

Impact Of Data Augmentation And Layer Freezing On CNN-Based Brain Tumor Classification Using MRI Scans

S. M. Zakariya^{1,*} and Mohammad Sarosh Umar²

¹Electrical Engineering Section, University Polytechnic, Aligarh Muslim University, Aligarh, Uttar Pradesh 202001, India

²Department of Computer Engineering, ZHCET, Aligarh Muslim University, Aligarh, Uttar Pradesh 202001, India

*smzakariya.ubp@amu.ac.in

ABSTRACT

Brain tumors pose significant diagnostic challenges and demand efficient, automated solutions. This research investigates the application of Convolutional Neural Networks (CNNs), focusing on the VGG16 architecture and its frozen-layer variant, for brain tumor detection from MRI scans. A total of 7,023 MRI images were utilized, preprocessed using the Multivariate Fast Iterative Filtering (MFIF) technique, and further enhanced through data augmentation methods including rotation and flipping. The models were trained with the Adamax and SGD optimizers, set to learning rates of 0.001 and 0.0001, respectively. The frozen-layer VGG16 model attained a training accuracy of 99.98% and a test accuracy of 98.12%, along with AUC scores of 1.0 and 0.9994. Evaluation metrics such as Accuracy, Precision, Recall, F1-score, and ROC analysis confirmed that both data augmentation and layer freezing contribute significantly to improving classification outcomes, with an overall average accuracy of 98%.

Keywords: Brain Tumor, Convolution Neural Network, VGG16 Model, VGG16 Frozen Layer, Data Augmentation, Performance Evaluation.

1. INTRODUCTION

Detecting and classifying brain tumors is a vital area of research, where numerous machine learning and deep learning methods are being developed to improve diagnostic precision. Accurate and early identification of brain tumors remains a key objective in medical imaging. This initiative plays a pivotal role in combating neurological diseases. Among the available imaging techniques, Magnetic Resonance Imaging (MRI) stands out as one of the most effective and widely used methods. It excels in capturing detailed anatomical information with unmatched clarity [1], [2]. Deep learning techniques can automatically learn complex patterns from large datasets of MRI images, enabling them to detect brain tumors with high accuracy. This opens up an exciting opportunity to make brain tumor diagnosis faster and more accurate [3]. The primary objective of this research is to investigate the application of deep learning methods for the detection and classification of brain tumors. The study concentrates on three common categories: gliomas, pituitary tumors, and normal brain tissue, using MRI scans as the basis for analysis [4].

Neurological disorders such as strokes, brain hemorrhage, multiple sclerosis, and brain tumors pose significant challenges to patients and healthcare professionals alike. The timely and exact diagnosis of these conditions is important for initiating appropriate treatment and improving patient outcomes. Brain tumors pose a serious threat to human health and are among the most challenging medical conditions. Timely and precise diagnosis is crucial for successful treatment and improved patient prognosis. MRI scans are considered the most reliable way to diagnose brain tumors [5].

Although significant progress has been made in imaging techniques and diagnostic methods, accurately identifying and classifying brain tumors—especially gliomas and pituitary tumors—continues to be a complex and demanding task. Traditional methods often depend on radiologists manually interpreting the scans. This process can be inefficient and subjective and lead to differences in opinions between experts [6]. In this regard, the rise of deep learning models offers significant potential to automate and streamline tumor detection and analysis, enabling earlier diagnosis and contributing to better clinical outcomes.

2. RELATED WORKS

Detecting brain tumors early and accurately is essential for better patient outcomes. Acting quickly can significantly boost survival rates and improve quality of life. Traditional diagnostic methods rely heavily on manual inspection of medical imaging data. Traditional methods are often time-intensive and susceptible to human error, prompting growing interest in advanced machine learning solutions. Among these, Convolutional Neural Networks (CNNs) have demonstrated strong potential in enhancing both the precision and accuracy of brain tumor detection [7]. CNNs like VGG16 have shown strong performance in image recognition, including medical imaging. This section highlights studies using VGG16, especially with frozen layers, for detecting brain tumors. For instance, the potential of a CNN-based approach for segmenting tumor diseases from MRI images is demonstrated in [8], achieving remarkable segmentation accuracy. For instance, CNN-based methods have demonstrated remarkable accuracy in segmenting gliomas from MRI scans [9]. Furthermore, deep learning models have been effectively utilized to differentiate between various types of brain tumors, underscoring the versatility of CNNs in medical diagnostics [10]. Moreover, the combination of AlexNet with other CNN architectures, such as VGGNet, ResNet, and InceptionNet, has been investigated to enhance detection performance through the ensemble method [11]. While results are encouraging, building reliable and general models for brain tumor detection is still challenging due to class imbalance, limited labeled data, and variation in medical images, calling for continued research and improvement [12]. Overcoming these challenges demands a holistic strategy that includes data augmentation, sophisticated regularization methods, and the integration of multi-modal imaging data [13]. This research paper focuses on analyzing the use of these models in detail. It highlights their strengths, limitations, and possible future applications. By bringing together insights from existing studies, the paper aims to improve diagnostic accuracy. It also seeks to support the growing use of artificial intelligence in medical practice [14]. Table 1 shows some of the results from the previous study on the multi-classification of brain tumor diseases.

Table 1: Survey of Multiclass Classification Approaches from 2017 Onward

S. No.	Model	Pros	Cons	Dataset	Year	Accuracy	Ref.
1.	CNN	Balanced analysis, critical thinking, structured argument	Time-consuming, depth of analysis, dataset limitation	Kaggle brain MRI, 3,064	2017	91.43%	[15]
2.	Deep CNN	High accuracy, segmented free approach, data augmentation,	Limited dataset, computational resources	T1-weighted, 3064	2019	96.13%	[16]
3.	VGG19	High accuracy using five-fold cross-validation, Innovative approach using transfer learning and block-wise fine tuning, Minimal preprocessing	lacks normal brain Images, dataset limitation, complex fine-tuning	CE-MRI T1-weighted, 3,064	2019	94.82%	[17]
4.	Alex and Google	Innovative approach, high accuracy, detailed methodology, use of transfer leaning,	model Potential overfitting, computational complexity, Lack of real-time application, future work needed, s	BRATS 2013, 2014, 2015, 2016, and ISLES 2018 1,300	2019	98.91%	[18]

5.	ResNet50 + Adam optimizer	High accuracy, comprehensive evaluation, data augmentation	Computational resources, limited dataset	Benchmark MRI, 3,064	2019	99%	[19]
6.	CNN+ SVM	High accuracy, effective combination, robust evaluation	Data limitation, computational complexity, classifier requires more tuning on real images	BRATS 2013, BRATS 2015, and OPEN, 3,064	2020	95.82%	[20]
7.	U-Net with ResNet50	High accuracy, advanced technique, Comprehensive approach.	Computational complexity, Generalizability, limited dataset	Figshare MRI, 3,064	2020	99.6%	[21]
8.	Densenet 201	High accuracy, innovative technique, comprehensive approach	Data limitation, computational cost, feature reduction limitation	BRATS2018 and BRATS2019, 620 cases	2021	95%	[22]
9.	DCNN	Multiscale approach, high accuracy, No preprocessing needed	Limited dataset, model architecture and training process are complex, instances of false positive	T1-CE MRI, 3,064	2021	97.3%	[23]
10.	ResNet50 and Densenet 201	Innovative approach, advanced techniques, moderate accuracy	The dataset is limited in size and requires the use of a single-step feature selection method without combining multiple features, challenging to implement, specific and small dataset	BRATS2019, 335	2021	87.8%	[24]
11.	CNN	Maximum accuracy, automated diagnosis, no segmentation needed	Small dataset, training deep CNN from scratch is complex, optimising the hyperparameters can further improve it.	Figshare, Radiopaedia, REMBRANDT, 3,295	2021	100%	[25]

12.	Deep ResNet	Novel approach, high accuracy, ant colony optimization technique, comprehensive evaluation	Limited scope, quite complex, generalization, It needs optimization and a larger dataset	Name and size of dataset is missing	2023	98.69%	[26]
13.	GAN, Transfer Learning	Innovative approach, high accuracy, data augmentation	Complexity, limited scope, small dataset, It needs optimization and a larger dataset	Open-source MRI, 3,064	2024	99.51%	[27]
14.	CNN	Innovative approach, Harris Hawks Optimization, high accuracy, edge detection	Limited scope, complexity, small dataset. It needs optimization and a larger dataset	Kaggle Brain MRI, 3,064	2023	98%	[28]
15.	Transfer learning, ResNet50	Modified ResNet50, Maximum accuracy, comprehensive evaluation	Complexity, limited scope, insufficient details about datasets	Kaggle Brain MRI, 3,064	2023	100%	[29]
16.	TL-based CNN	High accuracy, five-fold cross validation, efficiency, comprehensive evaluation	Complexity, limited scope, small dataset, optimising the hyperparameters can further improve it.	Figshare, T1-weighted MRI, 3,064	2023	99.02%	[30]
17.	DCNN	High accuracy, comprehensive evaluation, Innovative approach, used two datasets,	Complexity, limited scope, generalization, optimising the hyperparameters can further improve it.	Two public MRI dataset of 872 patients, Exact size of dataset not given	2023	98.90%, 99.29%	[31]
18.	3ACL (Attention-CNN-LSTM)	Novel 3D deep learning, high accuracy, used two datasets, practical application	Hardware requirements, complexity, generalization. The dataset lacks normal brain Images	BRATS 2015 and BRATS 2018, 620 cases	2023	98.90%, 99.29%	[32]
19.	DCNN	High accuracy, innovative approach,	Time-consuming I-HGS	Gold-standard	2023	99.89%, 99.72%,	[33]

