HIGH SCHOOL EDITION
Journal of Student Research
Volume 14 Issue 1 (2025)

Improving Brain Tumor Image Classification with Transfer Learning and Selective Augmentation

Anika Ganu

Archbishop Mitty High School, USA

ABSTRACT

Brain tumors are increasingly prevalent health concerns, with approximately 90,000 people diagnosed with a primary brain tumor every year [1]. Glioma tumors are the most lethal and aggressive type of brain tumors and therefore there is a need for increased efficiency in diagnosis. Currently, radiologists diagnose brain tumors through Magnetic Resonance Imaging (MRI) scans. With the advent of machine learning technologies, the process of preliminary image classification can be optimized. The convolutional neural network (CNN), an example of machine learning implementations, is widely used for image classification. In this project, a baseline CNN model was developed using 3,264 labeled MRI images to classify between 4 categories: gliomas, meningiomas, pituitary tumors, and no tumors. The model's initial overall accuracy was 74%. With image augmentation and transfer learning using EfficientNetB0, the accuracy improved to 80%. However, recall for the glioma category was only 33%, consistently lower than that of meningiomas, pituitary tumors and no tumors. By retraining the model with additional training images for only the glioma category, a method known as selective data augmentation, the recall for gliomas improved to 62%, and the model's overall accuracy increased to 94%. To further investigate the low recall for glioma category, identifying potentially mislabeled training data and removing those images was also evaluated. Overall, the results indicate that transfer learning applied to CNN models can benefit diagnostic image classification. Selective augmentation and identifying noisy training data can be used to further improve performance.

Introduction

Brain tumors pose a significant challenge in terms of diagnosis and treatment plans for the patient. Efficient diagnosis is crucial to providing patients with the care they need as soon as possible as the tumors can grow rapidly. The most dangerous type of brain tumor are gliomas, some of which can double in size in less than two weeks. The current methodology for diagnosis of gliomas can take over a month, delaying urgent treatment [2]. By making the process of diagnosis more efficient and getting images flagged as containing gliomas to a radiologist faster, patients can receive the critical care they need sooner. Machine learning models trained to detect various kinds of tumors, including gliomas, with high accuracy can thus be very useful supplementary tools to aid radiologists. This project explores how data augmentation and transfer learning can improve the performance for brain tumor detection.

Background

In recent years, there have been several applications of machine learning in health care specifically related to brain tumor segmentation and classification using images from MRI scans. Researchers at Stanford Medicine have developed a deep learning model to predict the aggressiveness of glioblastoma tumors in the diagnosis phase using histology images [3]. Machine learning techniques that require preprocessing, feature engineering and feature extraction such as random forest classifications, and support vector machines have also been used for tumor classification [4]. Convolutional neural networks have been shown to work well with images for brain tumor classification using convolution

and pooling to "scan" through the image and automatically extracting features by parameter tuning. A previous study has shown the comparison of various machine learning methods including CNNs for brain tumor detection [5]. CNN performance depends on the amount of available training data and with larger datasets, it may require more computation and resources for training. Transfer learning techniques have shown improvements in this area as they rely upon knowledge gained by pre training on different related data (such as detecting shapes, edge, corners), and applying it to a different image dataset. In a 2024 study, researchers applied transfer learning models to improve the accuracy of tumor classification from MRI images [6]. Their study compared the performance of baseline CNN model with transfer learning models such as Inception-v3, EfficientNetB0 and VGG19 and found that VGG19 yielded the best overall results amongst transfer learning models. In medical imaging, the number of available datasets and labelled images is quite limited, and that can impact the accuracy of the model performance when tested with images from external datasets. To improve the generalization of the model, data augmentation techniques that increase the amount of training data have been shown to be effective [7]. Data augmentation includes geometric transformations (such as rotation, zooming, scaling), adding synthetic data using Generative Adversarial Networks (GAN) and other techniques to enrich the original training data. The work in [6] did not include image augmentation and the hypothesis was that applying image augmentation techniques would further improve the performance of their model. Building upon the work done using transfer learning, improvements with image augmentation were evaluated in this paper.

