Journal of High School Science

Classification of brain tumors from MRI images using deep transfer learning

Kazemi A

Submitted: July 20, 2023, Revised: version 1, August 28, 2023
Accepted: October 27, 2023

Abstract

Each year, more than 100,000 people in the United States are diagnosed with a brain tumor. An early and accurate diagnosis is crucial in getting patients the necessary treatment and increasing survival rates. In recent years, machine learning algorithms have become increasingly popular in the medical field due to their ability to recognize complex patterns and reduce human errors. However, accurate diagnosis using deep learning algorithms requires a large amount of training data, which is not always available. Additionally, training a model from scratch can take a long time and requires vast amounts of computational power. As a solution, this study aims to utilize transfer learning, which uses the knowledge gained by the model on one dataset to aid in classifying the second dataset. In this study, a dataset of MRI images consisting of four classes of brain tumors (no tumor, pituitary tumor, meningioma, and glioma) were used. The performance of seven pre-trained models (ResNet18, ResNet50, VGG16, DenseNet, GoogLeNet, ShuffleNet, and MobileNet) were evaluated in order to see which would achieve the highest classification accuracy. Additionally, this study examined two different methods for the implementation of transfer learning. In the first method, the convolutional base of the pre-trained model was frozen and in the second method, the convolutional base was trained. The best performing models proved to be ResNet18 and ShuffleNet with the base trained, achieving an accuracy of 97.86%. The results also showed that the models with the convolutional base trained outperformed those with the convolutional base frozen.

Keywords

Brain tumor, Deep learning, Transfer learning, Magnetic Resonance Imaging, Image classification, Hyperparameters, Vanishing gradient, ResNet18, ShuffleNet, Convolutional base

---

Armita Kazemi, Century Senior High School, 2525 Viola Rd NE, Rochester, MN, USA.
armita.kazemi85@gmail.com

Introduction

Each year, hundreds of thousands of people are diagnosed with brain tumors across the world. In 2021, 83,570 individuals were diagnosed with brain tumors in the United States, with 24,530 being malignant and 59,040 being benign. Of those, 18,600 patients died of the disease.

An efficient and accurate brain tumor diagnosis is critical in determining the best treatment and increasing the chances of recovery. Brain tumors are commonly misdiagnosed, affecting the health of thousands of patients and costing millions of dollars. In recent years, machine learning—a type of artificial intelligence (AI)—has emerged as a powerful tool for improving medical diagnosis. Machine learning programs are trained on large amounts of data and then identify structures and patterns in that data. Then they use those patterns to predict answers to problems. Not only can machine learning greatly improve the accuracy of brain tumor diagnosis, but it can also aid countries with underdeveloped healthcare systems. In places like Sub-Saharan Africa with an average of 0.2 doctors per 1,000 people, machine learning can provide a way to diagnose people that would be impossible otherwise.

Deep learning has become increasingly popular for medical image detection. Many researchers have proposed different architectures and techniques to aid in medical diagnosis using image classification. Deep learning has been commonly used in the analysis of breast cancer and diagnosis of lung or skin cancer. Zuo et al. (1) proposed a deep learning algorithm by utilizing recurrent neural networks for human skin detection. Charron et al. (2) used a 3D convolutional neural network (DeepMedic) to detect and segment brain metastases on MRI.

The problem that arises with previous medical image classification methods is the lack of large, high-quality datasets. Machine learning models rely on large quantities of training data to achieve high classification accuracies. The MNIST dataset, one of the most popular datasets in machine learning research, contains 60,000 images for training. Many studies done using this dataset have yielded high accuracies (3, 4, 5). However, datasets of this scale are not available for medical image classification.

The solution to this problem lies in transfer learning, a new deep learning technique that has been dominating the studies on image classification. Transfer learning is a technique in which a Convolutional Neural Network (CNN) model trained on an initial task is reused to enhance the performance on a related task. More importantly, transfer learning has been gaining attention in relation to medical image classification. Rahaman et al. (6) used transfer learning to identify COVID-19 from chest X-ray images. Furthermore, Burdick et al. (7) applied transfer learning to skin lesion segmentation to diagnose melanoma. Transfer learning has also been used to detect malignant prostate lesions using MRI (8). Additionally, Lakhani (9) used transfer learning to classify endotracheal tube position.

Specifically, transfer learning has been used in applications related to neuro-oncology. Chelghoum et al. (10) used nine pre-trained models to classify brain tumors. Another study done by Swati et al. (11) applied a block-wise fine-tuned network based on transfer learning. Additionally, Deepak and Ameer (12) achieved

remarkable classification performance with deep transfer learning on their work with brain tumors. Toğaçar et al. (13) proposed a new convolutional neural network called BrainMRNet with an architecture built on attention modules and hypercolumn techniques. Sharif et al. (14) applied an active deep learning feature selection approach paired with an InceptionV2 pre-trained model to aid in classifying glioma grade.

An additional challenge is presented when classifying brain tumors into subtypes. The challenge is attributed to tumors having similar appearances, even from different pathological types. Cheng et al. (15) worked on a three-class brain tumor classification problem using MRI images. The authors proposed a method of augmenting the tumor region via image dilation and splitting the augmented region into fine ring-form subregions. Talo et al. (16) applied transfer learning to the diagnosis of brain abnormalities such as cerebrovascular and neoplastic diseases.

In this paper, we compare the performance of multiple pre-trained Convolutional Neural Network (CNN) models in distinguishing between four classes of brain tumors (no tumor, pituitary tumor, meningioma, and glioma). Furthermore, we compare the implementation of transfer learning with two different methods (freezing the convolutional base and training the convolutional base) to discover which yields the best results. Following testing, a complete evaluation of each model is conducted. The major contributions of this paper are; **1]** A transfer learning approach is applied to a four-class brain tumor classification task, **2]** CNN models are evaluated on the classification of brain tumor types, **3]** For the first time, the significance of trainable layers in transfer learning is discovered in relation to brain tumor classification, **4]** The proposed method achieves high accuracies compared to other methods while using a relatively small dataset

Materials and Methods

Dataset

The dataset used in this study was publicly available on Kaggle (17). It contains a total of 3264 brain MRI images. The images had a class distribution of 500 with no tumor, 901 with pituitary tumors, 937 with meningiomas, and the remaining 926 with gliomas. The images are available as .jpg files and the size of each image is 512×512. The tumor type labels were changed to numerical values by assigning 0 to no tumor, 1 to pituitary tumor, 2 to meningioma, and 3 to glioma.

In the data preprocessing stage, the images were resized to 224×224 and normalized to scale the intensity values between 0 and 1. The dataset was divided into 80% and 20% for training and testing, respectively. Then the dataset was converted to a PyTorch Tensor. The dataset was passed through the PyTorch DataLoader with a batch size of 32 and shuffling for the training set.