		comprehensive evaluation, three datasets used	algorithm, complexity, generalization, size of all datasets are missing	MRI, exact size not specified		99.88%	
20.	Deep transfer learning	High accuracy, innovative approach, use five-fold cross validation, practical application	Limited dataset and system efficiency, complexity, limited scope	Kaggle Brain MRI, 7,023	2024	99.75%	[34]
21.	DCNN	High accuracy, minimal preprocessing, hybrid approach	Limited dataset, complexity, generalization	T1-weighted MRI, 3,064	2024	99.75%, 97.98%, 94.75%	[35]
22.	Optimized ML algorithms	High accuracy, innovative approach, optimization technique, time consuming	complexity, generalization, computational complexity	Figshare, SARTAJ, and Br35H, 7,023	2024	97.15%	[36]
23.	Multipath CNN, SVM classifier	High accuracy, improved metrics, reduced complexity, robust validation	Limited scope, dependency on datasets, size of datasets missing.	Used two MRI datasets, exact size is not given	2024	98.3%, 98.2%, 99.1%	[37]
24.	Hybrid mechanism, VGG16 + Inception	High accuracy, optimization, Data augmentation, Public dataset	Computational Demands, Lack of Explainability, Generalizability, Integration Challenges, Dataset Limitations	T1-weighted MRI, 3,064 and 4,278 before and after balancing	2024	98.64%	[38]

3. METHODOLOGY IMPLIED:

Figure 1 illustrates the key stages of the proposed methodology.



Figure 1: Workflow of the Proposed Approach.

3.1 Brain Tumor Dataset Description:

This study uses an MRI dataset from Kaggle, combining Figshare, the SARTAJ dataset, and Br35H, totaling 7,023 brain scan images. The scans are classified into four groups: glioma, meningioma, pituitary tumor, and no tumor—the latter sourced only from Br35H. The data is organized into training and testing folders, each with four subfolders by class [39]. Figure 2 shows sample MRI images, and Table 2 outlines class-wise details.

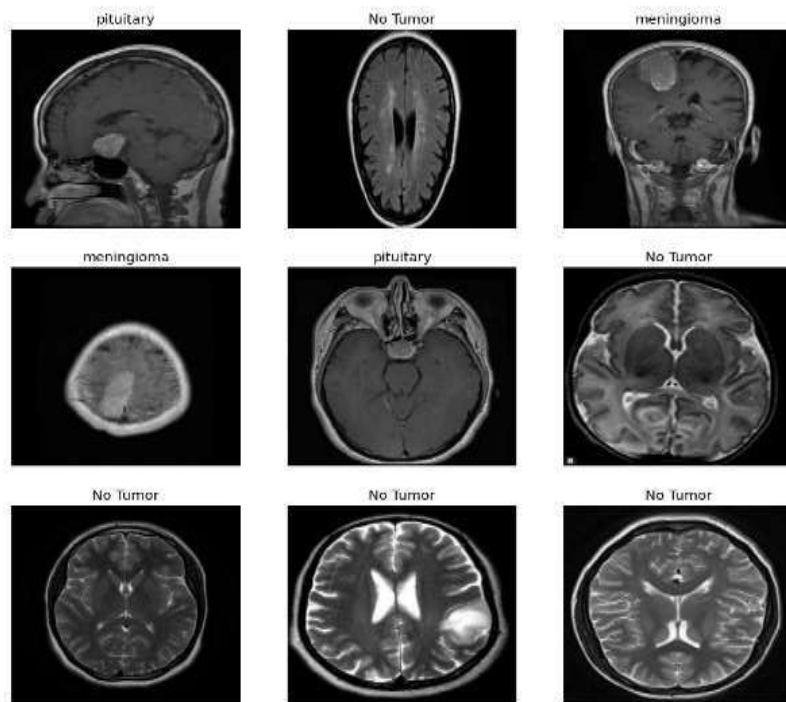


Figure 2: Brain Tumor MRI Dataset Samples

Table 2: Summary of Sartaj MRI dataset

S. No.	Sub-Class	No. of Samples	Description
1	Glioma	1,621	Affected
2	Meningioma	1,757	Affected
3	Pituitary	1,6451	Affected
4	No Tumor	2,000	Normal (Unaffected)

3.2 Data Preprocessing and Model Training Evaluation: All MRI images were resized to $240 \times 240 \times 3$ pixels for uniformity. To reduce complexity and improve processing speed, RGB images were converted to grayscale. The MFIF technique was applied during preprocessing to enhance image quality by reducing noise and sharpening important features [40]. The dataset was then split into training (80%), validation (10%), and testing (10%) sets. VGG16, a widely used CNN model for image classification, served as the core model in this study. Training was performed over several epochs with early stopping to prevent overfitting [41]. Model performance was monitored using validation data, assessing metrics such as accuracy, precision, recall, and AUC-ROC. This helped fine-tune the model and ensure it could generalize well to new data [42]. Finally, both the original and frozen-layer versions of VGG16 were evaluated on the test set using key metrics like loss, accuracy, precision, recall, and AUC-ROC. This step assessed how well the models classified brain tumor types and their potential for real-world clinical use [43]. Figure 3 presents an overview of the study's methodology.

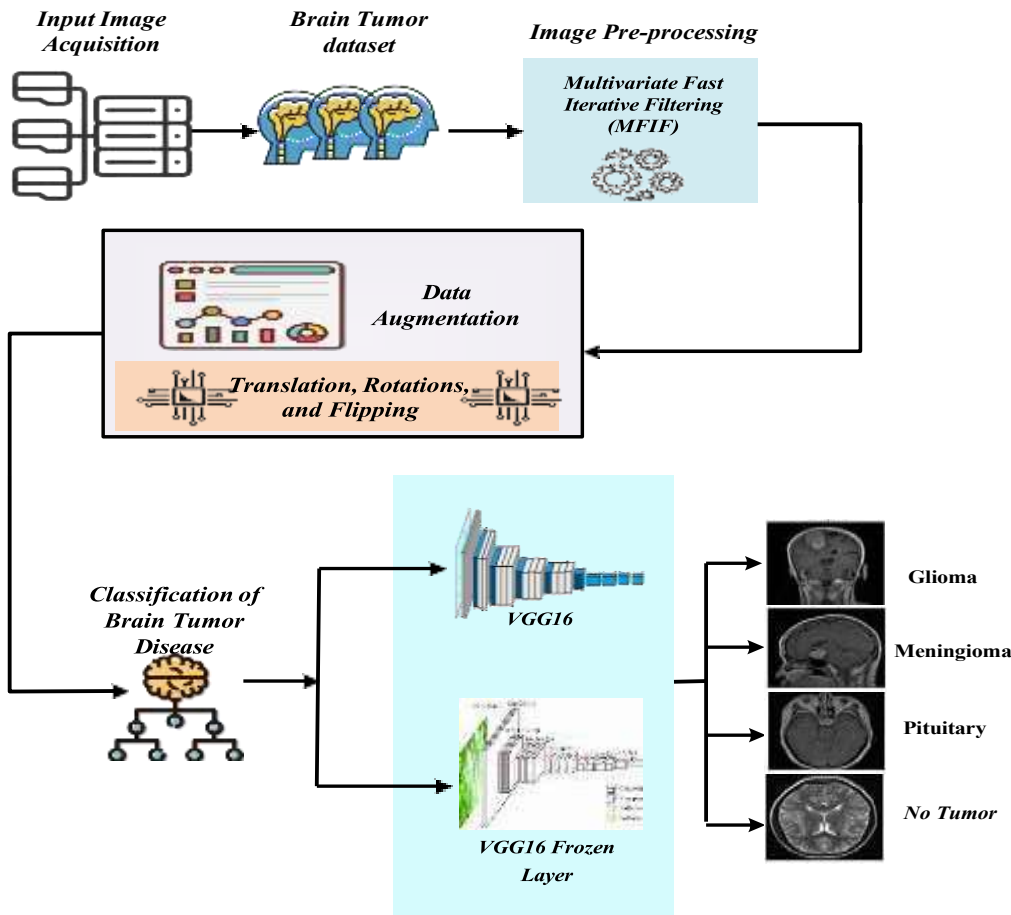


Figure 3: Flowgraph of the study

3.3 Training Configuration: The model was trained for 30 epochs with a batch size of 32. Adamax and SGD optimizers, using learning rates of 0.001 and 0.0001 respectively, were applied to update weights during training. ReLU activation added non-linearity to help the model learn complex features, while softmax was used in the output layer for class probability prediction. Core to this setup is the 2D convolution operation, a key element in CNNs widely used for processing image data [44]. Given an input image Im and a kernel Ka , the 2D convolution procedure can be represented scientifically as follows:

$$C(i, j) = (Im * Ka)(i, j) = \sum_m \sum_n Im(m, n) \cdot Ka(i - m, j - n) \quad (1)$$

Here: $C(i, j)$ signifies the value at position (i, j) in the output feature map. $Im(m, n)$ shows the pixel figure of the input image at position (m, n) . $(i - m, j - n)$ highlights the rate of the kernel at position $(i - m, j - n)$. The double summation is performed over all valid positions of the filter/kernel in the input image. '*' denotes the convolution operation.

