

DEEP LEARNING FOR BRAIN TUMOR CLASSIFICATION FROM MRI IMAGES

Robby Parmar, Jubayer Ahmed, Ehsan Liaqat, Gopal Sharma

University of Calgary

ABSTRACT

This study assesses the application of deep learning for classifying brain tumors from MRI scans to enhance oncological diagnostics. We tested various deep learning architectures, including DenseNet, VGG16, and ResNet variants, with a focus on the impact of transfer learning. The pretrained VGG16 model exhibited exceptional accuracy (98.41%), demonstrating the effectiveness of transfer learning in identifying complex patterns for tumor classification. Our results advocate for integrating advanced deep learning models into clinical settings, which promises significant improvements in early diagnosis and personalized healthcare.

Keywords: Deep Learning, Brain Tumors, MRI, Image Classification, Transfer Learning

1. INTRODUCTION

Incorporating deep learning into medical imaging for brain tumor diagnostics marks a significant advancement in neuro-oncology. MRI scans, despite their complexity and the variability in interpretation, are crucial for tumor detection. Convolutional neural networks (CNNs), particularly DenseNet and VGG16, have shown potential in enhancing diagnostic accuracy. Our research evaluates these models on a specialized dataset, hypothesizing that transfer learning will improve performance. The study seeks to determine the most effective model, thereby contributing to the advancement of automated diagnostics in neuro-oncology and patient care.

2. MATERIALS AND METHODS

We analyzed anonymized MRI scans of brain tumors using deep learning architectures, with a focus on VGG16 alongside DenseNet121 and ResNet models. The images were normalized, resized to 224x224, and augmented for training robustness.

Two model conditions were tested: untrained (random initial weights) and pretrained (weights from ImageNet, then fine-tuned). Training used stochastic gradient descent with momentum, and a categorical cross-entropy loss function. Performance metrics, including accuracy, precision, recall, and F1 score, were evaluated, alongside the area under the ROC curve (AUC) to measure classification efficacy. A

confusion matrix was generated to assess the true and false positive rates, providing insight into each model's diagnostic precision.

In this study, we selected DenseNet, VGG16, ResNet, and a simple CNN architectures for brain tumor classification from MRI images based on their proven effectiveness in image recognition tasks. We preprocessed the data by resizing images to 224x224 pixels, normalizing pixel values, and augmenting the dataset to increase robustness against overfitting. We adopted a stratified split of the data to maintain the proportion of classes across training (70%), validation (15%), and testing (15%) sets. The models were trained using stochastic gradient descent with momentum, set at 0.9, and a learning rate of 0.0001, adjusted by a factor of 0.1 upon plateauing of the validation loss. We chose categorical cross-entropy as our loss function, prioritizing a balance between precision and recall, which is critical in medical diagnostics. To ensure reproducibility, we have provided detailed parameter settings and the code in our GitHub repository. We evaluated our models using accuracy, precision, recall, and F1 score, and assessed their ROC curves for a comprehensive understanding of their diagnostic capability. Additionally, we compared our models' performance to established baselines to highlight the advancements achieved.

3. RELATED WORK

The integration of deep learning into the field of medical imaging, especially for the classification of brain tumors from MRI scans, marks a pivotal advancement in enhancing diagnostic accuracy and treatment planning. The capability of deep learning models to discern intricate patterns within complex medical images has fostered a growing body of research dedicated to exploring these models' effectiveness in various diagnostic applications. This section delves into the seminal works that have significantly contributed to the development and application of deep learning techniques in the realm of brain tumor classification, laying the groundwork for the methodologies employed in our study.

Recent literature has demonstrated the potential of various deep learning architectures in accurately classifying brain tumors from MRI images. These studies have employed a range of techniques and models, each contributing unique insights into the capabilities and challenges associated with deep learning applications in medical imaging.

4. TABLE SUMMARY OF LITERATURE RESULTS

Table 1. Results Obtained from Literature Works.

