**Brain Tumor Classification**

**CPSC 393 Final Project**

**Introduction**

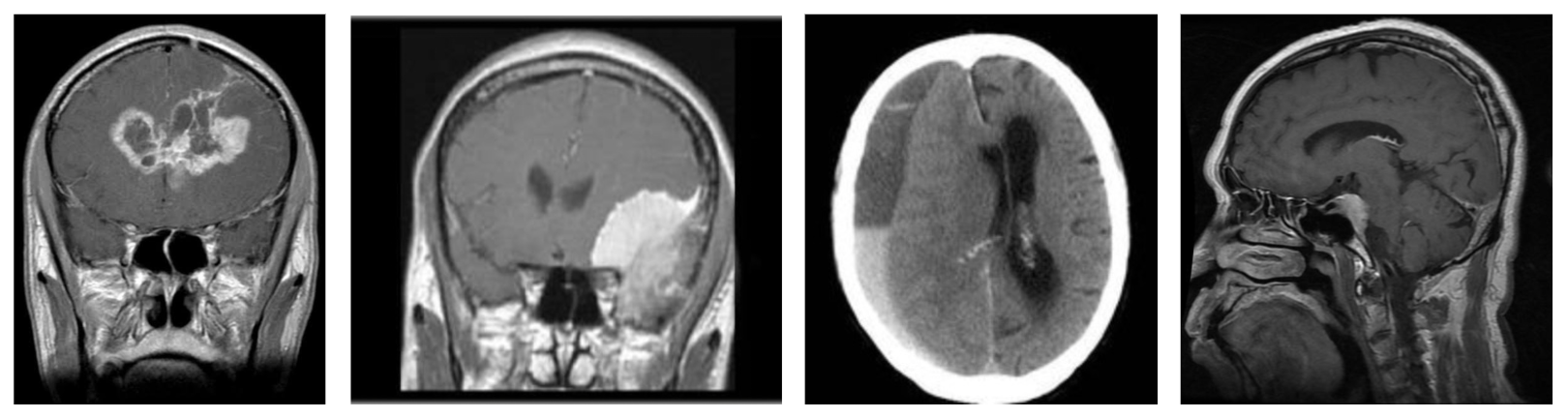
The ability to correctly classify brain tumors is very crucial. Brain tumors are extremely aggressive in children and adults and need to be identified accurately to be treated correctly. Every year, there are around 11,700 people that are diagnosed with some type of brain tumor, accounting for almost 85 to 90 percent of all primary central nervous system tumors. This means these types of tumors are very common, so doctors need to be able to identify and treat them accurately and efficiently. A few examples of the different types of brain tumors are glioma tumors, meningioma tumors, schwannoma tumors, and pituitary tumors. Brain tumors can be benign or malignant; differentiating between the two is also important as it helps determine the type of treatment.

There are many cases where a brain tumor can be misdiagnosed, mainly because a tumor can mimic the symptoms of other common diseases. This is where the issue of incorrect treatment arises, where physicians could prescribe patients with medications that are meant to treat another kind of disease. This can be very harmful to patients as the disease they have isn’t being treated, meaning it can spread to a different part of the body. Any kind of error with the diagnosis and medications can not only be harmful to the patient but can also be deemed medical malpractice. Physicians and doctors use CT/ MRI scans to ensure they accurately diagnose their patients; however, these can always be misinterpreted if not read correctly. According to a study conducted by Johns Hopkins University, approximately 250,000 people die each year from misdiagnoses and medical malpractice. This can be prevented if the interpretations of the CT and MRI scans are made more accurate.

There are some advancements in this area where new techniques are being implemented in order to improve tumor classifications. One promising approach is the use of advanced imaging techniques such as magnetic resonance spectroscopy (MRS) and diffusion tensor imaging (DTI) which analyze the metabolic activity and microstructure of brain tumors in order to classify them. These scans help provide more detailed information about a tumor that helps doctors make more accurate diagnoses. Another approach is to use machine learning algorithms that analyze imaging data and help doctors make accurate diagnoses. It consists of using large datasets of medical images of tumors that were diagnosed correctly to train the machine learning algorithm. The model then classifies new images based on their location, size, shape, and any other prominent features. These algorithms would mainly be used to assist doctors in diagnosing brain tumors correctly, as they act as a second eye for doctors. The use of a machine learning algorithm helps prevent physicians and doctors from making mistakes that are irreversible to fix since there are different plans to treat different types of brain tumors. It also helps patients feel more confident in their treatment and reassures them about their final diagnosis.

