ITM Web of Conferences **70**, 03014 (2025) https://doi.org/10.1051/itmconf/20257003014
*DAI 2024*

# Enhancing Brain Tumor Detection: A Comparative Study of CNN Architectures Using MRI Data

*Zhimeng* Zhu

School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, Hubei, 430074, China

**Abstract.** Deep learning models have become essential for automated medical image analysis in brain tumor detection. Existing Convolutional Neural Network (CNN) models like Visual Geometry Group 19 (VGG19), Residual Network 18 (ResNet18), and Residual Network 34 (ResNet34), despite their success in image classification and recognition, face challenges such as unclear boundary detection, limited generalization, and lower computational efficiency in detecting brain tumors. To address these issues, this study introduces VGG19-Brain's MRI for Tumor (BMT), an enhanced version of the classic VGG19 model. VGG19-BMT incorporates targeted optimizations, including adjustments to convolutional layers and improved feature extraction modules. A systematic comparative analysis of VGG19-BMT and traditional CNN models was conducted using the Kaggle dataset "Brain MRI Images for Brain Tumor Detection." The results demonstrate that VGG19-BMT outperforms conventional models' boundary recognition accuracy, generalization, and computational efficiency, providing a more effective solution for automated brain tumor detection. This advancement not only enhances diagnostic capabilities but also sets the stage for future model improvements and clinical applications in medical imaging.

## 1 Introduction

Brain tumor is initiated by the cells' abnormal growth that multiply uncontrollably. These tumors can develop from brain cells, meninges, nerves, or glands [1]. Brain tumors are among the most dangerous diseases, necessitating precise and early detection. Currently, most detection relies on neurospecialists and radiologists for image evaluation, and this way may be prone to errors made by humans and is often time-costing [2]. Detecting a brain tumor can be the difference in a person's life or death [3]. In the study of brain tumor detection, diagnosing patients is currently a labor-intensive process, relying on a limited number of experts who are often overburdened [4]. To make this process automatizing is difficult because of the significant variability in the tumor tissues' appearance across kinds of patients and in many situations their similarity to normal tissues [5]. Many models of deep learning, particularly CNNs, are becoming essential tools for automated medical image analysis.

Corresponding author: U202215216@hust.edu.cn

Automated detection methods have greatly enhanced the understanding of both normal and diseased conditions in medical research. They play a crucial role in treatment planning and diagnosis, especially as patients' number continues to rise [6]. However, despite their outstanding performance in image classification and recognition tasks, existing Convolutional Neural Network (CNN) models such as VGG19, ResNet18, and ResNet34 still face challenges when applied to brain tumor detection. These challenges include unclear boundary recognition, insufficient generalization ability, and relatively low computational efficiency. To solve these problems, this study proposes a better model—VGG19-Brain's MRI for Tumor (BMT), which builds upon the classic VGG19 architecture with targeted optimizations, including adjustments to the convolutional layers and improvements in the feature extraction modules. It conducted a systematic comparative analysis of VGG19-BMT and other traditional CNN models using the dataset from Kaggle: "Brain MRI Images for Brain Tumor Detection". The experimental results demonstrate that VGG19-BMT surpasses conventional models in both boundary recognition accuracy and generalization capability, while also improving computational efficiency without compromising detection accuracy. The findings of this study provide a more effective solution for automated brain tumor detection and lay the groundwork for further model improvements and real-world clinical applications.

# 2 Background

## 2.1 Medical image analysis: deep learning

Over the past decade, machine learning (ML) and computer vision (CV) have revolutionized various aspects of the world. A subset of machine learning called Deep Learning, has delivered exceptional results across multiple fields, particularly in the biomedical domain, because of its capability to analyze and process big volumes of data [7]. Tumor detection is a dire application zone in medical image analysis, involving the growth and application of various methods and technologies. Traditional tumor detection methods primarily rely on the manual interpretation of medical images, including MRI, CT, and X-rays by imaging experts, a process that is time-consuming and prone to human subjectivity. Deep learning algorithms, however, have demonstrated significant advantages in this field. For example, by analyzing medical imaging data, deep learning algorithms can accurately detect the location, size, and shape of tumors, providing crucial references for doctors in formulating treatment plans. Moreover, deep learning can also be used for organ segmentation, automatically identifying and precisely segmenting different organs in images, thereby facilitating surgical navigation and treatment planning.

