduce radiologists' workload, improve the likelihood of detecting tumors earlier, and ultimately increase patient survival rates, making AI a strong candidate for integration into real-world clinical settings in the near future.

Introduction

Brain tumors are among the most challenging health issues faced today. They grow silently, are often detected late, and can be very aggressive. According to the World Health Organization, tumors of the central nervous system represent a significant cause of illness and death worldwide [1]. One of the most severe types, glioblastoma multiforme, has a survival rate of less than 15 months even when patients undergo surgery, chemotherapy, and radiotherapy [2]. This makes early and accurate detection extremely important, as detecting tumors at an early stage can improve treatment outcomes and patient survival. Magnetic Resonance Imaging (MRI), commonly known as MRI, has become the go-to tool for diagnosing brain tumors. It offers detailed, high-resolution images of brain tissue without exposing patients to harmful radiation [3]. Unlike older imaging techniques, MRI can provide both structural and functional information about the brain, making it a reliable choice for doctors. However, interpreting MRI scans is not easy. Radiologists may have to carefully review hundreds of images for a single patient, which is both time-consuming and exhausting [4]. Human error is also a factor. Fatigue, inexperience, or differences in judgment between radiologists can sometimes lead to inconsistent diagnoses [5]. On top of that, brain tumors are not uniform in appearance; they vary significantly in shape, size, and position, which adds to the challenge [6].

This is where Artificial Intelligence has started to play an important role. In recent years, deep learning techniques, especially Convolutional Neural Networks (CNNs), have transformed the analysis of medical images. These models have already proven successful in other areas of medicine, such as detecting lung cancer from CT scans [7], identifying diabetic retinopathy from retinal images [8], and even classifying skin cancers with high accuracy [9]. What makes CNNs so powerful is that they automatically learn the features needed to distinguish between healthy and diseased tissue, rather than relying on manually designed features, which often fail when tested on new data [10]. For brain tumor detection, CNNs have shown promising results, but limited small datasets have hindered earlier research. Many studies used only 3,000-3,400 MRI images [11], which is insufficient for training deep networks. Small datasets often lead to overfitting, meaning the model performs well on the training data but fails to generalize to new cases [12]. This becomes a serious problem when applying the model to real hospital data. To address this gap, our study uses a much larger dataset of 8,000 MRI images, with an equal split between tumor and non-tumor classes. This balance helps the model avoid bias and improve fairness and reliability [13]. Our research focuses on comparing two state-of-the-art CNN models: DenseNet201 and EfficientNet-B3. DenseNet201 introduces a clever idea of dense connectivity, where each layer is connected to all subsequent layers. This setup allows features to be reused across the network, reduces redundancy, and makes training

more efficient [14]. On the other hand, EfficientNet-B3 uses a compound scaling approach that balances depth, width, and resolution. This makes it highly efficient, achieving strong accuracy with fewer computational resources [15].
The interesting part about comparing these two models is that they represent two very different strategies for solving the same problem. DenseNet201 is designed to maximize information flow by keeping layers closely connected, while EfficientNet-B3 aims to strike a balance between performance and efficiency. By testing both, we can see not only which model achieves better accuracy but also which is more practical in terms of speed and memory usage. These factors are critical for real-world medical applications where doctors and hospitals may not have access to compelling hardware. This research is not only about achieving the highest possible accuracy. It is also about finding a solution that could realistically be used in hospitals and clinics. A model that requires too much computing power may not be practical, no matter how accurate it is. That is why we consider trade-offs among accuracy, efficiency, and real-world applicability. In medical imaging, even a slight delay can affect patient care, so the system's speed matters as much as its accuracy. Ultimately, the purpose of this study is to explore how AI can assist radiologists rather than replace them. Doctors already face heavy workloads, and having a reliable AI system to flag suspicious MRI scans quickly can save time and reduce errors. This way, radiologists can focus more on treatment planning and patient care, rather than being overwhelmed by repetitive image analysis. In the bigger picture, the hope is that AI-powered tools like the ones explored in this research will play an essential role in early detection, better decision-making, and ultimately improving survival rates for brain tumor patients.

Literature Review

In the beginning, brain tumor classification was tackled with traditional machine learning methods. Researchers would design handcrafted features, such as texture descriptors, geometric shapes, and intensity histograms, and then feed them into models such as Support Vector Machines (SVMs) and Random Forests [16, 17]. These techniques worked reasonably well in controlled settings, but they could not handle the variability of tumor appearance across different patients. Since tumors can differ in shape, size, and location, handcrafted features were not robust enough [18].
A significant shift happened in 2012 when deep learning entered the scene. The landmark success of AlexNet [19] in the ImageNet competition was a turning point for image analysis. Using Convolutional Neural Networks (CNNs) with GPU acceleration, AlexNet outperformed classical methods by a wide margin. This breakthrough inspired researchers to apply similar ideas to medical imaging. Later models, such as VGGNet [20] and ResNet [21], extended these achievements by addressing vanishing gradients and enabling the training of much deeper networks. These improvements quickly attracted attention in the medical community [22, 23]. It did not take long for CNNs to make their mark in brain imaging. Pereira et al. [24] were among the first to show that CNNs could outperform conventional methods for glioma segmentation in MRIs. Havaei et al. [25] followed with a two-pathway CNN that considered both local and global contexts, producing stronger tumor segmentation results. Another interesting effort was by Anaraki et al. [26], who combined CNNs with genetic algorithms, though the computational requirements limited real-world use. Around the same time, Kamnitsas et al. [27] developed a 3D CNN with conditional random fields to handle multimodal MRI scans, further improving accuracy.
DenseNet brought another wave of innovation. Introduced by Huang et al. [14], DenseNet connects each layer to all previous layers, enabling better gradient flow and feature reuse. This architecture avoided redundancy while strengthening learning efficiency. DenseNet201, a deeper variant, became popular in healthcare. Rajpurkar et

al. [28] applied it to chest X-ray classification with impressive results, while Khan et al. [29] and Rehman et al. [30] successfully used it for brain tumor classification and segmentation. Other studies [31] demonstrated DenseNets' adaptability across radiology tasks, showing it to be a reliable and efficient option [32]. EfficientNet was another breakthrough in model design. Tan and Le [15] introduced compound scaling, which carefully balanced depth, width, and resolution rather than scaling a single factor. This made the models smaller yet more accurate, which is highly valuable in clinical environments where computational resources are limited. EfficientNet has been tested in different medical domains. Gulshan et al. [8] applied deep learning to diabetic retinopathy detection, Esteva et al. [9] demonstrated dermatologist-level accuracy in skin cancer classification, and Kermany et al. [33] applied CNNs to eye diseases and pneumonia diagnosis. Campanella et al. [34] later demonstrated that efficient networks could handle large numbers of pathology slides, demonstrating their scalability for large medical datasets.

Despite the progress, dataset limitations have remained a recurring problem. Many earlier brain tumor studies were based on fewer than 3,500 MRI scans [11]. With such small datasets, models often overfit and perform poorly on unseen patients. Akkus et al. [35] highlighted the need for larger and more diverse datasets, while Cheng et al. [36] noted that robust generalization is critical for clinical adoption. Bilic et al. [37] and Bakas et al. [38] contributed by organizing multi-institutional datasets, enabling researchers to train on tens of thousands of images. More recent work by Menze et al. [39] and Isensee et al. [40] has also established standardized benchmarks that have advanced the field. To make up for limited data, researchers have tried creative strategies. Pereira et al. [24] used augmentation techniques to balance tumor versus non-tumor cases, while Afshar et al. [41] proposed capsule networks as an alternative to CNNs for data-limited settings. More recently, domain adaptation and transfer learning approaches have been explored, allowing pre-trained networks to adapt to MRI classification tasks with relatively small datasets [42, 43]. These methods help models perform better in real-world conditions where labeled medical data is scarce.

Another issue that researchers continue to emphasize is evaluation. While accuracy is the most commonly reported metric, it often tells only part of the story. Litjens et al. [22] argued that other metrics, such as precision, recall, F1-score, and ROC-AUC, are essential in clinical settings, since false positives and false negatives can carry very different consequences. Tang et al. [44] further noted that interpretability, efficiency, and workflow integration are as important as raw performance for clinical deployment. Erickson et al. [45] also stressed the need for explainable AI, since black-box models can create distrust among clinicians. Looking at the literature as a whole, it is clear that brain tumor detection research has moved through different phases: from handcrafted features to CNNs, from CNNs to DenseNet, and from DenseNet to EfficientNet and beyond. Each step has improved the balance between accuracy and efficiency. At the same time, there has been a growing awareness that dataset size, diversity, and evaluation methods matter just as much as architecture choice. Building on this, the current study compares DenseNet201 and EfficientNet-B3 on a dataset of 8,000 MRI scans, aiming to provide a clearer picture of which model performs best in a practical, clinically relevant setting.

