MRI-based Brain Tumor Classification using Transfer Learning: A Comparative Analysis

Anil Poudyal
London Metropolitan University
Holloway Rd, London, UK

Niruta Devkota
London Metropolitan University
Holloway Rd, London, UK

Gayatri Budha
London Metropolitan University
Holloway Rd, London, UK

ABSTRACT
Diagnosis of brain tumours quickly and accurately is crucial for effective treatment and better patient outcome. Although the traditional diagnosis procedures like MRI, CT scan and biopsies are useful, they pose significant human inconsistencies. This research explores five of the most popular transfer learning techniques in CNN to find an optimal model for classification of brain tumours. Pre-trained models-VGG16, ResNet50, DenseNet121, InceptionResNetV2 and InceptionV3 have been used to find the optimal model for this task. The used dataset includes 7023 MRI images divided into four categories: glioma, meningioma, pituitary tumour, and no tumour. Experimental results highlight DenseNet121's superior performance, achieving a validation accuracy of 94%, precision, recall, and F1-score of 0.94, outperforming other models. This study shows that deep learning and transfer learning can significantly improve the accuracy and efficiency of medical image analysis, leading to better healthcare outcomes.

Keywords
Deep Learning, Medical Image Analysis, Brain Tumor Classification, Convolutional Neural Network (CNN), Transfer Learning, Diagnostic Accuracy, Pretrained CNN Models, VGG16, ResNet, DenseNet, InceptionResNetV2

1. INTRODUCTION
The most used methods for diagnosing tumors are MRI, CT scan and biopsy. The process is solely overseen by human which are not consistent at times. Machine learning can greatly assist in diagnosing tumors, making it easier to develop treatment plans and helping doctors detect hidden features [1]. Use of convolutional neural networks (CNN) for medical image analysis have shown significant results in the past. With the combination of transfer learning approach of model building the results are found to be more accurate due to large architectural size. Some of the most popular transfer learning techniques are ResNet50, LeNet and GoogleNet; use of these pre-trained models has shown significant improvements in the diagnostic process. With the introduction of computer aided diagnosis of medical conditions, efficiency rates and pace of diagnosis have been improved. And with the advancement in medical image analysis, the procedure of diagnosis has taken a leap forward. With the involvement of artificial intelligence in healthcare sector the efficiency of diagnosis has increased, and the cost of diagnosis has come down significantly [2]. The potential of machine learning and artificial intelligence is high in the healthcare system in the diagnosis process due to both economic and critical factors. The economic factor of computer aided diagnosis will help regions that are backward in economic perspective.

This study will deep dive into parameters, hyperparameters, and augmentation techniques to identify the optimal configuration for diagnosing brain tumors using MRI scans. Through the utilization of transfer learning, the study anticipates an improvement in diagnostic accuracy, positively impacting patient outcomes.

2. LITERATURE REVIEW

2.1 Evolution of Machine in CAD
The integration of machine learning and artificial intelligence into CAD brought a transformative force to the sector, witnessing a significant shift in the 21st Century. Neural networks became more prevalent, enabling CAD systems to manage more complex image analysis tasks. From this integration onwards, Convolution Neural Networks (CNNs) emerged as the strategic choice for most image analysis tasks, making significant contributions to various fields. The medical sector has been a prominent beneficiary, with major research efforts ongoing, especially in areas like detecting COVID-19 and MRI analysis [3].

2.2 Transfer Learning in MRI Analysis
Transfer learning has emerged as the powerful paradigm in the landscape of medical image analysis which is ever-changing and dynamic [4]. Transfer learning leverages the knowledge gained from pretraining on source tasks to optimize the performance of a model on targeted tasks. Transfer learning solves the biggest difficulty int this sector, i.e. lack of labeled dataset. Transfer learning operates on the principle that features learned from related tasks can be transformed and adapted to a target task. The knowledge acquired from general image recognition tasks is transferred to perform a specific task. This approach holds an outstanding position by enhancing the efficiency and accuracy of models.

Transfer learning brings improvement in generalization by enabling models to understand and adapt to different dynamics of data. The major advantage of transfer learning models is that they are trained with a diverse dataset which helps the models to capture generic features. One of the significant challenges in this sector is the scarcity of labeled datasets, and transfer learning effectively addresses this issue. This generalization capability of models helps to boost the performance [5].