Dataset

The dataset used in this paper was obtained from a publicly available Kaggle repository [8]. This dataset consists of 3,264 MRI images classified in four categories: glioma, meningioma, pituitary tumor, and no tumor. The data is already split into training and testing datasets, with the training data containing 2,870 images and the testing data containing 394 images. The training data contains 826 images of gliomas, 822 images of meningiomas, 827 images of pituitary tumors, and 395 images containing no tumor. In the testing data, there are 100 images of gliomas, 115 images of meningiomas, 74 images of pituitary tumors, and 105 images containing no tumor. Thus, there is a fairly uniform distribution of each labelled class, apart from the no tumor class in the training data. The entire training directory was used to train the model, and it was tested on the testing directory. Examples images of each tumor class are shown in Figure 1.

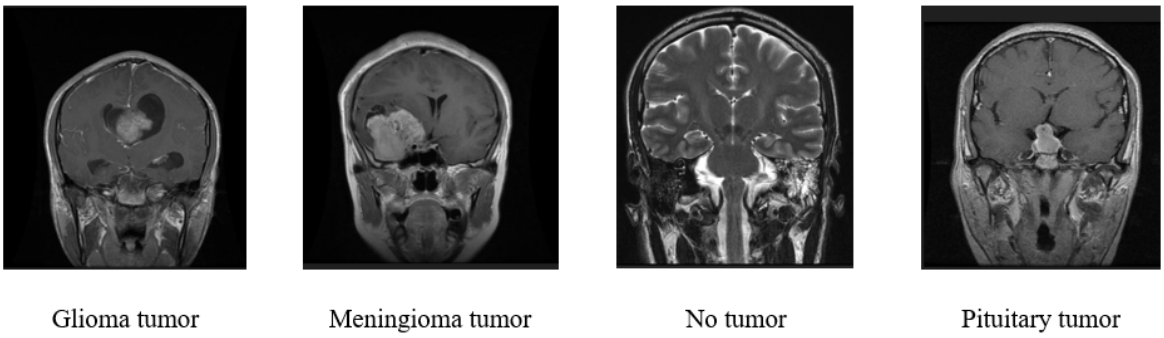

Figure 1. Sample images of each class from training dataset.

Data Pre-Processing

Cropping and Normalization

Prior to training, the images were resized to 256 x 256 pixels and normalized to pixel values between 0 and 1.

Image augmentation

To account for different orientations and border sizes, data augmentation techniques were used to create additional images for training by randomly flipping both horizontally and vertically, randomly rotating (0.2*2π), and changing the contrast of the pre-existing images (factor=0.9). After the augmentation step, the new training dataset contained 11,480 images.

Methods

All the models for baseline and transfer learning used the augmented training data set for learning and the separate test data set for evaluation.

Baseline Model

CNN models are known to have the best performance in supervised image classification. The base CNN model architecture was constructed with the same architecture and hyperparameters in prior studies [6]. The model consisted of three convolution layers with 32, 64 and 128 filters with a 3x3 kernel interspersed with max pooling. The activation function used was Rectified Linear Unit (ReLU) and the flattened output was connected to a 128-unit hidden layer followed by output using SoftMax activation to predict the probabilities of the 4 different classes of brain tumor. The pre-existing training data within the dataset was used to train and build the model, with 10% set aside for validation to prevent overfitting.

Transfer Learning

To compare against the CNN baseline model, two transfer learning methods VGG19 and EfficientNetB0 were evaluated with the image augmented training dataset.

Performance Metrics

For each model, the performance metrics of accuracy, recall, precision, and F1 score were calculated. In addition, a confusion matrix was used to identify the most common incorrect predictions.

Model Training

The model training was done using Google Colab running a mix of CPUs and A100 GPU for speed and efficiency.

Results

The results shown in Table 1 indicate that both transfer learning models VGG19 and EfficientNetB0 showed performance improvements as compared to the CNN based baseline model. Specifically, the overall accuracy improved to 80% with EfficientNetB0 with high F1-scores for the meningioma, no tumor and pituitary tumor classes.

Table 1. Performance comparison between baseline CNN model and transfer learning methods (VGG19 and EfficientNetB0), showing an improvement in accuracy compared to the baseline.