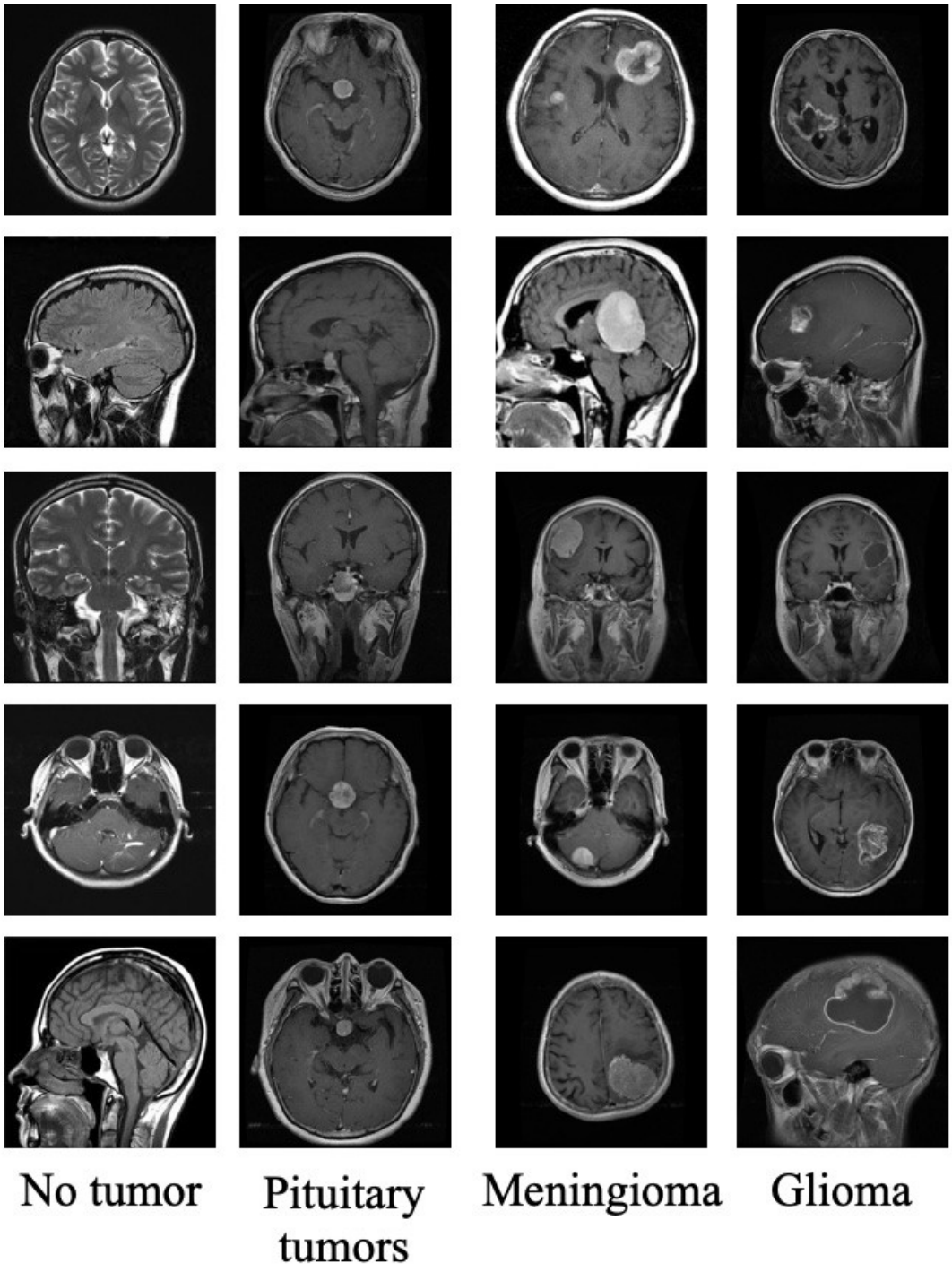

Figure 1: Sample Brain MRI Images of the Four Classes of Brain Tumors in This Study

Methodology

In this work, we applied different pre-trained machine learning models to the classification task via transfer learning.

Transfer learning is a machine learning technique where a model is initially trained on a large dataset, and then that pre-trained model is used as the starting point for a new task. The

pre-trained model has already learned general features like edges and shapes in an image. So, when the model is trained on the second dataset, it can focus on the features more specific to those images, as opposed to learning the basics as well (Figure 2). With transfer learning, there is no need to completely train a model from scratch, which eliminates the need for large datasets and computational power.

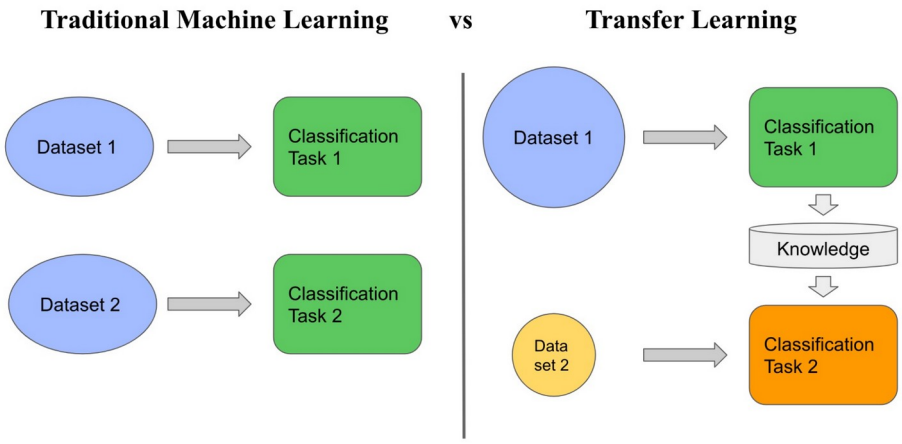

Figure 2: How Traditional Machine Learning Works Compared to Transfer Learning

The transfer learning models for this experiment were imported from PyTorch. All the models were pre-trained on the ImageNet dataset, which consists of over 14 million various images with 22,000 labeled categories and is intended for computer vision research. Each CNN model consists of a convolutional base and a final classifier layer (Figure 3). The convolutional base generally consists of multiple convolution layers, pooling layers, and activation functions. The classifier layer consists of a few fully connected layers and pooling layers. The last fully connected layer of each original model was removed and instead replaced with a new fully connected layer with an output size of four to fit the specifications of our classification task.

During the training and testing stages, Adam was chosen as the optimizer with a learning rate of 0.001. This learning rate was chosen as it was a reasonable value to keep constant throughout the various trials, and the models performed well on it. The loss was calculated using the Cross-Entropy Loss function. Each model was trained and tested over 100 epochs.

In this paper, we used two methods of implementing the pre-trained models in our classification task. The methods are described in Figure 3.