**Data Analysis**

The dataset being used to train the neural network model consists of a training set, which is used to train the model, and a testing set, which is used to measure how the model is performing. The training and testing sets were split using an 80-20 split. Four different types of images are found in each of the sets: glioma tumors, meningioma tumors, pituitary tumors, and no tumors. These four classifications will be used to train the model, and in each classification, there are a bunch of different MRI scans that can be used. The four classifications are shown in the image below in their respective order:

Figure 1: Tumor Classifications

The images were .jpg image files, so they had to be resized and converted to arrays in order for them to be correctly inputted into the model. These images were also from 2020, so the model will use slightly older data. This could be an issue if we apply our model on newer data since the accuracy score could turn out very different than what we would have expected. That’s because the model might not find some of the MRI scans familiar.

**Methods**

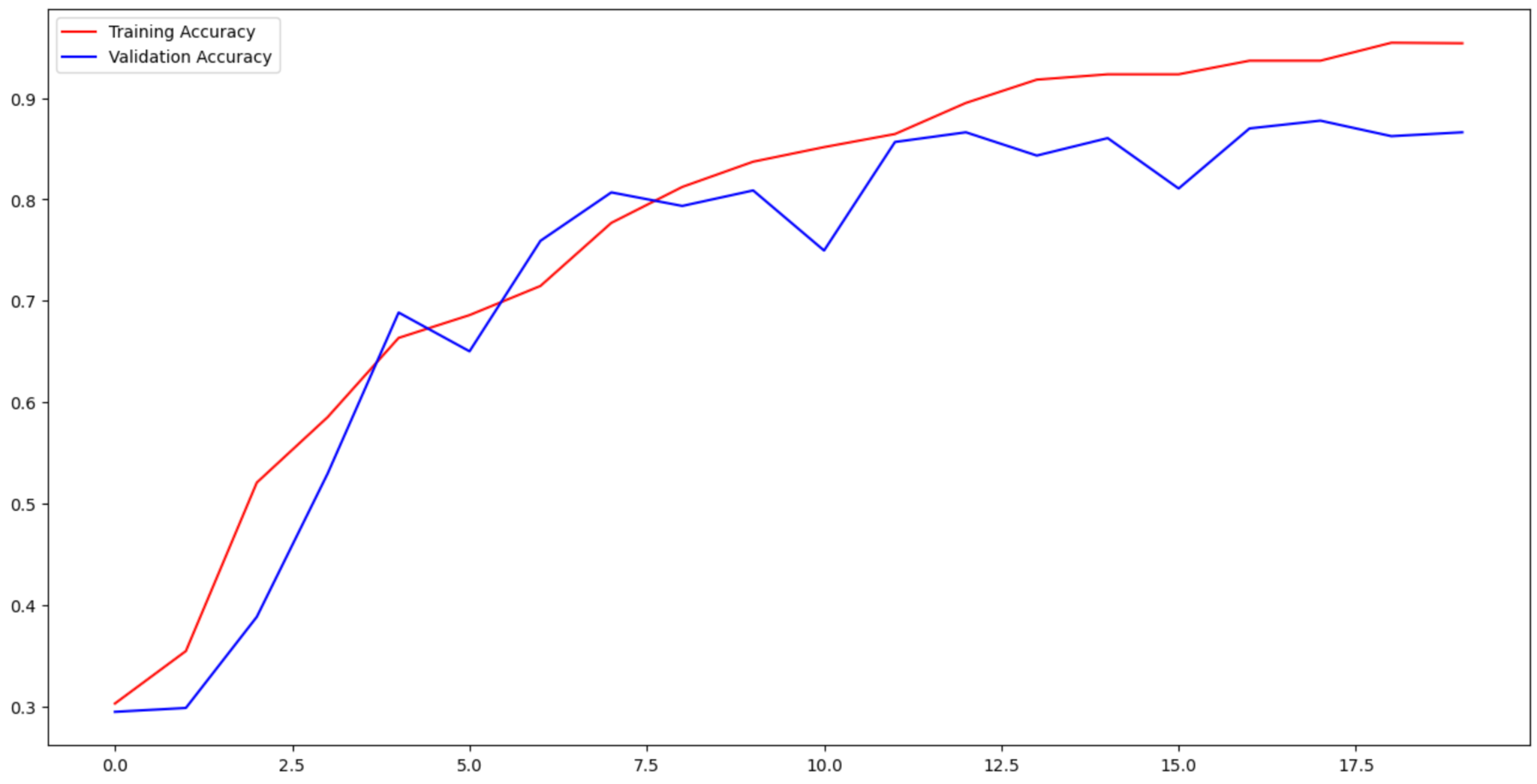
The model was built using a traditional Convolutional Neural Network (CNN), which has proven to be effective with image classification. The goal was to create an accurate CNN model that helps identify whether a patient has one of the three types of brain tumors mentioned above or none. The first CNN model consists of five sets of layers. The first four sets of layers have two convolutional layers with 64 and 128 hidden layers in each set and a kernel size of (3,3). Each of these layers uses the ReLu activation function because it has been proven to perform well for image classification. It is computationally efficient and helps tackle the vanishing gradient problem with deep neural networks.