	CNN (Baseline)	VGG-19	EfficientNetB0

Class	Precision	Recall	F1-Score		Precision	Recall	F1-Score		Precision	Recall	F1-Score
Glioma	1	0.13	0.23		1	0.14	0.25		1	**0.33**	0.5
Meningioma	0.59	0.77	0.67		0.65	0.85	0.74		0.68	1	0.81
No tumor	0.5	0.93	0.65		0.67	0.89	0.76		0.81	1	0.9
Pituitary	0.91	0.42	0.57		0.67	0.81	0.73		0.98	0.84	0.91
	Accuracy		**0.59**				**0.67**				**0.8**

However, the recall for glioma tumors remained strikingly low (33%) despite transfer learning. Note that recall is calculated as

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

This implied that of all images of gliomas in the testing set, only 33% were correctly identified by the model. This is particularly concerning, because although the precision for glioma was very high, the model did not perform well in distinguishing glioma images from other types of tumors. As gliomas are the most aggressive and harmful types of tumors, it is essential to reduce the number of false negatives, and thus improve the recall.

Improving Recall for Glioma Class

To further investigate the low recall for glioma class, we hypothesized that the low recall may be due to some glioma images in the testing dataset being very different from the ones used for training despite basic image augmentation. Since the datasets were publicly available without indication as to how they were sourced and divided into training and testing datasets, it is possible that the testing data set may be from a completely different source.

Selective Augmentation

Researchers have shown that selective augmentation can be used for improving results of image classification. In this work, we applied selective augmentation by moving 50 images from the testing data set for glioma class into the training dataset for glioma class. The geometric image augmentation (rotate, flip and contrast) was redone to increase the amount of training data specifically for gliomas. This resulted in a total of 11,680 images for training. This was done with the goal of improving the recall performance of the EfficientNetB0 model for the glioma class.

Table 2. Performance of EfficientNetB0 with selective augmentation compared to EfficientNetB0 only. The numbers in bold show improved recall for glioma and overall accuracy with selective augmentation.

	EfficientNetB0				**Selective augmentation**		
Class	Precision	Re-call	F1-Score		Preci-sion	Re-call	F1-Score
Glioma	1	**0.33**	0.5		1	**0.62**	0.77
Meningi-oma	0.68	1	0.81		0.87	1	0.93
No tumor	0.81	1	0.9		0.95	1	0.98
Pituitary	0.98	0.84	0.91		1	0.96	0.98

Accuracy	**0.8**			**0.94**

The results show a significant improvement in the recall for gliomas (62%) as well as the model's overall accuracy (94%), thereby suggesting that the quality and volume of the original training data may have been insufficient.

Investigation Into Anomalies in Training Data

The quality of the training data in the glioma class was then investigated in greater detail. Because manually inspecting and removing potentially mislabeled images was difficult without medical knowledge, a tool from CleanLab, was used to automate this process. This tool works by performing K-fold cross validation on the training dataset, comparing the predicted probabilities with the actual labels, and identifying potential anomalies in the data. Examples of some images identified as mislabeled glioma images by the tool are shown in Figure 2 along with an image that was identified as properly labelled by the tool in Figure 3.

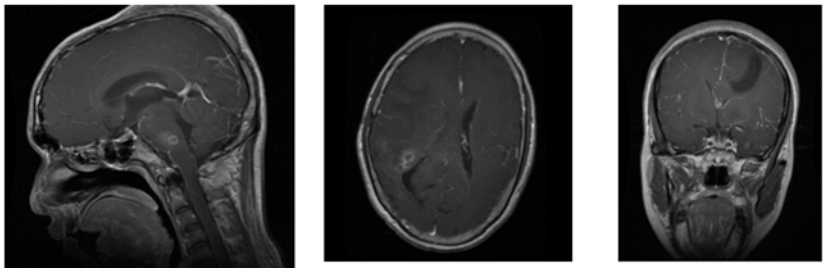

Figure 2. Glioma images identified as mislabelled by the tool.

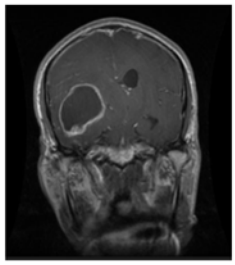

Figure 3. Glioma images identified as correctly labelled by the

The top 50 images in the training data identified as potential anomalies by the tool were removed from the dataset, and the model was retrained without selective data augmentation. This new model's performance was compared with the transfer learning model (EfficientNetB0) before selective data augmentation.