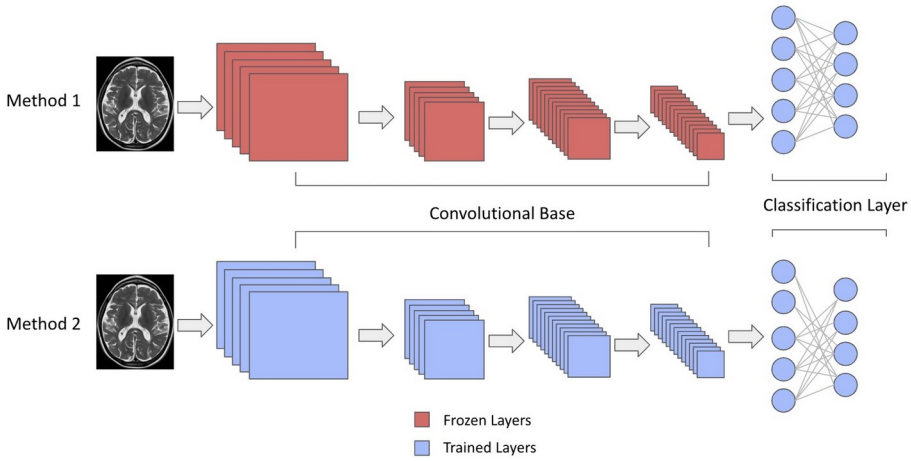

Figure 3: Different Implementation Methods of Transfer Learning

In method 1, the convolutional base of the pre-trained model was frozen, and only the classification layer was trained. This is because after each model is pre-trained on a dataset, it associates specific weights with each layer of the model. By freezing the convolutional base, the weights assigned from training on the ImageNet dataset are preserved. This method results in a decrease in training and testing time as only the top classification layer is trained.

In method 2, both the convolutional base and the classification layer were trained. This method deletes all the old weights associated with ImageNet and assigns new weights that correspond to the new dataset. This method can achieve higher accuracy rates as it is adapting the pre-trained features to the new data. Method 2 is preferred in scenarios where the source domain and target domain are different from each other. The drawback of this method is the increase in training and testing time for the model.

Models

We used the pre-trained models ResNet18, ResNet50, VGG16, DenseNet, GoogLeNet, ShuffleNet, and MobileNet.

ResNet, short for Residual Network, is a type of neural network that was introduced in 2015 by He, Zhang, Ren, and Kan Sun in their paper "Deep Residual Learning for Image Recognition" (18). The ResNet models were extremely successful, as can be ascertained from the fact that they won 1st place in the ILSVRC 2015 classification competition with an error of only 3.57%. VGG 16 was proposed by Simoyan and Zisserman in 2014 in their paper "Very Deep Convolutional Networks for Large-Scale Image Recognition" (19). This model won 1st and 2nd place in the 2014 ILSVRC challenge for object detection and classification. GoogLeNet was proposed by researchers at Google in 2014 in the research paper "Going Deeper with Convolutions" (20). This architecture won the LSVRC 2014 image classification challenge. The MobileNetV2 model was introduced by Google in 2018 with

the ability to run deep networks on personal mobile devices. The DenseNet model was introduced in the "Densely Connected Convolutional Network" paper in 2017 (21). ShuffleNet was introduced by Zhang, Zhou, Lin, and Sun in their paper "ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices" (22).

Results

Metrics

The classification of brain tumors from MRI images into four classes was evaluated on five performance metrics. The most common metric is the classification accuracy, which is the percentage of correctly classified samples from the total number of samples. However, final classification accuracy is most reliable when the dataset contains an equal number of samples from each class. The unbalanced nature of our dataset requires further evaluation with more performance metrics. For further evaluation of each model, a confusion matrix was generated using Scikit Learn. The confusion matrix shows the distribution of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). From the confusion matrix, the precision (1), recall (2), and F1 score (3) can be determined. The training and testing times per epoch were also recorded.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (3)$$

The results of all seven pre-trained models using both methods (freezing convolutional base and training convolution base) are shown in Table 1.

ResNet

As can be seen in Table 1, freezing the convolutional base (method 1) resulted in a classification accuracy of 89.13% by ResNet18 and 88.97% by ResNet50. ResNet18 had a shorter run time with an average of 6.59 seconds per epoch as opposed to ResNet50 with an average of 19.03 seconds per epoch. When the convolutional base was trained (method 2), ResNet18 achieved a classification accuracy of 97.86% and ResNet50 achieved 95.71%. The accuracies with method 2 were significantly higher than with method 1. However, the models ran longer with method 2 when compared to method 1. With method 2, ResNet18 had an average 11.59 seconds per epoch, while ResNet50 had 37.25 seconds per epoch.

Table 1: Results of Transfer Learning with Both Methods

Model	Accuracy (percent) Method 1	Accuracy (percent) Method 2	Best Accuracy (percent) Method 1	Best Accuracy (percent) Method 2	Average Time Per Epoch (seconds) Method 1	Average Time Per Epoch (seconds) Method 2
ResNet18	89.13	**97.86**	89.43	**98.62**	6.59	11.59
ResNet50	88.97	95.71	91.11	97.24	19.03	37.25
DenseNet	**91.58**	94.95	**92.19**	97.55	42.00	81.43
GoogLeNet	86.22	88.97	86.37	**98.62**	7.88	14.97
ShuffleNet	86.83	**97.86**	86.83	98.32	**4.04**	**6.58**
MobileNet	88.97	97.55	89.74	98.47	7.29	13.99
VGG16	88.97	86.52	90.51	86.83	28.67	53.30

Figures 4 and 5 show the confusion matrices for ResNet18, and ResNet50. ResNet18 had the highest accuracy for no tumor at 100% with method 2 and the lowest for meningioma at 80.11% with method 1. ResNet50 had the highest accuracy for pituitary tumors at 98.01% with method 2 and the lowest for glioma at 87.28% with method 1. The precision, recall and F1 scores are shown in Tables 2 and 3.

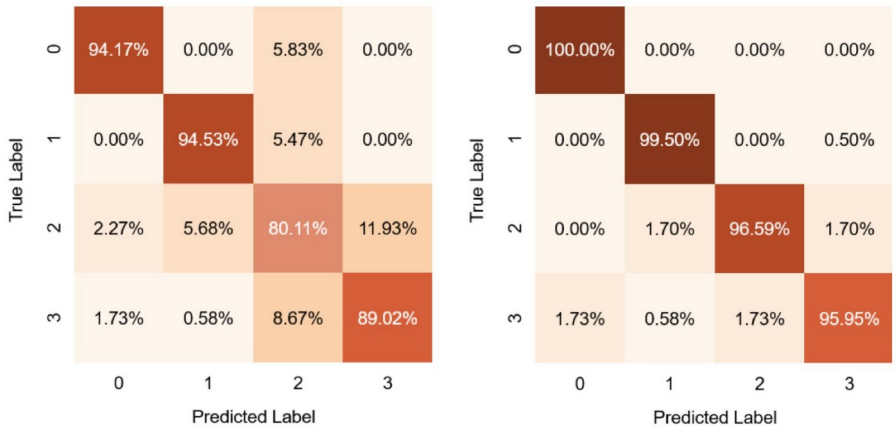

Figure 4: Confusion Matrices for ResNet18 with Method 1 (Left) and Method 2 (Right)