ReLU is a basic non-linear activation function defined as:

$$ReLU(y) = \max(0, y) \quad (2)$$

ReLU replaces negative inputs with zero while keeping positive values unchanged, adding non-linearity to help the network learn complex patterns. Its derivative is 1 for positive inputs and 0 for negatives.

Softmax is a widely used activation function in neural networks, especially for multi-class classification. It's typically applied in the output layer to convert raw model scores into probabilities for each class. Given an input vector z with K elements (where K denotes the total classes), it is defined as:

$$Softmax(z)_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad (3)$$

Here: $\text{Softmax}(z)_i$ represents the i th component of the output probability vector. e_i Euler's number (approximately 2.71828). z_i defines the i th component of the input vector. The denominator sums the exponential of all elements in the input vector.

3.4 VGG16 Architecture Layers: VGG16 consists of 13 convolutional layers and 3 fully connected layers, organized into blocks with max-pooling layers to downsample feature maps [45]. Table 3 outlines the layer-wise architecture of the VGG16 model.

Table 3: The VGG16-CNN layers description

Layer(s)	Output Shape	Param#
input_layer (InputLayer)	(--, 240, 240, 3)	0
1st_conv1 (Conv2D)	(--, 240, 240, 64)	1,792
1st_conv2 (Conv2D)	(--, 240, 240, 64)	36,928
1st_pool (MaxPooling2D)	(--, 120, 120, 64)	0
2nd_conv1 (Conv2D)	(--, 120, 120, 128)	73,856
2nd_conv2 (Conv2D)	(--, 120, 120, 128)	147,584
2nd_pool (MaxPooling2D)	(--, 60, 60, 128)	0
3rd_conv1 (Conv2D)	(--, 60, 60, 256)	295,168
3rd_conv2 (Conv2D)	(--, 60, 60, 256)	590,080
3rd_conv2 (Conv2D)	(--, 60, 60, 256)	590,080
3rd_pool (MaxPooling2D)	(--, 30, 30, 256)	0
4th_conv1 (Conv2D)	(--, 30, 30, 512)	1,180,160
4th_conv2 (Conv2D)	(--, 30, 30, 512)	2,359,808
4th_conv2 (Conv2D)	(--, 30, 30, 512)	2,359,808
4th_pool (MaxPooling2D)	(--, 15, 15, 512)	0
5th_conv1 (Conv2D)	(--, 15, 15, 512)	2,359,808
5th_conv2 (Conv2D)	(--, 15, 15, 512)	2,359,808
5th_conv2 (Conv2D)	(--, 15, 15, 512)	2,359,808
5th_pool (MaxPooling2D)	(--, 7, 7, 512)	0

VGG16 with Frozen Layer Algorithm:

```
VGG16Frozen Layer (input_shape, num_classes) Input (input_shape)
for layer in vggf.layers[:-5]: #Set 5th block as trainable print(layer.name)
layer.trainable = False #Othar than 5th as non-trainable return to model
```

The above algorithm iterates over all but the last 5 layers of the model. Prints the name of each layer being iterated over, which can help identify the layer architecture or confirm which layers are being frozen. The concept of freezing a layer in TensorFlow is given in web document. The following layers are frozen or non-trainable:

input_layer_1, 1st_conv1, 1st_conv2, 1st_pool, 2nd_conv1, 2nd_conv2, 2nd_pool, 3rd_conv1, 3rd_conv2, 3rd_conv3, 3rd_pool, 4th_conv1, 4th_conv2, 4th_conv3

3.5 Data augmentation: Data augmentation helps create a more diverse training dataset, which can improve the model's robustness and generalization. It is mostly valuable when the quantity of existing training data is restricted [46]. Table 4 presents the data augmentation methods applied in this study.

Table 4: Parameters for the data augmentation process used in this work.

S. No.	Param.	Value	S. No.	Param.	Value
1.	Horizontal_flip	True	5.	Width shift range	0.1
2.	Vertical_flip	True	6.	Height shift range	0.1
3.	Shear range	0.2	7.	Rotation	90°
4.	Zoom range	0.2			

3.6 Hyper-parameter (Fine-Tuning) used in these two models: Fine-tuning VGG16 for brain tumor classification involves several steps. This process typically includes preparing the dataset, adjusting the VGG16 architecture for the specific task, and then training the model. Figure 4 illustrates the block diagrams of the VGG16 and frozen-layer VGG16 models, including fine-tuning and applied hyperparameters [47]. Table 5 provides the details of the hyperparameter settings used during training.

Table 5: Hyperparameters and fine-tuning used in this implementation

Hyper-parameters	VGG16	VGG16 with Frozen Layer
Loss function	categorical_crossentropy	categorical_crossentropy
Optimizer	Adamax	SGD
Epochs	30	30
Batch_size	32	32
Learning_rate	0.001	0.0001
Momentum	Nil	0.9

A small learning rate, like 0.0001, allows for finer, more precise updates, which is beneficial when fine-tuning a model.

Momentum (0.9) helps accelerate convergence by allowing the optimizer to maintain the direction of previous gradients. This helps avoid getting stuck in local minima and smooths out oscillations during updates.

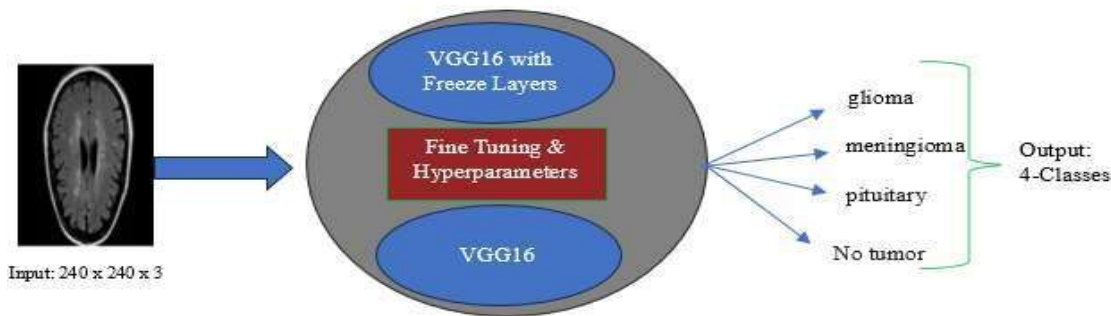


Figure 4: VGG16 and VGG16 with Freeze Layers models for the classification of brain tumour disease

3.7 Early Stopping and Performance Metrics: To avoid overfitting and improve training, early stopping and learning rate reduction were used [48]. Early stopping was triggered if validation accuracy didn't improve for 3 epochs. If accuracy plateaued for 2 epochs, the learning rate was reduced by 0.1 to help the model converge more smoothly. After training, the VGG16 model was evaluated using several key performance metrics [49], [50], including:

$$Accuracy = \frac{TruePositive + TrueNegative}{TruePositive + TrueNegative + FalsePositive + FalseNegative} \quad (4)$$

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \quad (5)$$

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \quad (6)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (7)$$

$$Loss = -\sum_{i=1}^c y_i * \log[\hat{y}_i] \quad (8)$$

Here, c represents the number of classes, y_i is the true probability for class i , and \hat{y}_i is the predicted probability. The AUC-ROC offers a summary of the model's ability to distinguish between classes across different thresholds [43].

$$True Positive Rate = \frac{TruePositive}{(TruePositive + FalseNegative)} \quad (9)$$

$$\text{False Postive Rate} = \frac{\text{FalsePositive}}{(\text{FalsePositive} + \text{TrueNegative})} \quad (10)$$

The ROC curve illustrates the balance between true positives and false positives.

4. RESULTS AND DISCUSSION

The experimental findings demonstrate the strong performance and reliability of both the standard VGG16 model and its frozen-layer variant in detecting brain tumors from MRI scans. Achieving a training accuracy of 99.97% and an AUC of 1.0, the model effectively learns and captures complex patterns from the training data. Table 6 summarizes the average training and validation accuracy, loss, and AUC values recorded at epoch 30. Notably, the model also generalizes well to unseen data, with a test accuracy of 98.12% and a test AUC of 0.9994.

Table 6: Average Training and Validation Accuracy, Loss, and AUC for CNN Models with and Without Data Augmentation

CNN-Model	Train			Test		
	Accuracy	Loss	AUC	Accuracy	Loss	AUC
VGG16	69.06%	0.8751	89.63%	67.76%	0.8947	88.93%
VGG16 with Augmentation	94.45%	0.1429	99.59%	89.25%	0.3549	98.14%
VGG16 Frozen Layers	99.80%	0.0204	99.99%	96.72%	0.0750	99.90%
VGG16 Frozen Layers with Augmentation	99.98%	0.0069	100.00%	98.12%	0.0571	99.94%

Figure 5 shows the average training performance of various CNN models based on VGG16 architecture, evaluated by accuracy, AUC, and loss. The VGG16 model with data augmentation achieved the highest performance, with 94.45% accuracy and a notable AUC improvement compared to the standard VGG16 model. Freezing layers in VGG16 maintained almost similar accuracy and AUC, while applying both augmentation and freezing layers yielded consistently high values for accuracy and AUC, demonstrating the effectiveness of augmentation in boosting model performance while minimizing loss.

