

Analyzing the Effects of Non-Generative Augmentation on Automated Classification of Brain Tumors

Adam Rolander

Received June 04, 2023

Accepted August 19, 2023

Electronic access September 15, 2023

Due to the devastating effects that brain tumors have on the body, early classification is crucial in reducing cancer mortality, improving quality of life, and developing a treatment plan. Although biopsies are often used for diagnosis, brain tumors can be classified using techniques like MRI, a process that can be automated. In this study, I developed a Convolutional Neural Network (CNN) model to classify four classes of brain tumors – gliomas, meningiomas, pituitary tumors, and no tumors. I then observed how different methods of data augmentation affected my model's capabilities when used individually and in combination with each other. This study sought to discover which method of augmentation was most effective at doing so. I tested several of the most commonly used methods of augmentation, including horizontal and vertical translations, reflections, rotations, and zooming, in different combinations over six trials. In doing so, I found that the model using no augmentation obtained a classification accuracy of 93.02%. The most successful trial, however, utilized random horizontal and vertical translations, which resulted in a classification accuracy of 95.80%. These results demonstrate the efficacy of augmentation in improving CNN models and show that translations were most successful at improving my model.

Introduction

The nervous system is a fundamental organ system in the human body whose primary constituent is the brain¹. The brain is responsible for understanding and responding to stimuli as well as controlling the body by transmitting electrical and chemical signals. Due to its significance in regulating bodily functions, abnormalities that affect the brain can have devastating impacts on one's overall health and quality of life. Among the most severe abnormalities that affect the brain are brain tumors, which the American Association of Neurological Surgeons defines as abnormal masses of tissue that grow uncontrollably within the brain¹. Brain tumors are classified as primary if they originate within the brain's tissue². Primary tumors are further categorized as benign (non-cancerous and typically not harmful) or malignant (cancerous)³. By contrast, metastatic brain tumors, or secondary brain tumors, are tumors originating outside of the brain that spread to the brain through the bloodstream. Metastatic tumors are always malignant, and they affect one in four people with cancer⁴.

Three of the four most common primary brain tumors are gliomas, meningiomas, and pituitary tumors⁴. Gliomas are malignant tumors, and because of their aggressive nature, early detection is a critical component of reducing mortality⁵. By contrast, meningiomas are typically benign tumors that develop in the membrane that surrounds the brain and spinal

cord. Finally, pituitary tumors are lumps that grow on the pituitary gland. Like meningiomas, pituitary tumors are typically benign; however, they can cause other medical issues due to their location, including issues with hormone production. **Figure 1** below shows the frequency of each tumor class in adult patients.

Because of the physiological changes that brain tumors cause, they can be identified with imaging technologies like magnetic resonance imaging (MRI), computerized tomography (CT), and positron emission tomography (PET)⁶. MRI is the most commonly used technology for imaging because it is non-invasive and provides the most detailed information about the size, shape, position, and type of tumors⁶. However, according to the Radiological Society of North America, human analysis of MRI images is a laborious process that provides ample opportunity for error⁷. Also, in addition to medical imaging, patients often undergo a surgical biopsy to classify their tumor, which is an invasive and inherently risky procedure.

Due to the potential for erroneous human analysis of MRIs and the risks of obtaining a biopsy analysis of brain tumors, automated image classification algorithms play an integral role in assisting physicians with less time-consuming, more accurate brain tumor diagnoses. In a 2015 study by Cheng et al., the researchers were able to automate the classification of gliomas, meningiomas, and pituitary tumors with an accuracy of 91.28% by augmenting the tumor regions of interest using image dilation⁸.

Tumor Class Frequency in Adult Patients

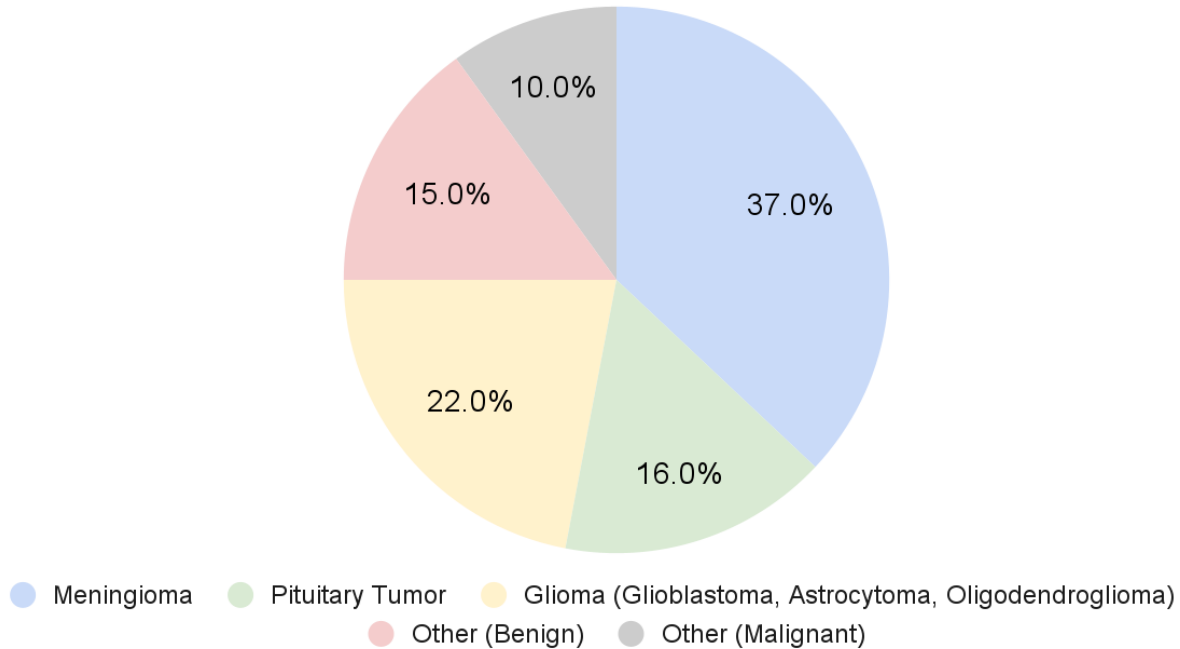


Fig. 1 Display of tumor class frequency in adult patients. Adapted from Lapointe et al.⁶

However, new technologies such as the Convolutional Neural Network (CNN) have proven to be even more successful. For example, a 2021 study by Francisco et al. achieved a classification accuracy of 97.3 % by using a multi-pathway CNN⁹. Furthermore, CNNs require less image preprocessing than other algorithms, allowing for greater efficiency and applicability to more fields¹⁰.

Despite their apparent benefits, Convolutional Neural Networks are limited in their abilities based on the size of the dataset used for training. This poses a problem for medical classification tasks because medical imaging datasets are often small due to rigorous standards for the acquisition of data. This issue presents the need for data augmentation in medical machine-learning tasks to improve the accuracy and efficiency of CNN models.