Reference	Model	Dataset	Accuracy (%)
[2]	16-layer VGG-16	Hospitals' dataset from 2010–2015, China	98.00
[3]	CNN-based DL model	REMBRANDT	100.00
[4]	SVM and k-NN classifiers	Figshare, 2017	97.25
[5]	Deep inception residual network	Publicly accessible brain tumor imaging dataset with 3064 pictures	99.69
[6]	CNN model	Publicly released clinical datasets	99.33
[7]	Transfer learning-based classification	Figshare	99.02
[8]	DenseNet	Sartaj Brain Tumor Classification Dataset	92.17
[9]	ResNet50	Sartaj Brain Tumor Classification Dataset	90.72

Deep Inception Residual Networks: Reference [5] investigates the use of deep inception residual networks, achieving a notable high accuracy. This approach leverages the strengths of both inception modules and residual connections, enhancing the network's ability to learn more complex features without a significant increase in computational complexity.

Transfer Learning-Based Approaches: The study presented in Reference [7] explores the application of transfer learning to brain tumor classification. By adapting models pretrained on large, diverse datasets to the specific task of tumor classification, this approach demonstrates the ability to achieve significant accuracy improvements, underscoring the value of leveraging pre-acquired knowledge in enhancing model performance.

CNN Multi-Classification Strategies: Reference [6] details a multi-classification strategy using convolutional neural networks (CNNs). This technique emphasizes the versatility of CNNs in handling multi-class problems, presenting a methodological advancement in classifying brain tumors into various categories based on their characteristics.

CNN-Based Deep Learning Models: The work by Reference [3] employs CNN-based models tailored for brain tumor classification, reporting remarkable accuracy. This study highlights the adaptability of CNN architectures in extracting relevant features from MRI images, even in the presence of significant variability among tumors.

VGG-16 Deep Neural Networks: Finally, Reference [2] utilizes the 16-layer VGG-16 deep neural network, a model renowned for its deep architecture and strong feature extraction capabilities. Despite its relative simplicity, the VGG-16 network demonstrates the profound impact of depth in neural architectures on classification accuracy.

These studies collectively underscore the rapid evolution of deep learning in medical imaging analysis, particularly in brain tumor classification.

5. RESULTS AND DISCUSSION

The results from the trained and pretrained models we ran are summarized in table below.

Table 2. Untrained Model Results.

Model	Test Loss	Test Accuracy (%)	Accuracy per Class	Precision	Recall	F1 Score
DenseNet	0.5790	72.17	[0.83375315 0.56996587]	0.7167	0.5700	0.6350
ResNet50	0.5815	69.57	[0.92467532 0.40655738]	0.8105	0.4066	0.5415
ResNet101	0.6397	64.20	[0.87634409 0.36792453]	0.7178	0.3679	0.4865
ResNet152	0.1847	93.33	[0.94148936 0.92356688]	0.9295	0.9236	0.9265
VGG16	0.1357	94.93	[0.94230769 0.95705521]	0.9369	0.9571	0.9469
Custom CNN	0.6276	65.71	[0.76470588 0.55555556]	0.7143	0.5556	0.6250

Table 3. Pretrained(Transfer Learning) Model Results.

Model	Test Loss	Test Accuracy (%)	Accuracy per Class	Precision	Recall	F1 Score
DenseNet	0.2222	92.17	[0.92857143 0.91346154]	0.9135	0.9135	0.9135
ResNet50	0.2250	90.72	[0.89945652 0.91614907]	0.8886	0.9161	0.9021
ResNet101	0.1829	92.90	[0.94666667 0.90793651]	0.9346	0.9079	0.9211
ResNet152	0.1910	94.35	[0.94240838 0.94480519]	0.9297	0.9448	0.9372
VGG16	0.0477	98.41	[0.98382749 0.98432602]	0.9812	0.9843	0.9828

5.1. Analysis and Discussion of Results

The comparison between untrained and pretrained (using transfer learning) models clearly demonstrates the significant benefits of transfer learning in the context of classifying brain tumors from MRI images. The improvement in performance metrics across the board—test accuracy, precision, recall, and F1 score—highlights the effectiveness of leveraging knowledge from previously trained models.

