

Brain Tumor Classification

ECS 170 Final Project Report

Maggie Chen, Isaac Craig, Parinita Gupta, Chelsea Huffman, Kirin Kapoor,
Arvind Tawker, Jack Thomson

June 12, 2024

1 Introduction

The goal of this project was to create a neural network that can classify brain MRI scans and determine what type of brain tumor is present in the image (Meningioma, Glioma, Pituitary). Our aim was to learn more about image classification using convolutional neural networks and the applications of machine learning to the field of medicine and brain-related diseases. Models of this sort have the potential to improve efficiency in diagnosing brain cancer in the future, as methods and accuracy improve over time.

2 Background

2.1 Brain Tumors

Cells have a process that allows them to replicate, and when the new cells turn old, they are programmed to die as the other cells replicate to take their place. It is an ordered process, but when the old cells don't die and continue to replicate uncontrollably, a mass is formed. These masses are known as tumors, which are usually caused by gene mutations, abnormal chromosomes, or external environmental factors that may disrupt the cell's processes.

The 3 different types of tumors that our model will be able to detect are meningioma, glioma, and pituitary tumor. Meningiomas are tumors that grow on the brain or the spinal cord, as it targets the primary central nervous system. They're the most common type of brain tumor, and can be classified into 3 different grades.¹ In grade 1, they grow slowly and are more easily treatable (usually through surgery), but Grades 2 and 3 are more rigorous tumors, which grow faster and have a higher chance of relapse even after removal. Meningiomas at grade 1 are more common in women, but higher grades are more common in men and in non-Hispanic white people. Symptoms include vision changes, loss of hearing or smell, confusion, seizures, and headaches that are worse in the morning.

Gliomas are also a type of cancer that occurs in the central nervous system, and happens when glial cells uncontrollably grow in the brain or spinal cord. The three most common types of gliomas

¹"Meningioma: Diagnosis and Treatment." NCI, www.cancer.gov/rare-brain-spine-tumor/tumors/meningioma. Accessed 10 June 2024.

are astrocytomas, oligodendrogliomas, and ependymomas, each originating from their respective cell types.² Lower grades of these tumors are more easily removable, as surgical intervention is the suggested treatment by the WHO. Initially, the most common presenting symptoms of gliomas are headaches, as the growth places pressure on the surrounding microvasculature, leading to edema (swelling caused by excess fluid trapped in the tissue). Other symptoms include nausea, vomiting, seizures, and potentially personality changes (only for more severe cases).

The last tumor we are classifying with our model is pituitary tumors. At the base of the skull, there is a small gland called the pituitary gland, and it is directly connected to the hypothalamus, linking the brain to the endocrine system. Because the endocrine system is where hormones are made and controls the levels of hormones made in the other endocrine glands of the body, pituitary tumors can affect more than just the pituitary gland itself. Most of the pituitary tumors start in the anterior pituitary (the larger, front part of the gland), and it can affect growth, the thyroid, cortisol, estrogen/testosterone, and prolactin hormone levels. If it begins in the posterior pituitary (the smaller, back part of the gland), it may affect vasopressin and oxytocin levels.³ Most pituitary tumors are not cancerous, meaning they don't spread, but it may still cause health problems as the mass may place pressure on surrounding parts of the brain, invade the areas around it, and produce excess hormones.

Our model does not differentiate between the different grades within each of the types of tumors, but it is a great stepping point in case we decide to do so in the future. When classifying which type of tumor it is, 1 represents a meningioma, 2 represents a glioma, and 3 represents a pituitary tumor. We are able to classify these tumors because of their difference in location of the brain and also in appearance. The dataset we are using includes axial (top), coronal (back), and sagittal (side) views of the three different brain tumors, which allows our model to be well trained with multiple types of MRI scan samples, increasing accuracy of the model.

2.2 Why Our Model is Important

As artificial intelligence continues to be developed and improved upon, it has become an important resource that helps medical professionals diagnose different types of diseases. Using our model is important because it would help neurologists accurately identify what tumor is present in the patient's brain. Identifying brain tumors early and accurately is important so the proper treatment can be provided before one's condition begins to worsen and affect their cognitive function. With most tumors, when caught at a low grade, they are usually benign and the intervention is more likely to result in permanent removal of all the cancerous cells. If caught later, at a higher grade, the tumor is more likely to be malignant and is tougher to treat. This emphasizes the importance of our model, as being able to classify whether a tumor is there or not and what type it is allows doctors to quickly come up with a specified treatment plan instead of spending time analyzing the scans. According to ABC News, the Johns Hopkins Hospital reported that 1 out of every 71 cases was a misdiagnosis, and 1 out of 5 cancer cases was a misclassification.⁴ This is why our model is important, as it will help reduce time spent on making a diagnosis and reduce the number of misdiagnoses, resulting in better outcomes for the patient.