Image data augmentation is the process by which an image dataset is diversified by modifying the original images through various methods. In classification tasks, commonly used augmentation methods include horizontal and vertical translations, horizontal reflections, random rotations, and random zooms¹¹.

Horizontal and vertical translations allow for an image to be shifted up, down, left, or right by a specified or random

amount, and horizontal reflections enable the images to be randomly flipped across their y-axis.

Rotation augmentations rotate images by a random degree value in a specified interval, and zooms cause an image to appear larger or smaller¹². Such methods play a crucial role in classification tasks, as they help to diversify a dataset and prevent overfitting, which can increase a model's accuracy. Each of the above methods was employed in this study.

In this study, I developed a custom Convolutional Neural Network model to assess the possibility of automating brain tumor classification with deep learning, and then I analyzed how different methods of data augmentation impacted my model's capabilities when used separately and in coalition with one another. The objective of this research was to discover how accurately brain tumors could be diagnosed through automated image classification, and to determine if different data augmentation strategies had the ability to improve the classification results.

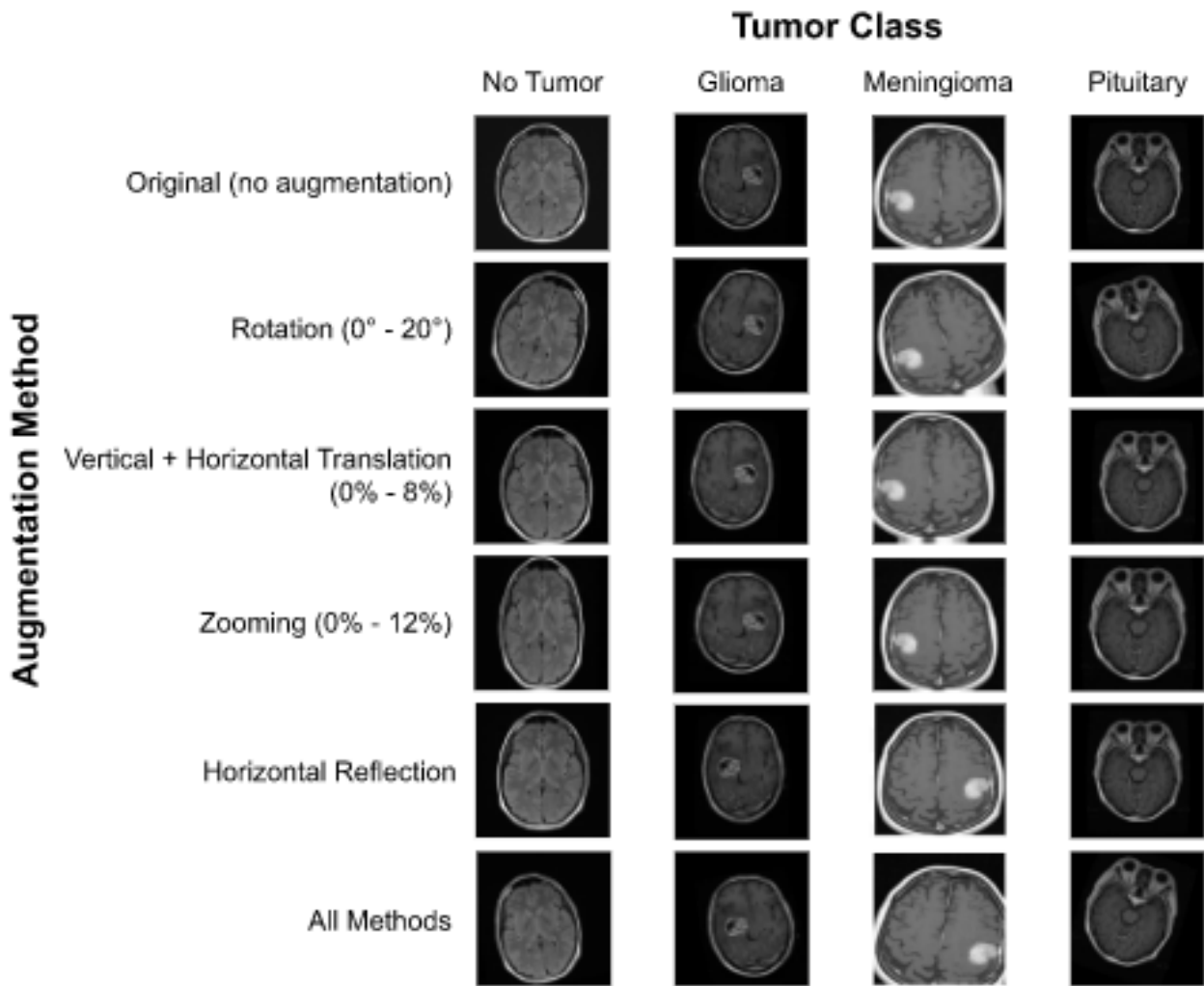


Fig. 2 Examples of no tumors (left column), gliomas (left-middle column), meningiomas (right-middle column), and pituitary tumors (right column) in the axial plane with different methods of data augmentation, as were implemented in the various CNN trials. The original image for each tumor class is shown in the top row. These images were generated and hand-selected outside of the CNN trials. They are included solely for the sake of visualization.

Results

Data Augmentation

I hypothesized that my CNN model would obtain the highest classification accuracy when utilizing each method of data augmentation simultaneously because this would provide the most diverse training dataset and most significantly decrease the chances of overfitting.

Several methods of data augmentation were applied to the images in the dataset in order to prevent overfitting and increase the classification accuracy of my CNN model. Exam-

ples of augmented images using each method of data augmentation for each tumor class in the axial plane are shown in **Figure 2**. These images were all generated according to the parameters outlined in the Methods section.