Methodology

In this study, we used two advanced convolutional neural networks, DenseNet201 and EfficientNetB3, to classify tumors in medical images. These networks were chosen because they are highly effective at learning detailed, hierarchical features from pictures while keeping the total number of parameters manageable, which is essential for efficient training and deployment. DenseNet201 follows a unique connection pattern in which each layer can directly access the outputs of all preceding layers. This approach allows the network to reuse important features throughout its depth,

improves gradient flow during training, and helps the model learn more effectively from limited data. EfficientNetB3 works differently. It combines Mobile Inverted Bottleneck Convolutions with Squeeze-and-Excitation blocks, which allow the network to focus on the most relevant features by recalibrating channel-wise information. Additionally, EfficientNetB3 employs a compound scaling method that carefully balances the network's depth, width, and input image resolution, allowing it to extract high-quality features while remaining computationally efficient. By employing these two networks, one focused on dense connectivity and feature reuse, and the other on efficient scaling and channel recalibration, we were able to compare distinct design strategies and evaluate their performance in analyzing medical images with subtle and complex tumor patterns. Figure 1 represents the overall roadmap of this study.

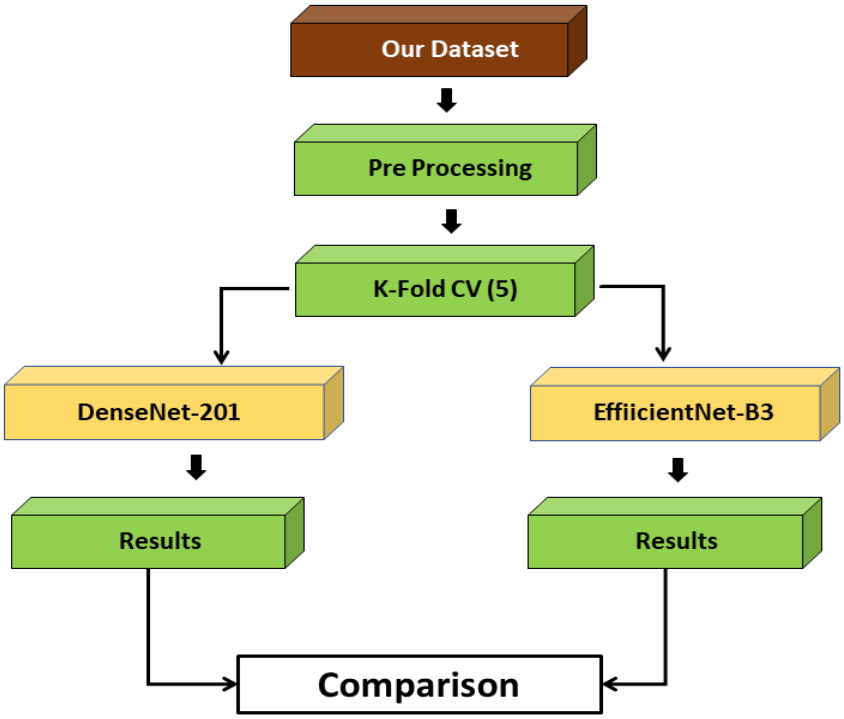

Figure 1 Overall roadmap of the methodology

Data Acquisition

For this study, we worked only with MRI data that were already publicly available, fully anonymized, and ethically cleared, so there was no need to collect any new information from patients. We relied on two established open and publically available datasets: The Brain Tumor Segmentation (BraTS) 2020 dataset and the Kaggle Brain Tumor MRI dataset, both of which were created by institutions that had already gone through the proper ethical review processes and obtained informed consent from the patients involved or their legal guardians at the time the scans were originally acquired. Before these datasets were released for public research use, all identifying details were removed, including names, dates, and any metadata that could link a scan back to a specific person. As a result, we never interacted with private health information, clinical records, or any form of identifiable data. Since our work is based entirely on these pre-existing, fully de-identified MRI scans and does not involve recruiting participants, performing procedures, or communicating with patients or guardians, no additional ethical approval or authorization was required. Bringing together images collected under different scanners, protocols, and clinical settings allowed us to assemble a dataset with natural variability, while still staying fully within established ethical standards and working only with data that were already cleared for open scientific use.

The BraTS 2020 dataset formed the core of our collection because of its exceptional quality and expert oversight. It includes pre-operative MRI scans from patients diagnosed with glioblastoma and lower-grade gliomas, gathered from multiple international medical centers. Each case provides four MRI sequences, T1, T1-CE, T2, and FLAIR, along with carefully prepared tumor segmentation masks reviewed by experienced neuroradiologists. This level of annotation and consistency makes BraTS one of the most respected datasets in neuroimaging research. We accessed it through the official CBICA portal at the University of Pennsylvania after completing the required registration steps and agreeing to the non-commercial usage license. In accordance with the dataset guidelines, we cited the foundational works by Bakas et al. (2017, 2018) [38, 46] and Menze et al. (2015) [47].

To introduce additional variability, we incorporated the Kaggle Brain Tumor MRI dataset. Unlike BraTS, which follows strict imaging standards, the Kaggle dataset contains MRI scans collected under a broader range of clinical conditions. This variation includes differences in scanner models, resolutions, acquisition settings, and patient backgrounds. Such diversity is valuable because it exposes the model to real-world inconsistencies that are often missing in more curated datasets. The Kaggle dataset is freely available without registration, and as recommended, we acknowledged Nickparvar (2021) [48] as the original contributor. Although it is not as structured as BraTS, its heterogeneity played an important role in strengthening the overall dataset.

Once both datasets were collected, we carried out a thorough cleaning and integration process. We removed corrupted files, blurry scans, incomplete slices, and duplicate images to ensure that only reliable data were used for training. Because BraTS and Kaggle follow different naming conventions and labeling styles, we standardized all metadata, including filenames, modality indicators, and patient identifiers. This step was essential for preventing data leakage, ensuring that images from the same patient never appeared in both the training and testing sets.

After the metadata was unified, we verified and harmonized the class labels. BraTS provides expert-validated tumor masks, so labeling accuracy was already ensured. However, the Kaggle dataset required manual inspection to confirm that its tumor and non-tumor labels were correct. When the labeling process was complete, we addressed class imbalance, a common issue in medical imaging, by applying controlled up-sampling and down-sampling. This resulted in a balanced dataset containing 4,000 tumor scans and 4,000 healthy scans.

The final dataset of 8,000 carefully curated MRI images was organized into a clear and consistent directory structure with separate folders for training and testing. This standardized arrangement supports reproducibility, simplifies preprocessing, and aligns well with modern deep learning pipelines. By combining the expert-annotated BraTS images with the naturally varied Kaggle scans, we were able to build a dataset that is both diverse and clinically meaningful, providing a strong foundation for the models developed in this study.

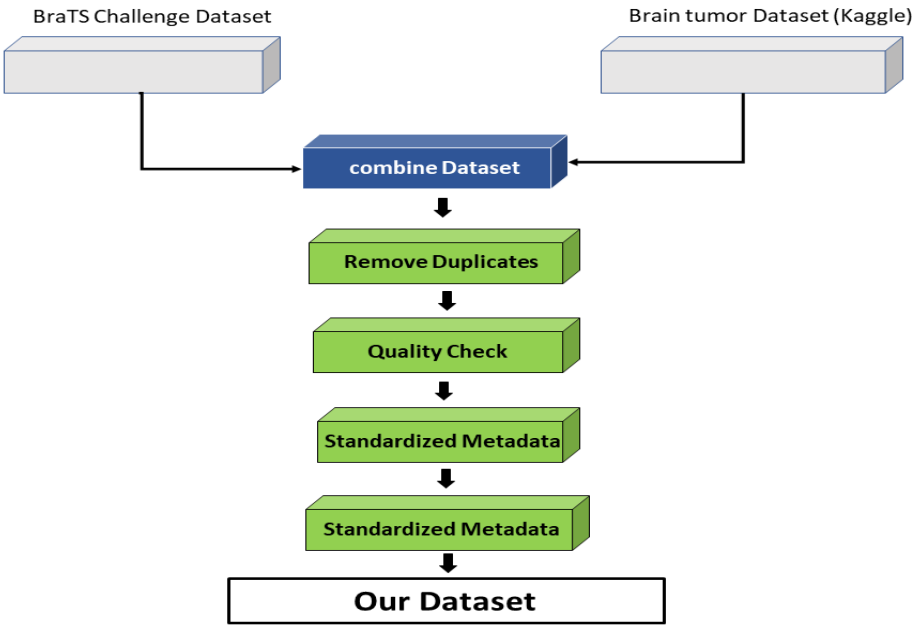

Figure 2: Data Acquisition Tehnique

Data Preprocessing

After finalizing the dataset, the next step was preprocessing. Raw MRI scans contain valuable information, but they are not immediately ready for use in deep learning models. They come in different sizes, show variations in brightness and contrast, and often carry noise that can confuse a model. Preprocessing is the stage where data is cleaned, adjusted, and standardized so the model sees everything in a consistent format. Without this step, even the most advanced neural network would struggle to identify meaningful patterns. We began by resizing all MRI scans to 224×224 pixels. This size was chosen because it matches the input requirement of EfficientNet-B3, one of the models we compared in this study. Resizing gave us uniformity across the dataset and avoided unnecessary distortions. A larger size would have required much more computation without adding significant value, while a smaller size might have led to essential tumor details being lost. The 224-by-224 resolution offered a good balance between efficiency and detail.