Advanced deep learning models, particularly CNN-based models, are widely used for tumor detection. CNN excels at extracting features from images, effectively capturing tumor morphology, edges, and texture, leading to more accurate detection. Pre-trained models like VGG16, ResNet, and EfficientNet have demonstrated superior performance in brain tumor detection, surpassing traditional methods in both accuracy and robustness when dealing with complex tumor morphologies and boundaries. However, challenges remain in clinical applications, such as model interpretability, data bias, and variations in images caused by different equipment and environments.

## 2.2 Classic CNN models

CNNs are widely used in CV tasks and are composed of several essential components: convolutional layers, activation functions, pooling mechanisms, and fully connected layers.

https://doi.org/10.1051/itmconf/20257003014

The convolutional layer is a core element, employing learnable filters (kernels) that move across the image to identify features like textures, edges, and patterns. Pooling layers are responsible for down-sampling feature maps, emphasizing crucial features while minimizing the risk of overfitting. Non-linear activation functions allow the model to recognize complex patterns, and fully connected layers are used for making the final predictions. Figure 1 depicts the structure of a CNN. With advancements like model stratification, it has become possible to train more complex models within the CNN framework. CNNs ares extensively used in tasks, for example, image classification, semantic segmentation, and object recognition.

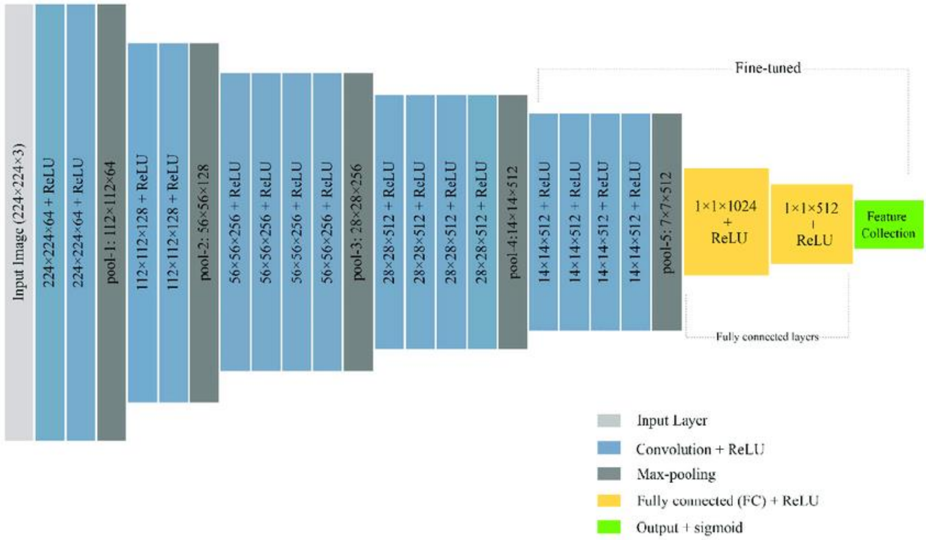


**Fig. 1** Neural layer structure of a CNN network (Photo/Picture credit: Original).

### 2.3 Advanced deep learning models

As a core technology in the field of deep learning, CNNs have evolved significantly, leading to the development of many classic models that have achieved remarkable results in tasks, for example, object detection, semantic segmentation, and image classification. In 1998, Yann LeCun and colleagues proposed the LeNet-5 model, the first widely used CNN model for handwritten digit recognition. Its structure contains multiple convolutional layers, pooling layers, and fully connected layers, laying the foundation for modern CNN architecture. In 2012, Alex Krizhevsky and his team developed AlexNet, a model that achieved groundbreaking success in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). AlexNet innovated CNN performance in image classification tasks by incorporating ReLU activation functions, dropout regularization, and GPU-accelerated training. Building on this success, Oxford University has a group called the Visual Geometry Group introduced the VGG16 and VGG19 models, which utilized very deep network architectures (16 or 19 layers) to further improve image classification accuracy. These models employed numerous small convolutional kernels (3x3), enabling robust feature representation while maintaining spatial information.