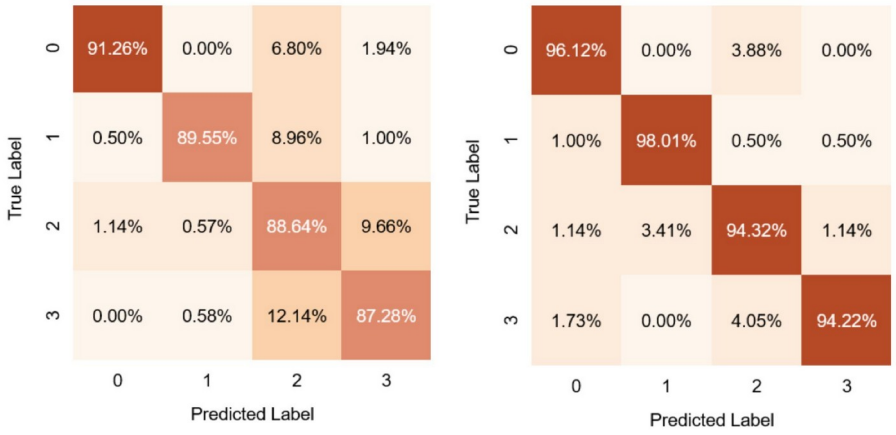

Figure 5: Confusion Matrices for ResNet50 with Method 1 (Left) and Method 2 (Right)

Table 2: Class-specific Evaluation of ResNet18

Method 1			
Tumor Type	Precision	Recall	F1 Score
No Tumor	93	94	94
Pituitary	95	95	95
Meningioma	82	80	81
Glioma	88	89	89

Method 2			
Tumor Type	Precision	Recall	F1 Score
No Tumor	97	100	99
Pituitary	98	100	99
Meningioma	98	97	97
Glioma	98	96	97

Table 3: Class-specific Evaluation of ResNet50

Method 1			
Tumor Type	Precision	Recall	F1 Score
No Tumor	97	91	94
Pituitary	99	90	94
Meningioma	77	89	83
Glioma	88	87	88

Method 2			
Tumor Type	Precision	Recall	F1 Score
No Tumor	93	96	95
Pituitary	97	98	98
Meningioma	93	94	94
Glioma	98	84	86

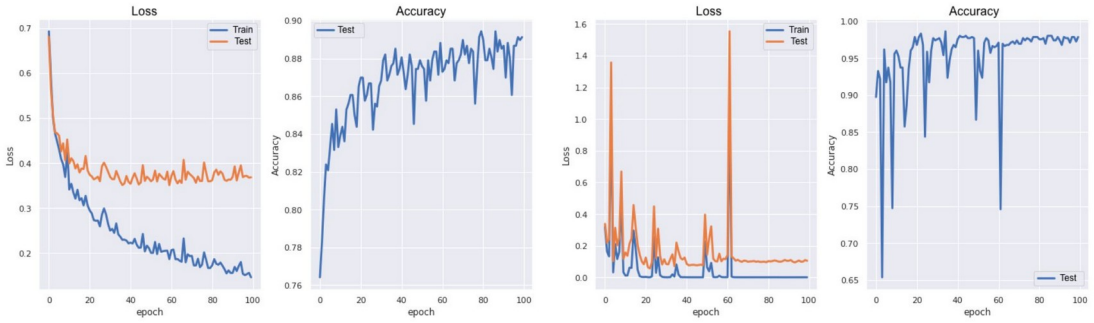

Figure 6: Loss and Accuracy vs Epoch Graphs for ResNet18 with Method 1 (Left) and Method 2 (Right)

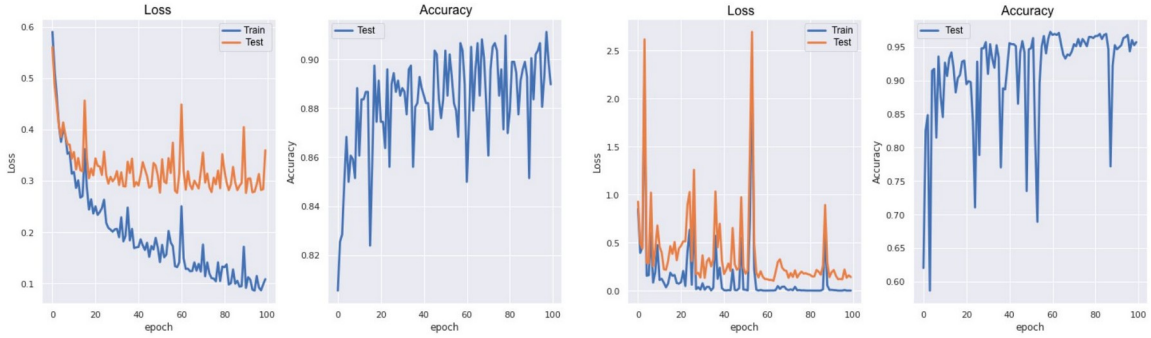

Figure 7: Loss and Accuracy vs Epoch Graphs for ResNet50 with Method 1 (Left) and Method 2 (Right)

DenseNet

DenseNet achieved the best classification accuracy out of all the models run using method 1 at 91.58% (Table 1). With method 2, DenseNet achieved an accuracy of 94.95%. DenseNet took the longest time to run with both methods. With method 1, DenseNet averaged 42.00 seconds per epoch and with method 2 it averaged 81.43 seconds per epoch. These times are significantly longer than all the other models. The longer run time of DenseNet can be attributed to the enormous output that results from all the layers being connected in its architecture.

Figure 8 shows the confusion matrices for DenseNet. With method 1, the model had the highest accuracy for pituitary tumors at 99.41% and the lowest for meningioma at 84.07%. With method 2, the model had the highest accuracy for pituitary tumors at 100% and the lowest for glioma at 91.98%. The precision, recall and F1 scores are shown in Table 4.

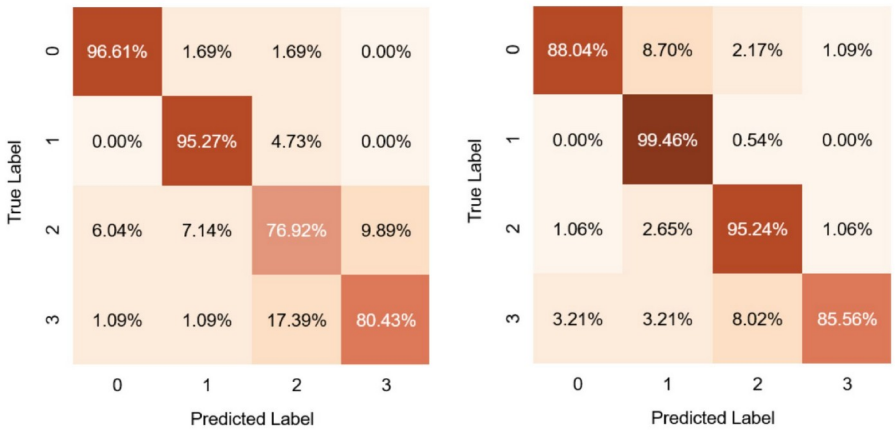

Figure 8: Confusion Matrices for DenseNet with Method 1 (Left) and Method 2 (Right)

Table 4: Class-specific Evaluation of DenseNet

Method 1			
Tumor Type	Precision	Recall	F1 Score
No Tumor	98	94	96
Pituitary	95	99	97
Meningioma	88	84	86
Glioma	87	90	89

Method 2			
Tumor Type	Precision	Recall	F1 Score
No Tumor	90	100	95
Pituitary	98	98	98
Meningioma	97	93	95
Glioma	93	92	92