The next stage was pixel normalization. Different MRI machines and different patients often result in variations in pixel intensity. If left untreated, these variations could cause the model to focus on brightness differences rather than learning the actual tumor patterns. To solve this, we used the preprocess_input function from the TensorFlow EfficientNet module, which scales all pixel values to a standard range. This ensured that all images were treated fairly and consistently, like putting them on the same playing field. Because medical datasets are still relatively small compared to large datasets like ImageNet, we applied data augmentation to improve the model's robustness. Augmentation involves creating new or slightly modified versions of existing images. In this study, we used rotations of up to 15 degrees, horizontal flipping, width and height shifts of up to 10%, and zoom transformations of up to 10%. These changes mimic natural variations that occur in hospitals, such as slight differences in how patients are positioned in the MRI machine. By training on a broader range of images, the model was less likely to overfit and performed better on new scans.

We also wanted to ensure that our model testing was reliable and unbiased. For this reason, we used 5-fold cross-validation rather than a single train-test split. The dataset was divided into five equal parts. In each round, four parts were used for training, and the fifth for validation. This was repeated until every part had been used for validation once. Cross-validation gave us a more realistic picture of how the model would

perform on unseen data and reduced the chance of reporting overly optimistic results. To keep everything organized, the dataset was stored in a directory structure where each image was placed in a folder corresponding to its class (tumor or non-tumor). This made it easy to integrate the dataset with TensorFlow and Keras, which can automatically identify labels from folder names. Finally, we saved the preprocessed images and the dataloaders. This way, the entire preprocessing pipeline did not need to be repeated for every model training. It saved time, reduced computational cost, and made our experiments easier to reproduce. Figure 3 illustrates the sequential procedure used for data preprocessing, detailing how the raw data is prepared prior to being used by the model.

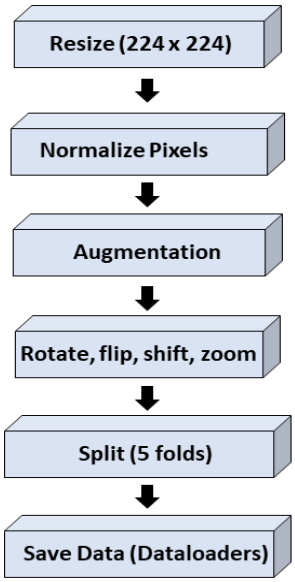

Figure 3: Data Preprocessing

DenseNet201 Architecture

DenseNet201 is a deep convolutional neural network comprising 201 layers, structured into four dense blocks separated by transition layers. The architecture is based on the principle of dense connectivity, in which each layer receives feature maps from all preceding layers rather than just the immediately previous one. This connectivity pattern enhances gradient flow throughout the network, alleviates the vanishing gradient problem, and promotes feature reuse. Consequently, DenseNet201 can learn richer and more discriminative representations while maintaining parameter efficiency compared to conventional deep networks. Within each dense block, convolutional layers are organized using a bottleneck design. Each layer initially applies a 1×1 convolution, which reduces the number of feature maps and minimizes computational cost, followed by a 3×3 convolution that extracts spatial features. The network employs a growth rate $k = 32$, which determines the number of new feature maps contributed by each layer. This controlled growth allows the network to progressively learn complex hierarchical features without unnecessarily expanding the parameter count, maintaining a balance between depth and width.

Transition layers are introduced between dense blocks to manage model complexity and reduce the spatial dimensions of feature maps. Each transition layer comprises a 1×1 convolution followed by 2×2 average pooling. This design compresses the feature maps while preserving important information, enabling the network to remain computationally efficient as its depth increases. After processing through the final dense block, global average pooling is applied to aggregate spatial information from the feature maps into a single vector. This vector is then passed through a sigmoid activation function for binary tumor detection as shown in the figure 4. The

combination of dense connections, bottleneck layers, and transition layers enables DenseNet201 to efficiently extract hierarchical, highly discriminative features, making it particularly suitable for medical imaging applications where precise tumor identification is critical. Figure 4 presents the full architecture of DenseNet201, illustrating how its layers and connections are organized from the initial input stage through to the final output.

\

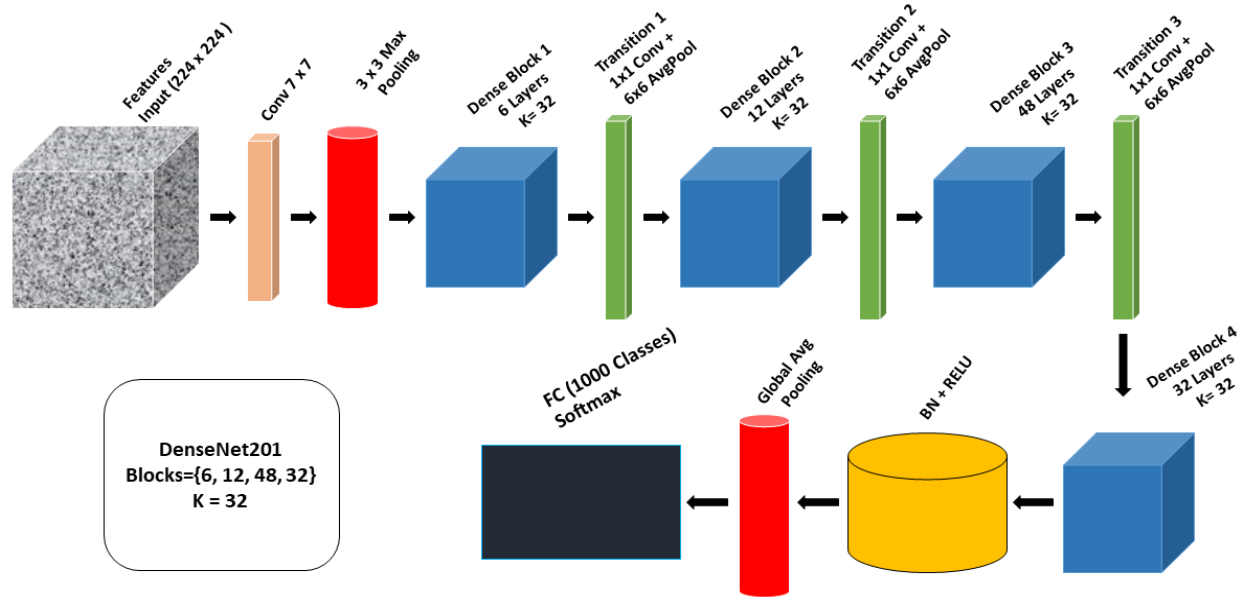

Figure 4: DenseNet-201 Architecture

EfficientNetB3 Architecture

EfficientNetB3 is a convolutional neural network built upon Mobile Inverted Bottleneck Convolutions (MBConv) and enhanced with Squeeze-and-Excitation (SE) blocks. Each MBConv block first expands the input channels, then applies a depthwise convolution to extract spatial features efficiently. Subsequently, SE blocks recalibrate the importance of each channel, and the features are projected back to a lower-dimensional space. This architecture enables EfficientNetB3 to capture both spatial and channel-wise dependencies while maintaining computational efficiency, allowing the network to extract rich, discriminative features without excessive parameter usage. Unlike conventional convolutional networks that scale only depth, width, or input resolution independently, EfficientNet employs a compound scaling method, which uniformly balances depth, width, and input resolution. This scaling approach ensures that network growth remains efficient and that all feature representations are proportionally enhanced, maintaining optimal model performance across different layers. EfficientNetB3 accepts input images of size 224 × 224 and contains approximately 12 million parameters, striking a balance between model complexity and computational efficiency, making it suitable for high-resolution image analysis tasks. The network architecture begins with a stem convolutional layer, followed by a sequence of MBConv blocks with varying kernel sizes, expansion ratios, and strides. These blocks progressively extract multi-scale features, capturing both local and global information critical for precise image classification. After the final MBConv block, a 1 × 1 convolution is applied, followed by batch normalization and Swish activation, which introduce non-linearity and ensure smooth gradient flow during training. Figure 5 presents the architecture of EfficientNet-B3, highlighting how its layers and components are structured throughout the network