In 2014, Google proposed the Inception model (also known as GoogLeNet), which introduced the Inception module. This module improves network efficiency and accuracy by using multiple convolutional and pooling operations in parallel at each layer. In 2015, Kaiming He and colleagues introduced the Residual Network (ResNet), which used residual connections to solve the vanishing gradient problem in deep networks. ResNet activated the construction of very deepper networks, such as ResNet-50 and ResNet-101, which performed exceptionally well in ImageNet and other challenges. In 2019, Google launched the EfficientNet series of models, which balanced depth, width, and resolution through a compound scaling method, achieving higher accuracy and efficiency. The EfficientNet-B0 to B7 versions have performed excellently across various computer vision tasks. Pre-trained Model

### *2.3.1 EfficientNet-B2*

The key innovation of EfficientNet lies in its method of scaling width and resolution to optimize both accuracy and efficiency. This scaling approach enables EfficientNet to achieve higher accuracy compared to previous state-of-the-art models, all while using less parameters and requiring less computational power [8]. EfficientNet models range from versions B0 to B7, with each version growing in size and complexity through a technique called compound scaling. This approach is characterized by scaling the model's dimensions—depth, width, and resolution—according to specific formulas, allowing for a balanced increase in capacity and performance across the versions.

$$Depth: d = \alpha^{\Phi} \tag{1}$$

$$Width: w = \beta^{\Phi} \tag{2}$$

$$Resolution: r = \gamma^{\Phi} \tag{3}$$

Where: $\alpha, \beta$ correspond to depth, width, and $\gamma$ means resolution.

$\Phi$ determines the overall scaling factor, a user-defined parameter.

EfficientNet-B2, along with other versions in the EfficientNet family, leverages compound scaling, which optimally balances these dimensions and significantly improves performance compared to scaling methods that focus solely on width, depth, or resolution.

### *2.3.2 ResNet-18*

The ResNet is a deep network architecture widely used, and it is known for its ability to achieve high-accuracy results in various image processing tasks [9]. ResNet-18 was introduced by Kaiming He and colleagues in 2016. It is a CNN architecture, which is recognized for its use of residual blocks. These blocks are a crucial innovation that tackles the vanishing/exploding gradient issues, allowing the CNN to be deeper without experiencing performance degradation. ResNet-18's basic residual block is represented by the following:

$$y = F(x, \{W_i\}) + x \tag{4}$$

Where: $x$: the residual block's input.

$F(x, \{W_i\})$: operations' sequence within the block, including convolutional layers, ReLU activations, and batch normalization.

$y$: the residual block's resulting output.

### *2.3.3 VGG models: VGG16 and VGG19*

The VGG family of CNN, introduced by the Visual Geometry Group (VGG) in 2014, includes two notable architectures: VGG16 and VGG19. Figure 2 shows the differences between the two models. Both models are renowned for their deep convolutional layers and have demonstrated exceptional performance on the ImageNet dataset.

VGG16: This architecture's feature has 3 fully connected layers and 13 convolutional layers. It was a significant advancement in deep learning at the time of its release, achieving high accuracy on ImageNet. However, VGG16 is characterized by a large number of parameters—approximately 138 million—which can contribute to extended training times and cause the number of computational requirements to increase.

VGG19: VGG-19 is a pre-trained model, and it has been used in many tasks, such as classifying different grades of gliomas and distinguishing between high-grade and low-grade gliomas [10]. Building upon the VGG16 architecture, VGG19 adds 4 additional convolutional layers, making a total of 3 fully connected layers and 16 convolutional layers. This deeper network enhances the model's ability to learn and represent more complex features. Like VGG16, VGG19 also has a substantial number of parameters, which similarly impact training time.

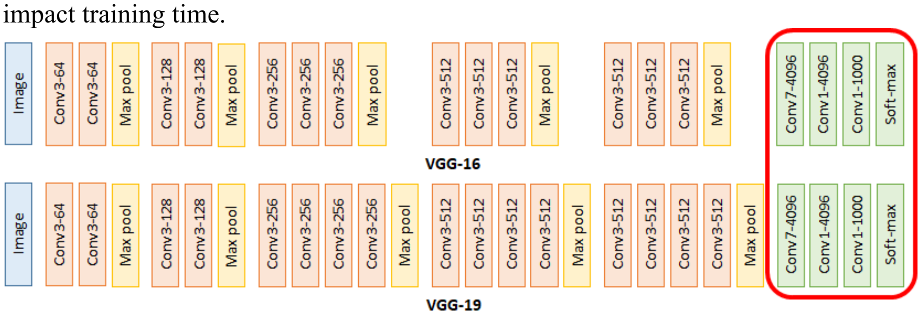


**Fig. 2** The differences between the two models (Photo/Picture credit: Original).

# 3 Methodology

## 3.1 Dataset

This study utilizes the dataset, publicly available on Kaggle: "Brain MRI Images for Brain Tumor Detection". The dataset comprises a series of labelled brain MRI images, categorized into 2 classes: tumor, and non-tumor. This dataset is widely used for the research and development of automated brain tumor detection algorithms due to its variety of tumor types and extensive sample size, making it a dream resource to train deep learning models. Each sample is a high-resolution MRI scan of the brain, providing rich detail that helps models learn and identify complex pathological features.