Researchers in Indonesia successfully utilized the VGG16 architecture for tuberculosis detection, achieving an impressive accuracy of 99.76%. The model was trained and validated using chest X-rays. Notably, the researchers identified the pivotal role of batch size in enhancing accuracy, determining an optimal batch size of 50. Any increase beyond 50 resulted in decreased performance, suggesting a potential bias in the learning process [6]. Although the figures were impressive for the chosen dataset the developed model was notably slow, this slowness was a result of the number of parameters and depth of architecture.

ResNet50 has shown significant results in medical image analysis particularly in tasks requiring complex feature extraction and pattern recognition. Studies have shown that leveraging pre-trained ResNet50 weights enhances the performances of models in various medical image-related tasks [7]. A 2022 study showed significant results in classification tasks on COVID-19 (iNat2021 Mini Dataset) images. The Fine-tuned model averaged around 99% on each accuracy figure including F1-Score [8].

InceptionNet was developed by Google in the year 2014 and has been in use for face recognition, object detection and image classification tasks since then. InceptionNet has been proven to have excelled at capturing hierarchical features of radiological images, enabling the network to identify subtle details within the scans [9]. The action of reducing dimensions before larger convolution has been proven to improve efficiency. This ultimately reduces the computational load marking it one of the resource efficient models. InceptionNet can address challenges such as computational efficiency, model depth, and multi-scale features [10].

DenseNet emerges as a formidable candidate in the realm of medical image analysis as its unique architectural design and inherent advantages align with the demands of efficient analysis of medical images. It has collected significant attention in the field of deep learning for its performance in medical image analysis tasks [11]. DenseNet has been used in various research areas of medical image analysis but faces significant challenges due to massive amounts of medical data and their intricate details like labels with lesion conditions, and clinical diagnosis reports cannot be fully utilized. To add to the advantage of DenseNet is its depth the pretrained model of DenseNet has depth of 242 making it the model with the most depth in this study Table 1. A study done in 2024 has shown that there is slight edge to the models that do have deeper depth [12].

Table 1: Performance comparison of pre-trained models

Model	Top-5 Accuracy	Parameters	Depth
VGG16	90.10%	138.4 M	16
ResNet50	92.10%	25.6 M	107
InceptionV3	93.70%	23.9 M	189
DenseNet121	92.30%	8.1 M	242
InceptionResNetV2	95.30%	55.9M	449

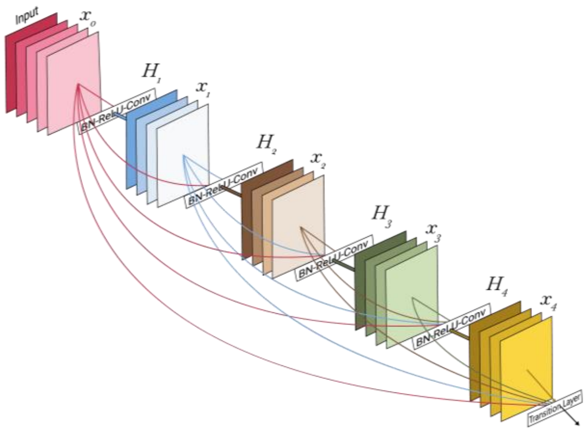

Figure 1: DenseNet Architecture

3. METHODOLOGY

3.1 Data preprocessing

The dataset used in this study data is collected from Kaggle's "Brain Tumor MRI Dataset" which was carefully curated by Masoud Nickparvar. The dataset is split into training and testing datasets comprising 7023 human brain MRI images in total with 5712 instances of images in training and 1311 in testing. The images are classified into four classes: glioma, meningioma, no tumour, and pituitary. Each class has 1300 to 1500 images which is well balanced for training and testing machine learning models without requiring extensive data balancing techniques. The images in the dataset were of diverse sizes. Therefore, all the images were resized to 224x224 pixels during the training of the models. Different augmentation techniques like rescaling, rotation, shearing, zooming, and horizontal flipping were applied to increase the dataset size and create variation. The sample images for each class are presented in Figure 2.