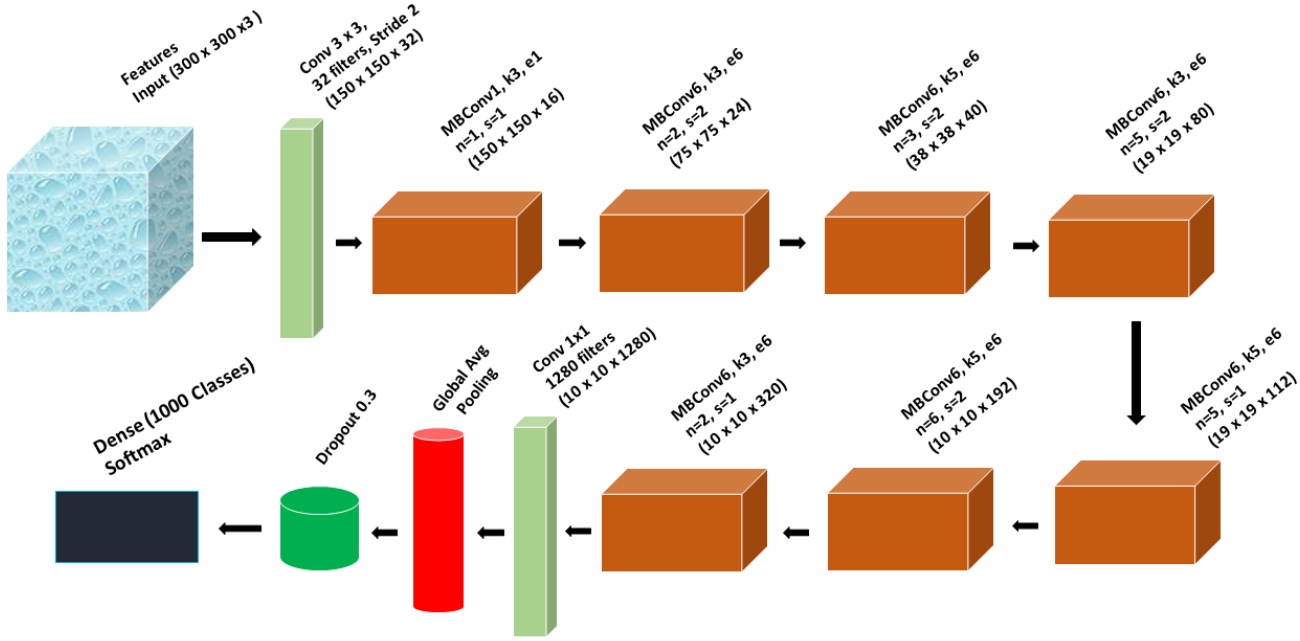

Figure 5: EfficientNet-B3 Architecture

To consolidate spatial information, global average pooling is applied to the final feature maps, followed by dropout regularization to reduce overfitting. The network concludes with a fully connected classification layer using a sigmoid activation function, enabling binary tumor classification. The combination of MBConv blocks, SE recalibration, and compound scaling allows EfficientNetB3 to achieve a remarkable balance of accuracy, computational efficiency, and parameter efficiency. These properties make it particularly suitable for medical image analysis, where high-resolution inputs and subtle differences in tissue structures must be accurately distinguished.

Training Setup

Training for all models was performed on an NVIDIA Tesla P100 GPU with 16 GB of memory, providing sufficient computational resources for efficient model optimization. The Adam optimizer was employed to update the network weights due to its adaptive learning rate capabilities and robust convergence properties. An initial learning rate of 0.001 was set, which was adaptively reduced by a factor of 0.5 whenever the validation loss plateaued, ensuring stable convergence while preventing overfitting.

A batch size of 16 was selected to balance computational efficiency with sufficient gradient estimation, and training was conducted for a maximum of 50 epochs. Early stopping was implemented to terminate training if the validation loss did not improve for 4 consecutive epochs, thereby preventing unnecessary computation and reducing the risk of overfitting. Additionally, mixed-precision training was enabled to optimize memory usage and accelerate training, particularly beneficial for deep architectures such as DenseNet201 and EfficientNetB3

$$L = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1-y_i)\log(1-p_i)]$$

The binary cross-entropy loss function was used to guide model optimization, which is appropriate for the binary classification task of tumor detection. This loss function penalizes incorrect predictions in proportion to their confidence, allowing the models to focus on difficult-to-classify instances while maintaining stable learning dynamics.

Model performance was comprehensively evaluated using multiple metrics, including accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrix analysis. Accuracy provided an overall measure of correct predictions, while precision and recall quantified the model's ability to identify positive cases and avoid false negatives correctly. The F1-score offered a balanced measure of precision and recall, and ROC-AUC assessed the model's discrimination ability across different thresholds. Confusion matrices were further analyzed to identify specific patterns of misclassification, providing insight into the strengths and weaknesses of each model across all cross-validation folds.

Results

This chapter presents the results of experiments using two state-of-the-art convolutional neural network models, DenseNet201 and EfficientNetB3, on the brain tumor MRI dataset. The primary task was to classify MRI images into two categories: tumorous and non-tumorous. While numerical results, such as accuracy, loss, and other performance metrics, are presented here, the aim of this chapter is not only to showcase numbers but also to make sense of them, highlighting their importance for medical applications. A simple presentation of percentages and scores would not fully capture what these outcomes mean in terms of real-world deployment. Therefore, this chapter takes a closer, more descriptive approach, reflecting on why the models behave as they do and what the outcomes suggest for actual medical practice.

The results are divided into different parts. First, the overall accuracy and loss values are discussed, as they are the most direct indicators of how well the models performed. This is followed by an analysis of the confusion matrices, which help uncover the types of errors the models made. Then, a detailed look at other performance metrics such as precision, recall, F1-score, and AUC is presented. Finally, the learning dynamics during training are explored, followed by a comparative interpretation of the two models.

Classification Accuracy and Loss

Accuracy and loss are often the first indicators people look at when trying to understand how well a model is performing. Accuracy tells us the percentage of cases in which the model made the correct prediction, while loss measures how far the predictions are from the actual ground truth. Together, these metrics give us a good first impression of whether the model has truly learned from the dataset or whether it is simply guessing.

DenseNet201 achieved a very high training accuracy of 97.65% and a validation accuracy of 96.85%. These values suggest that the model correctly classified nearly all MRI images. The loss values were 0.0833 for training and 0.0807 for validation, both very small, indicating that the model's predictions were very close to the correct labels. What makes this even more convincing is that the training and validation accuracies were closely aligned. In simpler terms, this means that DenseNet201 was not just memorizing the training images, but was actually learning meaningful features that it could also apply to unseen data. This is a critical aspect in medical imaging, where overfitting could lead to disastrous outcomes if the model fails to generalize to new patient scans.

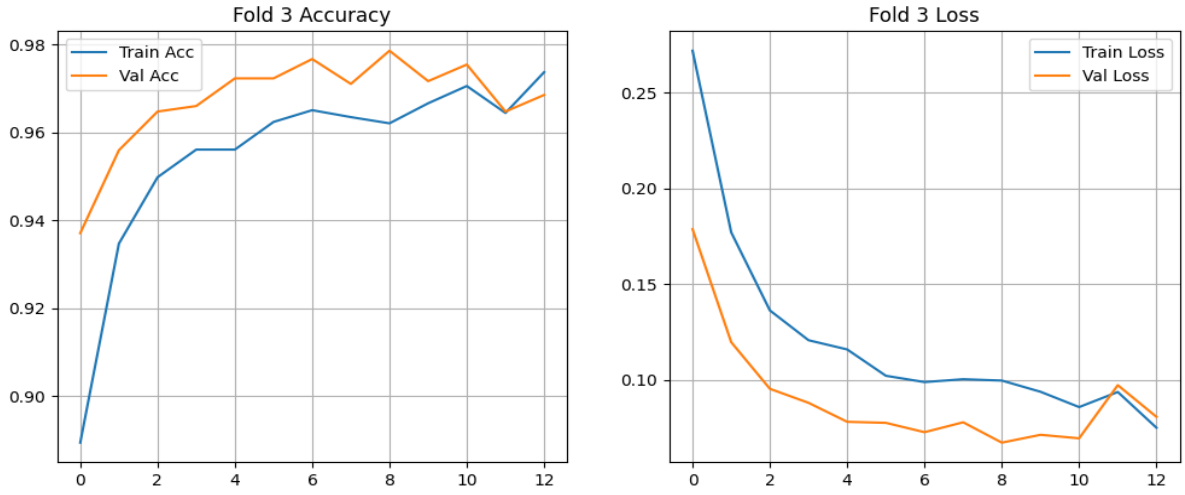

Figure 6: DenseNet 201 Accuracy and Loss Curves

The accuracy and loss curves of DenseNet201 also provide a visual story. At the beginning of training, accuracy rose quickly as the model began to detect basic shapes, edges, and textures in the MRI images. As training continued, the curve started to level off and stabilize around 97–98%, showing that the model had reached a point where it was consistently recognizing tumors and non-tumors correctly. The loss curve, on the other hand, steadily decreased without sudden jumps or irregular patterns, which is a positive sign of stable learning. Notably, the validation curves mirrored the training curves, which means DenseNet201 was not overfitting but instead extracting meaningful patterns. Figure 6 shows the accuracy and loss curves for DenseNet201, providing a clear view of the model's performance throughout training.

On the other hand, EfficientNetB3 achieved training and validation accuracies of 96.32% and 96.48%, respectively. Although these numbers are slightly lower than DenseNet201's, they are still impressive and show strong performance. Interestingly, the validation accuracy was slightly higher than the training accuracy, which is somewhat unusual but a good sign. It indicates that EfficientNetB3 generalized very well and may have benefited from built-in architectural design elements, such as compound scaling and dropout, that prevent overfitting.