## 3.2 Image augmentation and preprocessing

To reduce overfitting, particularly with smaller training datasets, image augmentation and preprocessing techniques are employed. Images are randomly flipped horizontally and vertically with a 50% probability, then rotated by ±10 degrees. After these transformations, images are resized to 224 by 224 pixels to maintain consistency during CNN batch processing. Since the dataset images are already square, no cropping is needed. The resized images are then converted into tensors with pixel values scaled to [0, 1], followed by normalization to prepare them for effective model training.

## 3.3 LearningRateandEpoch

Training began with a learning rate of 0.001, under the application of a learning rate scheduler. To ensure proper convergence, this scheduler reduces the learning rate by a factor of 0.1 every 8 epochs. The number of epochs is decided by the validation dataset's accuracy; if there is no improvement in validation accuracy for 8 consecutive epochs, training is halted. This approach helps prevent overfitting in the model.

## 3.4 Proposed model: VGG19-BMT

This paper implemented a series of deep learning models to compare their performance in the brain tumor detection task. Among these models, the proposed VGG19-BMT represents a significant advancement, leveraging the foundation of VGG19 with several key modifications tailored for improved performance in medical image analysis.

VGG19-BMT utilizes the VGG19 architecture as its base, known for its deep convolutional layers and robust feature extraction capabilities. The original VGG19 model is enhanced with additional layers and modifications to better suit the brain tumor classification task.

### *3.4.1 Advanced image augmentation*

The VGG19-BMT model employs a comprehensive image augmentation strategy through a code block named `augmentation_generator`. This strategy includes rescaling pixel values to the range of 0 to 1, applying horizontal and vertical shifts of 10%, performing shearing and zooming by up to 10%, rotating images by up to 30 degrees, horizontally flipping images, and adjusting brightness within a range of 0.5 to 1.0. The augmented images are then prepared using `flow_from_directory` with a batch size of 64, facilitating effective training on a dataset of 253 images across two classes.

### *3.4.2 Adjustment of dynamic learning rate*

Learning Rate Scheduler: Based on the training accuracy (acc) a ReduceLROnPlateau callback is integrated to dynamically adjust the learning rate. If the accuracy plateaus for 2 epochs, the learning rate will be reduced by a factor of 0.3, with the smallest learning rate of 0.0000001. This approach helps in achieving stable convergence and improved model performance.

### *3.4.3 Customized classifier design*

The VGG19-BMT model features a custom classifier designed to optimize performance for brain tumor detection. It includes a fully connected layer with a 20% dropout rate to mitigate overfitting, ensuring the model generalizes well to new data. Afterward, a fully connected linear layer with 1,000 output features is applied, providing a rich representation of the learned features. Another dropout layer with a 20% probability is added before the final linear layer, further preventing overfitting and enhancing model robustness. The final linear layer, with four output classes, is tailored specifically for the classification task, enabling the model to effectively learn and distinguish complex patterns associated with brain tumors.

## 3.5 Performances measure

To assess the performance of different models, it employed several key metrics, including accuracy, precision, recall, average loss and F1-score. For evaluating classification tasks, these metrics are vital.

$$\text{Accuracy} = \frac{\boldsymbol{TN + TP}}{\boldsymbol{TP + FP + TN + FN}} \tag{5}$$

$$Precision = \frac{\boldsymbol{TP}}{\boldsymbol{TP + FP}} \tag{6}$$

$$Recall = \frac{\boldsymbol{TP}}{\boldsymbol{FN + TP}} \tag{7}$$

$$F1\ Score = \frac{2 \cdot \boldsymbol{Precision} \cdot \boldsymbol{Recall}}{\boldsymbol{Recall + Precision}} \tag{8}$$

The corresponding equations, the calculations for these metrics are based on the following values:

TP: True Positive
FP: False Positive
FN: False Negative
TN: True Negative
To calculating the average loss, cross-entropy was utilized as the loss function.

$$L = -\sum_{i=1}^{N} y_i \log(\hat{y}_i) \tag{9}$$

N: represents the number of classes
$\hat{y}_i$: predicted probability for class $i$
$y$: true label.
The criterion selected for all tests in this paper was cross-entropy, as it is extensively used in classification tasks. To measure differences between actual outcomes and predicted probabilities accurately, it serves as a reliable metric for evaluating model performance.