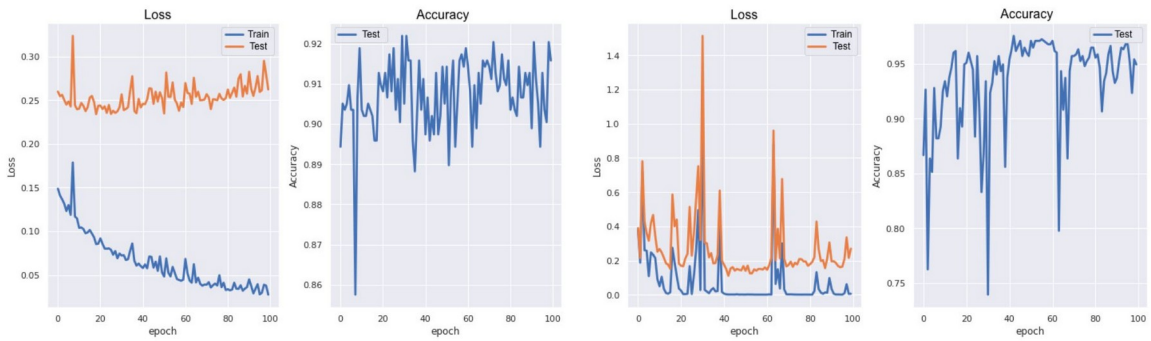

Figure 9: Loss and Accuracy vs Epoch Graphs for DenseNet with Method 1 (Left) and Method 2 (Right)

GoogLeNet

GoogLeNet had the lowest accuracy of all the models run using method 1 at 86.22% (Table 1). With method 2, the accuracy increased to 88.97%, although this increase was not as significant as with other models. GoogLeNet had an average runtime of 7.88 seconds per epoch with method 1 and 14.97 seconds per epoch with method 2.

Figure 10 shows the confusion matrices for GoogLeNet. With method 1, the model had the highest accuracy for no tumors at 96.61% and

the lowest for meningioma at 76.92%. With method 2, GoogLeNet had the highest accuracy for pituitary tumors at 99.46% and the lowest for glioma at 85.56%. The precision, recall and F1 scores are shown in Table 5.

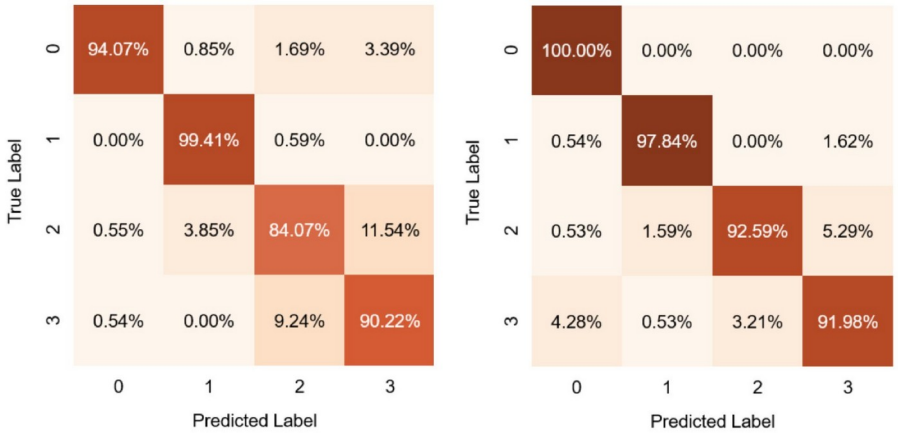

Figure 10: Confusion Matrices for GoogLeNet with Method 1 (Left) and Method 2 (Right)

Table 5: Class-specific Evaluation of GoogLeNet

Method 1			
Tumor Type	Precision	Recall	F1 Score
No Tumor	90	97	93
Pituitary	90	95	93
Meningioma	77	77	77
Glioma	89	80	85

Method 2			
Tumor Type	Precision	Recall	F1 Score
No Tumor	91	88	90
Pituitary	91	99	95
Meningioma	91	95	93
Glioma	98	86	91

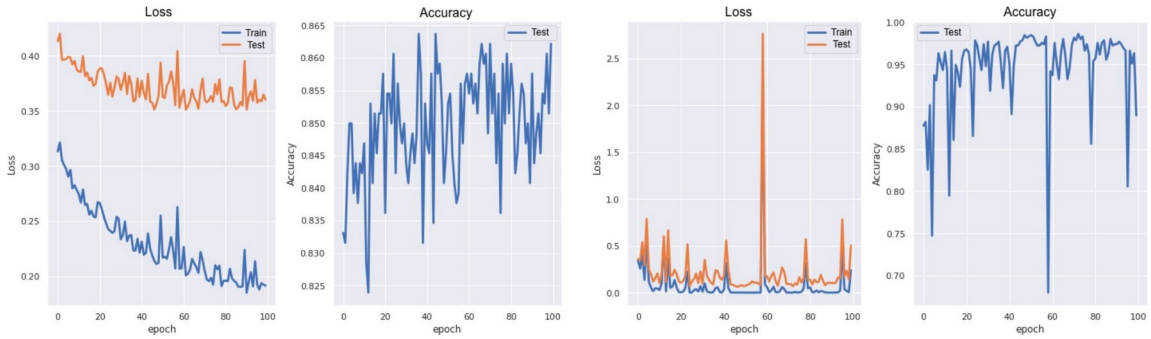

Figure 11: Loss and Accuracy vs Epoch Graphs for GoogLeNet with Method 1 (Left) and Method 2 (Right)

ShuffleNet

ShuffleNet achieved a classification accuracy of 86.83% with method 1 (Table 1). The model improved significantly with method 2, achieving the highest classification accuracy of all the models with 97.86%. ShuffleNet proved to be the fastest model regardless of the method used. It ran at 4.04 seconds per epoch with method 1 and 6.58 seconds per epoch with method 2.

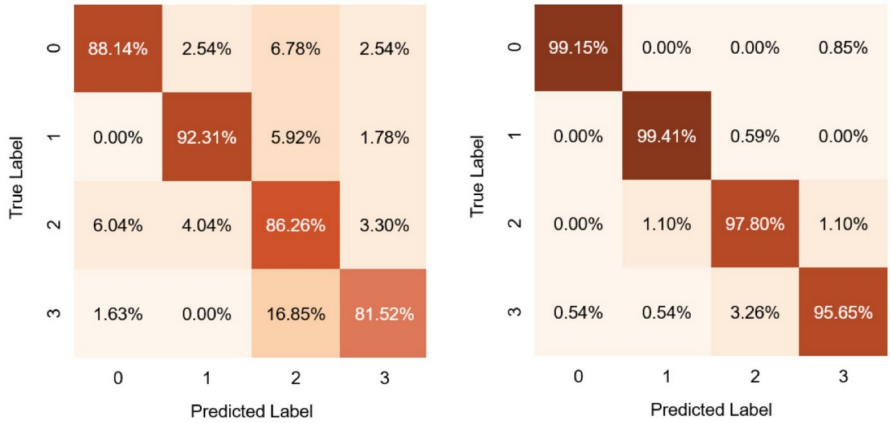

Figure 12: Confusion Matrices for ShuffleNet with Method 1 (Left) and Method 2 (Right)

Figure 12 shows the confusion matrices for ShuffleNet. With method 1, the model had the highest accuracy for pituitary tumors at 92.31% and the lowest for glioma at 81.52%. With method 2, the model had the highest accuracy for pituitary tumors at 99.41% and the lowest for glioma at 95.65%. The precision, recall and F1 scores are shown in Table 6.