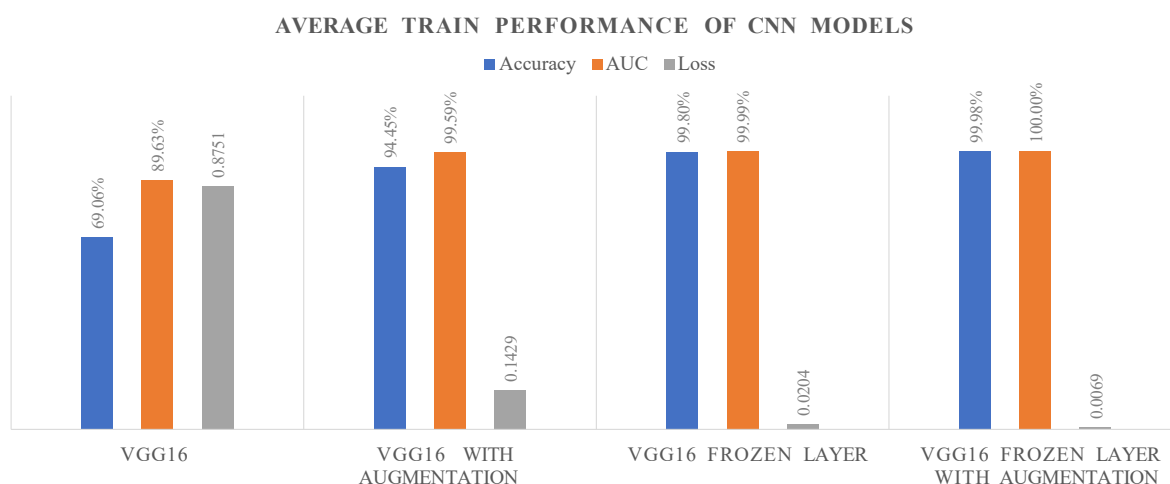


Figure 5: Average training performance of the VGG16 and VGG16 with Freeze Layers models for the classification of brain tumour disease

Similarly, Figure 6 presents the average validation performance of CNN models based on the VGG16 architecture, evaluated by accuracy, AUC, and loss. The VGG16 model with augmentation achieved the highest validation accuracy (89.25%) and AUC, outperforming the standard VGG16 model and confirming

the benefit of data augmentation in generalizing model performance. Models with frozen layers also showed stable accuracy and AUC, indicating that freezing layers, especially when combined with augmentation, contribute to maintaining strong generalization with reduced loss.

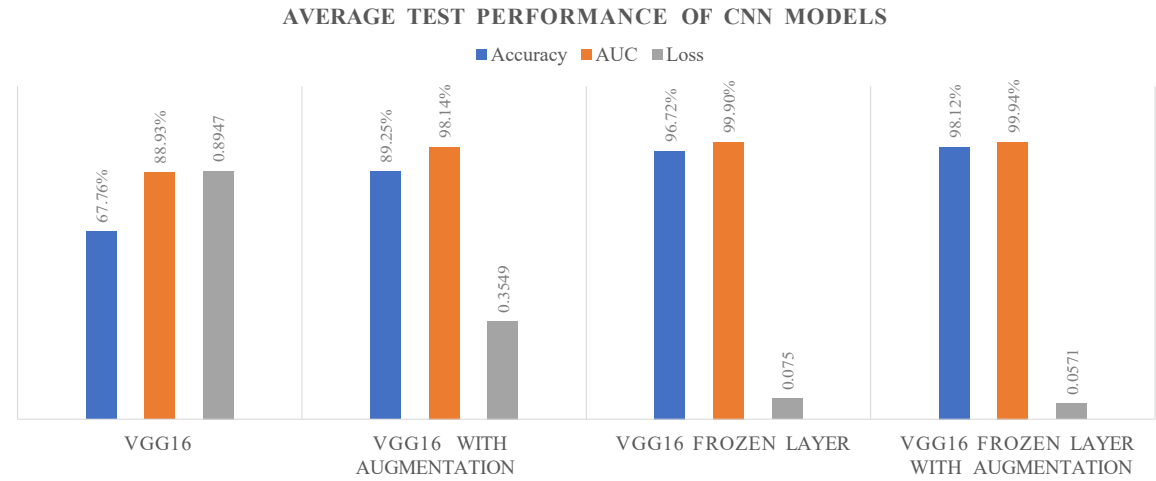



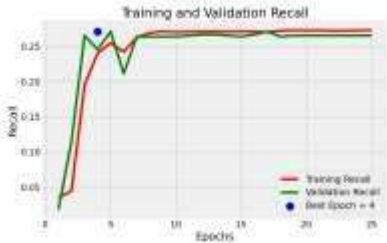


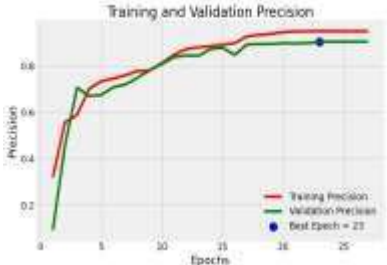



Figure 6: Average test performance of the VGG16 and VGG16 with Freeze Layers models for the classification of brain tumour disease

Table 7 displays the training and validation metrics—loss, accuracy, precision, and recall—for both CNN models: VGG16 and its frozen-layer version, tested with and without data augmentation. For the VGG16 model, several key trends are observed: the top-left graph shows a clear drop in both training and validation loss, suggesting effective learning with minimal signs of overfitting. The top-right graph highlights a steady rise in accuracy, with validation accuracy leveling off around 0.70. Precision, shown in the bottom-left graph, initially varies but eventually settles close to 0.9, indicating improved consistency in correct predictions. Recall climbs quickly and stabilizes around 90%, with training and validation values remaining closely matched—showing the model maintains strong sensitivity across epochs. Similar patterns were observed in the other models, as reflected in the table.

Table 8 presents the classification report, confusion matrix, and ROC curve for each brain tumor category—glioma, meningioma, no tumor, and pituitary—across all models. These evaluation tools offer a comprehensive view of model performance by detailing precision, recall, F1-score, and accuracy for each class. The ROC curve helps visualize how well the model distinguishes between true and false positives. The confusion matrix further breaks down the predictions, showing how often the model correctly or incorrectly labeled each class, offering valuable insight into its strengths and weaknesses. Table 9 provides a comparison of the CNN models based on their average accuracy and ROC scores across all four categories.

Table 7: Performance of training progress and validation: Loss, Accuracy, Precision, and Recall of CNN-Model

CNN-Model	Training Progress performance vs. validation Accuracy, Loss, Precision and Recall	
VGG16		
		