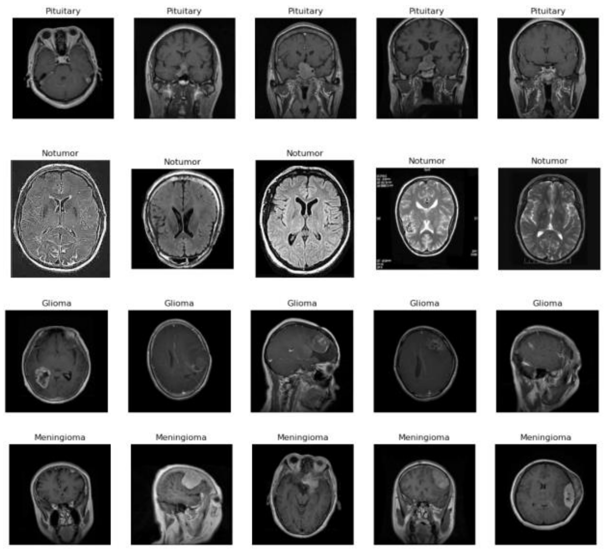

Figure 2: Sample Images

3.2 Implementation Details

This study has explored five of the most popular pre-trained models i.e., VGG16, ResNet50, DenseNet121, InceptionResNetV2, and InceptionV3 as base models for the classification task. The image size of 224x224x3 has been used.

The initial layers of pre-trained models are frozen for retaining the featured learned from the ImageNet dataset and allowing fine-tuning only the later layers to adapt to the specific task of classification of brain tumors. Introduction of batch normalization is done immediately after inputs are fed to the pre-trained base model to enhance the training stability. Furthermore, the dense layer with 128 neurons has been introduced with L1 and L2 regularization for the prevention of overfitting. Across all the models Adam optimizer with varied learning rate and categorical cross-entropy as loss function has been used. The performance of models has been evaluated after 50 epochs of training with early stopping.

The models have been trained on machine with following specifications.

Table 2: Specifications Detail

Title	Description
Platform	Windows (11)
Processor	i7-9750H
Memory	16 GB
GPU	Nvidia GTX G-force 1660Ti

3.3 Evaluation Metrics

The evaluation of models for classification task is done through accuracy, prevision, recall and F1-score. Following are the mathematical representations of respective evaluation metrices:

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 - Score = \frac{2 * Precision * Recall}{Precision + Recall}$$

Here,

TP= True Positive *TN= True Negative*

FN = False Negative *FP = False Positive*

4. Experimental Results

The five most popular pre-trained models were evaluated with the dataset of 7023 human brain MRI images with 5712 instances of images in training and 1311 in testing. All five models were trained in similar configurations for better comparisons. Validation Accuracy, Validation Loss, Precision, Recall and F1-Score were used as performance metrics for comparison. These accuracy metrices allowed to assess the effectiveness thoroughly. And ensured that models did perform well in both identifying and minimizing errors during classification.

The DenseNet model achieved the highest validation accuracy of 0.94 and its precision, recall and F1-Score were all 0.94. DenseNet also had the lowest validation loss of 0.27. This indicates that DenseNet is the most effective model for classifying brain tumors in the experiment done using "Brain Tumor MRI Dataset". Following DenseNet, InceptionNet achieved a validation accuracy of 0.91 with corresponding precision, recall and F1-scores of 0.91.

Number of neurons in dense layers was seen to have a significant impact on the model's performance. Model's validation loss was seen to be improved significantly. The performance improvement was seen to be improved when number of neurons were increased from 64 to 128 but on further increment, models' performance metrics did not indicate any significant improvement. This result was observed when the model was trained over 40 epochs with a variable learning rate of 0.001 and drop rate of 0.25.

Based on experimental results, DenseNet was found to be the most effective model on this dataset. While other 3 pre-trained models had comparable performance, ResNet50 did not perform well enough in this setting.