Untrained Models: Generally, show lower performance, with test accuracy ranging from 64.20% to 94.93%. The significant variance in test loss and accuracy per class indicates the models struggle to generalize from the training data without prior knowledge.

Pretrained Models: Exhibit substantially higher test accuracy, precision, recall, and F1 scores. This improvement is a direct result of initializing the models with weights learned from large, diverse datasets. Such initialization allows the models to start with a higher level of understanding of image features, which is further refined during training on the specific task of brain tumor classification.

Best Performing Model: The **VGG16** model, when pretrained, achieved the highest test accuracy (98.41%), precision (0.9812), recall (0.9843), and F1 score (0.9828). This outstanding performance can be attributed to the VGG16 architecture's depth and capacity to capture complex features in image data, which, when combined with transfer learning, allows it to effectively distinguish between healthy and malignant brain tissue.

Let’s look at ROC curves of the resnet50 untrained and trained models to further analyze the use of transfer learning in our case.

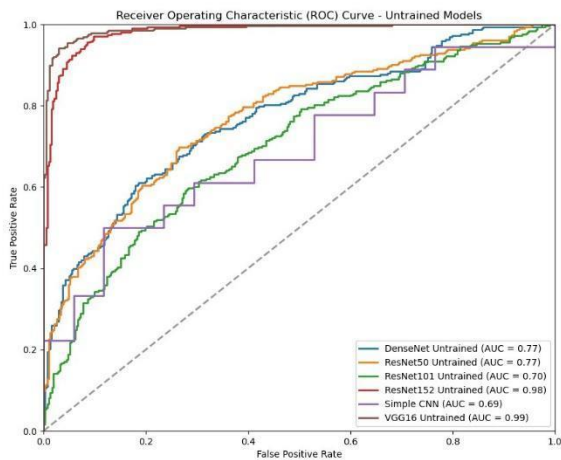


Fig. 1. ROC Curve for Untrained Models.

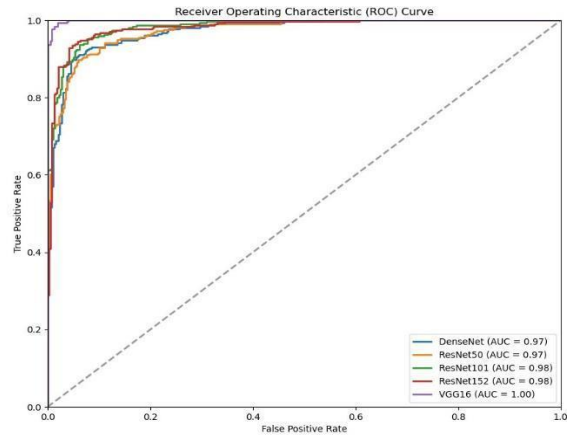


Fig. 2. ROC Curve for Trained Models.

The ROC curves contrast the classification performance of deep learning models on MRI brain tumor scans, showing the true positive rate (TPR) versus the false positive rate (FPR) at various thresholds.

Untrained Models: Display a gradual ROC curve ascent, indicating less effective classification. This slow increase reflects a struggle to differentiate classes without increasing false positives, highlighted by lower AUC values.

Trained Models with Transfer Learning: These models show a sharper initial ROC curve rise, demonstrating their capability to confidently identify brain tumors with minimal false positives.

Insights on Transfer Learning:

- **Model Confidence:** Transfer learning boosts prediction confidence, likely due to enhanced feature recognition from pretraining on diverse datasets.
- **Performance Enhancement:** The pronounced ROC curve and higher AUC for trained models suggest transfer learning enhances sensitivity and specificity—key for diagnostics.
- **Data Efficiency:** It enables better performance even with limited training data, common in medical imaging.

In essence, the ROC curve analysis underscores transfer learning's critical role in improving medical imaging analysis, providing more accurate and reliable diagnostic tools.