VGG16 with Augmentation		
		

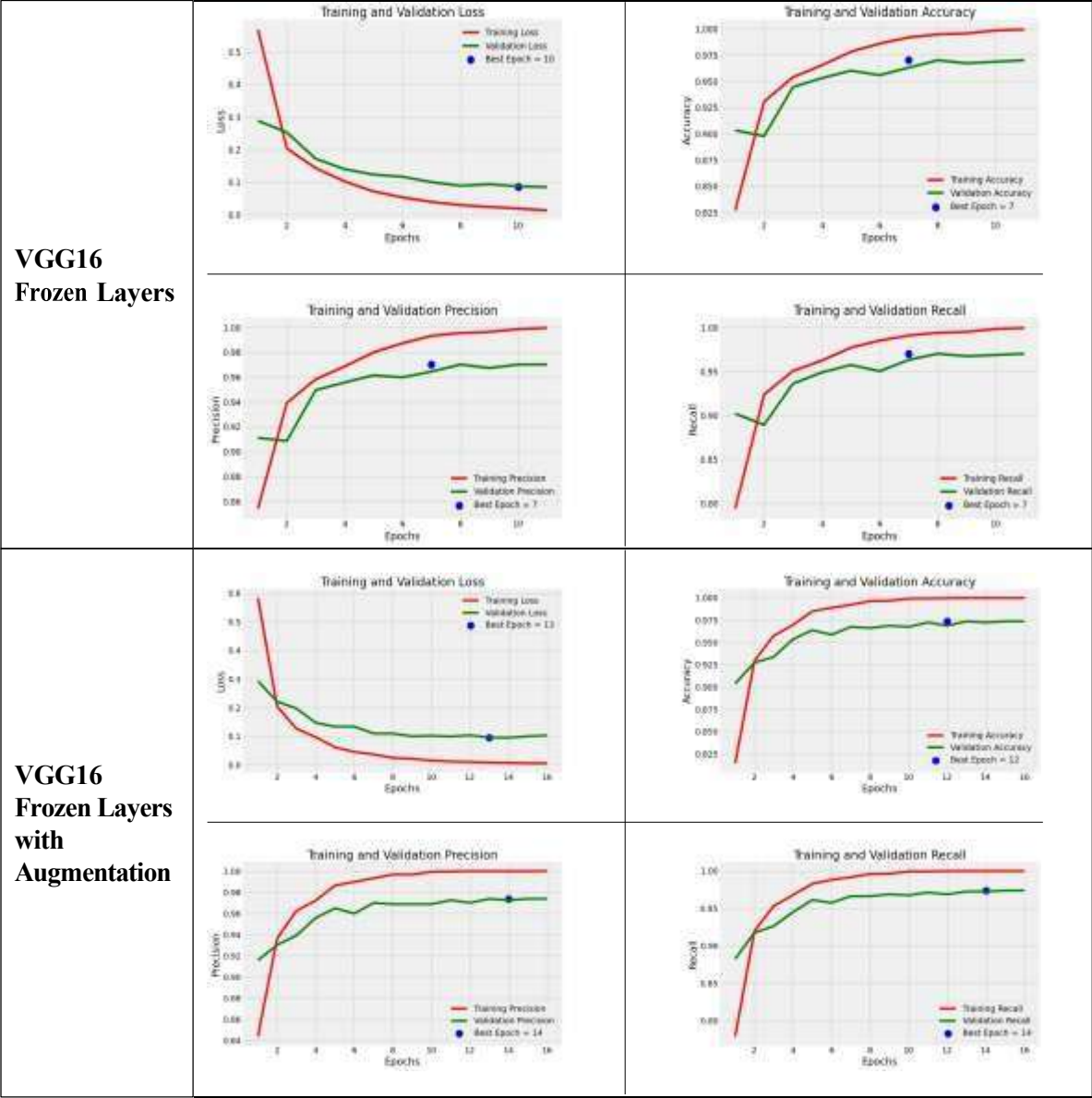
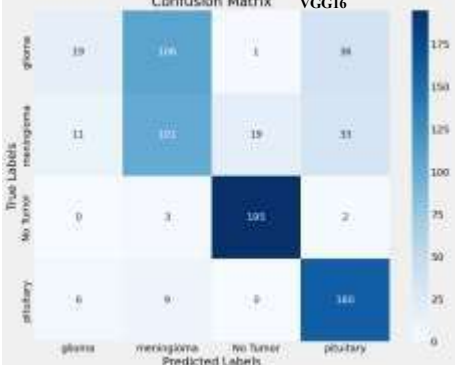
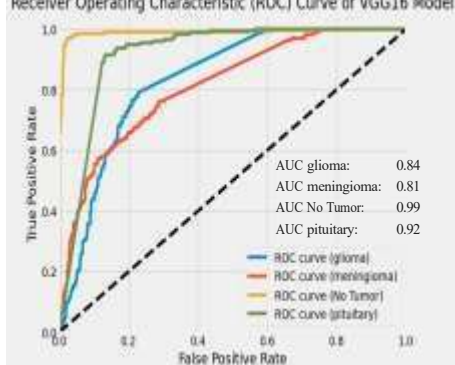

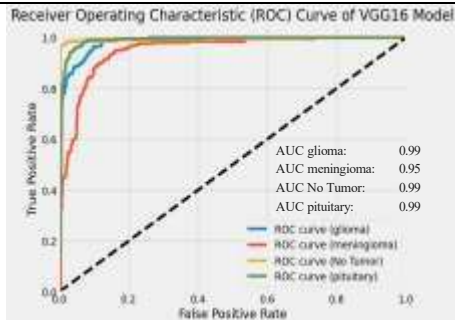


Table 8: Classification report, Confusion Matrix and ROC curve of the different CNN-Models

CNN-Model	Classification Report, Confusion Matix and ROC Curve				
VGG16	Classification Report				
		precision	recall	f1-score	support
	glioma	0.53	0.12	0.19	162
	meningioma	0.46	0.62	0.53	164
	No Tumor	0.91	0.97	0.94	200
	pituitary	0.69	0.91	0.79	175
	accuracy			0.68	701
VGG16	macro avg	0.65	0.66	0.61	701
	weighted avg	0.66	0.68	0.63	701
	Confusion Matrix			ROC Curve	

	<div><div><div>Confusion Matrix</div><div>VGG16</div></div><div><div>Receiver Operating Characteristic (ROC) Curve of VGG16 Model</div></div></div>																																								
VGG16 with Augmentation	<div><div>Classification Report</div><table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>glioma</td><td>0.90</td><td>0.85</td><td>0.88</td><td>200</td></tr><tr><td>meningioma</td><td>0.78</td><td>0.82</td><td>0.80</td><td>200</td></tr><tr><td>No Tumor</td><td>0.98</td><td>0.96</td><td>0.97</td><td>200</td></tr><tr><td>pituitary</td><td>0.91</td><td>0.93</td><td>0.92</td><td>200</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.89</td><td>800</td></tr><tr><td>macro avg</td><td>0.89</td><td>0.89</td><td>0.89</td><td>800</td></tr><tr><td>weighted avg</td><td>0.89</td><td>0.89</td><td>0.89</td><td>800</td></tr></tbody></table></div> <div><div>Confusion Matrix</div><div>VGG16 Augmented</div></div> <div><div>ROC Curve</div><div>Receiver Operating Characteristic (ROC) Curve of VGG16 Model</div></div>		precision	recall	f1-score	support	glioma	0.90	0.85	0.88	200	meningioma	0.78	0.82	0.80	200	No Tumor	0.98	0.96	0.97	200	pituitary	0.91	0.93	0.92	200	accuracy			0.89	800	macro avg	0.89	0.89	0.89	800	weighted avg	0.89	0.89	0.89	800
		precision	recall	f1-score	support																																				
	glioma	0.90	0.85	0.88	200																																				
	meningioma	0.78	0.82	0.80	200																																				
	No Tumor	0.98	0.96	0.97	200																																				
	pituitary	0.91	0.93	0.92	200																																				
	accuracy			0.89	800																																				
	macro avg	0.89	0.89	0.89	800																																				
	weighted avg	0.89	0.89	0.89	800																																				
	VGG16 Frozen Layer	<div><div>Classification Report</div><table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>glioma</td><td>0.97</td><td>0.93</td><td>0.95</td><td>162</td></tr><tr><td>meningioma</td><td>0.92</td><td>0.96</td><td>0.94</td><td>164</td></tr><tr><td>No Tumor</td><td>1.00</td><td>0.98</td><td>0.99</td><td>200</td></tr><tr><td>pituitary</td><td>0.98</td><td>0.98</td><td>0.98</td><td>175</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.97</td><td>701</td></tr><tr><td>macro avg</td><td>0.97</td><td>0.97</td><td>0.97</td><td>701</td></tr><tr><td>weighted avg</td><td>0.97</td><td>0.97</td><td>0.97</td><td>701</td></tr></tbody></table></div> <div><div>Confusion Matrix</div><div>ROC Curve</div></div>		precision	recall	f1-score	support	glioma	0.97	0.93	0.95	162	meningioma	0.92	0.96	0.94	164	No Tumor	1.00	0.98	0.99	200	pituitary	0.98	0.98	0.98	175	accuracy			0.97	701	macro avg	0.97	0.97	0.97	701	weighted avg	0.97	0.97	0.97
		precision	recall	f1-score	support																																				
glioma		0.97	0.93	0.95	162																																				
meningioma		0.92	0.96	0.94	164																																				
No Tumor		1.00	0.98	0.99	200																																				
pituitary		0.98	0.98	0.98	175																																				
accuracy				0.97	701																																				
macro avg		0.97	0.97	0.97	701																																				
weighted avg		0.97	0.97	0.97	701																																				