²Mesfin FB, Al-Dhahir MA. Gliomas. [Updated 2023 May 20]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK441874/>

³"What Is a Pituitary Tumor?" What Is a Pituitary Tumor? — American Cancer Society, www.cancer.org/cancer/types/pituitary-tumors/about/what-is-pituitary-tumor.html. Accessed 10 June 2024.

⁴ABC News. "Misdiagnosed Cancer Not Uncommon." ABC News, ABC News Network, abc-news.go.com/WNT/story?id=131047page=1. Accessed 10 June 2024.

2.3 MRI Technology

Magnetic resonance imaging (MRI) is a non-invasive medical imaging technique that is capable of providing detailed images of the body’s internal structures, and is widely used to diagnose brain cancer. This method uses powerful magnets that produce a strong magnetic field, which forces protons present in the body to align with that field.⁵ A radiofrequency current that is pulsed through the patient causes the protons to then strain against the magnetic field. Once the current is turned off, the MRI sensors detect the energy released by the protons as they realign with the magnetic field. The time taken to realign and the energy released by the protons indicate different types of tissues, and allow for the differentiation between healthy brain tissue and cancerous tumors.

2.4 Convolutional Neural Networks for Image Classification

Convolutional neural networks (CNNs) are a machine-learning method commonly used for computer vision tasks, such as image classification. Neural networks consist of layers of nodes, or neurons, where each node is connected to another and has an associated weight and threshold. If the output of a node meets or exceeds the threshold value, the next node is activated.⁶

The classification of images by a CNN involves convolutional layers, where input images are represented as a matrix of pixels. A feature detector moves across the image, checking for certain features of the image. The filter shifts over the image, outputting a dot product of the input pixel values and the filter values. The final result is a feature map that represents the presence of certain features in a specific region of the image. Pooling layers then take this feature map, and reduce the dimensionality to retain the most important features. Fully-connected layers perform the classification based on the features extracted through previous layers and their respective parameters, and produce a probability that indicates the contents of the image defined by the user. In our case, the input images are the MRI scans, and the output of the network will indicate the probability that a tumor is present.

3 Methodology

3.1 Dataset

The data for our model came from a research paper called “Enhanced Performance of Brain Tumor Classification via Tumor Region Augmentation and Partition.” The collection of brain scans and tumor types were acquired from 2 hospitals in China in between the years of 2005 and 2010. Our dataset contains 3064 images from 233 patients with three types of brain tumors. There are 708 images containing meningioma, 1426 images containing glioma, and 930 images containing pituitary tumors. Each image contains a label indicating the type of tumor (1 for meningioma, 2 for glioma, 3 for pituitary tumor), a patient ID, a vector that contains the coordinates of the tumor border, and a tumor mask, or binary image with 1s indicating the tumor region. (We stored this data in a python dictionary due to the flexibility it provides).⁷

⁵ “Magnetic Resonance Imaging (MRI).” National Institute of Biomedical Imaging and Bioengineering, U.S. Department of Health and Human Services, www.nibib.nih.gov/science-education/science-topics/magnetic-resonance-imaging-mri. Accessed 10 June 2024.

⁶ “What Are Convolutional Neural Networks?” IBM, 6 Oct. 2021, www.ibm.com/topics/convolutional-neural-networks.