The first four sets of layers also consist of a max pooling layer and a dropout layer. The max pooling layer uses a (2x2) filter, which helps downsize the image before being inputted into the next layer. The dropout layer uses a value of 0.3, which means that around 30% of the data points are randomly dropped during each epoch to ensure the model doesn’t become overfit for the data. The final part flattens the data to output a 1-dimensional vector, which is then passed through additional dense layers and a dropout layer before finally being passed through a dense layer using the softmax activation function that helps with multiclass classification. The final output of this model is the classification of the type of brain tumor identified. The model is then compiled using the categorical cross-entropy loss function, which is used for multi-class classification as it measures the difference between the predicted probability distribution and the true probability distribution in the target classes of the model.

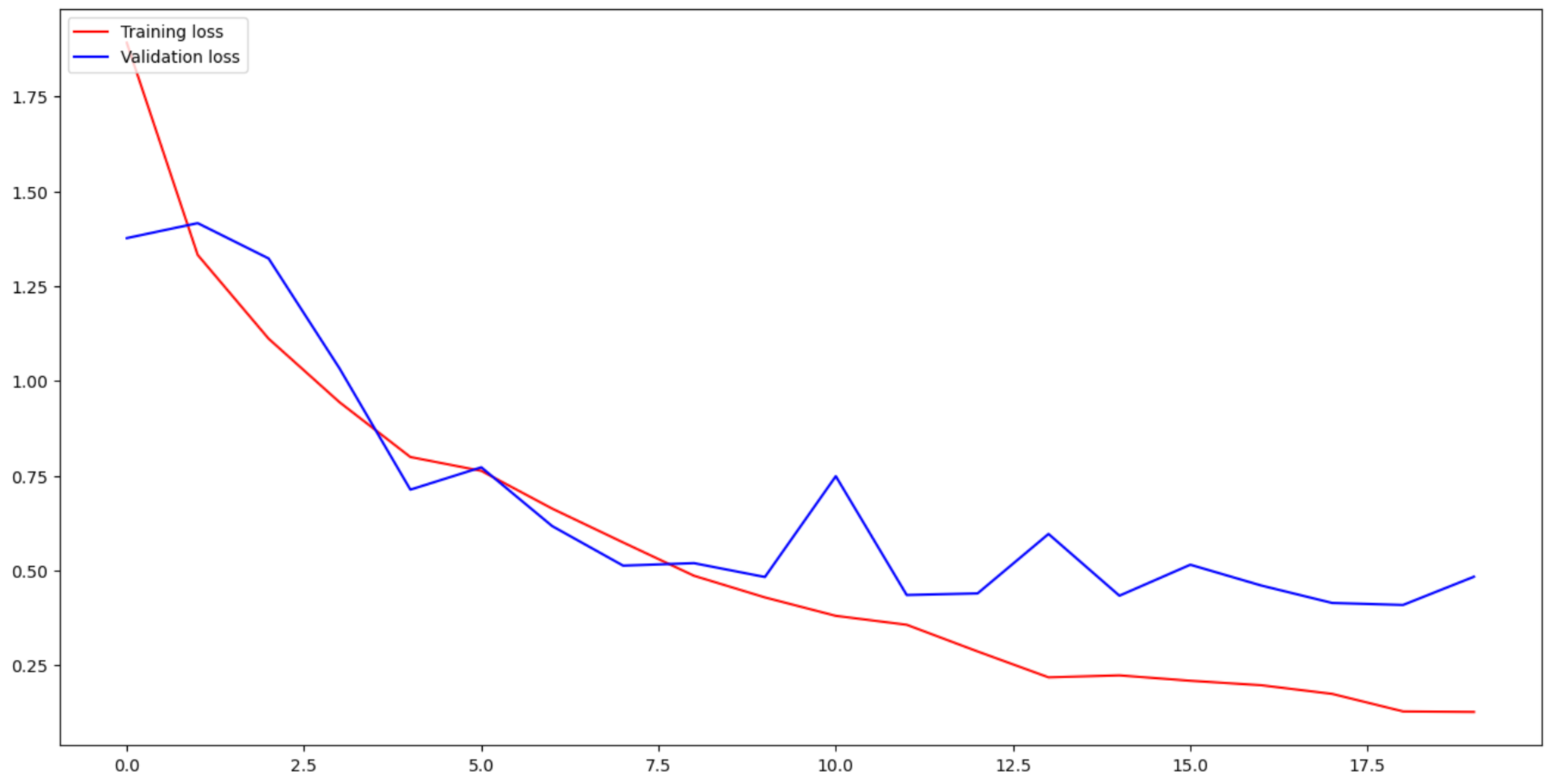
The second model consists of three sets of layers. The first two sets of layers have two convolutional layers with 64 and 128 hidden layers in each set and a kernel size of (3,3). They also have a max pooling layer that uses a (2x2) filter and a dropout layer with a 0.3 value to make sure the model isn’t overfitted for the given data. The sets of layers are very similar to those of the first model. The final part has a layer that flattens the data into a 1-dimensional vector and then two dense layers that also use the ReLu activation function. The final layer of the model is a dense layer that uses the softmax activation function to generate an output that classifies the type of tumor. This model also uses the categorical cross-entropy loss function to help identify how the model is performing.