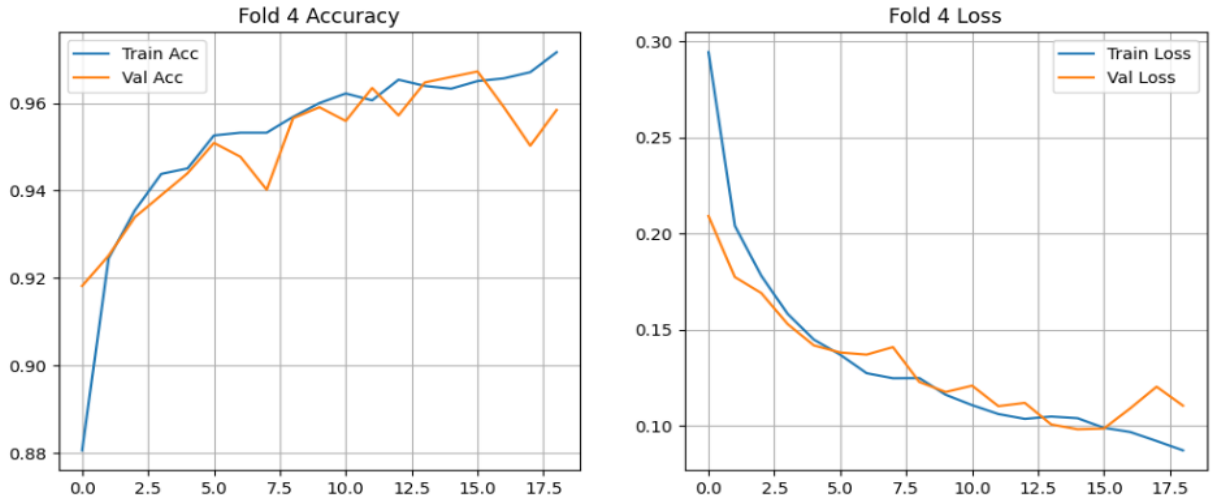

Figure 7: Efficient-Net B3 Accuracy and Loss Curves

The accuracy and loss curves for EfficientNetB3 showed a similar trend to DenseNet201, though there were subtle differences. The accuracy climbed rapidly in the early epochs, suggesting that the model learned the dataset's dominant features quickly. However, the plateau occurred a little earlier and was slightly lower than that of DenseNet201, which explains its marginally lower accuracy. The loss curve also decreased smoothly, but the final loss remained higher than DenseNet201's,

suggesting that EfficientNetB3 found it more challenging to capture very subtle differences in tumor features. Figure 7 shoss the accuracy and loss curves for EfficientNet-B3, offering a detailed view of the model's performance during training.
In summary, both models delivered excellent accuracy, with validation performance above 96%. DenseNet201 performed slightly better at capturing subtle features, while EfficientNetB3 proved very stable and efficient.

Confusion Matrix Analysis

Accuracy gives us the overall picture, but sometimes it hides important details. In medical contexts, not all mistakes are equal. Misclassifying a healthy patient as having a tumor (false positive) is inconvenient and stressful, but misclassifying a patient with cancer as healthy (false negative) could be life-threatening. To better understand these differences, confusion matrices were analyzed.

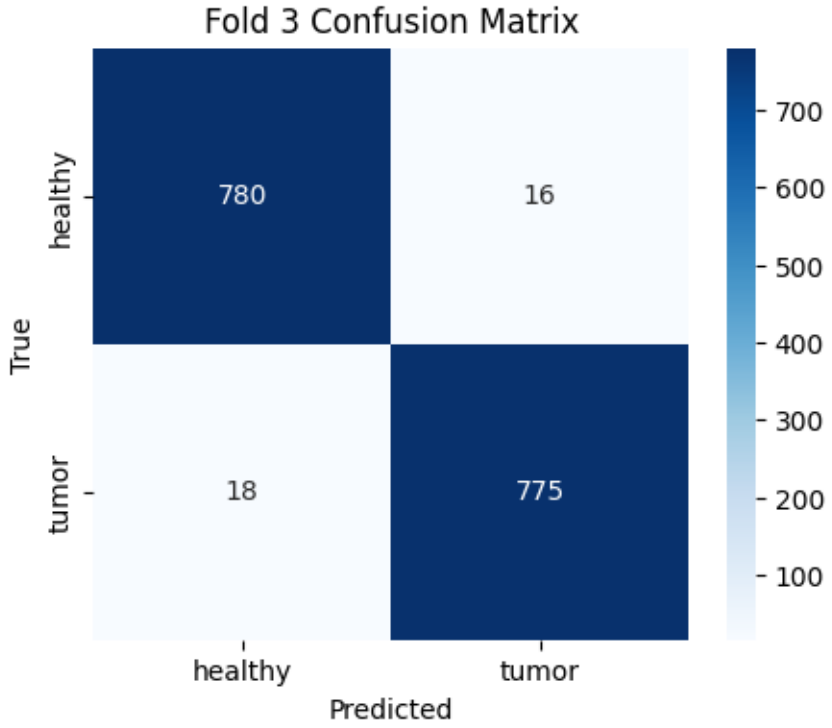

Figure 8: Dense-Net 201 Confusion Matrix

For DenseNet201, the confusion matrix showed that the majority of predictions were correct, with very high counts of true positives and true negatives. In simpler terms, this means the model was highly successful in correctly identifying both tumorous and non-tumorous cases. What was particularly important was the minimal number of false negatives. DenseNet201 was excellent at detecting tumors, rarely missing them. This is a critical strength for medical usage, as missing a tumor can delay treatment and lead to worse outcomes for the patient. Figure 8 shows the confusion matrix for DenseNet201, illustrating how the model's predictions compare with the true class labels.

DenseNet201 produced a few false positives, labeling some healthy images as tumorous. While these errors are not ideal, they are less harmful than false negatives. A patient who receives a false-positive result would undergo additional tests, which may cause stress but ultimately ensure safety. In this sense, DenseNet201 is very much aligned with a "safety-first" principle. It is better to raise a false alarm than to miss an actual case of a tumor.

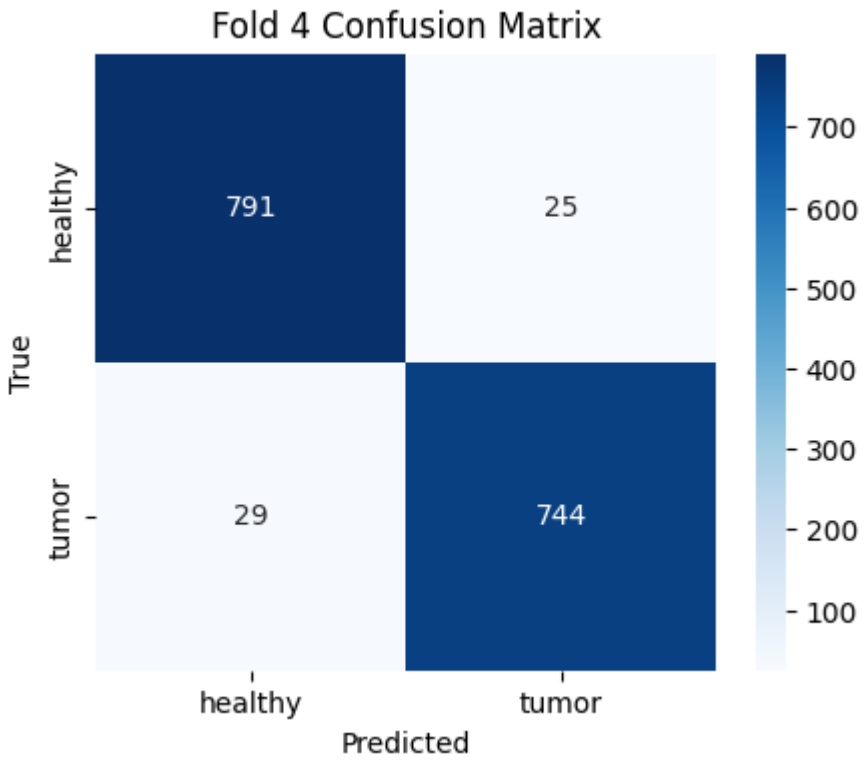

Figure 9: Efficient-Net B3 Confusion Matrix

For EfficientNetB3, the confusion matrix showed a slightly different balance. While it also achieved high true positives and true negatives, the number of false negatives was a bit higher than that of DenseNet201. This means that EfficientNetB3 missed a few more tumor cases. On the positive side, it produced fewer false positives than DenseNet201, meaning it was less likely to alarm patients without tumors incorrectly. This reflects higher precision but slightly lower recall. Clinically, this trade-off can be concerning because while patients without tumors are spared from extra testing, there is a greater risk of overlooking patients who actually need treatment. Figure 9 shows the confusion matrix for EfficientNet-B3, showing the relationship between the model's predicted labels and the actual class labels.

Thus, the confusion matrix analysis clarifies the distinction between the two models. DenseNet201 leaned toward sensitivity, detecting nearly all tumors, while EfficientNetB3 leaned toward precision, minimizing false alarms but sometimes missing subtle tumors.