Table 6: Class-specific Evaluation of ShuffleNet

Method 1			
Tumor Type	Precision	Recall	F1 Score
No Tumor	88	88	88
Pituitary	93	92	93
Meningioma	76	86	81
Glioma	93	82	87

Method 2			
Tumor Type	Precision	Recall	F1 Score
No Tumor	99	99	99
Pituitary	98	99	99
Meningioma	96	98	97
Glioma	98	96	97

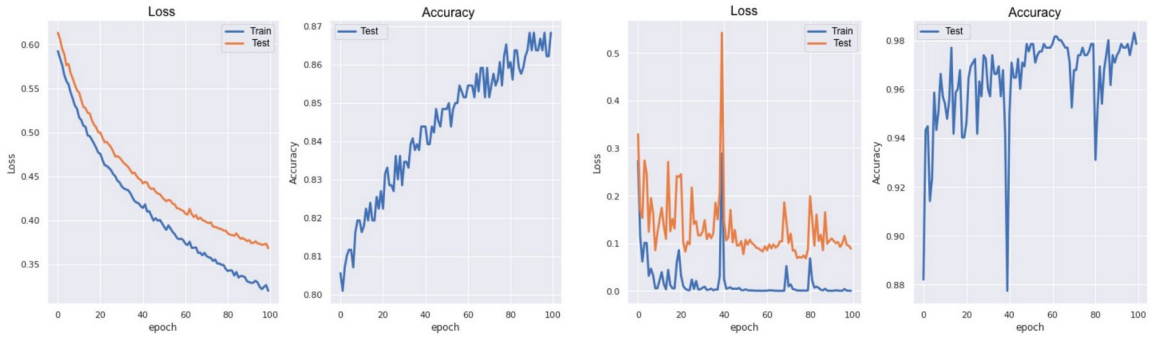

Figure 13: Loss and Accuracy vs Epoch Graphs for ShuffleNet with Method 1 (Left) and Method 2 (Right)

MobileNet

MobileNet achieved a classification accuracy of 88.97% when run with method 1 (Table 1). However, the model improved greatly when run with method 2, achieving the second highest accuracy of 97.55%. MobileNet was one of the fastest models, averaging 7.29 seconds per epoch with method 1 and 13.99 seconds per epoch with method 2.

Figure 14 shows the confusion matrices for MobileNet. With method 1, the model had the highest accuracy for pituitary tumors at 98.82% and the lowest for meningioma at 73.08%. With method 2, the model had the highest accuracy for glioma at 98.37% and the lowest for meningioma at 96.15%. The precision, recall and F1 scores are shown in Table 7.

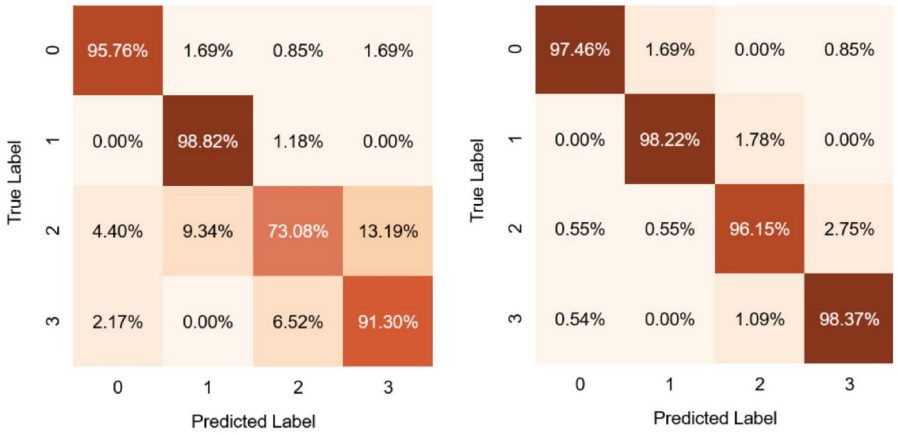

Figure 14: Confusion Matrices for MobileNet with Method 1(Left) and Method 2 (Right)

Table 7: Class-specific Evaluation of MobileNet

Method 1			
Tumor Type	Precision	Recall	F1 Score
No Tumor	90	96	93
Pituitary	90	99	94
Meningioma	90	73	81
Glioma	87	91	89

Method 2			
Tumor Type	Precision	Recall	F1 Score
No Tumor	98	97	98
Pituitary	98	98	98
Meningioma	97	96	97
Glioma	97	98	98

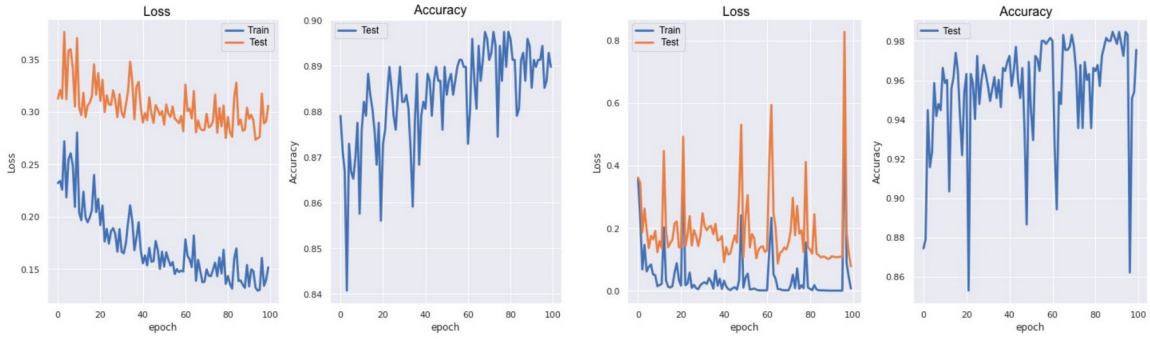

Figure 15: Loss and Accuracy vs Epoch Graph for MobileNet with Method 1 (Left) and Method 2 (Right)

VGG16

VGG16 performed with an accuracy of 88.97% when run with method 1, but the accuracy dropped to 86.53% with method 2 (Table 1). This is the opposite of what was seen with all the other models. VGG16 was also the second slowest model, with 28.67 seconds per epoch with method 1 and 53.30 seconds per epoch with method 2.