**Results**

Both models performed fairly well based on their accuracy when the testing set was used. For the first model, the training set had an accuracy score of 0.96, and the testing set had an accuracy score of 0.89. Both sets had really high accuracy scores, which shows that they are performing really well with the data, but since the training set had a higher accuracy score than the testing set, the model is slightly overfit. The following graphs show the accuracy and loss values for the training and validation sets.



Graph 1: Accuracy for training and validation sets in model 1



Graph 2: Loss for training and validation sets in model 1

The accuracy and loss values for both sets of data are similar enough to say the model is performing well and isn’t too overfit on the given data. The results of the classification can also be plotted using a confusion matrix which also helps show model performance. It is a matrix that shows the predicted classification against the true value.

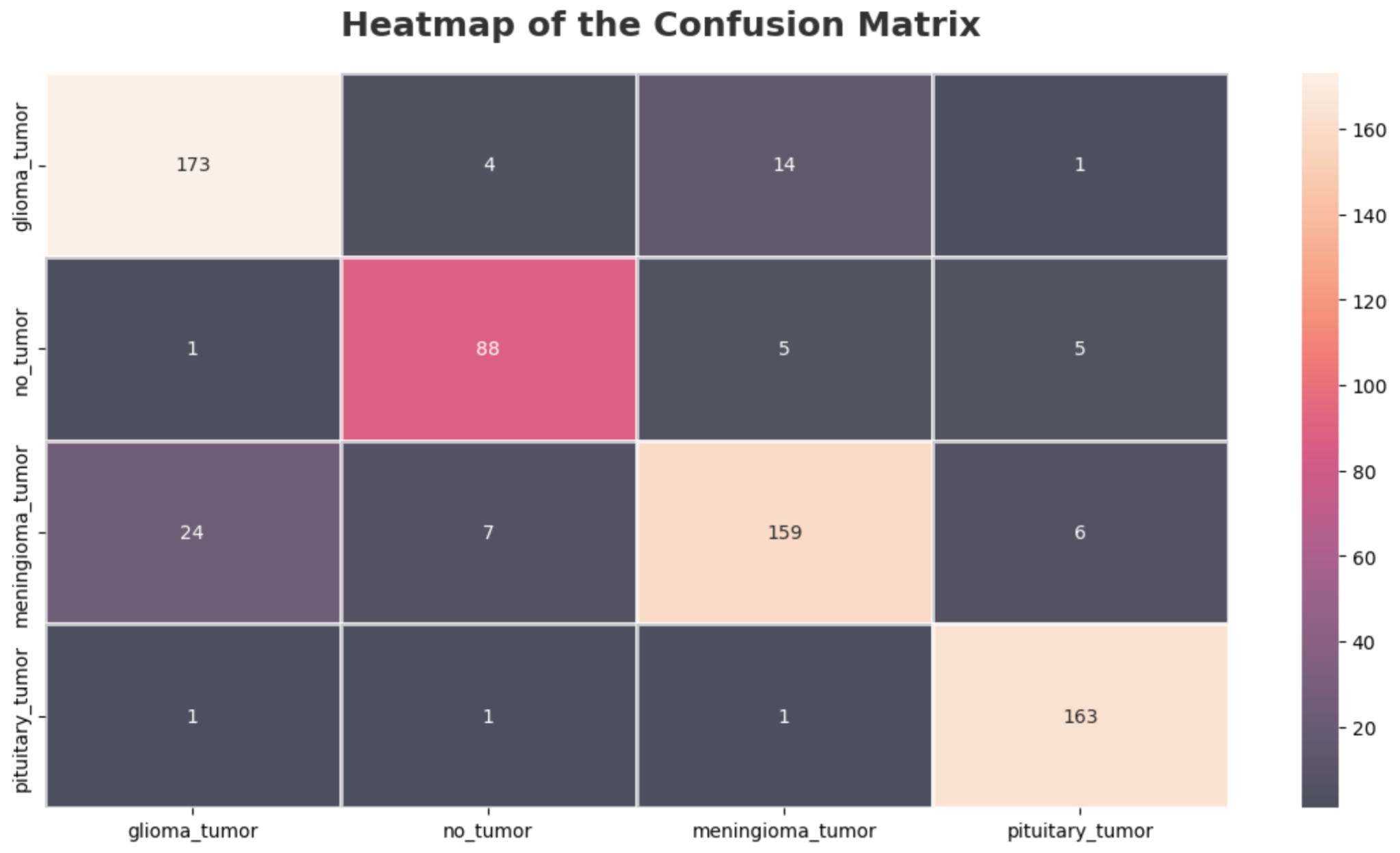
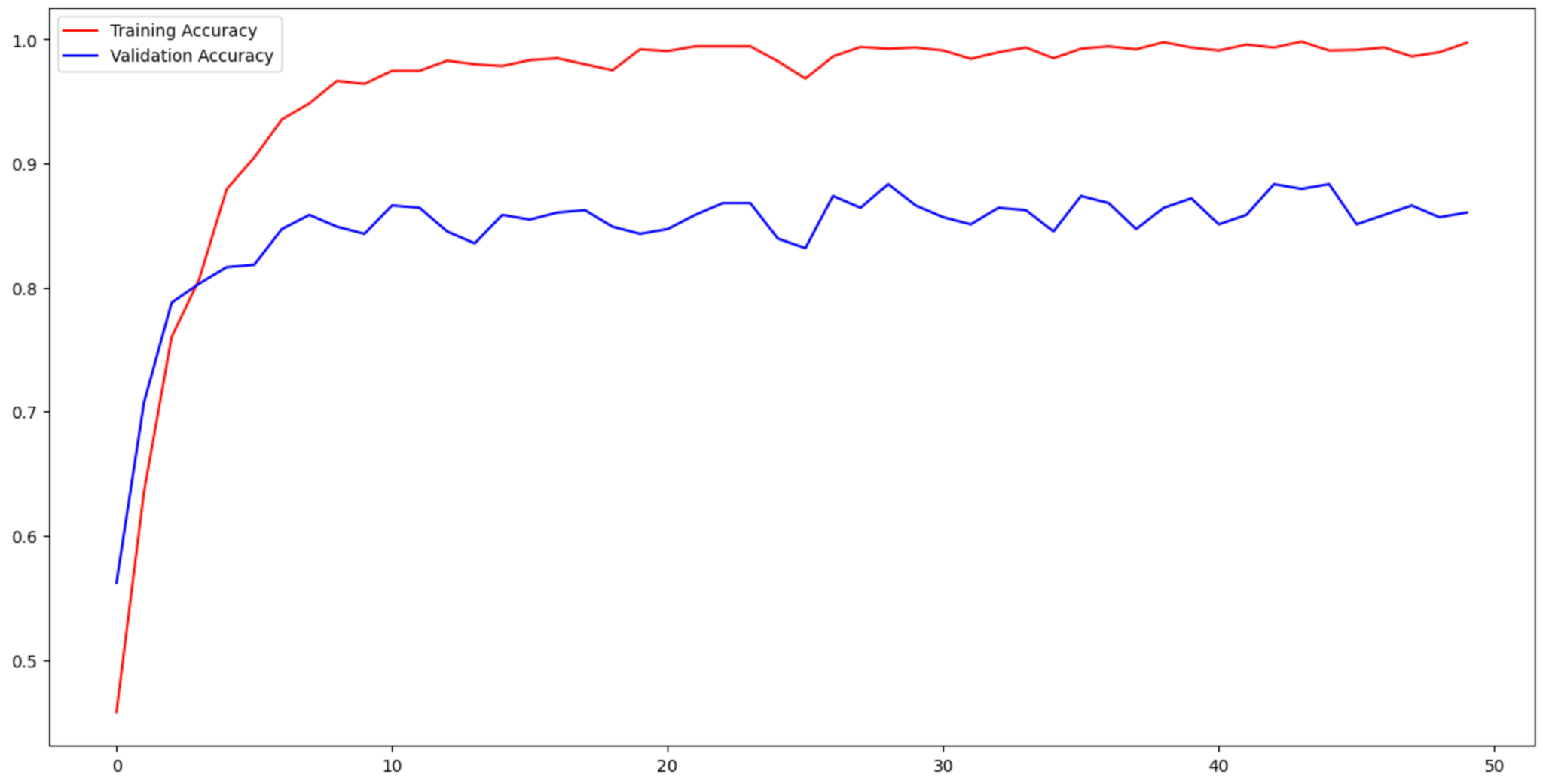