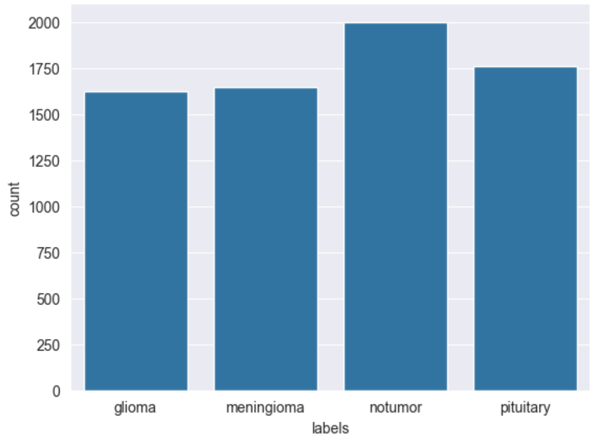

Figure iii: Distribution of Dataset

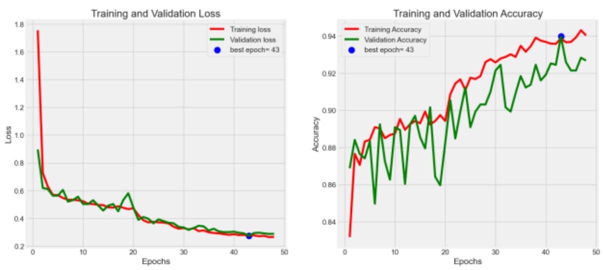

Figure iv: Training and Validation Loss and Accuracy for DenseNet Model over 50 Epochs.

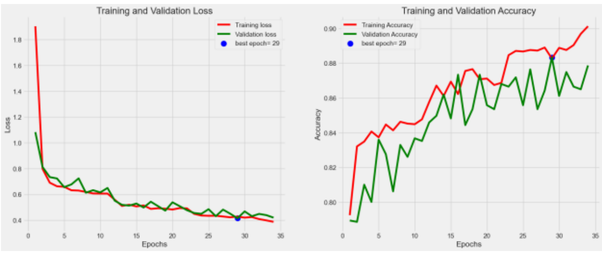

Figure v: Training and Validation Loss and Accuracy for VGG16 Model over 50 Epochs.

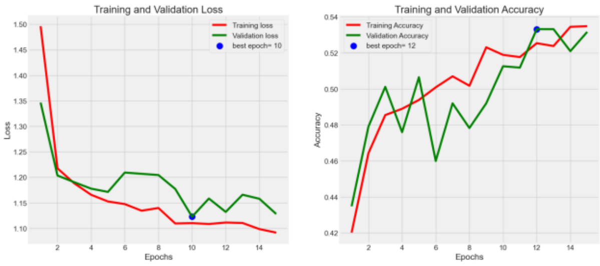

Figure vi: Training and Validation Loss and Accuracy for ResNet50 Model over 50 Epochs.

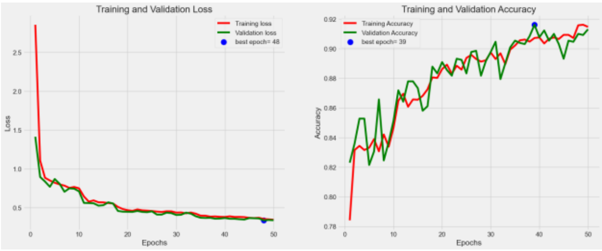

Figure vii: Training and Validation Loss and Accuracy for InceptionV3 Model over 50 Epochs.

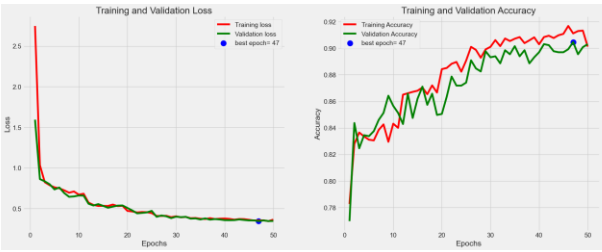

Figure viii: Training and Validation Loss and Accuracy for InceptionResNetV2 Model over 50 Epochs.

5. CONCLUSION

This study conducted a comparative analysis of five popular pre-trained models—VGG16, ResNet50, DenseNet121, InceptionResNetV2, and InceptionV3—using transfer learning for the classification of brain tumors based on MRI dataset. Among these models, DenseNet121 demonstrated superior performance with the highest validation accuracy (94%) and precision, recall, and F1-scores of 0.94, making it the optimal choice for brain tumor classification. The results obtained showed that the deep learning techniques have greater efficacy advantage on medical image analysis over other traditional machine learning techniques.