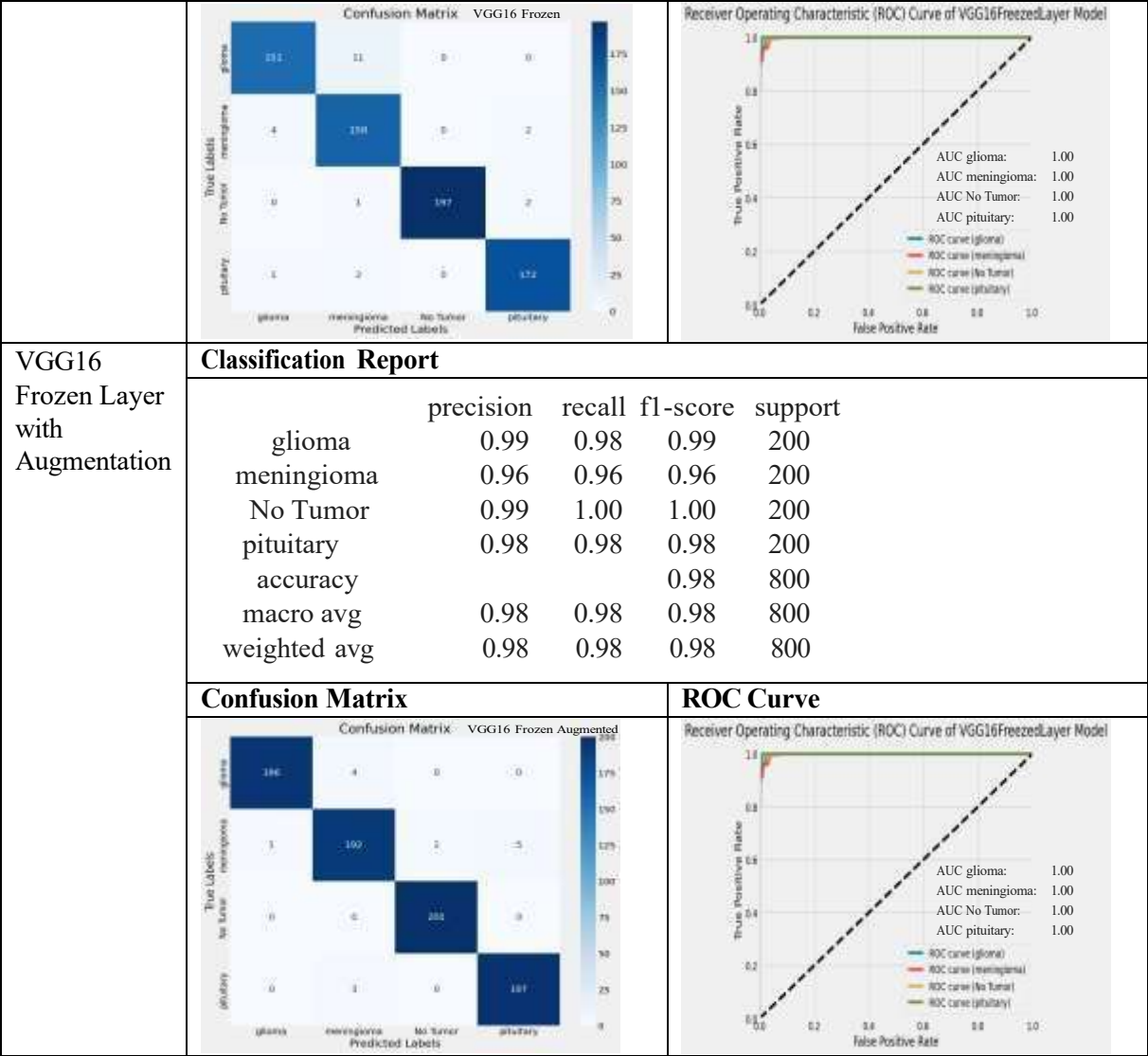


Table 9: Comparison of Different CNN Models Based on Average Accuracy and ROC across All Four Classes

CNN-Models	Average Accuracy	Average ROC
VGG16	68%	0.89
VGG16 with Augmentation	89%	0.98
VGG16 Frozen Layer	97%	1.00
VGG16 Frozen Layer with Augmentation	98%	1.00

Figure 7 compares different CNN models based on average accuracy and ROC across all four classes.

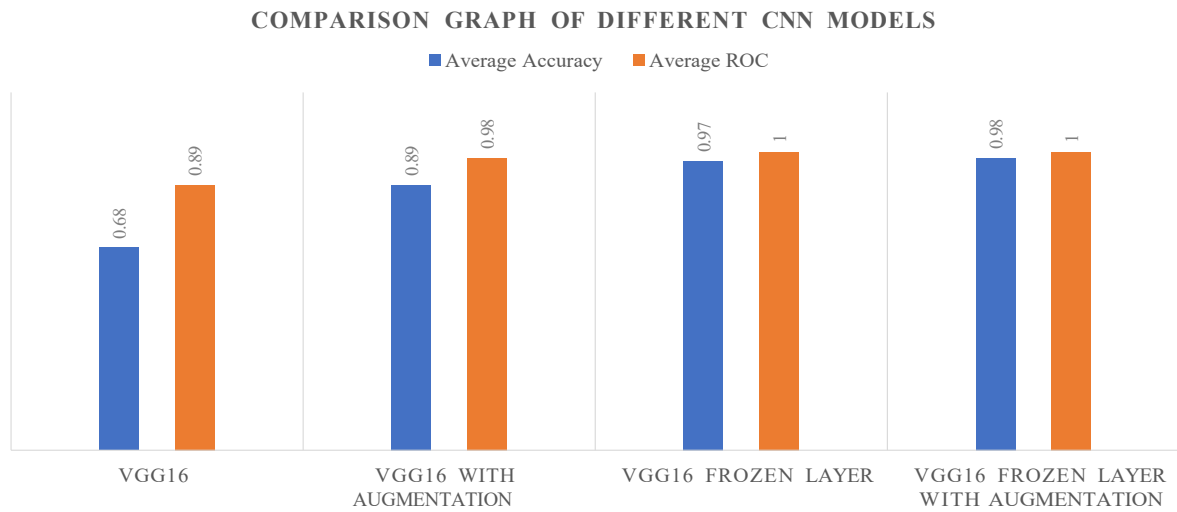


Figure 7: Comparison graph of different CNN models based on average accuracy and ROC curve across all four classes

Comparative Analysis:

A detailed comparative evaluation of various deep learning and hybrid approaches applied to the Kaggle Brain MRI dataset [39]. The comparison is based on key attributes, including the choice of CNN architectures, optimization techniques, dataset size, batch size, and average accuracy. Table 10 provides a summary of these findings.

Table 10: Comparative analysis of the proposed method against recent techniques: A detailed evaluation of performance metrics and methodologies.

Method	CNN-Model	Optimizer	Dataset	Batch Size	Year	Avg. Accuracy	Ref.
Deep Learning and Transfer Learning	ResNet152	Adam (learning rate = 0.001)	Kaggle Brain MRI, 7,023	128	2024	98.5%	[34]
	VGG19					96%	
	DenseNet169					96.75%	
	MobileNetv3					96%	
Hybrid CNN with ML	DenseNet201-SVM	Adam (learning rate = 0.001)	Kaggle Brain MRI, 7,023	32	2024	96.87%	[36]
	Inceptionv3-SVM					95.3%	
	ResNet50-SVM					96.3%	
	Proposed hybrid					97.15%	
Deep Learning	Optimized CNN	Adam (learning rate = 0.001)	Kaggle Brain MRI, 7,023	32	2024	95%	[51]
Image Enhancement and CNN	VGG16	Adam (learning rate = 0.001)	Kaggle Brain MRI, 7,023	32	2024	95%	[52]
	ReNet50					94.75%	
	VGG19					94.83%	
	Proposed-CNN					97.84%	
Multi-fused CNN with Auxiliary Layer	MFR-CNN	Adam (learning rate = 0.00001)	Kaggle Brain MRI, 7,023	16	2024	94%	[53]
CNN with Fine-tune	VGG16 with and without augmentation	Adamax (learning rate = 0.001)	Kaggle Brain MRI, 7,023	32	--	68% (without Augmentation) 89% (with Augmentation)	In this paper

CNN with transfer learning	VGG16 Frozen Layer with and without augmentation	SGD (learning rate = 0.0001	Kaggle Brain MRI, 7,023	32	--	97% (without Augmentation) 98% (with Augmentation)	In this paper
----------------------------	--	-----------------------------	-------------------------	----	----	---	---------------

Figure 8 presents sample prediction outcomes from the VGG16 model with frozen layers. Four input images, each representing one category, display prediction probabilities close to 100% for their respective classes. Each subplot shows the input image, its actual label, and the model's predicted probability distribution across the four possible classes. The input image is labeled as “pituitary”. The model correctly classifies it with 100% confidence in the pituitary class, indicating strong certainty in its prediction. The input image is labeled as “No tumor”. The model predicts “No tumor with 100% confidence, accurately identifying the absence of a tumor. The input image is labeled as “glioma”. The model classifies it as “glioma” with 100% certainty, showing a highly accurate prediction. The input image is labeled as “meningioma”. The model predicts “meningioma” with 99.99% confidence, nearly perfect in its prediction.

Overall, this figure demonstrates the model’s high accuracy and confidence in predicting the correct tumor class for each sample, with minimal ambiguity across different tumor types.