⁷ figshare.com/articles/dataset/brain_tumor_dataset/1512427?file=7953679

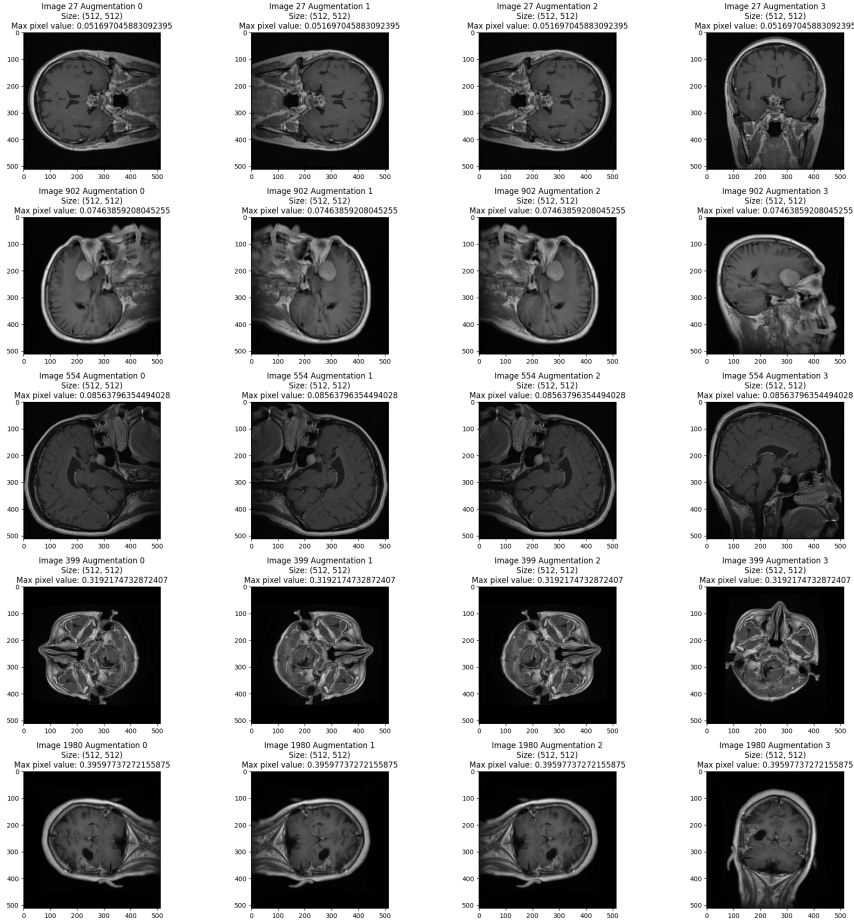


Figure 1: Example of a batch of augmented data, where the image’s orientation is changed to create more training data and tests on unseen examples.

3.2 Data Pre-Processing

Our CNN required data to be provided as input for the model, meaning we needed to ensure that the information we provided the model was in a consistent format so the model can have the utmost accuracy. All the images were in grey-scale, but as part of our data preprocessing, we rescaled the images to be 512x512 pixels. We also normalized the images so that the pixel values would be between 0 and 1.

We shuffled the data before splitting it up into training, validation, and testing sets to reduce potential biases in the ordering of information. Our model was trained on 2144 training samples, which is 70% of our overall dataset. Then we validated the model using 15% of our dataset and tested our model on the last 15% of the data.

Before running our CNN, we also created comprehensive augmentation sets so the model is forced to find underlying patterns in tumors beyond the absolute position in the brain (Figure 1).

3.3 Libraries

Libraries used: Numpy, matplotlib, scipy.io, h5py, os, cv2, random, Keras

3.4 Network Structure

We utilized Keras to compile our model and train it with our training data. The resulting convolutional neural network has 9 total layers:

Input Layer: (512, 512, 1)

- Specify that the input to the network is an image with the dimensions 512x512 pixels and is greyscale.

First Convolutional Layer: (32, (3, 3))

- ReLU activation function
- 32 filters of size 3x3.
- strides = 2

Max Pooling 2D Layer: (2, 2)

- 2x2 pooling window
- strides = 2

Second Convolutional Layer: (64, (3, 3))

- ReLU Activation Function
- 64 filters of size 3x3.
- strides = 2

Max Pooling 2D Layer: (2, 2)

- 2x2 pooling window
- strides = 2

Third Convolutional Layer: (128, (3, 3))

- 128 filters of size 3x3
- ReLU Activation Function
- strides = 2

Flatten Layer

- Flattens the 3D output of the last layer to a 1D vector.

First Dense Layer

- 128 units
- ReLU activation function

Output Dense Layer

- 3 Units for the number of classes
- Softmax Activation
- Produces a probability distribution over 3 classes (for each tumor type).

Model Compilation:

- Adam optimizer
- Sparse categorical cross-entropy
- Metrics: Accuracy

We used the Adam optimizer, which iteratively updates the weights of the network in a similar fashion to stochastic gradient descent. We decided to use the sparse categorical cross-entropy loss function because we are categorizing our input data into 3 different classes.