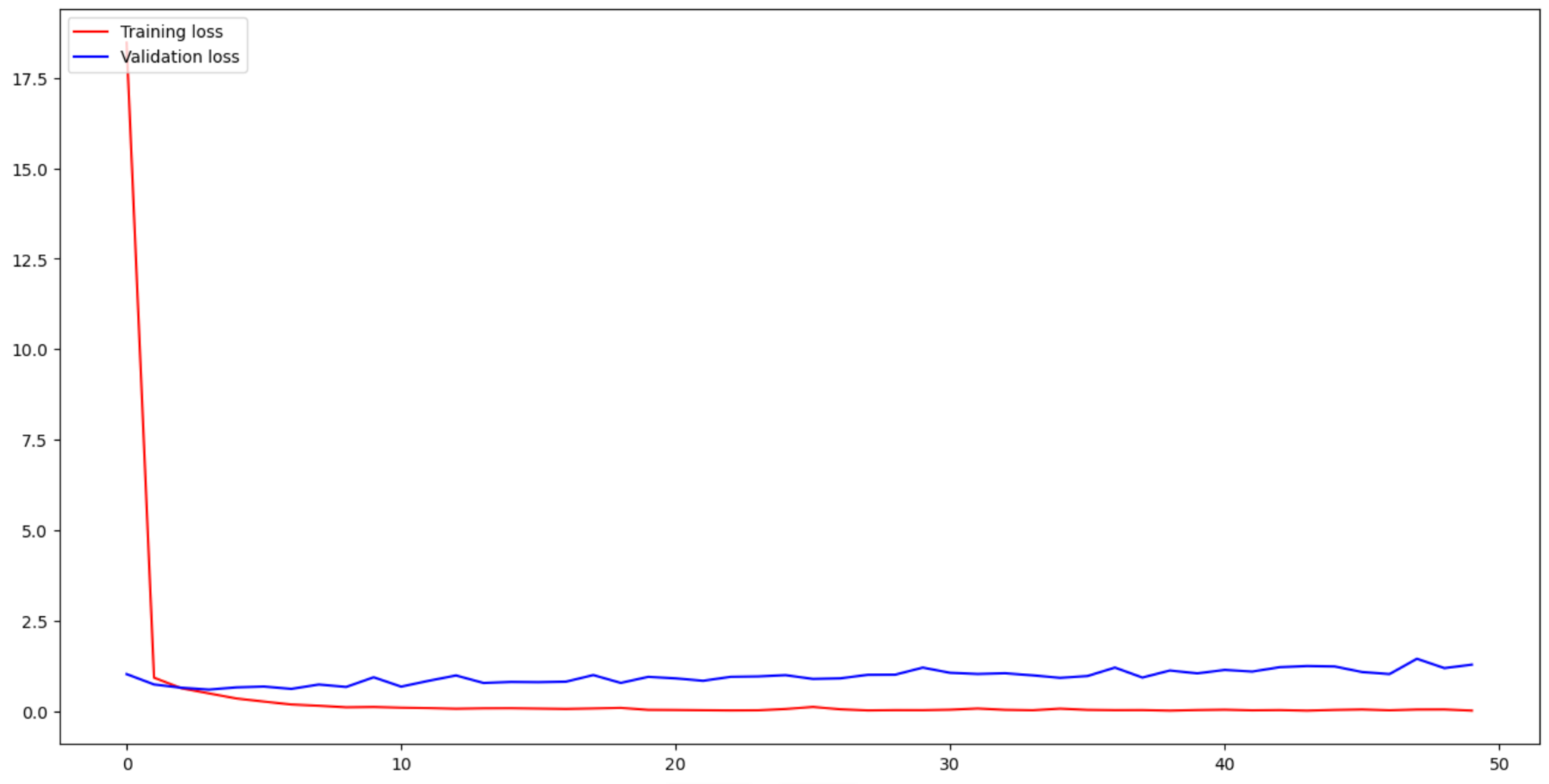
Figure 2: Heatmap for Model 1

The higher values are all diagonal, showing that the model mostly classifies correctly. The main misclassification is between the glioma and meningioma tumors. An important aspect to check for is false positives, which tell us how often the model classifies a scan as a tumor when it doesn't have a tumor. These values are really low, but since this form of misclassification exists, the output of the model needs to be used carefully. Another important aspect to check for is false negatives, which tells us how many times the model classifies a scan as no tumor when the patient actually has a tumor. These values are also low, but it’s a dangerous form of misclassification as it results in the patient not getting the treatment they require.

For the second model, the training set had an accuracy score of 0.97, while the testing set had an accuracy score of 0.88. Both sets, yet again, had high accuracy scores, but since the training set had a higher accuracy score, there needs to be a concern for slight overfitting. The following graphs show the accuracy and loss values for the training and testing sets.



Graph 3: Accuracy for training and validation sets in model 2



Graph 4: Loss for training and validation sets in model 2

The accuracy for the training set gets higher than the validation set, which shows that the model could be overfit on the data it’s given. This could lead to more misclassifications of data the model hasn't seen. The loss values are pretty low, so the model isn’t performing too poorly. The predictions can be checked using a confusion matrix.