# 4 Evaluation and discussion of results

## 4.1 Performance

**Table 1.** Comparison of models on the classification of brain tumor

| Name Of Model | Accuracy | Recall | Precision | F1-Score |
|---|---|---|---|---|
| VGG19 | 0.9458 | 0.9435 | 0.9503 | 0.9445 |
| VGG16 | 0.9001 | 0.9033 | 0.9332 | 0.9187 |
| EfficientNet-B2 | 0.8900 | 0.7800 | 1.0000 | 0.8764 |
| ResNet-18 | 0.8600 | 0.9677 | 0.8333 | 0.8955 |
| **VGG19-BMT** | **0.9951** | **0.9634** | **0.9563** | **0.9498** |

## 4.2 Pre-trained models' performance

Through Table 1, five models were evaluated on the dataset: CNN, VGG19, VGG16, EfficientNet-B2, and ResNet-18. The models achieved the following performance metrics: accuracies of 0.8600, 0.9458, 0.9001, 0.8900, and 0.8600; precision scores of 0.8333, 0.9503, 0.9332, 1.0000, and 0.8333; recall scores of 0.9677, 0.9435, 0.9033, 0.7800, and 0.9677; and F1-scores of 0.8955, 0.9445, 0.9187, 0.8764, and 0.8955. Among these, VGG16 has the largest number of parameters, approximately 138 million. Using transfer learning, all models were implemented, with only the last layer being fine-tuned. While the accuracy and F1 scores are promising, further optimization and fine-tuning may enhance the results.

