Brain Tumor Classification Using a Convolutional Neural Network

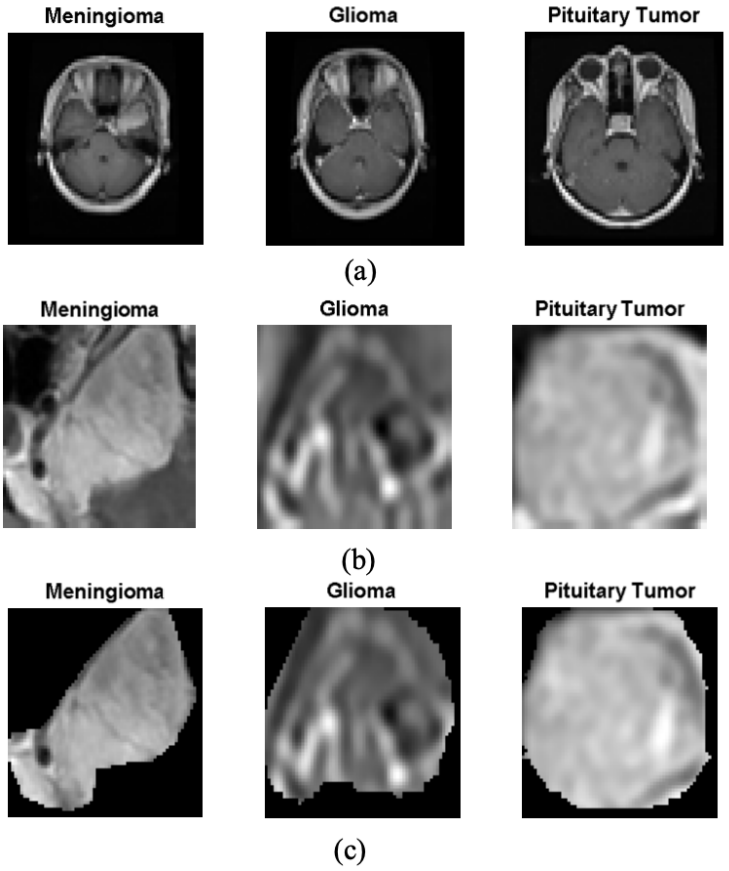
Toshiki Kato, Raphael Li, Alexander Zheng

1. Introduction:

This project is explores the research by Ali Mohammad Alqudah et al.[1]. Deep learning is a tool that is becoming more and more commonly used within the medical field when dealing with complex issues that require high accuracy and precision. Brain tumors are some of the most aggressive diseases that rapidly decrease quality and longevity of life. Early detection and classification of such tumors are critical to creating beneficial treatment plans for the patients. In general, there are three types of brain tumors that are separated based on the primary affected area: meningioma, glioma, and pituitary. We reimplemented a paper's architecture and used a Convolutional Neural Network (CNN) in order to build a deep learning model that can determine the type of the brain tumor given input magnetic resonance (MR) images in order to create a generic classifier for brain tumors.

2. Methodology:

2.1 Dataset: The brain tumor dataset came from Cheng, Jun, et al. [2] which is available through this following link https://figshare.com/articles/dataset/brain_tumor_dataset/1512427/5. The dataset contained 3064 images of three types of brain tumors(glioma, meningioma, and pituitary tumor). We preprocess the image using 3 processes(uncropped, cropped, segmented). Within each process, we downsample to 3 sizes(32x32, 64x64, 128x128).



2.2 Model: The model architecture is a CNN architecture with 18 layers and 4 2D convolution layers with ReLU activations, maxpools, and batch normalizations. We end with 2 fully connected dense layers with a softmax activation between them. We trained our model with the Adam optimizer with a learning rate of 1e-3 for 600 epochs. Below is a diagram of the