4 Results

We trained our model over 8 epochs. This yielded promising results, with the training accuracy reaching about 99%, and the validation accuracy reaching about 92.5% by the 8th epoch (Figure 2). Figure 3 shows the performance by class, using the following methods:

- precision (ratio of correctly predicted positive observations to the total predicted positives)
- recall (ratio of correctly predicted positive observations to all observations in the actual class)
- f1-score (harmonic mean of the precision and recall scores)
- support (number of true instances for each class in the dataset)

We found that the overall accuracy of the model was 91%, with a weighted precision of 91%, weighted recall of 91%, and weighted F1-score of 91%. Per class, we can see differences between the model's ability to classify different types of tumors. For class 1 (meningioma), the precision of the predictions was 84%, with recall being 80%. This suggests that the model has more difficulty correctly identifying these instances. For class 2 (glioma), the model performs well, with 90% precision and 93% recall, indicating strong accuracy in regards to glioma tumors. The model performs best in predicting class 3 (pituitary), with its precision and recall both coming in at 98%.

Figure 3 shows the confusion matrix, and provides a breakdown of true positives, false positives, and false negatives per class. For class 0 (Meningioma in this instance), 1400 instances were correctly predicted as class 0, while 360 were incorrectly predicted as class 0 and 264 instances were incorrectly predicted as class 1 (glioma) or class 2 (pituitary tumor). 3316 instances were correctly predicted as class 1, while 264 were incorrectly predicted as class 1 and 352 instances were

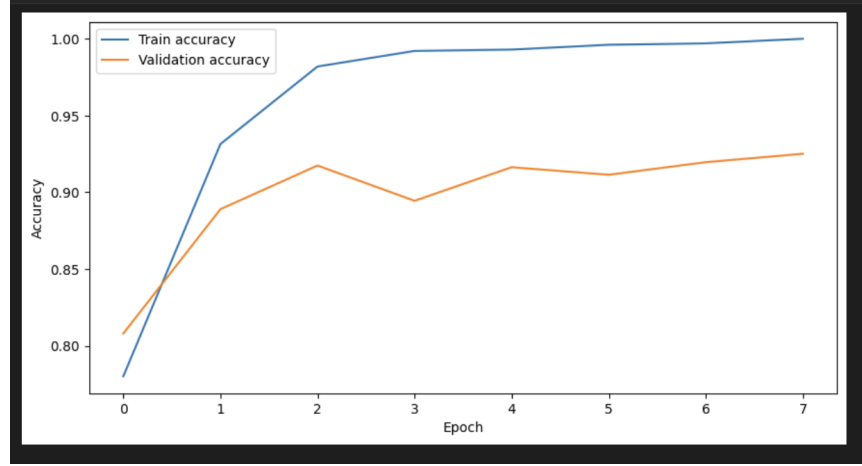


Figure 2: Training and Validation accuracy curve

	precision	recall	f1-score	support
1	0.84	0.80	0.82	1760
2	0.90	0.93	0.92	3568
3	0.98	0.98	0.98	1984
accuracy			0.91	7312
macro avg	0.91	0.90	0.90	7312
weighted avg	0.91	0.91	0.91	7312

Figure 3: Precision, Recall, f1-score, and support for each class. Values for accuracy, macro averages, and weighted averages are displayed as well

incorrectly predicted as class 0 or 2. 1936 instances were correctly predicted as class 2, with 32 instances being incorrectly predicted as class 2 and 48 instances incorrectly predicted as class 0 or 1.

The Receiver Operating Characteristic (ROC) curves plot the true positive rate against the false positive rate. The area under the curve (AUC) represents the model's ability to discriminate between classes. For class 0 (meningioma), the AUC is 0.96. For class 1 (glioma), it is 0.97, and for class 2 (pituitary tumor) it is 1.00. These all indicate a strong ability to distinguish between the classes.

4.1 Results from the addition of k-fold cross-validation

To prevent overfitting and ensure the model generalizes to new data, we decided to implement k-fold cross-validation. This entails dividing the data set into k (in our case, 4) equally sized subsets.