### 4.3 Visual analytics

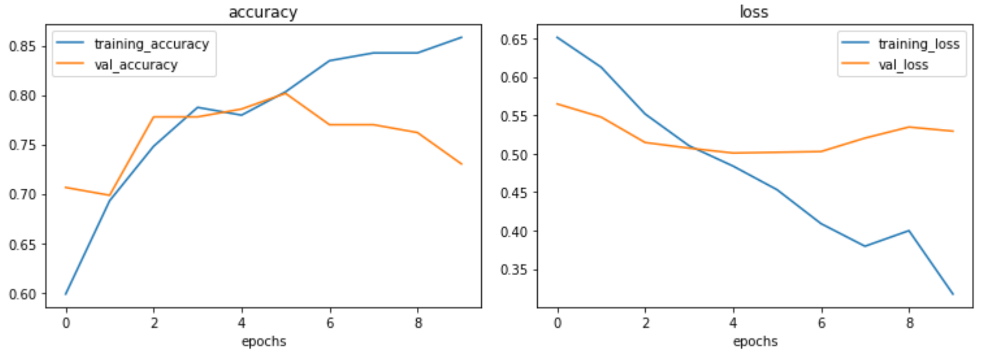


**Fig. 3** Resnet-18 Accuracy and Loss (Photo/Picture credit: Original).

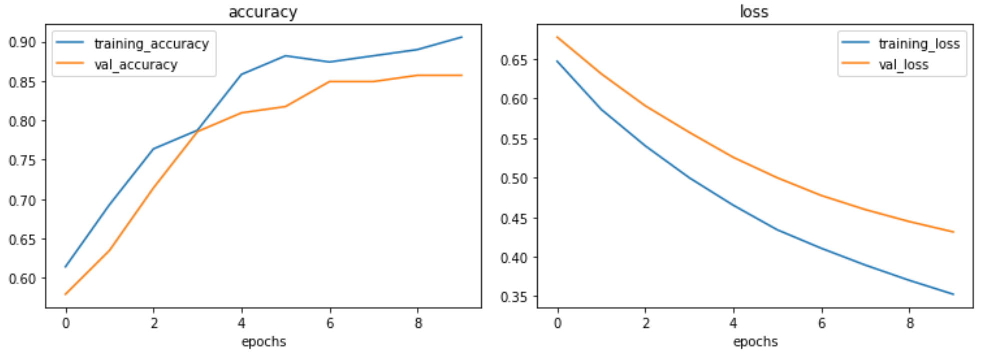


**Fig. 4** Efficient-B2 Accuracy and Loss (Photo/Picture credit: Original).

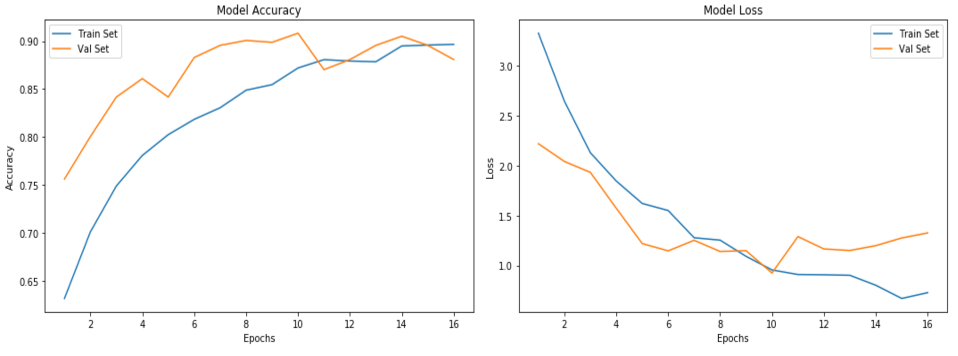


**Fig. 5** VGG16 Accuracy and Loss (Photo/Picture credit: Original).

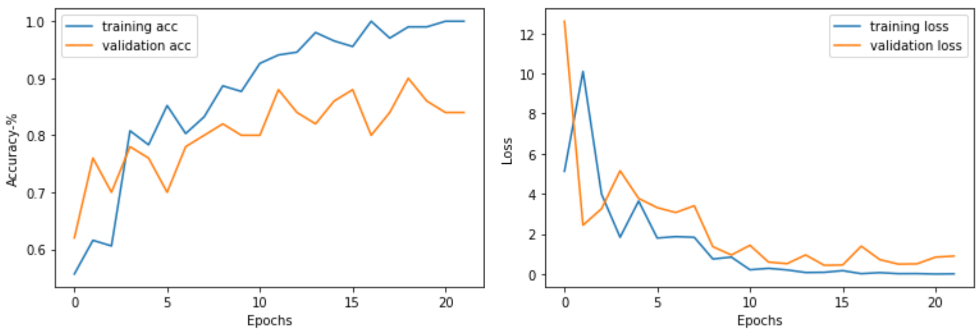


**Fig 6.** VGG19 accuracy and loss (Photo/Picture credit: Original).

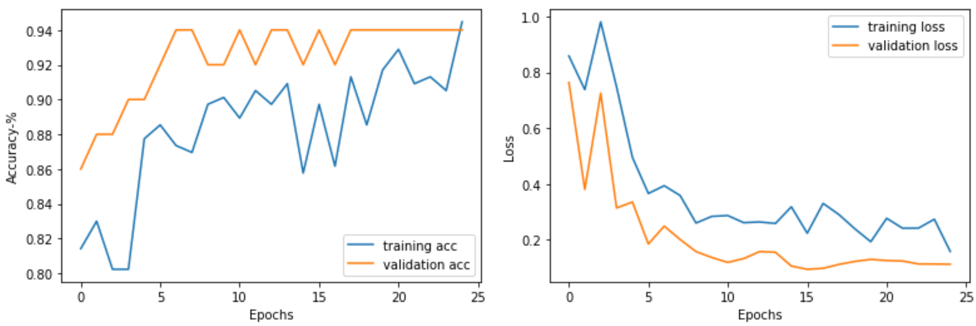


**Fig. 7** VGG19-BMT accuracy and loss (Photo/Picture credit: Original).

When comparing the performance of various models for brain tumor classification, multiple metrics are taken into account. For the pre-trained models, the accuracy rates are as follows: VGG19 achieves 0.9458, VGG16 reaches 0.9001, EfficientNet-B0 stands at 0.8900, and ResNet-18 is at 0.8600. The corresponding F1 scores for these models are 0.9445, 0.9187, 0.8764, and 0.8955. However, the custom VGG19-BMT model outperforms all of these pre-trained models, attaining an F1 score of 0.9498 and an accuracy of 0.9951. And the substantial improvement in medical image classification performance by the VGG19-BMT model can be attributed to advanced image augmentation techniques and dynamic learning rate adjustments. Detailed results are provided in Table 1.