CNN Classification Results

No Augmentation

I implemented and evaluated the CNN model I developed over six trials, each of which used different methods of data augmentation in order to assess their independent and collective

| Tumor Class | Recall% | Precision % | F1-Score % | Accuracy% |
|-------------|---------|-------------|------------|-----------|
| Glioma | 86.49 | 99.34 | 92.47 | - |
| Meningioma | 94.83 | 80.21 | 86.91 | - |
| No Tumor | 94.89 | 97.20 | 96.06 | - |
| Pituitary | 95.90 | 97.44 | 96.66 | - |
| Average | 93.03 | 93.55 | 93.03 | 93.02 |

Table 1 CNN classification results for no data augmentation CNN trial

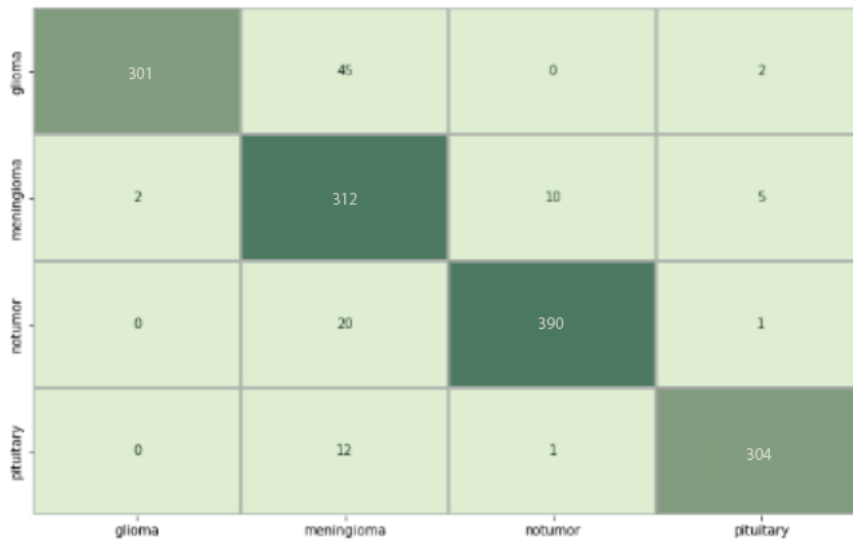


Fig. 3 Confusion matrix for no augmentation CNN model. Rows denote the actual number of images in a class, and the columns denote the predicted number of images in a class

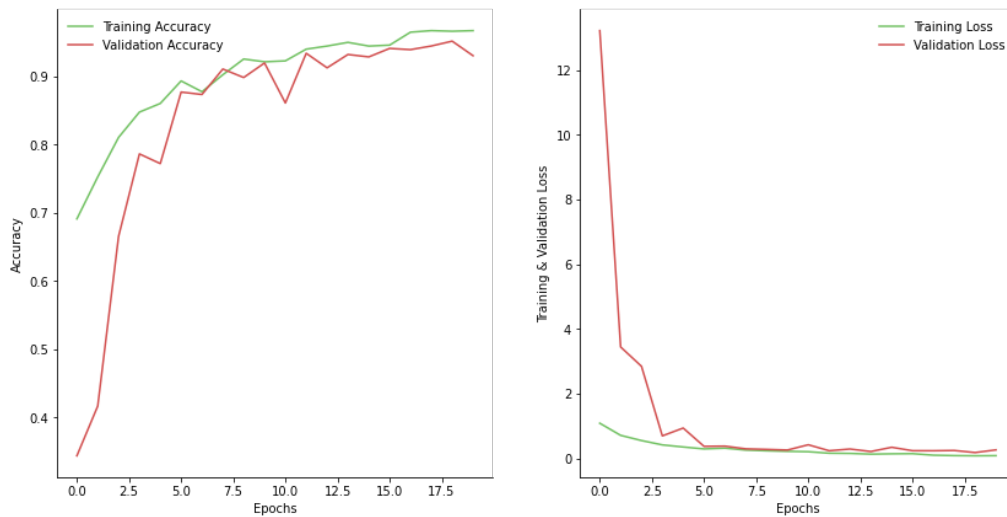


Fig. 4 Training and validation accuracy vs epochs for the no-augmentation CNN model are displayed on the left, and training and validation loss vs epochs are displayed on the right

efficacy. The first CNN trial was tested with no data augmentation in order to establish a control. Its classification results are displayed in Table 1 below.

Despite utilizing no data augmentation, the CNN model used in this study was able to classify brain tumors correctly 93.02% of the time and obtained an average F1 score of 93.03%. This is an impressive result, most likely attributed to the relatively large size of the dataset, but the implementation of data augmentation was able to improve CNN's classification accuracy significantly.

Figures 3 and 4 below display the CNN model's confusion matrix and accuracy/loss plots generated from the testing dataset.

Rotational Augmentation

Table 2 displays the classification results for the CNN model utilizing rotational data augmentation. Examples of images after rotational data augmentation are displayed in **Figure 2** above.

When utilizing rotational data augmentation, which used the same number of images in the training and testing datasets, my CNN model's metrics decreased by a slight margin compared to the no-augmentation model.

In the trial, rotational data augmentation appears to have been successful in some aspects when classifying no tumor and pituitary tumor images. However, the model struggled to classify gliomas and meningiomas successfully when rotational augmentation was used, which ultimately decreased the overall success and resulted in an overall classification accuracy of 91.17% and an average F1 score of 90.99%.

Figures 5 and 6 display the CNN model's confusion matrix and accuracy/loss plots generated from the testing dataset.

Translational Augmentation

Table 3 displays the classification results for the CNN model utilizing both horizontal and vertical translations. Examples of images after translational data augmentation are displayed in **Figure 2** above.

When utilizing translations as data augmentation, which used the same number of images for training and testing, my CNN model's metrics increased significantly from a classification accuracy of 93.02% and an average F1 score of 93.03% in the no-augmentation to an accuracy of 95.80% and an average F1 score of 95.66%. In the trial, translational augmentation increased the model's capabilities in nearly every category, especially for no tumor and pituitary tumor images. This implementation resulted in the most successful classification metrics in terms of overall accuracy and average F1 score, even outperforming the CNN model utilizing all methods of augmentation.

Figures 7 and 8 display the CNN model's confusion matrix and accuracy/loss plots generated from the testing dataset.

Zooming Augmentation

Table 4 displays the classification results for the CNN model utilizing zooming. Examples of images after zooming for data augmentation are displayed in **Figure 2** above.

When utilizing zooming for data augmentation, the CNN's metrics decreased slightly when compared to the no-augmentation model. In the trial, zooming caused the model to struggle to classify images in each class as successfully as the no-augmentation model, which resulted in a slight decrease in overall accuracy and average F1 score. For this trial, the model obtained an accuracy of 92.81% compared to 93.02% in the no-augmentation model and an average F1 score of 92.66% compared to 93.03% in the no-augmentation model.

Figures 9 and 10 display the CNN model's confusion matrix and accuracy/loss plots generated from the testing dataset.

Horizontal Reflection Augmentation

Table 5 displays the classification results for the CNN model utilizing random horizontal reflections (occurring with a theoretical probability of 50%) for data augmentation. Examples of images after being horizontally reflected are displayed in **Figure 2**.

When utilizing random horizontal reflections as data augmentation, the CNN achieved similar but slightly increased results compared to the model utilizing no data augmentation. In the trial, horizontal reflections increased the model's overall accuracy from 93.02% to 93.17% but caused a slight decrease in the average F1 score from 93.03% to 92.85%. As with many of the previous methods of data augmentation, horizontal reflections caused the model to be more successful at classifying no tumor and pituitary tumor images but with similar or slightly less success when dealing with MRIs of gliomas and meningiomas.

Figures 11 and 12 display the CNN model's confusion matrix and accuracy/loss plots generated from the testing dataset.