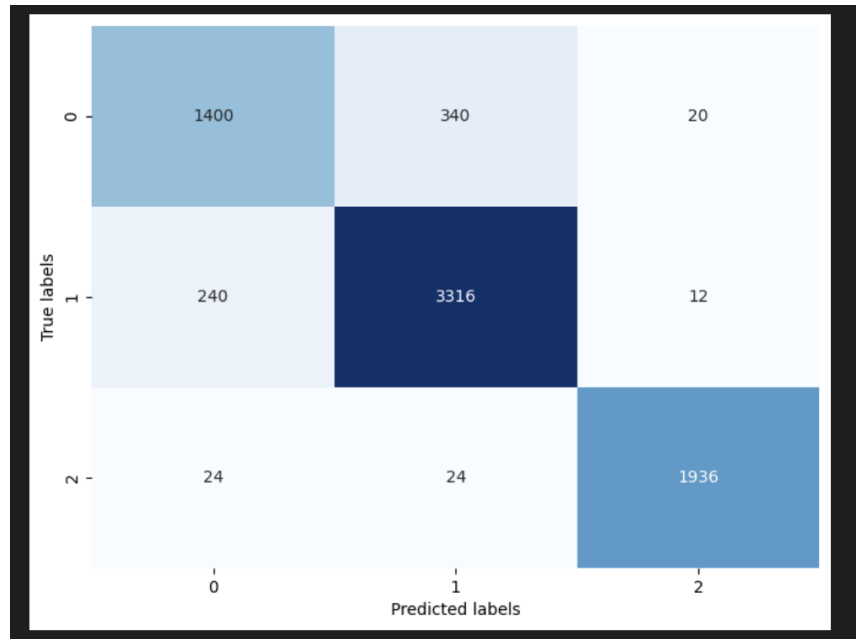


Figure 4: Confusion matrix for the 3 class predictions (before k-fold cross-validation)

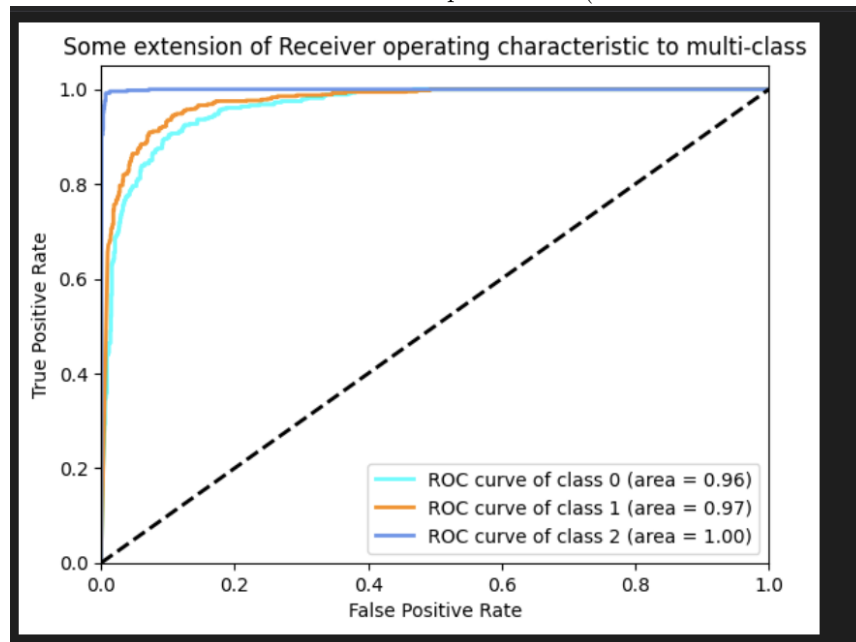


Figure 5: ROC curve for each class

	precision	recall	f1-score	support
1	0.99	0.99	0.99	1344
2	1.00	1.00	1.00	4672
3	1.00	1.00	1.00	2896
accuracy			1.00	8912
macro avg	1.00	1.00	1.00	8912
weighted avg	1.00	1.00	1.00	8912

Figure 6: Performance metrics after implementing k-fold cross-validation (k=4)

For each iteration, one fold is held out as the validation set, while the remaining folds are used as the training set. Performance metrics were recorded for each fold.

As indicated by Figure 6, k-fold validation greatly increased precision, recall, f1-score, and accuracy. Most significantly, the precision in classifying meningiomas increased from 84% to 99% and its recall increased from 80% to 99%. For gliomas, the precision increased from 90% to 100%, and its recall increased from 93% to 100%. Finally, the precision for pituitary tumors also increased from 98% to 100%, and its recall improved from 98% to 100%. Overall, the weighted average precision, recall, and f1-score of the entire model all reached 100%. The confusion matrix (figure 7) reflects the summary of true positives, false positives, and false negatives for each class.