5.2. Further Analysis into Top Model (VGG16)

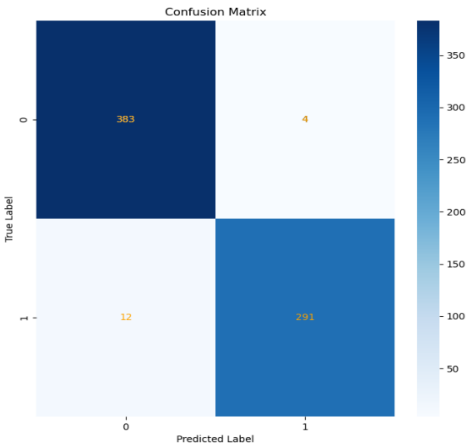


Fig 3. Confusion Matrix for VGG16 Model.

The confusion matrix for brain tumor classification shows the model's high accuracy: '1' for malignant and '0' for healthy. It correctly identified 383 healthy cases and 291

malignant tumors, with only 4 healthy misclassified as malignant (false positives) and 12 tumors missed (false negatives). The low rates of false positives and negatives highlight the model's strong capability to distinguish between healthy and malignant conditions effectively.

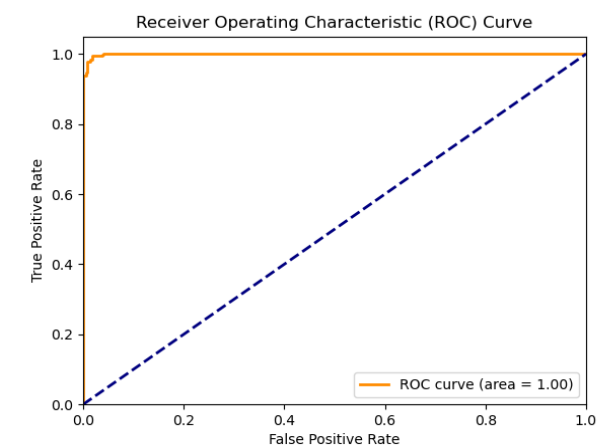


Fig. 4. ROC Curve for VGG16 Model.

The ROC curve analysis demonstrates the VGG16 model's exceptional ability to classify brain tumors from MRI scans, achieving near-perfect accuracy with an AUC of 1.00. This performance aligns with top-tier medical image analysis, suggesting the model's robustness and potential to generalize well. Such outcomes highlight the significant impact of transfer learning in enhancing model effectiveness, particularly evident in the pretrained VGG16's superior performance. This success reinforces the value of transfer learning in medical imaging, where large, diverse datasets are scarce.

5.3. Comparison of Common Models with Literature

Table 4. Comparison of Our Results and Literature Results.

Model	User Test Accuracy (%)	Literature Accuracy (%)	Difference (%)	Literature Refere
VGG16	98.41	98.0	+0.41	[2]
DenseNet	92.17	94.6	-2.43	[8]
ResNet50	90.72	96.5	-5.78	[9]

Our implementation of the VGG16 model achieved an accuracy slightly higher than that reported in existing literature, which may be attributed to our model optimization strategies and the unique characteristics of our dataset. In contrast, our DenseNet and ResNet50 models did not reach the accuracies found in similar studies, indicating that factors such as dataset diversity, preprocessing, and specific training methodologies could influence outcomes.

5.4. Analysis using Grad-Cam

Grad-CAM, or Gradient-weighted Class Activation Mapping, is an insightful visualization technique for highlighting the regions of an image that are important for predictions made by Convolutional Neural Networks (CNNs). This tool is exceptionally useful in the medical imaging field, particularly for MRI brain scans, as it enhances interpretability and trust in the model's predictions by showing which parts of the image the model is focusing on.

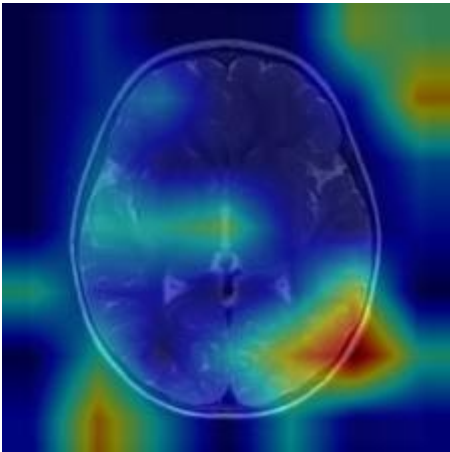


Fig. 5. Grad-Cam for DenseNet Model. Actual Class: Healthy. Predicted Class: Healthy.