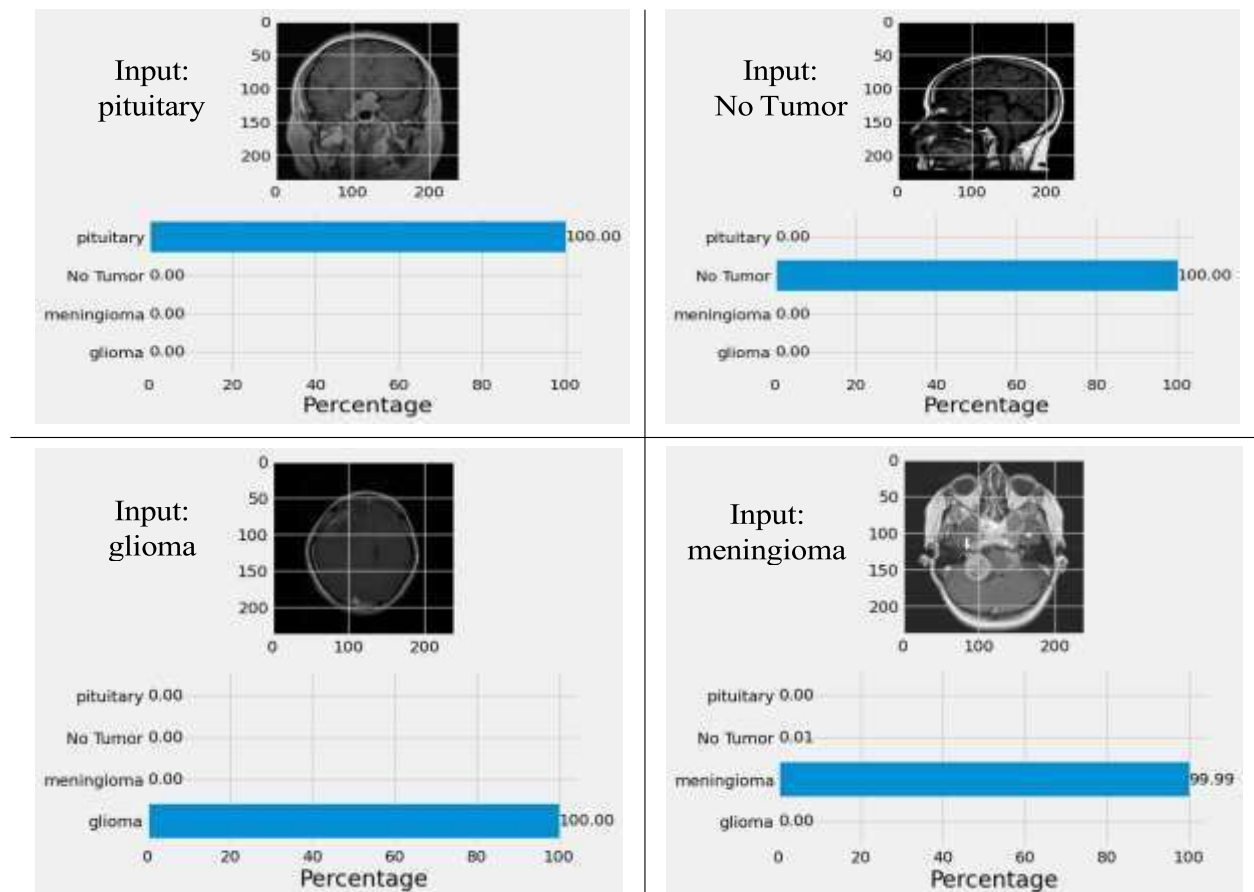


Figure 8: The prediction results for the VGG16 model with frozen layers, without any data augmentation, are provided for each input class. The image displayed represents the input to the model.

5. CONCLUSION WITH FUTURE DIRECTION:

This study employed a dataset of 7,023 MRI images classified into four brain tumor categories and evaluated the performance of VGG16 and VGG16 with frozen layers. Models were trained using 30 epochs and optimized with Adamax and SGD. The frozen-layer VGG16 with augmentation achieved the highest performance, with 99.98% training accuracy and 98.12% test accuracy, along with perfect AUC scores. Data augmentation and layer freezing significantly boosted accuracy—up to 30.74% in training and 28.96% in

validation—demonstrating the effectiveness of these techniques. The findings affirm the potential of CNN-based models, especially VGG16 variants, for robust and automated brain tumor detection. Future work may explore larger datasets, advanced augmentation strategies, and other CNN architectures for further improvement.

REFERENCES:

- [1] A. Chattopadhyay and M. Maitra, “MRI-based brain tumour image detection using CNN based deep learning method,” *Neuroscience Informatics*, vol. 2, no. 4, p. 100060, Dec. 2022, doi: 10.1016/j.neuri.2022.100060.
- [2] Md. S. I. Khan et al., “Accurate brain tumor detection using deep convolutional neural network,” *Comput Struct Biotechnol J*, vol. 20, pp. 4733–4745, 2022, doi: 10.1016/j.csbj.2022.08.039.
- [3] S. Maqsood, R. Damaševičius, and R. Maskeliūnas, “Multi-Modal Brain Tumor Detection Using Deep Neural Network and Multiclass SVM,” *Medicina (B Aires)*, vol. 58, no. 8, p. 1090, Aug. 2022, doi: 10.3390/medicina58081090.
- [4] H. H. Sultan, N. M. Salem, and W. Al-Atabany, “Multi-Classification of Brain Tumor Images Using Deep Neural Network,” *IEEE Access*, vol. 7, pp. 69215–69225, 2019, doi: 10.1109/ACCESS.2019.2919122.
- [5] I. Abd El Kader, G. Xu, Z. Shuai, S. Saminu, I. Javaid, and I. Salim Ahmad, “Differential Deep Convolutional Neural Network Model for Brain Tumor Classification,” *Brain Sci*, vol. 11, no. 3, p. 352, Mar. 2021, doi: 10.3390/brainsci11030352.
- [6] M. Z. Khaliki and M. S. Başarslan, “Brain tumor detection from images and comparison with transfer learning methods and 3-layer CNN,” *Sci Rep*, vol. 14, no. 1, p. 2664, Feb. 2024, doi: 10.1038/s41598-024-52823-9.
- [7] J. Amin, M. Sharif, A. Haldorai, M. Yasmin, and R. S. Nayak, “Brain tumor detection and classification using machine learning: a comprehensive survey,” *Complex & Intelligent Systems*, vol. 8, no. 4, pp. 3161–3183, Aug. 2022, doi: 10.1007/s40747-021-00563-y.
- [8] R. Ranjbarzadeh, A. Bagherian Kasgari, S. Jafarzadeh Ghouschi, S. Anari, M. Naseri, and M. Bendechache, “Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images,” *Sci Rep*, vol. 11, no. 1, p. 10930, May 2021, doi: 10.1038/s41598-021-90428-8.
- [9] S. Deepak and P. M. Ameer, “Brain tumor classification using deep CNN features via transfer learning,” *Comput Biol Med*, vol. 111, p. 103345, Aug. 2019, doi: 10.1016/j.combiomed.2019.103345.
- [10] A. M. Alqudah, “Brain Tumor Classification Using Deep Learning Technique - A Comparison between Cropped, Uncropped, and Segmented Lesion Images with Different Sizes,” *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no. 6, pp. 3684–3691, Dec. 2019, doi: 10.30534/ijatcse/2019/155862019.
- [11] M. Havaei et al., “Brain tumor segmentation with Deep Neural Networks,” *Med Image Anal*, vol. 35, pp. 18–31, Jan. 2017, doi: 10.1016/j.media.2016.05.004.
- [12] Z. Liu et al., “Deep learning based brain tumor segmentation: a survey,” *Complex & Intelligent Systems*, vol. 9, no. 1, pp. 1001–1026, Feb. 2023, doi: 10.1007/s40747-022-00815-5.
- [13] W. Chen, B. Liu, S. Peng, J. Sun, and X. Qiao, “S3D-UNet: Separable 3D U-Net for Brain Tumor Segmentation,” 2019, pp. 358–368. doi: 10.1007/978-3-030-11726-9_32.
- [14] D. Ueda et al., “Fairness of artificial intelligence in healthcare: review and recommendations,” *Jpn J Radiol*, vol. 42, no. 1, pp. 3–15, Jan. 2024, doi: 10.1007/s11604-023-01474-3.
- [15] J. S. Paul, A. J. Plassard, B. A. Landman, and D. Fabbri, “Deep learning for brain tumor classification,” in *Medical Imaging 2017: Biomedical Applications in Molecular, Structural, and Functional Imaging*, SPIE, Mar. 2017, p. 1013710. doi: 10.1117/12.2254195.
- [16] H. H. Sultan, N. M. Salem, and W. Al-Atabany, “Multi-Classification of Brain Tumor Images Using Deep Neural Network,” *IEEE Access*, vol. 7, pp. 69215–69225, 2019, doi: 10.1109/ACCESS.2019.2919122.
- [17] Z. N. K. Swati et al., “Brain tumor classification for MR images using transfer learning and fine-tuning,” *Computerized Medical Imaging and Graphics*, vol. 75, 2019, doi: 10.1016/j.compmedimag.2019.05.001.