Figures 8 - 11 (pg. 11) show the training and validation accuracies over the 4 training sessions of 4 epochs. You can see that the validation accuracy slowly increases, and eventually becomes greater than the validation accuracy over by the 3rd fold. This indicates that our model overcame the previously observed overfitting, as it is generalizing well to unseen data.

5 Discussion

Based on the results, our model performed well in classifying the different types of tumors in the dataset. The model initially seemed to classify gliomas and pituitary tumors the best, while having moderately high performance for meningioma. However, the difference between training accuracy and validation accuracy could indicate that the model had some overfitting, as it was not able to generalize as well to unseen data. With the addition of k-fold cross-validation, however, there was only a 0.01 difference in precision in the classification of meningiomas compared to gliomas and pituitary tumors.

We faced several challenges throughout this project. We had issues with loading the dataset, considering how large it is, and dealing with the format. Because each file came with several different categories of data, we had to problem solve in order to handle it correctly and efficiently. The brain scans were high-quality, large resolution images with a mix of 256x256 and 512x512 pixels, in a .mat file format. This made it difficult to convert with deprecated libraries (h5py), and increased the training times. We also faced issues when generating batches of augmented data, specifically handling the shape of the data and the generation of empty batches from the data generator. We

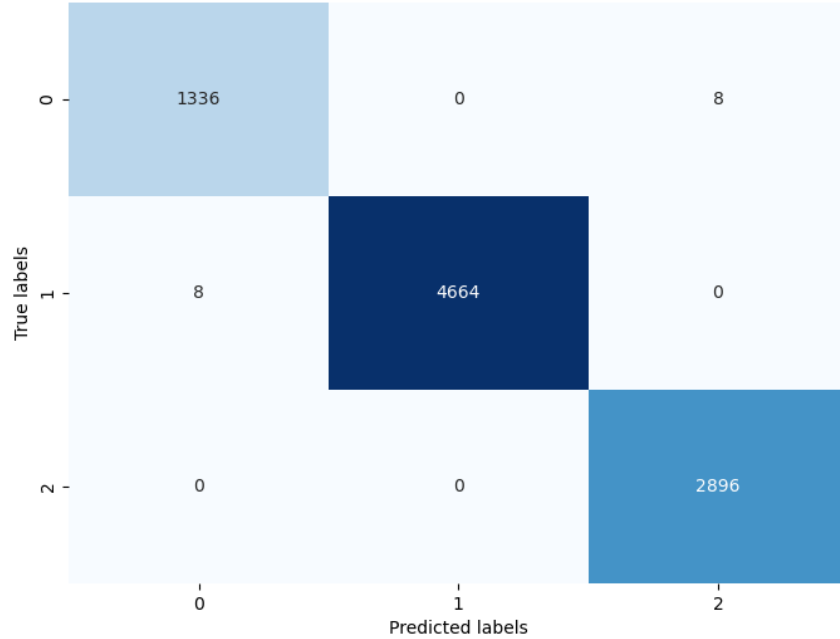


Figure 7: Final confusion matrix after the addition of k-fold cross-validation

learned a lot about data pre-processing, and how to prepare image data for classification. Lastly, our model struggled with overfitting. The initial model converged very quickly, within 2 epochs, and its training accuracy hit 100%, while test capped at 80%. We changed to test-train split and implemented k-cross fold validation and dropout layers.

This model also has some limitations. The model can only read MRI scans, but often a biopsy is necessary to confirm the diagnosis. This model should not be used as a sole method of diagnosis, but rather a tool. Additionally, if the tumor in the brain scan is less visible, our model is more likely to misclassify it.

Models like ours, with more work, could deliver promising results for the classification of brain tumors from MRI scans. This method of identifying brain tumors has the potential to change the landscape of brain cancer diagnosis. Although it would take more refining, it could serve as a supplementary tool when classifying a patient’s brain tumor.