Combined Data Augmentation

Table 6 displays the classification results for the CNN model utilizing horizontal and vertical translations, zooming, rotation, and horizontal reflection. Examples of images after rotational data augmentation are displayed in the bottom row of **Figure 2**

When utilizing all of the described methods of data augmentation, the CNN's abilities increased significantly, albeit not as significantly as I observed when using translational data

| Tumor Class | Recall% | Precision % | F1-Score % | Accuracy% |
|-------------|---------|-------------|------------|-----------|
| Glioma | 81.61 | 98.61 | 89.31 | - |
| Meningioma | 87.23 | 80.39 | 83.67 | - |
| No Tumor | 98.30 | 92.03 | 95.06 | - |
| Pituitary | 96.53 | 95.33 | 95.93 | - |
| Average | 90.92 | 91.59 | 90.99 | 91.17 |

Table 2 CNN classification results for rotational augmentation CNN trial



Fig. 5 Confusion matrix for rotational augmentation CNN model

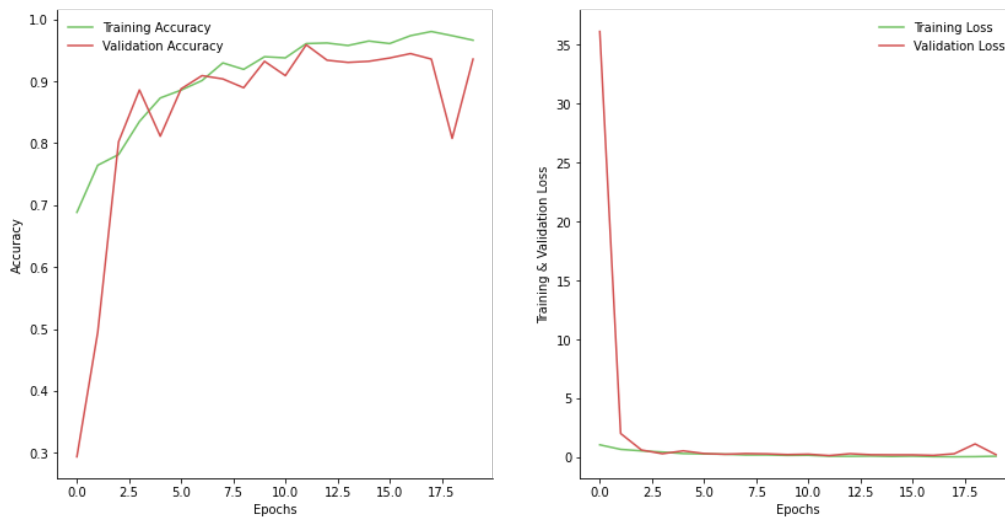


Fig. 6 Training and validation accuracy vs epochs for the rotational augmentation CNN model are displayed on the left, and training and validation loss vs epochs are displayed on the right

| Tumor Class | Recall% | Precision% | F1-Score% | Accuracy% |
|-------------|---------|------------|-----------|-----------|
| Glioma | 95.97 | 94.35 | 95.15 | - |
| Meningioma | 89.36 | 93.33 | 91.30 | - |
| No Tumor | 98.78 | 97.36 | 98.06 | - |
| Pituitary | 98.42 | 97.81 | 98.11 | - |
| Average | 95.63 | 95.71 | 95.66 | 95.80 |

Table 3 CNN classification results for translational CNN trial



Fig. 7 Confusion matrix for translational augmentation CNN model

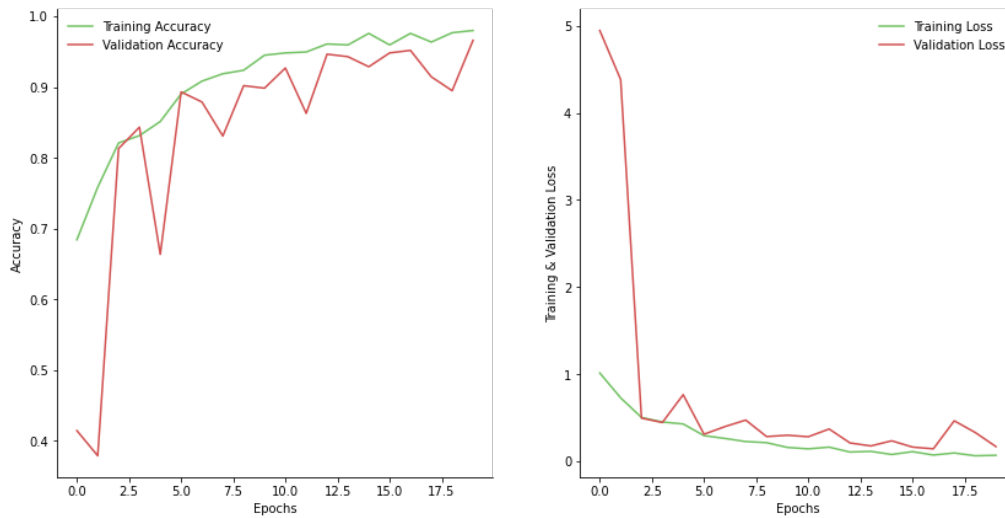


Fig. 8 Training and validation accuracy vs epochs for the translational augmentation CNN model are displayed on the left, and training and validation loss vs epochs are displayed on the right.

| Tumor Class | Recall% | Precision% | F1-Score % | Accuracy% |
|-------------|---------|------------|------------|-----------|
| Glioma | 86.21 | 97.40 | 91.46 | - |
| Meningioma | 87.84 | 85.00 | 86.38 | - |
| No Tumor | 99.27 | 92.52 | 95.78 | - |
| Pituitary | 96.85 | 97.15 | 97.00 | - |
| Average | 92.54 | 93.02 | 92.66 | 92.81 |

Table 4 CNN classification results for zooming CNN trial



Fig. 9 Confusion matrix for zooming-based augmentation CNN model

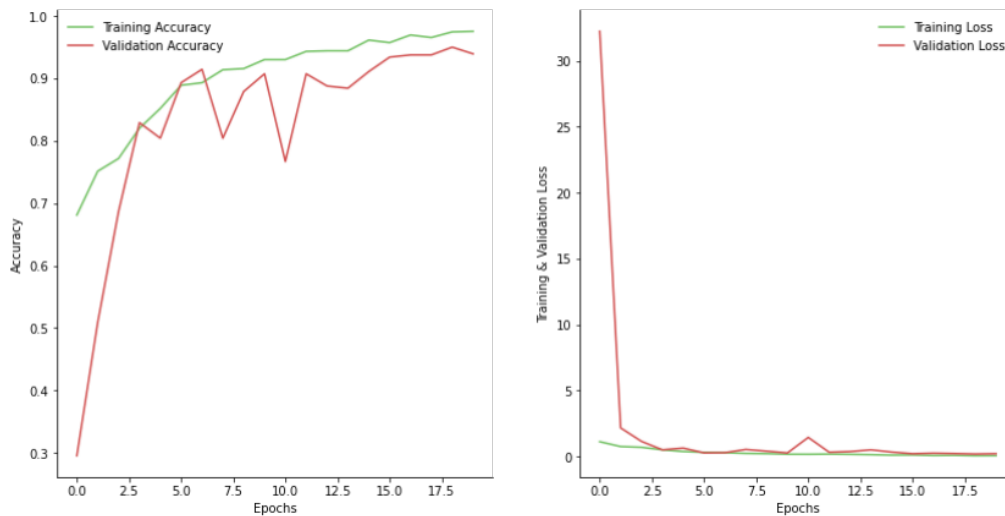


Fig. 10 Training and validation accuracy vs epochs for the zooming-based augmentation CNN model are displayed on the left, and training and validation loss vs epochs are displayed on the right