While pre-trained models are effective in medical image classification, the disparities in dataset size and image type between medical and general images demand modifications to the models. Performance in this domain can be greatly improved by customizing the architecture of pre-trained models and fine-tuning their layers. For example, VGG19-BMT achieved an accuracy of 99.51%, whereas the standard VGG19 model attained 94.45%. This underscores the necessity of further customizing and fine-tuning existing models. Such efforts are crucial for achieving notable improvements in medical image classification performance.

Figures 3-7 display the loss and accuracy across epochs. VGG19-BMT achieved its optimal performance at the 23rd epoch. At this point, there was no substantial discrepancy in accuracy between the validation and training sets finally, indicating that this model is well-optimized and effectively trained, with no significant issues of overfitting or underfitting. The loss value decreased rapidly, with a more notable reduction in validation loss compared to the training set, indicating the model's superior performance. Additionally, the F1 score was higher than that of other models, underscoring its effectiveness in capturing complex patterns and accurately differentiating between tumor and non-tumor cases.

## 5 Conclusion

In this study, VGG19-BMT demonstrates its superior performance over several well-established pre-trained models in brain tumor detection. The VGG19-BMT model, with its customized architecture, advanced image augmentation strategies, and dynamic learning rate adjustment, achieved remarkable improvements in accuracy, F1 score, and boundary recognition. These improvements emphasize the critical role of model customization and fine-tuning. This is especially true in the medical image classification field, where the subtle details of image data require more specialized and tailored approaches.

VGG19-BMT is well-suited for clinical applications, offering high accuracy and robustness across various MRI scans. Additionally, the architecture and methodologies used in VGG19-BMT could be applied to various other medical imaging tasks, including the detection of lung nodules, screening for breast cancer, or the analysis of retinal images. Future work will explore these possibilities, aiming to refine and extend the application of VGG19-BMT across different medical domains.

However, despite these improvements, the VGG19-BMT model also has certain limitations. The model's increased complexity and the additional layers introduced for feature extraction may cause longer training times and make computational demands higher. Additionally, while the model excels in detecting complex tumor boundaries, it might still struggle with certain edge cases, such as highly irregular tumor shapes or very small tumors that are hard to identify from normal tissue

## References

1. J. Amin, M. Sharif, M. Yasmin, S. L. Fernandes. A distinctive approach in brain tumor detection and classification using MRI. Pattern Recognition Lett. 139, 118-127 (2020).
2. H. E. M. Abdalla, M. Y. Esmail. Brain tumor detection by using artificial neural network. In: 2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), Khartoum, Sudan, 2018, pp. 1-6.
3. W. M. Salama, A. Shokry. A novel framework for brain tumor detection based on convolutional variational generative models. arXiv:2202.09850v1 (2022).
4. J. Martinů, J. Novotný, K. Adámek, P. Čermák, J. Kozel, D. Školoudík. A survey of feature detection methods for localisation of plain sections of axial brain magnetic resonance imaging. Biomed. Signal Process. Control 82, 104611 (2023).
5. S. Roy, S. Nag, I. K. Maitra, S. K. Bandyopadhyay. A review on automated brain tumor detection and segmentation from MRI of brain. arXiv:1312.6150 (2013).
6. V. Pandiyaraju, S. Venkatraman, A. Abeshek, S. A. Aravintakshan, P. S. Kumar, S. Madhan. An integrated deep learning framework for effective brain tumor localization, segmentation, and classification from magnetic resonance images. arXiv preprint arXiv:2409.17273(2024).
7. M. Nazir, S. Shakil, K. Khurshid. Role of deep learning in brain tumor detection and classification (2015 to 2020): A review. Computerized Medical Imaging and Graphics 91, 101940 (2021).
8. R. S. S. Anwar. EfficientNet Algorithm for Classification of Different Types of Cancer. Artificial Intelligence and Applications (2023).
9. L. Zhang, H. Schaeffer. Forward Stability of ResNet and Its Variants. arXiv:1811.09885 (2018).
10. J. Amin, M. Sharif, A. Haldorai, et al. Brain tumor detection and classification using machine learning: a comprehensive survey. Complex Intell. Syst. 8, 3161–3183 (2022).