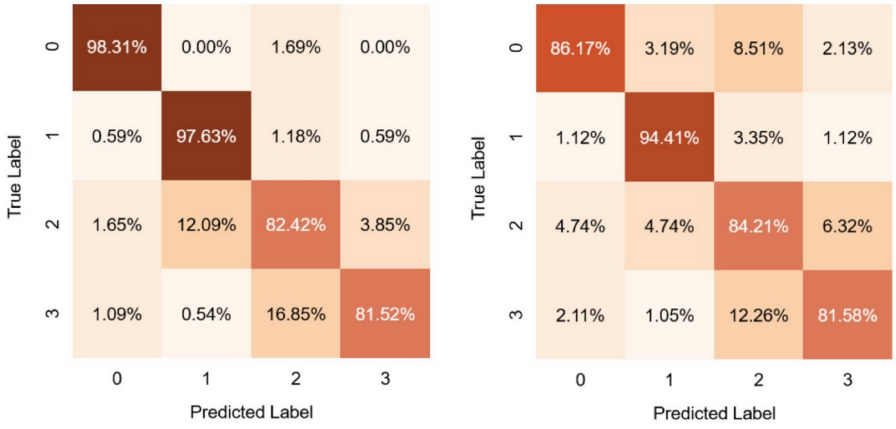

Figure 16: Confusion Matrices for VGG16 with Method 1(Left) and Method 2 (Right)

Figure 16 shows the confusion matrices for VGG16. With method 1 the model had the highest accuracy for no tumors at 98.31% and the lowest for glioma at 81.52%. With method 2 the model had the highest accuracy for pituitary tumors at 94.41% and the lowest for glioma at 81.58%.

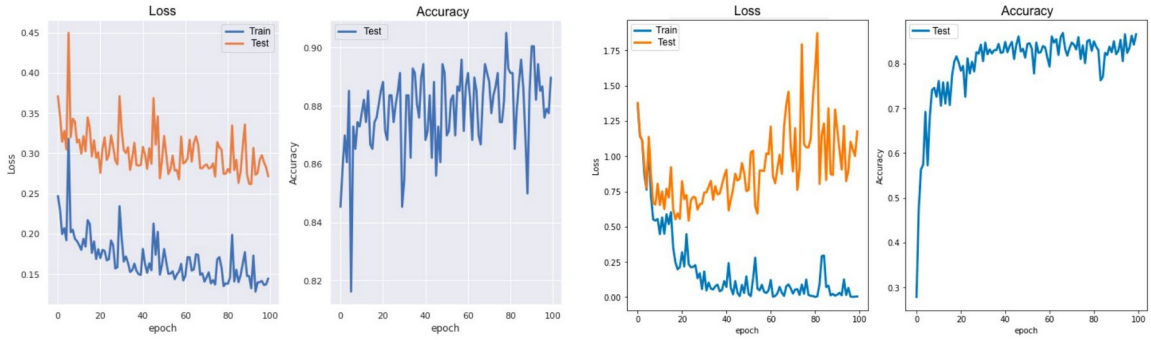

Figure 17: Loss and Accuracy vs Epoch Graphs for VGG16 with Method 1 (Left) and Method 2 (Right)

Discussion

The highest classification accuracy was achieved by ResNet18 and ShuffleNet at 97.86%. Both of these accuracies were achieved when the models were run using method 2, which was training the convolutional base. However, ShuffleNet ran almost twice as fast as ResNet18. The precision scores for ResNet18 and ShuffleNet were the same for pituitary tumors and glioma, and slightly different for no tumor and meningioma. ShuffleNet had a higher precision score for no tumors at 99% but ResNet18 had a higher recall score at 100%. ResNet18 had a higher precision score for meningioma, but it had a slightly lower recall than ShuffleNet. The F1 scores for both models were identical. High precision and recall scores are especially important when evaluating models on a medical classification problem. A high precision score indicates that most of the positives detected were true positives. A high precision score is critical in diagnosing brain tumors because it ensures that patients won't undergo unnecessary treatment due to false positives. On the other hand, a high recall score indicates that the model was able to correctly detect most occurrences of that brain tumor. A high recall score is especially important in diagnosing possibly malignant brain tumors because it ensures that there are very few false negatives. If there was a high false negative rate, it would mean that patients who actually had a brain tumor were falsely told they didn't, which could lead to delayed treatment and increased progression of the tumor.

On average, the models overwhelmingly had the highest classification accuracy for pituitary tumors and the lowest accuracy for meningioma or glioma. One explanation for the lower accuracies is that meningioma and glioma were often misclassified as each other. The factor most likely contributing to this misclassification was the similar appearance of meningioma and glioma in the MRI images.

VGG16 had a lower accuracy with method 2 than method 1. A possible explanation is the vanishing gradient problem. This problem occurs when multiple layers are added to neural networks, which causes the gradient of the loss function to approach zero, decreasing the network accuracy and preventing it from converging. When the convolutional base was frozen (method 1), the network was not training many layers and therefore did not have the problem of the gradient vanishing. However, when the convolutional base was trained (method 2), the entire network was trained and the likelihood of the gradient vanishing increased.

On the other hand, both ResNet models showed significant increase in accuracy when using method 2. Even though more layers were being trained with method 2, the gradient doesn't vanish due to the use of skip connections in the ResNet architecture. These skip connections allow the gradient signal to bypass layers, letting it flow more easily through the network and preventing the gradient from approaching zero. The DenseNet architecture was also created with the goal of eliminating vanishing gradients by connecting all the layers. Each layer receives the feature maps from all the preceding layers and passes its output to all the subsequent layers.

With both methods, the fastest models were ShuffleNet, Resnet18, and MobileNet. The slowest model for both methods was DenseNet. The long run time of DenseNet can be attributed to the model architecture. The DenseNet architecture interconnects all the layers in the model. While this method resulted in high accuracies, the architecture is not practical for deep neural nets as it increases the run time and consequently the computational costs. Additionally, while method 2 resulted in higher accuracies, it also doubled the run time for each model.

In recent years, numerous studies have been conducted using different methods to classify brain tumors using MRI images. A comparison between the results of our study and other methods shows that our best performing transfer learning model outperformed the previous methods in related studies.

One proposed method was by Cheng et al. (15) to use tumor region augmentation and partition in classifying three classes of tumors. They used the feature extraction methods of intensity histogram, GLCM, and bag-of-words (BoW) with accuracies 78.18%, 87.54%, and 91.28%, respectively. In addition to achieving higher accuracies, the transfer learning models proposed in our study are fully automated and do not require any feature extraction or selection. Ismael et al. (23) proposed an algorithm for the classification of brain tumors using Discrete Wavelet Transform (DWT) and Gabor filter techniques. The algorithm was evaluated to have a total accuracy of 91.9%. A capsule neural network (CapsNet) was used by Afshar et al. (24) to utilize the spatial relations between tumors and the surrounding tissues. Their proposed CapsNet architecture attained an accuracy of 90.89%.

Deepak and Ameer (12) focused on using GoogLeNet with support vector machine and K-Nearest Neighbors and achieved an accuracy of 97.1% However, this study only evaluated the transfer learning accuracy of one model where our study looked at seven different pre-trained models to determine the best one for

brain tumor classification. We also evaluated the effect of training the base layers or freezing the layers on the classification accuracy and found that ResNet18 and ShuffleNet had a higher accuracy of 97.86%. Swati et al. (11) proposed a block-wise fine-tuned network based on transfer learning to classify the CE-MRI dataset. The proposed method achieved a 94.82% classification accuracy. Toğaçar et al. (13) proposed a new convolutional neural network called BrainMRNet with an architecture built on attention modules and hypercolumn techniques. Their network was successful with a 96.05% accuracy. However, they only classified the MRI images into two classes of tumor or no tumor, whereas our study distinguished between the different types of brain tumors and achieved similar accuracies.