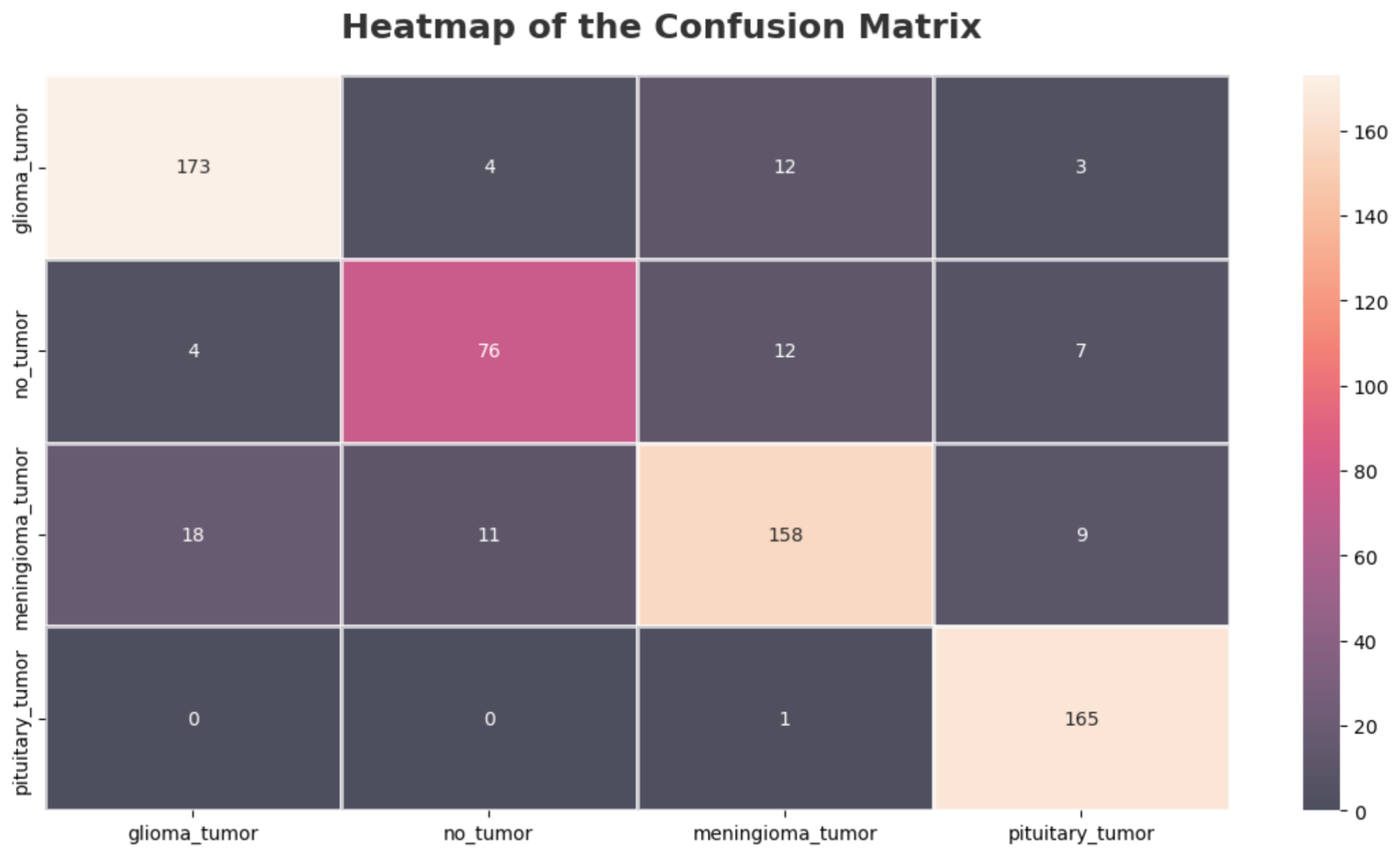
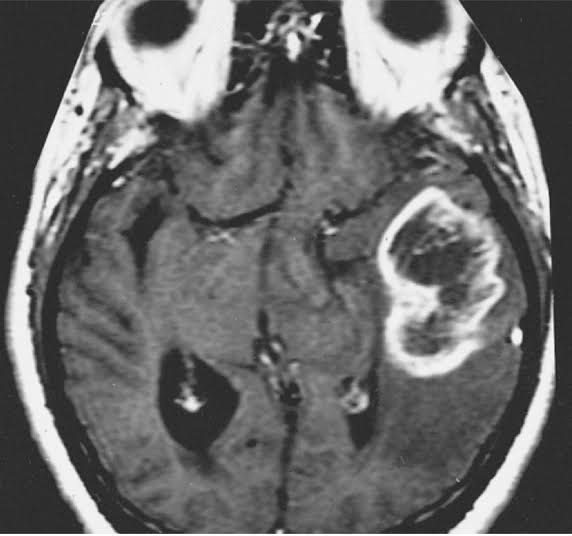
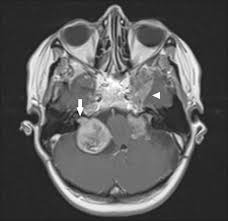
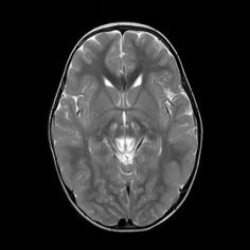
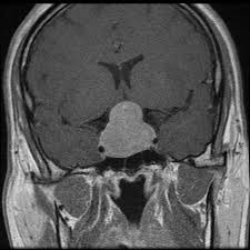