3. Results

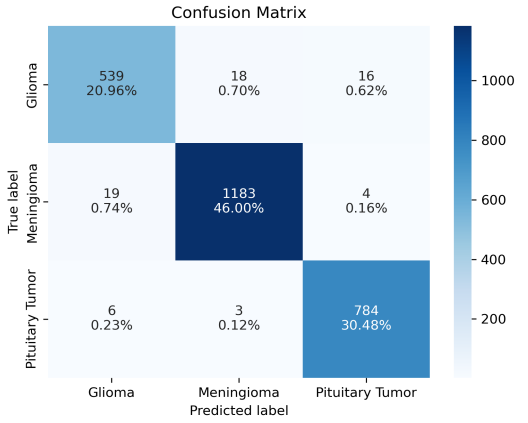
3.1 Metrics: For all cases and types of images, we created a confusion matrix based on the testing data in order to compare our results tested on some samples from the original dataset that were not a part of training or the validation dataset. Our primary metrics were 4 statistics: accuracy, sensitivity, specificity, and precision. These were all generated based on the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Since it was a 2D confusion matrix, we calculated each value for each type of brain tumor so that for glioma the true positives are the number of tumors that were correctly identified as glioma, true negatives were the number of tumors that were not glioma that were not classified as glioma, false positives are the number of tumors incorrectly identified as glioma, and false negatives are the tumors that were glioma but the classifier thought they were something else. The formulas for each metric are shown below.

Accuracy	$\frac{TP + TN}{TP + FP + TN + FN}$
Sensitivity	$\frac{TP}{TP + FN}$
Specificity	$\frac{TN}{TN + FP}$
Precision	$\frac{TP}{TP + FP}$

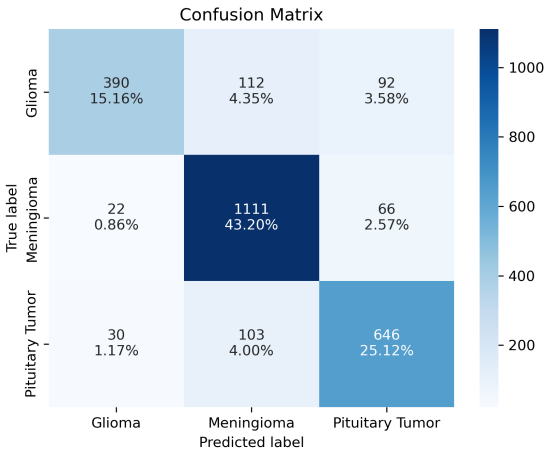
Accuracy is defined as the number of correct predictions over the total. Sensitivity is the percentage of tumors of a certain type (e.g. glioma) that were correctly identified. Specificity is the percentage of tumors that were not of a certain type that were correctly identified not as the type. Precision is the percentage of tumors identified as a certain type that were correctly classified.

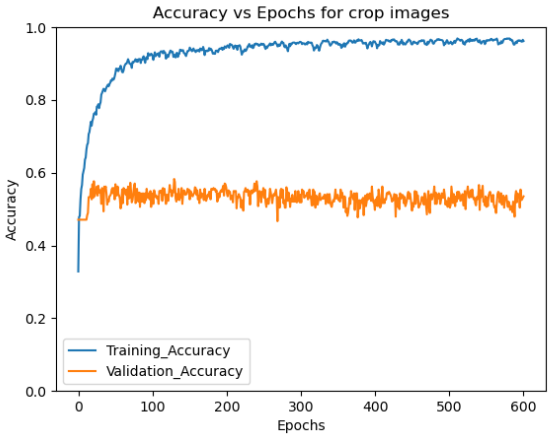
3.2 Results for 32x32 Images:

Uncropped 32x32:

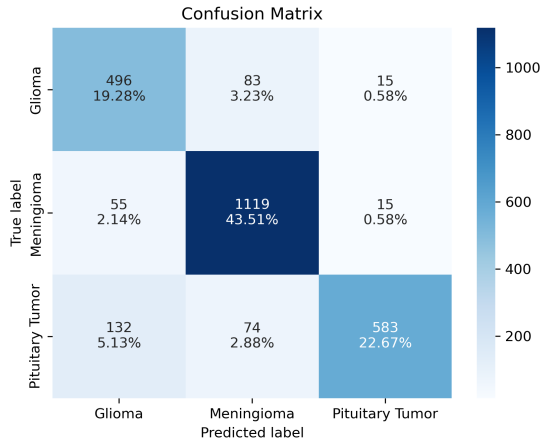


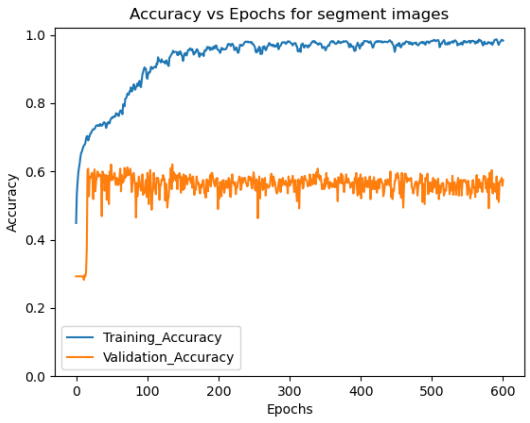
Cropped 32x32:





Segmented 32x32:

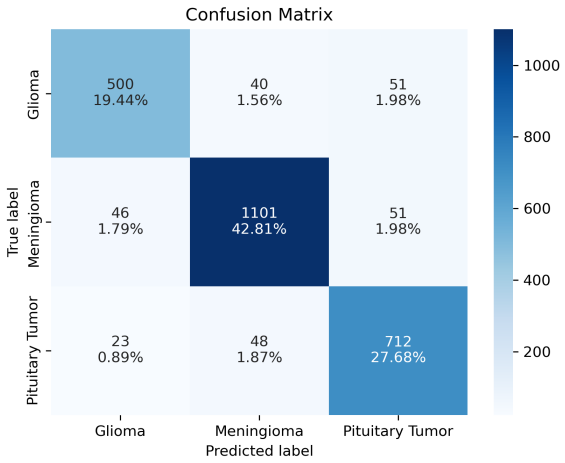




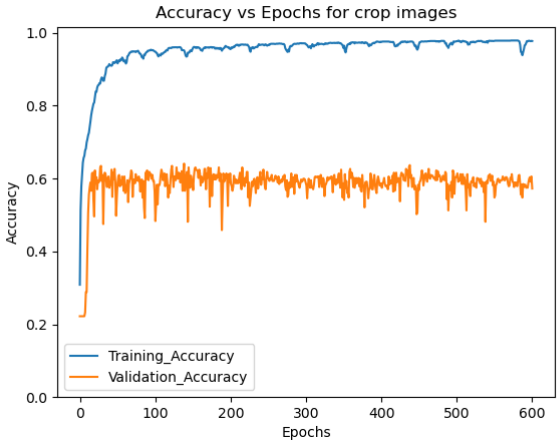
Uncropped 32x32:

3.3 Results for 64x64 Images:

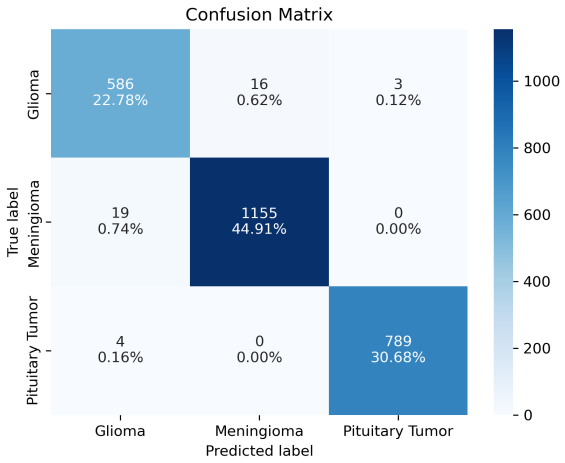
Uncropped 64x64:



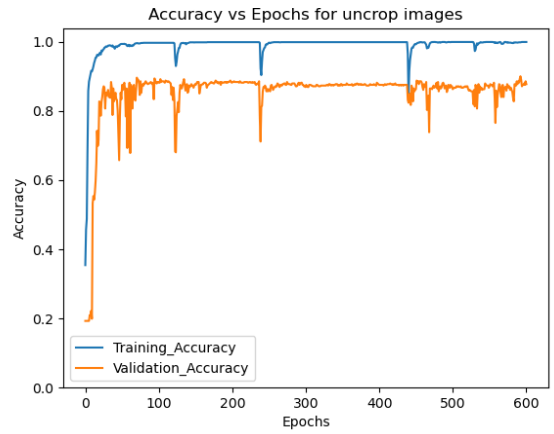
Uncroppped 64x64:



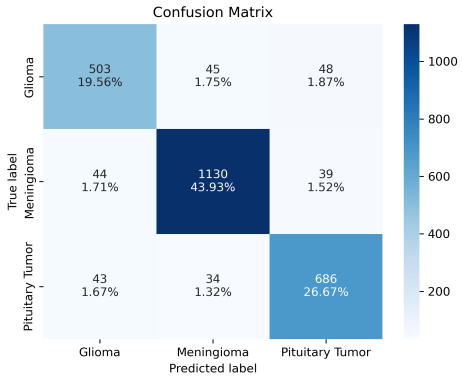
Cropped 64x64:



Cropped 64x64:



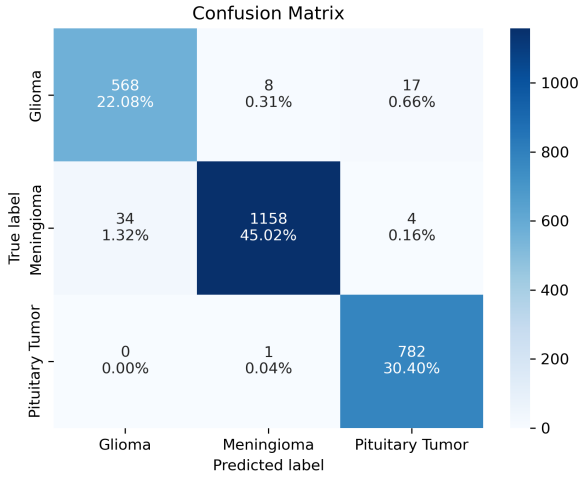
Segmented 64x64:



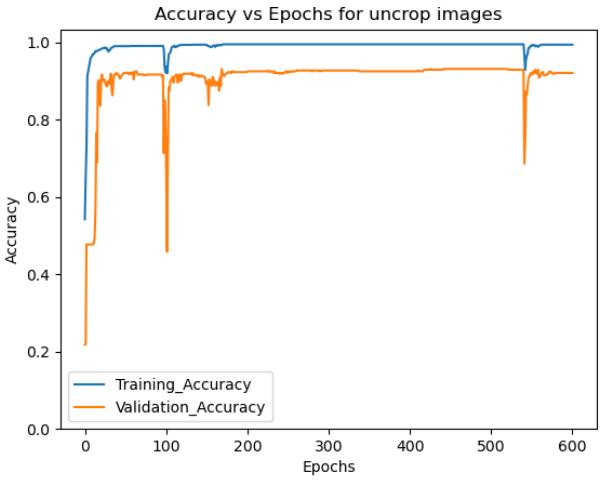
Segmented 64x64:

3.4 Results for 128x128 Images:

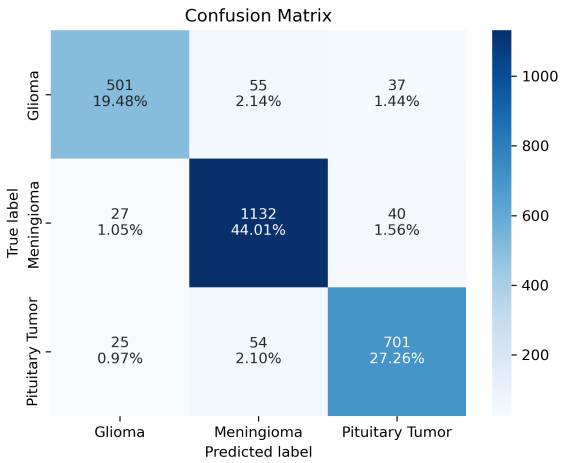
Uncropped:



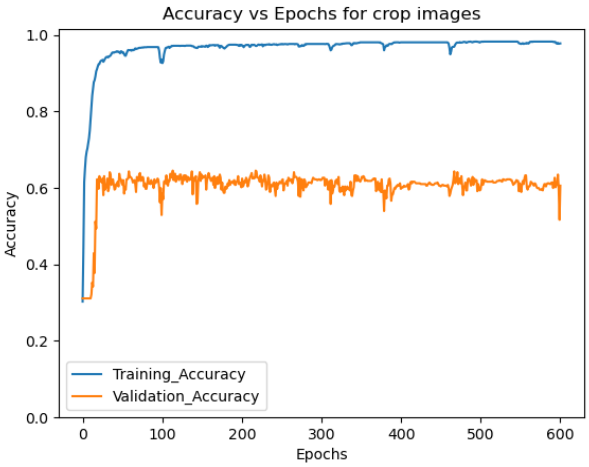
Uncropped:



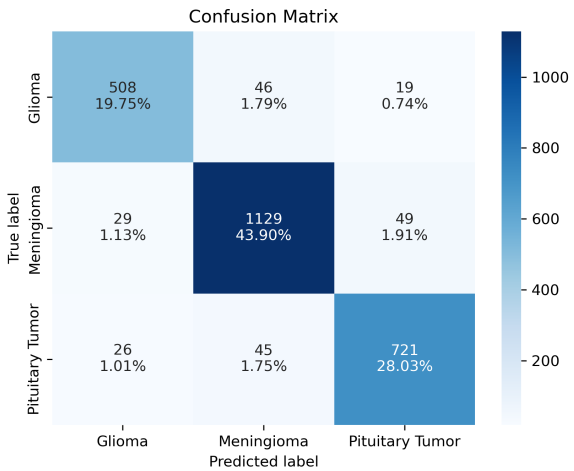
Cropped:



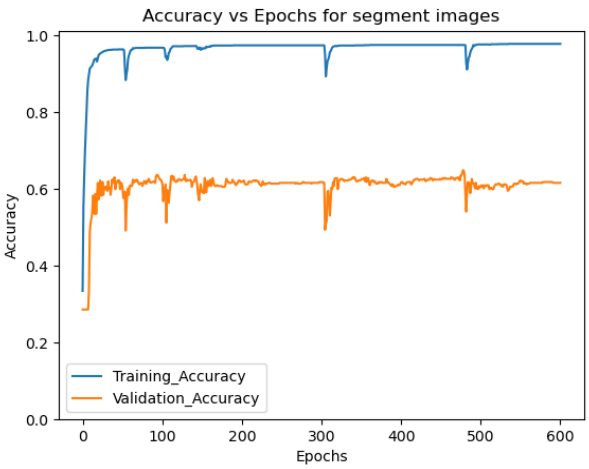
Cropped:



Segmented:



Segmented:



Challenges: A challenge with preprocessing is figuring out how to properly modify the data such that it is ready for the network. Specifically, the images were quite large and making sure that the downsampling didn't ruin the data was important. Additionally, masking out the tumor and cropping the images such that the aspect ratio was maintained was another challenge. At first, we were concerned that the upsampling and downsampling used to solve this problem may have degraded the data, but it actually might have worked to reduce overfitting.

We also ran into the problem of implementing the architecture and tuning the hyperparameters. The architecture provided in the replicating paper did not have much detail such as strides and filter size for the convolution layer. At most, they had a number with some layers which we assumed to be the number of filters the convolution layer ran. After the convolution layers, there is a dense layers with the number 2 before the classification layer. We thought this may be the output size, but found that the accuracy performance was quite low. After increasing the output size, the accuracy was much higher, but was something we were not expecting as we thought we were following the paper's architecture correctly.

Reflection: There were distinct limitations to our reimplementation of this paper. On one hand, the authors of the original paper ran the model on 128x128, 64x64, and 32x32 images, and there seemed to be less overfitting on the models with lower image qualities with many statistics receiving more accurate or precise results on lower image qualities. Further research could be done to replicate the image qualities that the researchers ran. Additionally, testing different image sizes whether that be smaller or larger than the range of qualities tested in the paper would give more insight into recommendations for improvements in this CNN architecture.

Another limitation is that there seemed to be a case of overfitting within the data, signified by the extremely high accuracy in the training data early on within the first 20 epochs of training. This high accuracy can be reduced by simplifying the CNN architecture but given the fact that we are reimplementing a paper, we wanted to keep consistent with the architecture presented in the paper. Additionally, the paper itself had no issues of extreme overfitting leading us to believe that it is an internal issue within our architecture caused by differences between them as it is not fully clear what their exact architecture they used was in the paper.

References:

1. Alqudah, A. M., Alquraan, H., Qasmieh, I. A., Alqudah, A., & Al-Sharu, W. (2019). Brain tumor classification using deep learning technique - A comparison between cropped, uncropped, and segmented lesion images with different sizes. *International Journal of Advanced Trends in Computer Science and Engineering*, *8*(6), 3684–3691. https://doi.org/10.30534/ijatcse/2019/155862019
2. Cheng, Jun, et al. Enhanced performance of brain tumor classification via tumor region augmentation and partition. PloS one, Vol 10, 2015. DOI: 10.1371/ journal.pone.0140381.