- [18] J. Amin, M. Sharif, M. Yasmin, T. Saba, M. A. Anjum, and S. L. Fernandes, "A New Approach for Brain Tumor Segmentation and Classification Based on Score Level Fusion Using Transfer Learning," *J Med Syst*, vol. 43, no. 11, Nov. 2019, doi: 10.1007/s10916-019-1453-8.
- [19] S. A. Abdelaziz Ismael, A. Mohammed, and H. Hefny, "An enhanced deep learning approach for brain cancer MRI images classification using residual networks," *Artif Intell Med*, vol. 102, Jan. 2020, doi: 10.1016/j.artmed.2019.101779.
- [20] V. K. Bairagi, P. P. Gumaste, S. H. Rajput, and K. S. Chethan, "Automatic brain tumor detection using CNN transfer learning approach," *Med Biol Eng Comput*, vol. 61, no. 7, pp. 1821–1836, Jul. 2023, doi: 10.1007/s11517-023-02820-3.
- [21] T. Sadad et al., "Brain tumor detection and multi-classification using advanced deep learning techniques," *Microsc Res Tech*, vol. 84, no. 6, pp. 1296–1308, Jun. 2021, doi: 10.1002/jemt.23688.
- [22] M. I. Sharif, M. A. Khan, M. Alhussein, K. Aurangzeb, and M. Raza, "A decision support system for multimodal brain tumor classification using deep learning," *Complex and Intelligent Systems*, vol. 8, no. 4, pp. 3007–3020, Aug. 2022, doi: 10.1007/s40747-021-00321-0.
- [23] D. G.-O. F.J. Díaz-Pernas, M. Martínez-Zarzuela, M. Antón-Rodríguez, "A Deep Learning Approach for Brain Tumor Classification and Segmentation Using a Multiscale Convolutional Neural Network," *Healthcare*, vol. 9, no. 2, p. 153, 2021, doi: <https://doi.org/10.3390/healthcare9020153>.
- [24] A. Aziz et al., "An Ensemble of Optimal Deep Learning Features for Brain Tumor Classification," *Computers, Materials & Continua*, vol. 69, no. 2, pp. 2653–2670, 2021, doi: 10.32604/cmc.2021.018606.
- [25] W. Ayadi, W. Elhamzi, I. Charfi, and M. Atri, "Deep CNN for Brain Tumor Classification," *Neural Process Lett*, vol. 53, no. 1, pp. 671–700, Feb. 2021, doi: 10.1007/s11063-020-10398-2.
- [26] H. Mehnatkesh, S. M. J. Jalali, A. Khosravi, and S. Nahavandi, "An intelligent driven deep residual learning framework for brain tumor classification using MRI images," *Expert Syst Appl*, vol. 213, p. 119087, Mar. 2023, doi: 10.1016/j.eswa.2022.119087.
- [27] M. A. Kumaar, D. Samiayya, V. Rajinikanth, D. R. Vincent P M, and S. Kadry, "Brain Tumor Classification Using a Pre-Trained Auxiliary Classifying Style-Based Generative Adversarial Network," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 8, no. 6, p. 101, 2024, doi: 10.9781/ijimai.2023.02.008.
- [28] S. Z. Kurdi, M. H. Ali, M. M. Jaber, T. Saba, A. Rehman, and R. Damaševičius, "Brain Tumor Classification Using Meta-Heuristic Optimized Convolutional Neural Networks," *J Pers Med*, vol. 13, no. 2, p. 181, Jan. 2023, doi: 10.3390/jpm13020181.
- [29] A. K. Sharma et al., "Brain tumor classification using the modified ResNet50 model based on transfer learning," *Biomed Signal Process Control*, vol. 86, p. 105299, Sep. 2023, doi: 10.1016/j.bspc.2023.105299.
- [30] B. S. Abd El-Wahab, M. E. Nasr, S. Khamis, and A. S. Ashour, "BTC-fCNN: Fast Convolution Neural Network for Multi-class Brain Tumor Classification," *Health Inf Sci Syst*, vol. 11, no. 1, p. 3, Jan. 2023, doi: 10.1007/s13755-022-00203-w.
- [31] M. A. HAQ, I. KHAN, A. AHMED, S. M. ELDIN, A. ALSHEHRI, and N. A. GHAMRY, "DCNNBT: A NOVEL DEEP CONVOLUTION NEURAL NETWORK-BASED BRAIN TUMOR CLASSIFICATION MODEL," *Fractals*, vol. 31, no. 06, Jan. 2023, doi: 10.1142/S0218348X23401023.
- [32] F. Demir, Y. Akbulut, B. Taşcı, and K. Demir, "Improving brain tumor classification performance with an effective approach based on new deep learning model named 3ACL from 3D MRI data," *Biomed Signal Process Control*, vol. 81, p. 104424, Mar. 2023, doi: 10.1016/j.bspc.2022.104424.
- [33] M. M. Emam, N. A. Samee, M. M. Jamjoom, and E. H. Houssein, "Optimized deep learning architecture for brain tumor classification using improved Hunger Games Search Algorithm," *Comput Biol Med*, vol. 160, p. 106966, Jun. 2023, doi: 10.1016/j.compbiomed.2023.106966.
- [34] S. K. Mathivanan, S. Sonaimuthu, S. Murugesan, H. Rajadurai, B. D. Shivahare, and M. A. Shah, "Employing deep learning and transfer learning for accurate brain tumor detection," *Sci Rep*, vol. 14, no. 1, p. 7232, Mar. 2024, doi: 10.1038/s41598-024-57970-7.

- [35] S. E. Nassar, I. Yasser, H. M. Amer, and M. A. Mohamed, "A robust MRI-based brain tumor classification via a hybrid deep learning technique," *J Supercomput*, vol. 80, no. 2, pp. 2403–2427, Jan. 2024, doi: 10.1007/s11227-023-05549-w.
- [36] M. Celik and O. Inik, "Development of hybrid models based on deep learning and optimized machine learning algorithms for brain tumor Multi-Classification," *Expert Syst Appl*, vol. 238, p. 122159, Mar. 2024, doi: 10.1016/j.eswa.2023.122159.
- [37] S. Khoramipour, M. Gandomkar, and M. Shakiba, "Enhancement of brain tumor classification from MRI images using multi-path convolutional neural network with SVM classifier," *Biomed Signal Process Control*, vol. 93, p. 106117, Jul. 2024, doi: 10.1016/j.bspc.2024.106117.
- [38] M. Aljohani et al., "An automated metaheuristic-optimized approach for diagnosing and classifying brain tumors based on a convolutional neural network," *Results in Engineering*, vol. 23, p. 102459, Sep. 2024, doi: 10.1016/j.rineng.2024.102459.
- [39] Nickparvar and Msoud, "MRI dataset," Kaggle:<https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset>, Sep. 2021.
- [40] S. Sharma, A. Shedsale, and R. R. Sharma, "Multivariate Fast Iterative Filtering Based Automated System for Grasp Motor Imagery Identification Using EEG Signals," *Int J Hum Comput Interact*, vol. 40, no. 23, pp. 7915–7923, Dec. 2024, doi: 10.1080/10447318.2023.2280327.
- [41] "Train Test Split – How to split data into train and test for validating machine learning models?," <https://www.machinelearningplus.com/machine-learning/train-test-split/>, [Online]. Available: <https://www.machinelearningplus.com/machine-learning/train-test-split/>
- [42] A. A. Grigoryan, "Understanding VGG Neural Networks: Architecture and Implementation," <https://thegrigorian.medium.com/>.
- [43] Sarang Narkhede, "Understanding AUC - ROC Curve," <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5>, Jun. 2018.
- [44] D. Choi, C. J. Shallue, Z. Nado, J. Lee, C. J. Maddison, and G. E. Dahl, "On Empirical Comparisons of Optimizers for Deep Learning," Oct. 2019, [Online]. Available: <http://arxiv.org/abs/1910.05446>
- [45] Driss Lamrani, Bouchaib Cherradi, Oussama El Gannour, Mohammed Amine Bouqentar, and Lhoussain Bahatti, "Brain Tumor Detection using MRI Images and Convolutional Neural Network," (IJACSA) *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 7, pp. 452–460, 2022.
- [46] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J Big Data*, vol. 6, no. 1, p. 60, Dec. 2019, doi: 10.1186/s40537-019-0197-0.
- [47] A. A. Asiri et al., "Brain Tumor Detection and Classification Using Fine-Tuned CNN with ResNet50 and U-Net Model: A Study on TCGA-LGG and TCIA Dataset for MRI Applications," *Life*, vol. 13, no. 7, p. 1449, Jun. 2023, doi: 10.3390/life13071449.
- [48] E. Agliari, F. Alemanno, M. Aquaro, and A. Fachechi, "Regularization, early-stopping and dreaming: A Hopfield-like setup to address generalization and overfitting," *Neural Networks*, vol. 177, p. 106389, Sep. 2024, doi: 10.1016/j.neunet.2024.106389.
- [49] S. Zakariya and M. Jamil, "Unsupervised Content based Image Retrieval at Different Precision Level by Combining Multiple Features," *J Phys Conf Ser*, vol. 1950, p. 12059, Aug. 2021, doi: 10.1088/1742-6596/1950/1/012059.
- [50] SM Zakariya and M. S. Umar, "Self-Attention Augmented Wasserstein Generative Adversarial Network-based Detection of Brain Alzheimer Disease Using MRI," *International Research Journal of Multidisciplinary Scope*, vol. 06, no. 01, pp. 1317–1327, 2025, doi: 10.47857/irjms.2025.v06i01.02645.
- [51] S. Bansal, R. S. Jadon, and S. K. Gupta, "A Robust Hybrid Convolutional Network for Tumor Classification Using Brain MRI Image Datasets," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 4, pp. 576–584, 2024, doi: 10.14569/IJACSA.2024.0150459.
- [52] Z. Rasheed et al., "Brain Tumor Classification from MRI Using Image Enhancement and Convolutional Neural Network Techniques," *Brain Sci*, vol. 13, no. 9, Sep. 2023, doi: 10.3390/brainsci13091320.

- [53] A. J. Alkhatib et al., “Diagnosing Brain Tumors from MRI images through a Multi-Fused CNN with Auxiliary Layers,” Sustainable Machine Intelligence Journal, vol. 6, Mar. 2024, doi: 10.61356/SMIJ.2024.66102.