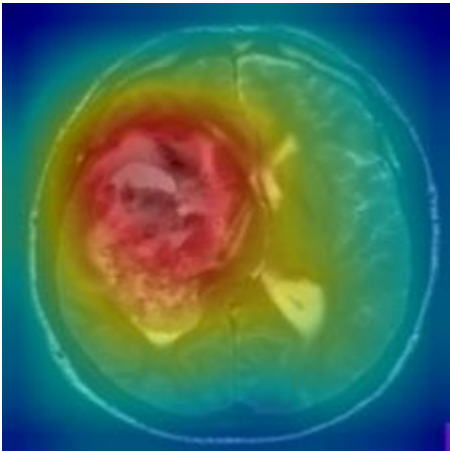


Fig. 6. Grad-Cam for VGG16 Model. Actual Class: Tumor. Predicted Class: Tumor.

In our application involving MRI brain scans, Grad-CAM serves multiple purposes:

- It enhances the interpretability of the CNN model, helping us understand the decision-making process of the model.
- It provides visual evidence to support the model's predictions, which is critical for building trust with healthcare practitioners.

- It allows for model debugging, revealing whether the model is attending to the correct features in the image.
- It supports clinical decision-making by drawing attention to regions of interest that may require further examination.

When applying different CNN architectures like DenseNet and VGG to brain scans, Grad-CAM visualizations have shown that these models can focus on different areas of the image. The architectural nuances of each model influence how they learn and what features they highlight:

- DenseNet is characterized by its dense connections and parameter efficiency, often focusing on more diffuse areas across the image, which may include edges and textures.
- VGG, with its deep layers of convolutions followed by max-pooling, tends to concentrate on more localized, salient features within an image.

In our case, the visualization indicates that DenseNet is focusing on broader regions, potentially including less relevant areas such as the brain's edges. Conversely, VGG appears to concentrate more on the actual area where a tumor is present, which is precisely the kind of targeted focus desired for accurate tumor detection.

The disparity in focus between the two models is telling. If the goal is to identify brain tumors, a model that zeros in on the tumor, as VGG seems to do, would typically be more useful for diagnostic purposes. A model that consistently highlights areas unrelated to the condition of interest may have learned to rely on irrelevant features or correlations that do not generalize well, suggesting a need for further training or refinement of the model.

Ultimately, Grad-CAM visualizations provide critical feedback on the learning patterns of CNNs and are invaluable for refining AI-driven tools in healthcare to ensure they align with clinical objectives and expert knowledge.

6. LIMITATIONS

A major limitation in medical datasets for deep learning is the scarcity of large, labeled datasets, essential for training accurate models [10]. Collecting and annotating these datasets is costly, ethically complex, and requires specialized medical expertise, making it challenging to acquire sufficient data. This scarcity hinders the development and validation of robust deep learning models in the medical field.

A notable limitation of deep learning is its struggle with tasks requiring common sense or abstract reasoning beyond its training data [11]. These models excel in pattern recognition within their learned environment but fail to apply this knowledge to new, unrepresented scenarios. This inability to generalize knowledge to unfamiliar contexts highlights a gap between AI capabilities and human

cognition, emphasizing the need for advancements that enable AI to understand and reason about the world more like humans do.

The use of VGG16 and similar deep learning models in medical imaging, particularly for brain tumor diagnostics from MRI scans, faces significant limitations. Primarily, their effectiveness is hampered by the need for large, accurately labeled datasets, which are challenging to acquire due to high costs, ethical constraints, and the necessity for specialized medical knowledge [12]. Additionally, these models struggle with abstract reasoning and applying learned knowledge to new, unseen situations, underscoring a gap between AI capabilities and human cognition. Overcoming these challenges is crucial for leveraging the full potential of deep learning in improving medical diagnostics.