| Tumor Class | Recall% | Precision % | F1-Score % | Accuracy% |
|-------------|---------|-------------|------------|-----------|
| Glioma | 94.25 | 93.18 | 93.71 | - |
| Meningioma | 82.07 | 90.30 | 85.99 | - |
| No Tumor | 96.84 | 96.84 | 96.84 | - |
| Pituitary | 98.74 | 91.25 | 94.85 | - |
| Average | 92.98 | 92.89 | 92.85 | 93.17 |

Table 5 CNN classification results for horizontal reflection CNN trial



Fig. 11 Confusion matrix for horizontal reflection augmentation CNN model

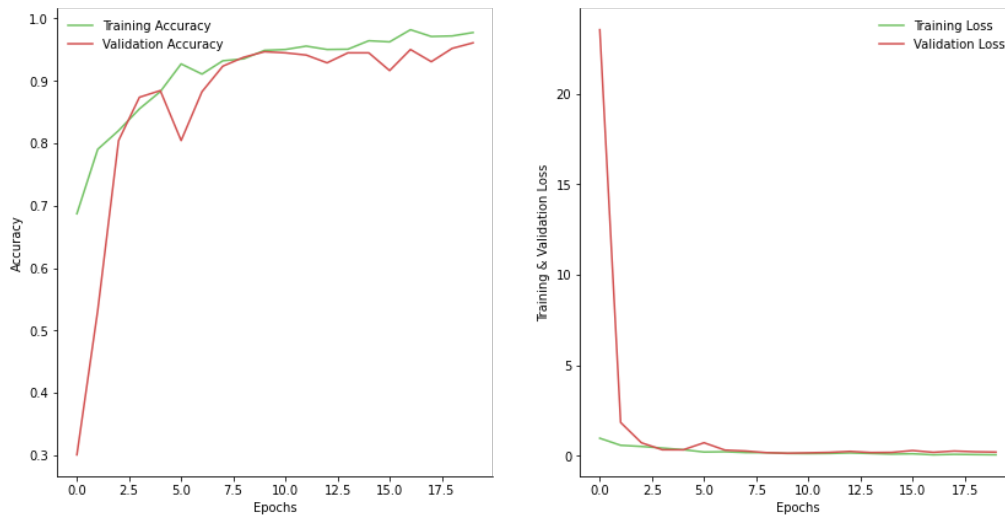


Fig. 12 Training and validation accuracy vs epochs for the horizontal reflection augmentation CNN model are displayed on the left, and training and validation loss vs epochs are displayed on the right

| Tumor Class | Recall% | Precision% | F1-Score% | Accuracy% |
|-------------|---------|------------|-----------|-----------|
| Glioma | 92.82 | 95.28 | 94.03 | - |
| Meningioma | 86.63 | 91.05 | 88.79 | - |
| No Tumor | 98.78 | 95.08 | 96.89 | - |
| Pituitary | 97.79 | 95.09 | 96.42 | - |
| Average | 94.01 | 94.13 | 94.03 | 94.23 |

Table 6 CNN classification results for combined augmentation CNN trial



Fig. 13 Confusion matrix for combined augmentation CNN model

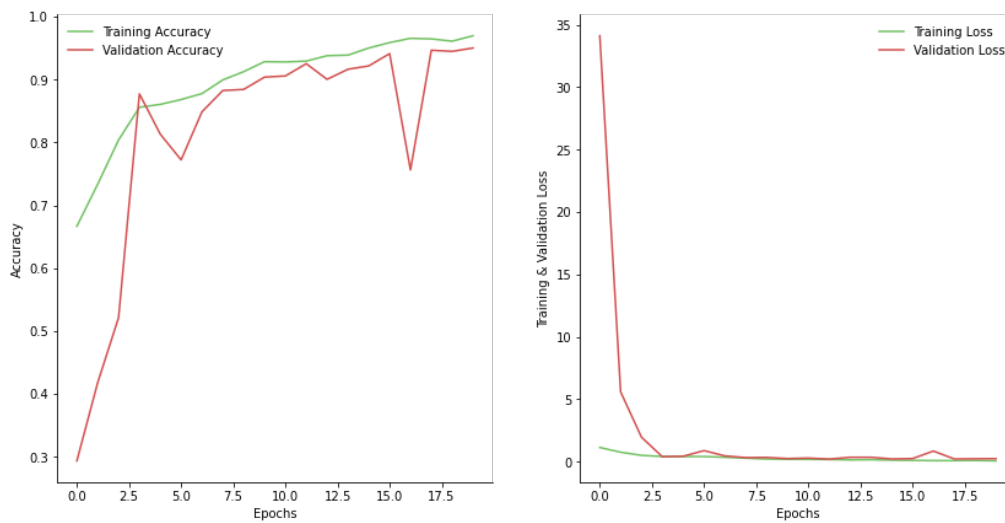


Fig. 14 Training and validation accuracy vs epochs for the combined augmentation CNN model are displayed on the left, and training and validation loss vs epochs are displayed on the right

augmentation by itself. In the trial, the combination of the data augmentation methods achieved the second most successful results in terms of average F1 score and overall classification accuracy. While the no augmentation model obtained an accuracy of 93.02% and an average F1 score of 93.03%, the trial utilizing all of the methods of data augmentation resulted in an accuracy of 94.23% and an average F1 score of 94.03%, suggesting that data augmentation was successful at preventing overfitting in the CNN trials which ultimately resulted in greater classification capabilities.

Figures 13 and 14 display the CNN model's confusion matrix and accuracy/loss plots generated from the testing dataset, respectively.

Cumulative Results

Table 7 displays the CNN classification results for all six trials.