Figure 3: Heatmap for Model 2

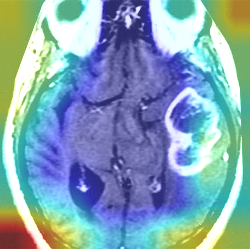
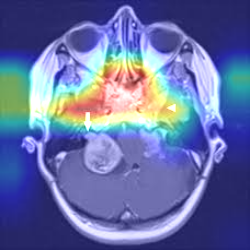
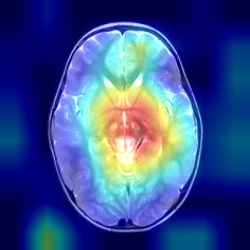
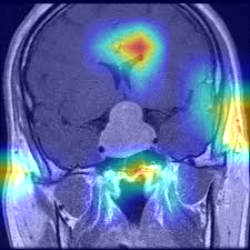
The higher values are diagonal, showing that the model mostly classifies the images correctly. The main misclassifications are between the glioma and meningioma tumor, as seen in the first model, and between the meningioma tumor and having no tumor, which is the more pressing misclassification. Even though the values are low, any sort of misclassification when the patient has no tumor could result in unnecessary treatment.

Overall, model 1 had better results and accuracy scores, so this model was used with the testing set. The model was given four random images from the training set and was able to classify all of them correctly. The model had a high accuracy score along with low false negative rates, which shows that the model is performing well with the given data. The model performed especially well with glioma and pituitary tumors, the most common types of brain tumors. The CNN model can be used to support physicians' and doctors' initial diagnoses. This results in fewer patients being misdiagnosed, which also allows them to get treated earlier with the right treatment plan. Patients will now be more confident in their treatment and can recover more quickly.

In order to try and improve the model so that it can reach the accuracy threshold to be used on current MRI scans, the hyperparameters of the more successful second model were adjusted. Creating model 3, the number of convolutional layers was decreased to 2, the dense layers were decreased to 1, and the number of epochs doubled to 100. These changes didn’t really affect the model, as it stayed at an accuracy score of 97 percent. To try and get the model up to the usage standard of the minimum accuracy score (99.9 percent), model 4 has been created. Model 4 went back to the original hyperparameters of model 1, with the 9 convolutional layers and 2 dense layers, while maintaining model 2’s 50 epochs. The results of model 4 increased the accuracy score to 98 percent, which, while better than any of the previous models, still isn’t enough to be used in industry.

To show what the model was looking at to determine the type of tumor, a heatmap was added onto each of the images that were used for predicting the tumor that was in each image.

These images, from left to right, are of a pituitary tumor, no tumor, a meningioma tumor, and a glioma tumor, all from the testing set. To get the heatmap, we used GradCAM to figure out what our model was looking at to classify each tumor. When the heatmap is applied over each image, we see that the heatmap is activated more so by other parts of the brain, rather than anything that is actually needed to identify the type of tumor. While these heatmaps did not do a great job of finding the tumor, the model still correctly identified which tumor was in each image.



GradCAM clearly finds that the other parts of the image are more important than the actual tumor, which backs up a point made by Rachel Lea Draelos and Lawrence Carin, who said that “Unfortunately, Grad-CAM does not have a location faithfulness guarantee, which means that Grad-CAM sometimes produces misleading explanations that highlight irrelevant locations”, and we believe that is the main reason as to why the tumor was not identified, even though all the models correctly decided which tumor is in each image provided.

**Further Improvements**

Although the models are performing fairly well, some improvements can still be made by adjusting the hyperparameters being used. Some changes can be made by testing these hyperparameters, such as kernel size, activation function, and number of epochs used. Different models can be trained using different hyperparameter values, and the accuracy scores of these models can be compared. The number of false negatives and false positives can also be used to check for the ideal hyperparameter value. In the medical world, the model will need to reach a high accuracy score of 99.9 percent in order to be trustworthy enough to be used on current brain MRI scans. There are also other interpretation techniques that can be used, such as saliency maps that can help understand how the model is making its predictions. This helps ensure that the model learns what we want it to. Another improvement that can be made is to incorporate other types of brain tumors on top of the three that were used in the model. Including more classifications, especially types of brain tumors that are less common, would help improve the accuracy of the model and allow it to be used in many more situations.

**Sources**

<https://github.com/lucrieffel/Brain-Tumor-Classification-DataSet/tree/master/Training/glioma_tumor>

<https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri>

<https://paulandperkins.com/brain-tumors/#:~:text=Brain%20tumors%20are%20most%20commonly,prescribe%20treatment%20for%20another%20ailment>.

<https://malpracticecenter.com/misdiagnosis/brain-tumor/>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7585345/>

<https://medium.com/analytics-vidhya/visualizing-activation-heatmaps-using-tensorflow-5bdba018f759>

<https://arxiv.org/pdf/2111.12215.pdf>

<https://arxiv.org/pdf/2011.08891.pdf>