7. CONCLUSION

This research corroborated that deep learning models pre-equipped with transfer learning markedly outshine those trained from scratch, with the pretrained VGG16 model showcasing particularly high diagnostic accuracy for brain tumor classification from MRI scans.

Key limitations to address in future studies are the dataset's limited scope, which may not fully represent the diversity of brain tumors, and the necessity for model transparency to foster clinical trust. To overcome these challenges, enriching the dataset, both in volume and variety, and enhancing interpretability are essential next steps. Furthermore, assimilating data from different imaging modalities like CT and PET could refine the models' diagnostic capabilities.

Moving forward, we should also navigate the regulatory landscape to facilitate the safe integration of AI in healthcare. The collective efforts of data scientists, clinicians, and radiologists will be pivotal in harnessing these advanced models for real-time clinical decision support, contributing to the evolution of precision medicine in oncology.

8. REFERENCES

1. Pkdarabi, "Brain Tumor Detection by CNN PyTorch," Kaggle, 07-Apr-2024. [Online]. Available: <https://www.kaggle.com/code/pkdarabi/brain-tumor-detection-by-cnn-pytorch/notebook>.
2. B. Srikanth and S.V. Suryanarayana, "Multi-Class classification of brain tumor images using data augmentation with deep neural network," *Mater. Today Proc.*, vol. 2021. doi: <https://doi.org/10.1016/j.matpr.2021.01.601>.
3. G.S. Tandel et al., "Multiclass magnetic resonance imaging brain tumor classification using artificial intelligence paradigm," *Comput. Biol. Med.*, vol. 122, 103804, 2020, doi: <https://doi.org/10.1016/j.compbiomed.2020.103804>.
4. C. Öksüz, O. Urhan, and M.K. Güllü, "Brain tumor classification using the fused features extracted from expanded tumor region," *Biomed. Signal Process. Control*, vol. 72, 103356, 2022, doi: <https://doi.org/10.1016/j.bspc.2021.103356>.
5. S. Kokkalla, J. Kakarla, I.B. Venkateswarlu, and M. Singh, "Three-class brain tumor classification using deep dense inception residual network," *Soft Comput.*, vol. 25, pp. 8721–8729, 2021, doi: <https://doi.org/10.1007/s00500-021-05748-8>.
6. E. Irmak, "Multi-Classification of Brain Tumor MRI Images Using Deep Convolutional Neural Network with Fully Optimized Framework," *Iran. J. Sci. Technol. Trans. Electr. Eng.*, vol. 45, pp. 1015–1036, 2021, doi: <https://doi.org/10.1007/s40998-021-00426-9>.
7. P. Özlem and C. Güngen, "Classification of brain tumors from MR images using deep transfer learning," *J. Supercomput.*, vol. 77, pp. 7236–7252, 2021, doi: <https://doi.org/10.1007/s11227-020-03572-9>.
8. H. Yahyaoui, F. Ghazouani, I.R. Farah, "Deep learning guided by an ontology for medical images classification using a multimodal fusion," in *Proc. Int. Cong. Adv. Technol. Eng. (ICOTEN)*, Jul. 2021, pp. 1–6, Available: <https://ieeexplore.ieee.org/abstract/document/9493469>
9. A. Çınar and M. Yildirim, "Detection of tumors on brain MRI images using the hybrid convolutional neural network architecture," *Med. Hypotheses*, vol. 139, 109684, 2020, doi: <https://doi.org/10.1016/j.mehy.2020.109684>.
10. J. Li et al., "A Systematic Collection of Medical Image Datasets for Deep Learning," 2021. [eess.IV]. Available: <https://arxiv.org/abs/2106.12864>.
11. B. Zohuri and M. Moghaddam, "Deep Learning Limitations and Flaws," *Modern Approaches on Material Science*, 2020, doi: <https://dx.doi.org/10.32474/MAMS.2020.02.000138>.
12. A. Younis et al., "Brain Tumor Analysis Using Deep Learning and VGG-16 Ensembling Learning Approaches," *Applied Sciences*, doi: <https://doi.org/10.3390/app12147282>.