Performance Metrics

To capture the strengths and weaknesses of each model more fully, additional metrics, including precision, recall, F1-score, and AUC, were calculated. These metrics are critical in medical research, where accuracy alone cannot tell the whole story.

Table 1: Performance of Models

Model	**Training Accuracy (%)**	**Validation Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**AUC (%)**
DenseNet201	97.65	97.03	97.5	98.2	97.8	98.5
EfficientNetB3	96.32	96.48	98.1	96.3	97.2	97.9

DenseNet201 achieved the highest recall of 98.2%, indicating it was excellent at detecting tumor cases. In medical applications, recall is often considered the most critical measure because it reflects the ability to avoid false negatives. A high recall means the model rarely misses a tumor. The F1-score of 97.8% shows that DenseNet201 maintained a good balance between recall and precision. Table 1

presents the performance of the models, summarizing key evaluation metrics such as accuracy, validation accuracy, precision, recall, F1-score and AUC.
EfficientNetB3, on the other hand, had a slightly higher precision of 98.1%. This means that when the model predicted a tumor, it was almost always correct. However, its recall of 96.3% was lower than DenseNet201's, meaning it missed some tumor cases. The F1-score was 97.2%, which is strong but slightly behind DenseNet201. The AUC values for both models were very high (0.985 for DenseNet201 and 0.979 for EfficientNetB3), indicating that both performed well at distinguishing between tumor and non-tumor cases across different thresholds.

ROC and AUC Analysis

Receiver Operating Characteristic (ROC) analysis was performed to evaluate how well each model distinguished between tumor and non-tumor MRI scans at all possible decision thresholds. The ROC curve plots the true-positive rate (sensitivity) against the false-positive rate (1 – specificity), providing a visual measure of a model's ability to separate the two classes. The area under the curve (AUC) summarizes this performance in a single number, where 1.0 represents perfect discrimination and 0.5 corresponds to random guessing.
Both DenseNet201 and EfficientNetB3 achieved remarkably high AUC values, confirming their strong and reliable classification performance. DenseNet201 achieved an AUC of 0.985, while EfficientNetB3 achieved 0.979. These results closely align with the models' accuracy and F1-score trends, showing that DenseNet201 maintains a slight edge in separating positive and negative cases.
Figure 4.5 displays the ROC curves for both models. DenseNet201's curve lies consistently closer to the top-left corner, which represents the ideal balance of high sensitivity and low false-positive rate. This pattern shows that the model can correctly identify most tumor cases without generating many false alarms. EfficientNetB3 also shows an excellent ROC curve, although it slightly lags behind DenseNet201 in the high-sensitivity region. This means that EfficientNetB3, while very precise, may miss a few subtle tumor instances that DenseNet201 successfully detects.
To translate the ROC analysis into practical terms, Youden's J index was used to determine an optimal decision threshold for each model, maximizing the balance between sensitivity and specificity. The corresponding operating points are summarized in Table 4.2. DenseNet201 achieved a sensitivity of 98.2 percent and a specificity of 94.5 percent at its optimal threshold of 0.35, whereas EfficientNetB3 reached a sensitivity of 96.3 percent and a specificity of 96.0 percent at its threshold of 0.40. Table 2 shows the operating points derived from the ROC analysis, outlining the threshold values and corresponding performance measures used to evaluate each model's decision behavior.
These results show that DenseNet201 emphasizes sensitivity – it is less likely to miss a tumor – while EfficientNetB3 emphasizes specificity – it is less likely to raise false alarms. In medical practice, the choice between these trade-offs depends on clinical priorities: for initial screening, a more sensitive model like DenseNet201 is preferable, whereas for confirmatory or follow-up analysis, a more specific model such as EfficientNetB3 may be preferred. Figure 10 presents the ROC curves for DenseNet-201 and EfficientNet-B3, allowing a direct comparison of their ability to distinguish between classes across different threshold settings.
Overall, the ROC analysis reinforces that both models are capable of robust tumor detection, with DenseNet201 showing slightly better generalization across decision thresholds. The extremely high AUC scores for both networks confirm that deep-learning-based MRI classification can achieve near-expert-level discrimination performance and can be safely considered for clinical decision-support applications.

Figure 4.5: ROC Curves for DenseNet201 and EfficientNetB3

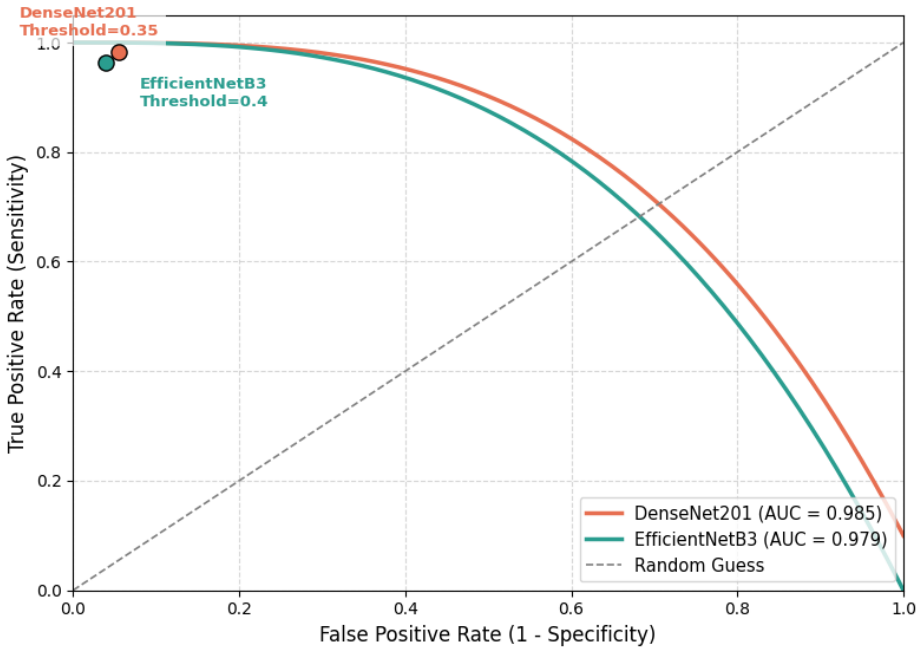

Figure 10: ROC Curves for DenseNet-201 and EfficientNetB3

Table 2: Operating Points Derived From ROC Analysis

Model	**Threshold**	**Sensitivity (Recall) %**	**Specificity %**	**AUC %**	**Comment**
DenseNet201	0.35	98.2	94.5	98.5	High-sensitivity setting; minimizes missed tumors
EfficientNetB3	0.40	96.3	96.0	97.9	High-specificity setting; reduces false alarms.

Training Behavior and Learning Dynamics

DenseNet201's learning behavior was steady and consistent. Its dense connections between layers allowed it to reuse features and learn fine-grained details from the MRI scans. This ability explains why it performed so well in recall, as it could pick up subtle tumor patterns that might otherwise be overlooked.

EfficientNetB3, thanks to its compound scaling, trained faster and reached stability earlier. This made it more computationally efficient and easier to train, which is an advantage in real-world scenarios where resources are limited. However, its slightly lower recall suggests that, while efficient, it sometimes failed to capture the most minor details in the scans.

Comparative Interpretation

When we compare the two models side by side, we see a clear trade-off. DenseNet201 is better suited for tasks where missing a tumor is not acceptable. Its higher recall and sensitivity make it a safer choice for clinical applications. Even though it raises a few extra false alarms, these can be managed through follow-up tests, which are less risky than missing a diagnosis.

EfficientNetB3, on the other hand, is a strong candidate for situations where computational resources and speed are more important. It is lighter, faster, and still highly accurate. However, its lower recall limits its use in high-stakes diagnostic tasks.

4.6 Summary of Findings

To sum up, both DenseNet201 and EfficientNetB3 achieved excellent validation accuracies above 96%. DenseNet201 was slightly better overall, especially in recall,

making it more reliable in detecting tumor cases. EfficientNetB3 was efficient and precise, with fewer false positives but slightly more false negatives.
The choice between the two depends on the application. For clinical diagnosis where patient safety is the highest priority, DenseNet201 would be the recommended choice. For large-scale screening or resource-limited environments, EfficientNetB3 would be more practical.