While DenseNet121 performed best in this study, there are several opportunities for future research and improvements. First, expanding the dataset to include more diverse MRI images and multi-class tumor classifications could further enhance the model's generalization and robustness. Additionally, integration with other diagnostic tools like CT scans or PET images could enable multi-modal analysis, improving diagnostic precision.

Exploring unsupervised or semi-supervised learning techniques may address the challenge of limited labeled medical datasets, enabling models to learn from unlabeled data. Another promising direction involves the development of lightweight models optimized for deployment in resource-constrained environments, particularly in regions with limited access to advanced medical facilities. Finally, incorporating explainability techniques such as Grad-CAM could provide visual explanations of the model's decisions, enhancing transparency and trust in AI-driven medical diagnoses.

By further refining these models and addressing current limitations, future research could revolutionize brain tumor diagnostics, making it faster, more accurate, and accessible worldwide.

Table 3: Experimental Results

Model Architecture	Validation Accuracy	Validation Loss	Precision	Recall	Precision
DenseNet121	0.94	0.27	0.94	0.94	0.94
InceptionNet	0.91	0.33	0.91	0.91	0.91
InceptionResNetV2	0.90	0.34	0.90	0.90	0.90
VGG16	0.88	0.41	0.88	0.88	0.88
ResNet50	0.51	1.12	0.51	0.51	0.49

6. REFERENCES

[1] A. K. Philip, B. A. Samuel, S. Bhatia, S. A. M. Khalifa and H. R. El-Seedi, "Artificial Intelligence and Precision Medicine: A New Frontier for the Treatment of Brain Tumors," *Life*, vol. 13, no. 1, p. 24, 2022.

[2] A. A. Kuwaiti, K. Nazer, A. Al-Reedy, S. Al-Shehri, A. Al-Muhanna, A. V. Subbarayalu, D. A. Muhanna and F. A. Al-Muhanna, "A Review of the Role of Artificial Intelligence in Healthcare," *Journal of Personalized Medicine*, vol. 13, no. 6, p. 951, 2023.

[3] M. K. A. Kumar, R. Akthar, M. S. Vignesh and M. Rahuman, "Covid-19 Detection using Deep Learning," *International Journal of Research Publication and Reviews*, vol. 4, no. 5, pp. 4723-4727, 2023.

[4] H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros and T. Ganslandt, "Transfer learning for medical image classification: a literature review," *BMC Medical Imaging*, vol. 22, no. 69, 2022.

[5] G. Blanchard, A. A. Deshmukh, Ü. Dogan, G. Lee and C. Scott, "Domain Generalization by Marginal Transfer Learning," *Journal of Machine Learning Research*, vol. 22, pp. 1-55, 2021.

[6] S. Aulia and S. Hadiyoso, "Tuberculosis Detection in X-Ray Image Using Deep Learning," *Jurnal Ilmiah Teknik*

Elektro Komputer dan Informatika (JITEKI), vol. 8, no. 2, pp. 290-297, 2022.

[7] S. Pattanaik, S. Behera, S. K. Majhi and P. K. Dwibedy, "Stacked BiLSTM with ResNet50 for Medical Image Classification," in *2022 IEEE Region 10 Symposium (TENSYMP)*, Mumbai, 2022.

[8] M. B. Hossain, S. M. H. S. Iqbal , M. M. Islam, M. N. Akhtar and I. H. Sarker, "Transfer learning with fine-tuned deep CNN ResNet50 model for classifying COVID-19 from chest X-ray images," *Informatics in Medicine Unlocked*, vol. 30, no. 2022, p. 100916, 2022.

[9] A. Agrawal, "Classification and Detection of Brain Tumors by Aquila Optimizer Hybrid Deep Learning Based Latent Features with Extreme Learner," in *ITM Web of Conferences*, New Delhi, 2023.

[10] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," *ArXiv,* vol. abs/1512.00567, 2015.

[11] T. Zhou, X. Ye, H. Lu, X. Zheng, S. Qiu and Y. Liu, "Dense Convolutional Network and Its Application in Medical Image Analysis," *BioMed Research International 2022,* vol. 2022, no. 1, p. 2384830, 2022.

[12] X. Li, "Exploring the effect of depth and width of CNN models on," in *Proceedings of the 4th International Conference on Signal Processing and Machine Learning*, Boston, 2024.