Discussion

In this study, I developed a Convolutional Neural Network (CNN) model to classify four classes of brain tumors (gliomas, meningiomas, no tumors, and pituitary tumors) from MRI images I obtained from the Kaggle database. The CNN model was trained on whole images, so the only preprocessing I performed was image resizing. I considered implementing skull stripping as further preprocessing but decided that doing so could introduce irregularity into the data, so I decided against it in the interest of maintaining more controls throughout trials. I then evaluated my CNN model over six trials when utilizing various methods of non-generative data augmentation, which was a crucial aspect of improving the CNN's classification accuracy because it diversified the dataset and may have helped to prevent overfitting. There is an inherent need for such capabilities in classification tasks because of the lack of large imaging data sets, as was the case with the dataset used in this study, which only contained 7023 images. The methods of augmentation I implemented included horizontal and vertical translations, horizontal reflections, rotations, and zooming.

I began by evaluating my CNN model without data augmentation to establish a control. Then, I tested each method of augmentation individually and together. That is, the second trial only used rotational augmentation, the third trial only used horizontal and vertical translations as augmentation, the fourth trial only used zooming augmentation, the fifth trial only used horizontal reflections as augmentation, and the final trial used all of the discussed methods of augmentation. After the trials, the CNN models were evaluated according to four metrics (accuracy, precision, recall, and F1 score), with the primary metrics being overall accuracy and average F1 score.

The no-augmentation model was run to establish a control, and it achieved an overall classification accuracy of 93.02% and an average F1 score of 93.03%. When utilizing rotational augmentation, my model achieved an overall classification accuracy of 91.17% and an average F1 score of 90.99%, suggesting that rotational augmentation hindered the model's abilities. I also observed a decrease in the model's capabilities with the implementation of zooming augmentation by itself, as it obtained an accuracy of 92.81% and an average F1 score of 92.66%.

When implementing horizontal reflections as augmentation, my model improved in overall accuracy to 93.17% but decreased in average F1 score to 92.85%. However, the most successful results were achieved with the implementation of translational augmentation by itself, as well as the combined model, which utilized every method of augmentation I discussed. The translational augmentation model obtained a classification accuracy of 95.80% and an average F1 score of 95.66%, the highest I observed. However, the model utilizing the combined methods of augmentation achieved similar but slightly less successful results, with an accuracy of 94.23% and an average F1 score of 94.03%. Although these results are less successful than those achieved by studies using transfer learning models, they constitute successful results for my custom CNN model and demonstrate the possibility of using Convolutional Neural Networks with data augmentation to assist physicians with more accurate, less time-consuming brain tumor diagnoses^{9, 13, 14, 15}.

The results I observed partially support my hypothesis that the combined augmentation model would be most successful at classifying MRI images. My hypothesis was correct in that the combined model showed a significant improvement in classification capability for each class compared to the no-augmentation model. However, the model using only horizontal and vertical translations for augmentation was the most successful. This was unexpected, as data augmentation is most commonly used as a combination of some or all of the methods I tested, but my results show that it was more effective to use translational augmentation by itself for my custom CNN model. I believe that the combined augmentation model was less successful than the translational augmentation due to the fact that both zooming and rotational augmentation decreased my CNN model's control-trial accuracy when used independently. Therefore, when zooming and rotational augmentation were used together with all the other augmentation methods in the combined model, the overall success of the model was hindered below the success of the translational model. I think that zooming and rotational augmentation were ultimately detrimental to the model's success because they created unrealistic alterations to the training data. For example, zooming augmentation might not have been useful because the vast majority of the testing images were equally zoomed out, so training

| Classification Trial | Recall% | Precision% | F1 – Score% | Accuracy% |
|---------------------------------------|---------|------------|-------------|-----------|
| Control (No Augmentation) | 93.03 | 93.55 | 93.03 | 93.02 |
| Rotational Augmentation | 90.92 | 91.59 | 90.99 | 91.17 |
| Translational Augmentation | 95.63 | 95.71 | 95.66 | 95.80 |
| Zooming Augmentation | 92.54 | 93.02 | 92.66 | 92.81 |
| Horizontal Reflection Augmentation | 92.98 | 92.89 | 92.85 | 93.17 |
| Combined Augmentation | 94.01 | 94.13 | 94.03 | 94.23 |

Table 7 Cumulative CNN classification results. The greatest accuracy achieved is bolded, and the lowest accuracy achieved is italicized.

the model on many different levels of zoom may have confused the model, leading to more harm than good.

Likewise, none of the testing images were rotated to a significant level, so training the model on rotated images may have led to confusion as well. On the contrary, I think the translational augmentation led to very accurate data transformations, as the tumor in each of the testing images was in a slightly different location. Therefore, horizontal and vertical translations of training images allowed the CNN to become more familiar with different possible tumor locations that it would encounter during testing. Horizontal reflection augmentation showed only a marginal increase in accuracy, and notably a slight decrease in average F1 score, so it is likely that this method of augmentation had little to no effect on the success of the CNN model.

Further research on the topic of data augmentation combinations should seek to vary and optimize the parameters of the methods I used. Additionally, more trials should be run that make use of different methods of augmentation to assess the efficacy of different combined alterations to training data. The goal of maximizing CNN classification capability through various augmentation strategies should be applied to different classification tasks and different CNN architectures, as results are likely to vary depending on the specific nuances of each research objective.

Methods

Data Acquisition

The imaging data used in this study consists of 7023 MRI images collected from the Kaggle Database and from Nanfang Hospital, Guangzhou, China and General Hospital, Tianjin Medical University, China¹⁶.

This dataset consists of 1621 glioma images, 1645 meningioma images, 2000 no tumor images, and 1757 pituitary tumor images. The images were collected in three planes: sagittal, axial, and coronal. Examples of the different tumors in their respective planes are displayed in **Figure 15**

Image Preprocessing

In order to preprocess the data used in this study, the MRI images in the dataset were resized to 224x224 pixels since they were of different sizes. This image size was chosen in order to limit the CNN model's training time while still providing adequate resolution to the neural network, and was the chosen size for many successful past studies, including Kang et al.¹³

Another method of image preprocessing, skull stripping, was considered but not applied. Skull stripping is the process of removing non-cerebral tissues from MRI images and is frequently used in neuroimaging tasks to improve the consistency of the data. However, I decided not to skull strip the data because doing so can be inaccurate, either unintentionally removing brain tissue from the image or not entirely removing the intended tissue, and I did not think it would significantly benefit the desired classification task¹⁷. The final task in data preprocessing was to split the dataset into testing and training sets. The frequency with which images are found in these sets is displayed in **Table 8**

Data Augmentation

In order to test how data augmentation affected my model's accuracy, I established a control by training my custom model without augmentation and observed how the model's suc-

| Dataset | Image Samples | Percentage of Original Data |
|--------------|---------------|-----------------------------|
| Training Set | 5618 | 79.99 |
| Testing Set | 1405 | 20.01 |
| Total | 7023 | 100.00 |

Table 8 Distribution of data among training and testing sets. The training and testing sets were split with 80% and 20% of the data, respectively. The dataset was concatenated and randomly shuffled prior to being split into training and testing sets, so the exact number of images from each tumor class in each set is unknown. However, in the training set, there were: 1297 glioma images, 1316 meningioma images, 1600 no-tumor images, and 1405 pituitary tumor images, on average. In the testing set, there were: 324 glioma images, 329 meningioma images, 400 no-tumor images, and 352 pituitary tumor images, on average.

cess changed when I implemented the augmentation strategies listed below, both separately and together. Also, it is important to note that the models utilizing augmented data were not trained on any of the original images, only the augmented variations. In this study, I augmented each image in the original dataset while maintaining the dataset’s original size for every experiment. This was accomplished with the following methods using the Image Data Generator from the Keras library¹². Images were rescaled in every trial.