Table 3. Performance of CleanLab anomaly removal with EfficientNetB0 compared to EfficientNetB0 only shows improved recall for glioma and overall accuracy (numbers in bold)

	EfficientNetB0			**Image cleanup**		
Class	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Glioma	1	**0.33**	0.5	1	**0.4**	0.77
Meningioma	0.68	1	0.81	0.82	1	0.93
No tumor	0.81	1	0.9	0.92	1	0.98
Pituitary	0.98	0.84	0.91	0.99	0.92	0.98
	Accuracy		**0.8**			**0.9**

The results show that the recall of glioma images rose from 33% to 40%, indicating the possibility of anomalies in the original training data. This indicates the importance of having high quality labeled datasets for building models for brain tumor image classification. These models can work well on external datasets and can be used by the medical community with confidence to improve diagnosis and speed up treatment plans for patients.

The source code for this paper can be found at https://github.com/anigan27/Brain-Tumor-Classification.

Conclusion

The goal of this paper was to develop a CNN based model to classify between major tumor classes, specifically gliomas, meningiomas, and pituitary tumors, to speed up tumor diagnosis. Image augmentation and transfer learning showed improved accuracy (around 80%) compared to baseline CNN model. To improve the low recall of the transfer learning model for glioma, selectively augmenting more images into the training data for glioma images significantly improved the recall to 62% and improving the overall accuracy of model to 94%. However, upon observation of the training images for gliomas and the consistent poor recall, it was determined that there was possibility that some images may have been mislabeled. To compensate for this, further research includes employing anomaly detection tools CleanLab to detect potentially mislabeled images and remove them from the training data. Early results show improvement in the recall for glioma.

For further investigation, there are several ideas to pursue including additional augmentation methods such as Generative Adversarial Networks (GANs) and alternate image processing techniques such as segmentation to improve the overall performance of the model. Additionally, techniques to improve the accuracy of the training data labeling will be considered.

Acknowledgments

I would like to thank Yanal Qushair for his support and guidance throughout this project.

References

1. American Brain Tumor Association. (2018). American Brain Tumor Association. https://www.abta.org/about-brain-tumors/brain-tumor-education/

2. Neurosurgeon Develops Artificial Intelligence Device to Diagnose & Treat Brain Cancer Faster. (2023, May 22). NYU Langone News. https://nyulangone.org/news/neurosurgeon-develops-artificial-intelligence-device-diagnose-treat-brain-cancer-faster

3. Zheng, Yuanning, et al. “Spatial Cellular Architecture Predicts Prognosis in Glioblastoma.” Nature Communications, vol. 14, no. 1, 11 July 2023, www.ncbi.nlm.nih.gov/pmc/articles/PMC10336135/, https://doi.org/10.1038/s41467-023-39933-0. Accessed 8 Feb. 2024.

4. Amin, Javeria, et al. “A Distinctive Approach in Brain Tumor Detection and Classification Using MRI.” Pattern Recognition Letters, Oct. 2017, https://doi.org/10.1016/j.patrec.2017.10.036.

5. Ali, Rawaa, et al. “Classification of Brain Tumors Using MRI Images Based on Convolutional Neural Network and Supervised Machine Learning Algorithms.” IEEE Xplore, 1 Oct. 2022, ieeexplore.ieee.org/document/9932690. Accessed 29 Dec. 2022.

6. Khaliki, M. Z., & Başarslan, M. S. (2024). Brain tumor detection from images and comparison with transfer learning methods and 3-layer CNN. Scientific Reports, 14(1), 2664. https://doi.org/10.1038/s41598-024-52823-9

7. Mumuni, Alhassan, and Fuseini Mumuni. “Data Augmentation: A Comprehensive Survey of Modern Approaches.” Array, vol. 16, Nov. 2022, p. 100258, https://doi.org/10.1016/j.array.2022.100258.

8. Bhuvaji, S. (2020). Brain Tumor Classification (MRI). www.kaggle.com. https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri/data

9. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

10. Tan, M. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946.

11. Xue, Yuan, et al. “Selective Synthetic Augmentation with HistoGAN for Improved Histopathology Image Classification.” Medical Image Analysis, vol. 67, Jan. 2021, p. 101816, https://doi.org/10.1016/j.media.2020.101816. Accessed 28 Mar. 2022.

12. Lok, W. J., & Mueller, J. (2022, April 21). Finding Label Issues in Image Classification Datasets. Cleanlab.