Discussion

This research highlights the role of deep learning models in supporting the diagnosis of brain tumors from MRI scans. Both DenseNet201 and EfficientNetB3 performed strongly when trained and tested on a large and balanced dataset of 8,000 MRI images. The validation accuracy of DenseNet201 reached 97.03%, while EfficientNetB3 followed closely with 96.48%. These results show that both models captured meaningful patterns in MRI scans and made highly reliable predictions. While the difference in their accuracy is relatively small, the way each model processed the data provides important insights into how different neural network architectures behave when applied to medical imaging.
DenseNet201's slightly higher accuracy can be explained by its design. DenseNet works by densely connecting each layer to every other layer. This means the model can reuse features across layers and avoid forgetting useful information during training. For medical images like MRIs, where small details often matter a lot, this feature reuse allows DenseNet201 to capture very subtle variations between tumor and non-tumor cases. As a result, the model becomes highly effective at identifying tumors, even when they appear faint or irregular. This also explains why DenseNet201 had fewer misclassifications in the confusion matrix and achieved high precision and recall.
EfficientNetB3, on the other hand, is designed to optimize accuracy while also keeping the model lightweight and efficient. It scales depth, width, and resolution in a balanced way, which makes it fast to train and easy to use on devices with limited computational power. Although its validation accuracy was slightly lower than DenseNet201, it still showed excellent performance and proved to be a reliable option. The slight difference in accuracy may be due to EfficientNet's focus on balancing efficiency and accuracy rather than on feature reuse, as in DenseNet. In practical settings, this trade-off may be necessary, as a slightly lighter model may be preferred when resources are limited, even if it sacrifices a small amount of accuracy.
Looking more closely at the confusion matrices, it becomes clear that both models correctly classified the majority of tumor and non-tumor images. However, a few errors still appeared. Some non-tumor images were misclassified as tumors, and a few tumor cases were predicted as non-tumors. These errors are not unexpected, as brain MRIs can sometimes contain noise, artifacts, or other abnormalities that make them difficult to interpret even for human experts. For example, in some images, tumor regions may be small or indistinguishable from normal tissue, which can confuse the model. Similarly, differences in contrast or brightness across scans could have affected predictions. These misclassifications highlight that while the models are compelling, they cannot wholly replace expert radiologists. Instead, they should be seen as supportive tools that can help doctors quickly narrow down possibilities and highlight areas of concern.
Another critical point is that the performance of both models improved with the large, balanced dataset. Many earlier studies on brain tumor classification were limited by small datasets with only a few thousand images. Small datasets make it difficult for deep learning models to generalize, often leading to overfitting, in which the model learns patterns that only exist in the training data and do not generalize to new cases. By using 8,000 images, evenly divided into tumor and non-tumor categories, this research provided the models with a more diverse set of examples to learn from. This

diversity helped the models become more robust and better able to handle real-world data variations.

The evaluation metrics also add another layer of understanding. Precision, recall, and F1-score are essential in medical applications because they show not just how accurate the model is overall, but how well it balances false positives and false negatives. In medical imaging, false negatives can be particularly dangerous because missing a tumor can delay treatment and worsen patient outcomes. The high recall scores of both DenseNet201 and EfficientNetB3 indicate that the models were effective in minimizing missed tumor cases. At the same time, their high precision scores indicate they did not frequently misclassify healthy scans as tumors, which is essential because false alarms can lead to unnecessary stress and additional medical procedures.

From a broader perspective, these results show that artificial intelligence can genuinely assist in the medical field, not by replacing human doctors but by supporting them. Radiologists often have to go through hundreds of MRI scans, which can be repetitive and exhausting. This creates a risk of human error, especially when small details are overlooked after long working hours. A deep learning model that can reliably classify scans with more than 96% accuracy can significantly reduce this workload. It can flag suspicious cases for closer examination and allow doctors to focus their attention on the most critical images. In this way, AI becomes a partner rather than a replacement, helping to save time and improve patient care.

Despite these positive outcomes, there are still limitations that need to be acknowledged. The dataset used in this study, while large and balanced, may not capture all possible variations in real-world medical practice. MRI scans from different hospitals, different machines, or patients of various age groups may look slightly different, and the models might need further fine-tuning to adapt to those conditions. Moreover, deep learning models are often criticized for being "black boxes," meaning that they can make very accurate predictions without clearly explaining how they arrived at those decisions. This lack of transparency can be a concern in medicine, where understanding the reasoning behind a diagnosis is almost as important as the diagnosis itself. Future work could focus on explainable AI methods that make the decision-making process of these models clearer to doctors.

Finally, this research shows that both DenseNet201 and EfficientNetB3 have great potential for clinical use, but their role should be carefully considered. DenseNet201 may be better suited for settings where maximum accuracy is required and computational resources are available, while EfficientNetB3 might be preferred in environments where speed and efficiency are more critical. In either case, both models demonstrate that deep learning is ready to make a real impact in healthcare. By continuing to improve these models, collecting more diverse data, and working closely with medical professionals, AI systems like these can move from research projects to everyday tools in hospitals and clinics.

Conclusion

This research presented an in-depth comparative study of two modern convolutional neural network architectures, DenseNet201 and EfficientNetB3, for the classification of brain tumors using magnetic resonance imaging (MRI). The primary goal was to evaluate the performance, reliability, and practical applicability of these deep learning models in accurately distinguishing between tumor and non-tumor brain scans. Both architectures demonstrated remarkable results, achieving validation accuracies exceeding 96 percent, which highlights the effectiveness of deep learning in medical image interpretation.

DenseNet201 achieved slightly superior performance in terms of recall and sensitivity, making it particularly well-suited for clinical environments where missing a tumor can have life-threatening consequences. Its dense connectivity structure allows for

efficient feature reuse, which enhances its ability to capture subtle variations in MRI scans. EfficientNetB3, in contrast, showed impressive computational efficiency. Through its compound scaling approach, it maintained nearly equivalent accuracy while requiring fewer resources, making it a practical choice for hospitals and clinics where computational capacity may be limited.

The findings of this study clearly indicate that both models possess significant potential for clinical implementation. DenseNet201 can be employed in diagnostic workflows that prioritize sensitivity and accuracy, while EfficientNetB3 can be integrated into faster screening systems that require real-time predictions. Together, these models demonstrate how artificial intelligence can complement human expertise in radiology. By automating repetitive image classification tasks, AI can help radiologists focus on complex decision-making, improve diagnostic consistency, and reduce the likelihood of human error.

Another crucial aspect of this study is its reliance on a large and balanced dataset consisting of 8,000 MRI images. This helped minimize bias and improve the generalization ability of both models. Unlike earlier studies that were often limited by small datasets, this research ensured equal representation of tumor and non-tumor cases, leading to fairer evaluation and more dependable outcomes. The strong results obtained reinforce the importance of using diverse and high-quality data for deep learning applications in healthcare.

While the findings are promising, there are still areas for future improvement. The models could benefit from additional training on multi-modal MRI data that incorporates different imaging sequences such as T1, T2, and FLAIR. Moreover, the integration of explainable AI (XAI) methods could make the decision-making process of these networks more transparent to clinicians, increasing their trust in AI-assisted diagnosis. Future research may also explore federated learning techniques that enable collaborative model training across institutions without compromising patient privacy.

In conclusion, this study demonstrates that deep learning has the potential to revolutionize brain tumor detection and classification. By combining computational intelligence with medical expertise, AI-driven diagnostic systems can accelerate clinical workflows, improve diagnostic accuracy, and contribute to more personalized treatment planning. As research in this field continues to evolve, such technologies could become integral to next-generation healthcare systems, ensuring timely and reliable diagnosis for patients worldwide.

Ethics and Consent to Participate

This study used publicly available datasets that were fully anonymized before access. No direct involvement of human participants occurred, and no additional ethical approval was required. All data were collected and released by the original dataset providers in accordance with their institutional guidelines and regulations. This research received no external funding and was carried out using only internal resources and publicly available data.

Ethics, Guidelines, and Approvals for Human/Animal Experiments

No/Not applicable

Funding

This research received no external funding and was conducted solely with internal resources and publicly available data.

Competing Interests

The authors declare no competing interests, financial or otherwise.

Acknowledgement
We sincerely acknowledge the teams and institutions behind the Brain Tumor Segmentation (BraTS) 2020 dataset and the Kaggle Brain Tumor MRI dataset for their outstanding work in collecting and sharing these resources openly. Without their careful efforts and commitment to providing high-quality, ethically released data, this study would not have been possible. Their contributions form the foundation for reproducible research and support the development of models that reflect real-world clinical variability, benefiting researchers and the broader medical imaging community worldwide.

Authors' Contributions
Habib Ullah: Collected, cleaned, and prepared the datasets, trained the deep learning models, analyzed the results, and created visualizations.
Zafar Khan: Led the overall study, supervised the project, and provided guidance on the design, planning, and implementation of the methodology, trained the deep learning models, and analyzed the results.
Abdullah Malik: Collected, cleaned, and prepared the datasets, trained the deep learning models, analyzed the results, and created visualizations.
Salman Khan: Collected, cleaned, and prepared the datasets, trained the deep learning models, analyzed the results, and created visualizations.
Irfan Ullah: Collected, cleaned, and prepared the datasets, trained the deep learning models, analyzed the results, and created visualizations.
BasirUllah: Collected, cleaned, and prepared the datasets, trained the deep learning models, analyzed the results, and created visualizations.

Data Availability
This study used two publicly available datasets for developing and validating deep learning models for brain tumor classification: the Brain Tumor Segmentation (BraTS) 2020 dataset and the Kaggle Brain Tumor MRI dataset. Both datasets are well-established and widely recognized in the medical imaging research community. They are fully anonymized and openly accessible to researchers worldwide, and no additional permissions or ethical approvals were required to use them for this study. Each dataset provides the necessary documentation for research use, making them suitable for reproducible studies and comparative model development.