6 Conclusion

In this project, we successfully created a convolutional neural network that yielded promising results, with high accuracy in classifying the different types of tumors present in our dataset. While there is still work to be done in reaching higher levels of accuracy in specific tumor types, we were able to reach substantial results with our current model. Further steps include increasing the size of the dataset, extracting the tumor to be more specific in our identification, and including additional brain tumor types and healthy brain scans. This project revealed that machine learning has great

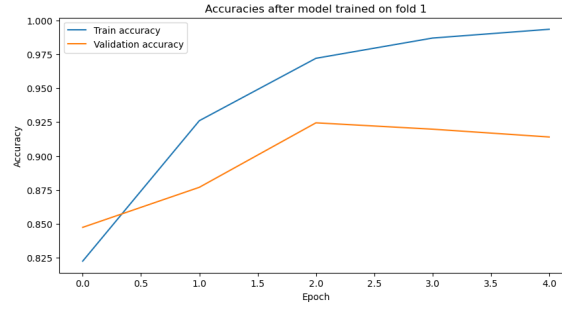


Figure 8: Training and validation accuracies after training on fold 1

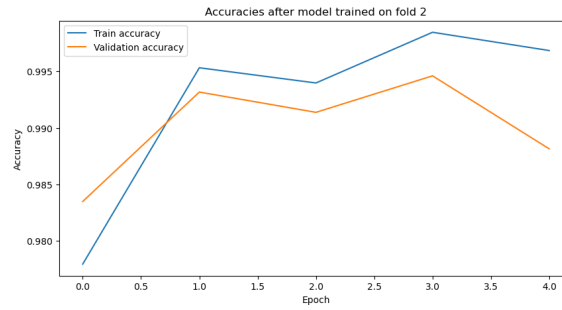


Figure 9: Training and validation accuracies after training on fold 2

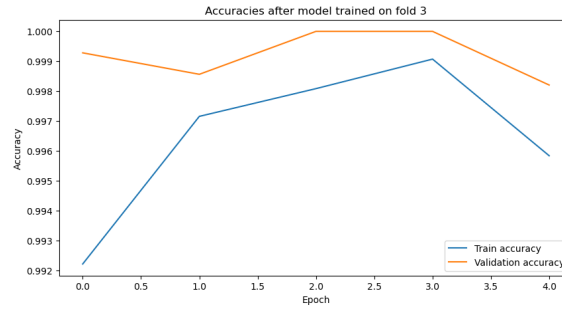


Figure 10: Training and validation accuracies after training on fold 3

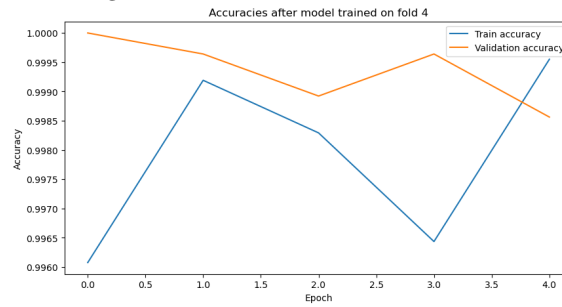


Figure 11: Training and validation accuracies after training on fold 4

potential to aid and accelerate the field of healthcare and the diagnosis process.

7 Contribution

- Maggie: Designed and deployed dynamic website on Heroku. Setup code for the 3d model, uploading .mat files, and using HTTP methods to connect to backend. Helped with early preprocessing.
- Isaac: Created and set up the GitHub pages repository for our website. Implemented a WEB3 template and tailored it to fit our project. Helped contribute to slides.
- Parinita: Brought robust background knowledge about the neurobiology aspect and the project's medical impact. Conducted thorough research on tumors, classification, and datasets, along with other necessary technical information. Contributed to the slideshow and final report, and assisted with preprocessing the dataset and training the model.
- Chelsea: Helped with early data exploration and extracted negative tumor images from a separate dataset. Also assisted in technical research, writing up the final report and creating the presentation slides.
- Kirin: Trained and hyperparameter-tuned the model. Introduced k-cross fold validation, managed overfitting. Provided compute power. Set up virtual environments for multiple group members and mounted a central Google Drive to manage tensorflow and package version conflicts.
- Arvind: Exploratory data analysis and visualization. Data preprocessing, such as tumor isolation and image augmentation to prevent overfitting. Sourced high-quality, high-resolution MRI images. Configured model layers from scratch and evaluated results
- Jack: Conducted research on preliminary datasets. Assisted in writing early pre-processing code and debugging of data generation of augmented data. Conducted early model training, and wrote up final report in LaTeX.