1 Images were rotated, either clockwise or counterclockwise, by a randomly generated value between 0 and 20 degrees, inclusive.

2 Images were horizontally translated, either left or right, by a randomly generated value between 0% and 8% of the image’s total width.

3 Images were vertically translated, either up or down, by a randomly generated value between 0% and 8% of the image’s total height.

4 Images were zoomed, either in or out, by a randomly generated value between 0% and 12%.

5 Finally, images in the dataset were randomly reflected over their y-axis, resulting in a horizontal flip with a theoretical frequency of 50/50%.

During certain augmentations, primarily horizontal and vertical transformations, the images were shifted, which added sections of empty space to their background¹². This issue was resolved by filling the empty space with black pixels, which did not affect the images because the original MRI images had black backgrounds as well.

Convolutional Neural Network (CNN) Model CNN Architecture

In order to classify the MRI Images used in this study, I developed a custom Convolutional Neural Network model that intakes and feeds an image through Feature Learning and Dense Layer Blocks until it is classified as glioma, meningioma, no tumor, or pituitary tumor⁸. The Feature Learning Block consisted of two components, Convolutional Layers, and Max

Pooling Layers, and was repeated three times before the Dense Layer Block was implemented. Each Convolutional Layer used a kernel size of (3, 3), strides of (1, 1), valid padding, and the RELU activation function. The first Convolutional Layer learned from 128 filters, the second layer learned from 64 filters, and the third learned from 32 filters.

Following the Convolutional Layer in each iteration through the Feature Learning Block was a pooling layer, which created subsamples of the previous layer by down-sampling the input’s width and height. Each Pooling Layer used a pool size of (2, 2), strides of (2, 2), and valid padding. This resulted in an output half the size of the input. Batch Normalization followed each Pooling Layer as well.

After the Feature Learning Block, the Concatenated Dense Layers Block was implemented. This block consisted of a Flattening layer which consolidated the final Pooling layer’s four-dimensional array output into a two-dimensional array with a shape of (32, 21632). This layer preceded two Fully Connected (FC) layers that used the RELU activation function. The first FC layer was followed by a Dropout Layer utilizing 20% dropout. After the second FC layer, a final FC layer using the Softmax activation function was used to determine a classification output for the original image. The Classification Output consisted of four possible options, glioma, meningioma, no tumor, or pituitary tumor, which are represented by the values 0, 1, 2, and 3, respectively.

In each Convolutional Layer, the kernel size of (3, 3) was chosen because the input images were fairly small (224 x 224 px), and a larger kernel size may have too broadly represented the input data during convolutions. While there are many valid options for kernel size, I chose (3, 3) in order to reduce the chances of underfitting the model during training. For the stride parameter in the Convolutional Layers, I chose strides of the minimum size, (1, 1), to dictate the number of pixels the kernel passes over during each convolution. Increasing the stride length would have resulted in a smaller output matrix at the end of each layer. While this would have reduced complexity, it would have also reduced the amount of input information available to each subsequent layer, possibly lim-

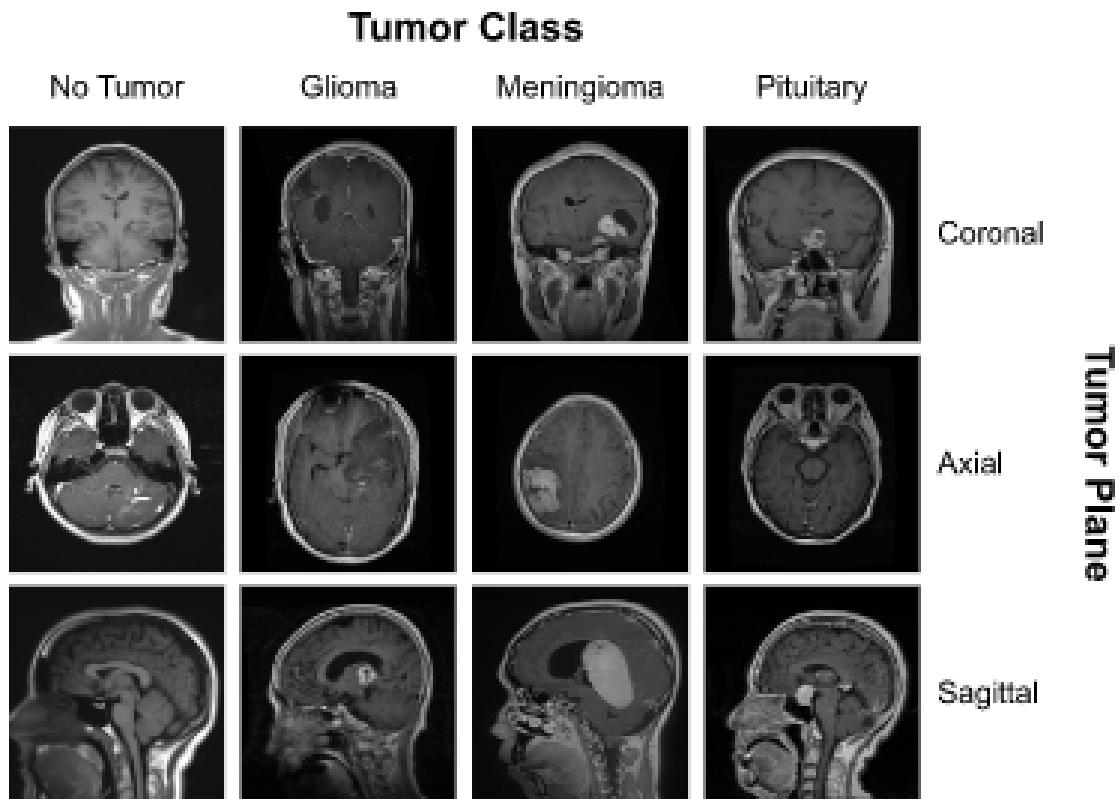


Fig. 15 MRI imaging of no tumors, gliomas, meningiomas, and pituitary tumors in different planes.

iting model success. For padding, I chose to use the default parameter, valid padding, which does not add padding to images following convolution.