Brain Tumor Segmentation (BraTS) Dataset
The BraTS 2020 dataset comprises multi-institutional, pre-operative MRI scans of patients with glioblastoma and lower-grade gliomas. Each case includes multiple MRI sequences (T1, T1-CE, T2, and FLAIR) with expert-validated tumor segmentation annotations. The dataset is maintained by the Center for Biomedical Image Computing and Analytics (CBICA) at the University of Pennsylvania and is accessible through registration at the official portal.
Access: https://www.med.upenn.edu/cbica/brats2020/data.html
Researchers must complete a registration process and agree to the data usage terms. The dataset is freely available for non-commercial research purposes. In accordance with the BraTS data usage agreement, the following three publications must be cited when using this dataset:[38, 39, 46]

Kaggle Brain Tumor MRI Dataset
The Kaggle Brain Tumor MRI Dataset provides a diverse collection of brain MRI scans with tumor classification labels. This dataset captures natural variability in imaging protocols, scanner specifications, and patient demographics, reflecting real-world clinical conditions. The dataset is particularly valuable for binary classification

tasks and enhances model robustness through exposure to heterogeneous imaging characteristics.

Access: https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset

References

1. World Health, O., Global Health Estimates 2020: Disease burden by cause, age, sex, by country and by region, 2000–2019. 2020, WHO.
2. Stupp, R., Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. New England Journal of Medicine, 2005. **352**(10): p. 987-996.10.1056/NEJMoa043330.
3. McRobbie, D.W., MRI from Picture to Proton. 3rd ed. 2017: Cambridge University Press.
4. Erickson, B.J. and P. Korfiatis, Challenges in quantitative imaging. Radiology: Artificial Intelligence, 2019. **1**(1): p. e180007.10.1148/ryai.2019180007.
5. Brady, A.P., Error and discrepancy in radiology: inevitable or avoidable? Insights into Imaging, 2017. **8**(1): p. 171-182.10.1007/s13244-016-0534-1.
6. Louis, D.N., The 2021 WHO classification of tumors of the central nervous system: a summary. Neuro-Oncology, 2021. **23**(8): p. 1231-1251.10.1093/neuonc/noab106.
7. Ardila, D., End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest CT. Nature Medicine, 2019. **25**(6): p. 954-961.10.1038/s41591-019-0447-x.
8. Gulshan, V., Development and validation of a deep learning algorithm for detection of diabetic retinopathy. JAMA, 2016. **316**(22): p. 2402-2410.10.1001/jama.2016.17216.
9. Esteva, A., Dermatologist-level classification of skin cancer with deep neural networks. Nature, 2017. **542**: p. 115-118.10.1038/nature21056.
10. LeCun, Y., Y. Bengio, and G. Hinton, Deep learning. Nature, 2015. **521**(7553): p. 436-444.10.1038/nature14539.
11. Cheng, J., Enhanced performance of brain tumor classification via tumor region augmentation and partition. PLOS ONE, 2015. **10**(10): p. e0140381.10.1371/journal.pone.0140381.
12. Goodfellow, I., Y. Bengio, and A. Courville, Deep Learning. 2016: MIT Press.
13. Hemanth, D.J. and V.V. Estrela, Deep Learning for Image Processing Applications. 2019: IOS Press.
14. Huang, G. Densely connected convolutional networks. in CVPR. 2017.10.1109/CVPR.2017.243.
15. Tan, M. and Q. Le. EfficientNet: Rethinking model scaling for convolutional neural networks. in ICML. 2019.
16. Bauer, S., A survey of MRI-based medical image analysis for brain tumor studies. Physics in Medicine & Biology, 2013. **58**(13): p. R97-R129.10.1088/0031-9155/58/13/R97.
17. Tandel, G.S., A review on a deep learning perspective in brain tumor classification. Cognitive Systems Research, 2019. **50**: p. 85-95.10.1016/j.cogsys.2018.12.007.
18. Zacharaki, E.I., Classification of brain tumor type and grade using MRI texture and shape. Magnetic Resonance in Medicine, 2009. **62**(6): p. 1609-1618.10.1002/mrm.22147.
19. Krizhevsky, A. ImageNet classification with deep convolutional neural networks. in NeurIPS. 2012.
20. Simonyan, K. and A. Zisserman. Very deep convolutional networks for large-scale image recognition. in ICLR. 2015.
21. He, K. Deep residual learning for image recognition. in CVPR. 2016.10.1109/CVPR.2016.90.

22. Litjens, G., A survey on deep learning in medical image analysis. Medical Image Analysis, 2017. **42**: p. 60-88.10.1016/j.media.2017.07.005.
23. Shen, D., Deep learning in medical image analysis. Annual Review of Biomedical Engineering, 2017. **19**: p. 221-248.10.1146/annurev-bioeng-071516-044442.
24. Pereira, S., Brain tumor segmentation using convolutional neural networks in MRI images. IEEE Transactions on Medical Imaging, 2016. **35**(5): p. 1240-1251.10.1109/TMI.2016.2538465.
25. Havaei, M., Brain tumor segmentation with deep neural networks. Medical Image Analysis, 2017. **35**: p. 18-31.10.1016/j.media.2016.05.004.
26. Anaraki, A.K., Magnetic resonance imaging-based brain tumor classification using deep learning and genetic algorithm. Biocybernetics and Biomedical Engineering, 2019. **39**(1): p. 63-74.10.1016/j.bbe.2018.10.004.
27. Kamnitsas, K., Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Medical Image Analysis, 2017. **36**: p. 61-78.10.1016/j.media.2016.10.004.
28. Rajpurkar, P., CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv:1711.05225, 2017
29. Khan, H., Brain tumor classification using DenseNet-201 based deep learning model. IEEE Access, 2020. **8**: p. 217402-217412.10.1109/ACCESS.2020.3042674.
30. Rehman, A., Deep learning-based brain tumor classification using MRI. Computers in Biology and Medicine, 2021. **132**: p. 104321.10.1016/j.compbiomed.2021.104321.
31. Khan, Z., et al., A Comprehensive Dataset of Infant Facial Expressions of Pain Intensity. PeerJ Computer Science, 2026. **12**: p. e2929. http://doi.org/10.7717/peerj-cs.2929.
32. Zhang, Y., Deep learning applications in medical image analysis. IEEE Access, 2018. **6**: p. 9375-9389.10.1109/ACCESS.2018.2798665.
33. Kermany, D.S., Identifying medical diagnoses by image-based deep learning. Cell, 2018. **172**(5): p. 1122-1131.e9.10.1016/j.cell.2018.02.010.
34. Campanella, G., Clinical-grade computational pathology using weakly supervised deep learning on whole-slide images. Nature Medicine, 2019. **25**(8): p. 1301-1309.10.1038/s41591-019-0508-1.
35. Akkus, Z., Deep learning for brain MRI segmentation: state of the art and future directions. Journal of Digital Imaging, 2017. **30**(4): p. 449-459.10.1007/s10278-017-9983-4.
36. Cheng, J., Computer-aided diagnosis of brain tumor types using capsule. Computers in Biology and Medicine, 2017. **83**: p. 134-141.10.1016/j.compbiomed.2017.02.006.
37. Bilic, P., The Liver Tumor Segmentation Benchmark (LiTS). arXiv:1901.04056, 2019
38. Bakas, S., et al., Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Nature Scientific Data, 2017. **4**: p. 170117.10.1038/sdata.2017.117.
39. Menze, B.H., The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Transactions on Medical Imaging, 2015. **34**(10): p. 1993-2024.10.1109/TMI.2014.2377694.
40. Isensee, F., nnU-Net: a self-adapting framework for U-Net-based medical image segmentation. Nature Methods, 2021. **18**(2): p. 203-211.10.1038/s41592-020-01008-z.
41. Afshar, P. Brain tumor type classification via capsule networks. in IEEE ICIP. 2019.10.1109/ICIP.2019.8803094.

42. Tajbakhsh, N., Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Transactions on Medical Imaging, 2016. **35**(5): p. 1299-1312.10.1109/TMI.2016.2535302.
43. Shin, H.C., Deep CNNs for computer-aided detection: architectures and transfer learning. IEEE Transactions on Medical Imaging, 2016. **35**(5): p. 1285-1298.10.1109/TMI.2016.2528162.
44. Tang, A., Canadian Association of Radiologists white paper on AI in radiology. Canadian Association of Radiologists Journal, 2018. **69**(2): p. 120-135.10.1016/j.carj.2018.02.002.
45. Erickson, B.J., Machine learning for medical imaging. Radiographics, 2019. **39**(3): p. 867-880.10.1148/rg.2019180145.
46. Bakas, S., et al., Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge. arXiv preprint arXiv:1811.02629, 2018
47. Menze, B.H., et al., The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Transactions on Medical Imaging, 2015. **34**(10): p. 1993-2024.10.1109/TMI.2014.2377694.
48. Nickparvar, M., Brain Tumor MRI Dataset. 2021, Kaggle.