A limitation of this work was the amount of computational power available. All the code for this work was run on Google Colaboratory, which provides a limited 8 GB of RAM. This resulted in longer run times, as well as having to rerun the code due to occasional crashes. Additionally, a limitation of transfer learning is that the efficiency of the original model can carry over to the subsequent models. If the initial model has suboptimal hyperparameters, the data is overfitted, or an inadequate number of epochs are used to train the initial dataset, then these deficiencies will also apply to the second dataset. However, training the convolutional base can correct some of these problems.

Future extensions of this work should explore implementing machine learning to classify subtypes of each brain tumor type. Additionally, future works could expand on the machine learning methods used in this study by implementing Support Vector Machines (SVM) or K-Nearest Neighbors (KNN) to improve the model accuracies.

Conclusion

Transfer learning can be used to accurately classify brain tumors using a small dataset of MRI images. Training the convolutional base of the CNN models yields higher accuracies compared to freezing the convolutional base. However, it takes longer to run and requires more computational power. The models with the highest classification accuracies are ResNet18 and ShuffleNet when the convolutional base is trained. ResNet18 has the highest recall scores, while the precision and F1 scores are the same for both models. The fastest model with both methods is ShuffleNet.

Acknowledgement

I would like to thank my mentor, Oscar O'Rahilly, for his guidance, support, and encouragement throughout this research project.

References

1. Zuo, H., Fan, H., Blasch, E., & Ling, H. (2017). Combining convolutional and recurrent neural networks for human skin detection. *IEEE Signal Processing Letters*, *24*(3), 289–293. https://doi.org/10.1109/lsp.2017.2654803

2. Charron, O., Lallement, A., Jarnet, D., Noblet, V., Clavier, J.-B., & Meyer, P. (2018). Automatic detection and segmentation of brain metastases on multimodal MR images with a deep convolutional neural network. *Computers in Biology and Medicine*, *95*, 43–54. https://doi.org/10.1016/j.compbiomed.2018.02.004

3. Li, Y., Song, K., Sun, Y., & Zhu, L. (2021). FrequentNet : A novel interpretable deep learning model for image classification. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3895462

4. Cheng, K., Tahir, R., Eric, L. K., & Li, M. (2020). An analysis of generative adversarial networks and variants for image synthesis on MNIST dataset. *Multimedia Tools and Applications*, *79*(19-20), 13725–13752. https://doi.org/10.1007/s11042-019-08600-2

5. Kadam, S. S., Adamuthe, A. C., & Patil, A. B. (2020). CNN model for Image Classification on Mnist and fashion-mnist dataset. *Journal of Scientific Research*, *64*(02), 374–384. https://doi.org/10.37398/jsr.2020.640251

6. Rahaman, M. M., Li, C., Yao, Y., Kulwa, F., Rahman, M. A., Wang, Q., Qi, S., Kong, F., Zhu, X., & Zhao, X. (2020). Identification of COVID-19 samples from chest X-ray images using Deep learning: A comparison of transfer learning approaches. *Journal of X-Ray Science and Technology*, *28*(5), 821–839. https://doi.org/10.3233/xst-200715

7. Burdick, J., Marques, O., Weinthal, J., & Furht, B. (2017). Rethinking skin lesion segmentation in a convolutional classifier. *Journal of Digital Imaging*, *31*(4), 435–440. https://doi.org/10.1007/s10278-017-0026-y

8. Chen, Q., Hu, S., Long, P., Lu, F., Shi, Y., & Li, Y. (2019). A transfer learning approach for malignant prostate lesion detection on multiparametric MRI. *Technology in Cancer Research & Treatment*, *18*, 153303381985836. https://doi.org/10.1177/1533033819858363

9. Lakhani, P. (2017). Deep convolutional neural networks for endotracheal tube position and X-ray image classification: Challenges and opportunities. *Journal of Digital Imaging*, *30*(4), 460–468. https://doi.org/10.1007/s10278-017-9980-7

10. Chelghoum, R., Ikhlef, A., Hameurlaine, A., & Jacquir, S. (2020). Transfer learning using convolutional neural network architectures for brain tumor classification from MRI images. *IFIP Advances in Information and Communication Technology*, 189–200. https://doi.org/10.1007/978-3-030-49161-1_17

11. Swati, Z. N., Zhao, Q., Kabir, M., Ali, F., Ali, Z., Ahmed, S., & Lu, J. (2019). Brain Tumor Classification for MR images using transfer learning and fine-tuning. *Computerized Medical Imaging and Graphics*, *75*, 34–46. https://doi.org/10.1016/j.compmedimag.2019.05.001

12. Deepak, S., & Ameer, P. M. (2019). Brain Tumor Classification using Deep CNN features via transfer learning. *Computers in Biology and Medicine*, *111*, 103345. https://doi.org/10.1016/j.compbiomed.2019.103345

13. Toğaçar, M., Ergen, B., & Cömert, Z. (2020). BrainMRNet: Brain tumor detection using magnetic resonance images with a novel convolutional neural network model. *Medical Hypotheses*, *134*, 109531. https://doi.org/10.1016/j.mehy.2019.109531

14. Sharif, M. I., Li, J. P., Khan, M. A., & Saleem, M. A. (2020). Active deep neural network features selection for segmentation and recognition of brain tumors using MRI images. *Pattern Recognition Letters*, *129*, 181–189. https://doi.org/10.1016/j.patrec.2019.11.019

15. Cheng, J., Huang, W., Cao, S., Yang, R., Yang, W., Yun, Z., Wang, Z., & Feng, Q. (2015). Enhanced performance of Brain Tumor Classification via tumor region augmentation and partition. *PLOS ONE*, *10*(10). https://doi.org/10.1371/journal.pone.0140381

16. Talo, M., Yildirim, O., Baloglu, U. B., Aydin, G., & Acharya, U. R. (2019). Convolutional neural networks for multi-class brain disease detection using MRI images. *Computerized Medical Imaging and Graphics*, *78*, 101673. https://doi.org/10.1016/j.compmedimag.2019.101673

17. Bhuvaji, S., Kadam, A., Bhumkar, P., Dedge, S., & Kanchan, S. (2020, May 24). *Brain tumor classification (MRI)*. Kaggle. Retrieved February 10, 2023, from https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri?select=Testing

18. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/cvpr.2016.90

19. Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv*. https://doi.org/arXiv:1409.1556

20. Szegedy, C., Wei Liu, Yangqing Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/cvpr.2015.7298594

21. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected Convolutional Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/cvpr.2017.243

22. Zhang, X., Zhou, X., Lin, M., & Sun, J. (2018). ShuffleNet: An extremely efficient convolutional neural network for mobile devices. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/cvpr.2018.00716

23. Ismael, M. R., & Abdel-Qader, I. (2018). Brain Tumor Classification via statistical features and back-propagation neural network. *2018 IEEE International Conference on Electro/Information Technology (EIT)*. https://doi.org/10.1109/eit.2018.8500308

24. Afshar, P., Plataniotis, K. N., & Mohammadi, A. (2019). Capsule Networks for brain tumor classification based on MRI images and coarse tumor boundaries. *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. https://doi.org/10.1109/icassp.2019.8683759