Padding can be useful to help models detect features near the borders of input images, but since the MRI images in this study did not contain valuable data near the borders, padding was not added. Regarding pool parameters in the Pooling Layers, a pool size of (2, 2) and the Max Pooling operation are very common, and were chosen based on success in studies by Badža et al., and Kang et al.^{10, 13}. Finally, 20% dropout was implemented in the Dense Layers to randomly eliminate 20% of nodes in order to reduce the chances of them becoming codependent. An overview of the CNN architecture used in this study is displayed in **Figure 16**.

CNN Implementation

The discussed CNN model was tested over six trials, as described in **Table 9**

My CNN model was implemented in all its trials using the Adam optimizer (no relation) with a learning rate of 1e-3 and was trained on 20 epochs with a batch size of 32 and had

| Classification Trial | Data Augmentation Used |
|----------------------|-------------------------------------|
| 1 | None (Control Trial) |
| 2 | Rotation |
| 3 | Horizontal and Vertical Translation |
| 4 | Zooming |
| 5 | Horizontal Reflection |
| 6 | All Methods |

Table 9 Description of augmentation methods used in six CNN classification trials

2,874,276 parameters. The CNN model was trained in Google Colaboratory on a personal computer with a single GeForce RTX 2070 SUPER graphics processing unit (GPU) and an 8-core Intel Core i7-9700F central processing unit (CPU).

Model Evaluation

The custom CNN was evaluated in its various trials according to accuracy, precision, recall, and F1 score. Precision, recall, and F1 scores were calculated for each tumor class and

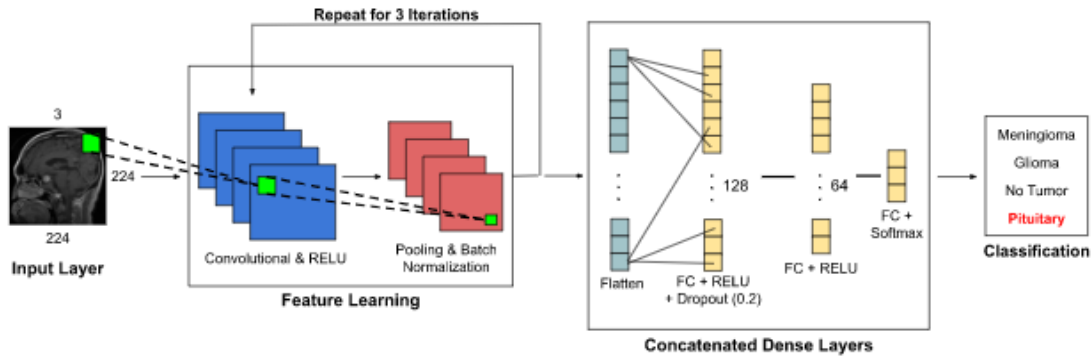


Fig. 16 Visual representation of Convolutional Neural Network Architecture, which consisted of the Input Layer, a Feature Learning Block, a Concatenated Dense Layers Block, and a Classification Block. The example image in the input layer is a Pituitary Tumor, as indicated by the output classification result in the Classification Block.

then averaged with equal weight, whereas accuracy was determined for the combined results. The formulas for these metrics are outlined in Equations (1 - 4) below^{10, 18, 19}.

$$\text{Accuracy} = \frac{\text{correct predictions}}{\text{Total predictions}} \quad (1)$$

$$\text{Precision} = \frac{\text{correct class predictions}}{\text{Total class predictions}} \quad (2)$$

$$\text{Recall} = \frac{\text{correct class predictions}}{\text{Total image in class}} \quad (3)$$

$$F1 - \text{Score} = 2 \times \frac{\text{Predictions} \times \text{Recall}}{\text{Predictions} + \text{Recall}} \quad (4)$$

Acknowledgments

I would like to acknowledge the Summer STEM Institute, its instructors, and Mr. Caswell for their instruction and support during my research.

References

- 1 American Association of Neurological Surgeons Editors, *A Neurosurgeon's Overview the Brain's Anatomy*, <https://www.aans.org/en/Patients/Neurosurgical-Conditions-and-Treatments/Anatomy-of-the-Brain>.
- 2 M. Markman, *Metastatic Brain Cancer*, <https://www.cancercenter.com/cancer-types/brain-cancer/types/metastatic-brain-cancer>.
- 3 MedlinePlus Editors, *Benign: MedlinePlus Medical Encyclopedia*, <https://medlineplus.gov/ency/article/002236.htm>.

- 4 American Association of Neurological Surgeons Editors, *Brain Tumors - Classifications, Symptoms, Diagnosis and Treatments*, <https://www.aans.org/en/Patients/Neurosurgical-Conditions-and-Treatments/Brain-Tumors>.
- 5 World Health Organization Editors, *Cancer*, <https://www.who.int/news-room/fact-sheets/detail/cancer>.
- 6 Cancer.Net Editorial Board, *Brain Tumor: Diagnosis*, <https://www.cancer.net/cancer-types/brain-tumor/diagnosis>.
- 7 R. Dargan, *Human Factors Drive Radiology Error Rates*, <https://www.rsna.org/news/2022/march/human-error-in-radiology>.
- 8 J. Cheng, W. Huang, S. Cao, R. Yang, W. Yang, Z. Yun, Z. Wang and Q. Feng, *PLoS one*, 2015, **10**, e0140381.
- 9 F. J. Díaz-Pernas, M. Martínez-Zarzuela, M. Antón-Rodríguez and D. González-Ortega, *Healthcare*, 2021, p. 153.
- 10 M. M. Badža and M. Č. Barjaktarović, *Applied Sciences*, 2020, **10**, 1999.
- 11 A. Antoniou, A. Storkey and H. Edwards, *arXiv preprint arXiv:1711.04340*, 2017.
- 12 J. Brownlee, *Machine Learning Mastery*, 2019, **5**, year.
- 13 J. Kang, Z. Ullah and J. Gwak, *Sensors*, 2021, **21**, 2222.
- 14 A. Çinar and M. Yildirim, *Medical hypotheses*, 2020, **139**, 109684.
- 15 P. Saxena, A. Maheshwari and S. Maheshwari, *Innovations in Computational Intelligence and Computer Vision: Proceedings of ICICV 2020*, Springer, 2020, pp. 275–285.
- 16 K. Dataset, *Accessed on Apr*, 2022.
- 17 Brain Suite Editors, *Skull-stripping*, <https://brainsuite.org/processing/surfaceextraction/bse/>.
- 18 B. Shmueli, *Part III: the Kappa Score (aka Cohen's Kappa Coefficient)*, *Towards Data Science*, [Online], Available: <https://towardsdatascience.com/multi-class-metrics-made-simple-the-kappa-score-aka-cohenskappa-coefficient-bdea137af09c>, [Accessed: June, 2021], 2019.
- 19 B. Shmueli, *Towards Data Science*, 2019, **3**, year.