Applied Machine Learning Intensive (AMLI), Summer 2021
Capstone Report

TUMOR DETECTION IN BRAIN MRI SCANS

July 29, 2021

Alexis Mercado, Gerardo Moreno, Julio Sibrian
Team 7
University of Arkansas
aimercad@uark.edu, gdmoreno@uark.edu, jcsibria@uark.edu

1 Abstract

In this project we will be using machine learning to classify weather a brain has a tumor or not. We first start with Exploratory data analysis. In this step we will be searching through our data set to either remove or fix our data. With this we will better understand our data which will help us tune our model better.

The data set we chose to work with has only 253 images in it so in order to increase our sample size we used data augmentation. We flip, rotate 180 degrees, and mirror our images to increase our image sample size to over 1000 images! With this we can better train our model. After this we will mask our data to help normalization. We will mask our data by contouring our images which will outline our brain image and make the pixels surrounding our brain 0. This will help with normalizing because the background of our image will have a lesser effect on the overall image. During this process we will also resize the image so that each image are the same dimensions.

Next, we will build our model using Convolutional Neural Network or CNN. We use the CNN model because it is a very reliable image classifier. CNN is a deep learning neural network designed for processing structured arrays of data such as images. The loss function we use for our model is binary cross entropy. We use this because we will determining if a tumor has a tumor or not, so our output will be either yes or no which is perfect for binary cross entropy. for our optimizer function we used the Adam optimizer algorithm. The Adam optimization algorithm extends stochastic gradient descent and has recently seen broader adoption for deep learning applications in computer vision and natural language processing.

Finally, we will conduct a model analysis so we can see if our model is efficient. We do this by creating a classification report that shows precision, recall, f1-score, and support. We achieved a 97 precision score for non tumors and a 92 precision score for tumors which is exceptional! With this we can for sure assist doctors, but certainly not replace them.

2 Introduction

There are around 700,000 people currently living with a brain tumor and the survival rate for them is about 75%. Additionally, there will be another 80,000-90,000 people that will be diagnosed with a tumor by the end of 2021. Sometimes these tumors may be difficult to detect because of the fact that symptoms may not be shown. Tumors are a time sensitive issue and every minute counts when detecting a tumor and providing adequate treatment.

The goal of this project is to address the issues of tumors and be able to diagnose them through machine learning. The benefits of this project would be the automation of a process that is typically done in person by a professional. We aim to create a program that could potentially speed up the process and even free up the professionals time in order for them to tend to more patients.

For this project, the role of program manager is undertaken by Alexis Mercado as well as developing the model. Christian Sibrian is in charge of organizing our code and notebook in terms of providing comments that explain our code adequately, as well as making the narrative of the notebook flow smoothly. Christian will also work on cleaning the data set and assuring that it will be usable for our program. Gerardo Moreno is responsible for all aspects of the presentation as well as the ethical report. Gerardo will ensure that our model is functioning by testing and using additional data.

3 Data Analysis

The data is comprised solely of images.These images contain MRI scans of the brain and will be used to train the machine learning model to detect a tumor in a MRI scan of the brain.The main data set that we will be using resides within the Kaggle domain. This data set is perfect for our model because it contains images of scans with tumors as well as without tumors in order to prevent any bias. Another reason why we chose this data set is because the images are very similar in regards to the angle taken and the overall structure of the image. The data was also very consistent in the aspect that we did not have to do much data cleaning. All images were downloaded accurately. This made it fairly simple to load the data into the Colab directory. The original data set was fairly small, so had to do plenty of data augmentation. For this, we resized, flipped, mirrored, and rotated the images to increase our data set. We also played around with the image filter by converting between RGB, gray scale, and other formats just to increase our data set. The data augmentation that we constructed vastly improved our model accuracy. The data was definitely the most vital part of the entirety of our notebook.

4 Project Implementation

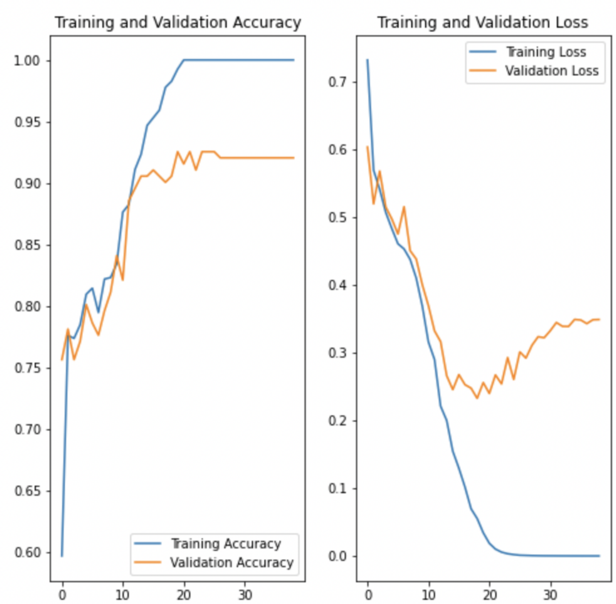
As mentioned, the purpose of our machine learning model is to be able to predict if a brain MRI scan has a tumor or not. Since we are working with images, we used a Convolutional Neural Network to build our model. We also used a binary loss function since we are simply classifying between tumors and non tumors. The model was fit using our data after augmentation and splitting between training and validation sets. When fitting, we also used a stop loss method that monitored our validation loss and the epochs would usually run around 35-45. Most of our challenges that came up from working on this capstone revolved around the issue of working with so many images and finding a proper variable to manipulate and insert into our model. Once we got passed this point, the model was fairly simple to implement. The images also became an issue when trying to feed the model our data. The way that we went about the data was by applying them all into a batch data set variable after iterating through the directories.

5 Experimental Results

Our data set started with 253 images and once we used data augmentation we ended up with 1009 images. Once we became familiar with the data set we started experimenting with different loss functions. We first started with binary-cross-entropy and we stayed with it because it yielded the best precision score. some other loss functions we experimented are hinge-loss and square hinge-loss. These yielded good precision scores but never where better than the binary-cross-entropy loss function. The following is a quantitative analysis approach in seeing our accuracy. This is from using binary-cross-entropy.

classification report:

	precision	recall	f1-score	support
0	0.96	0.86	0.90	83
1	0.91	0.97	0.94	118
accuracy			0.93	201
macro avg	0.93	0.91	0.92	201
weighted avg	0.93	0.93	0.92	201



The charts above are some qualitative analysis approaches in seeing our training and validation accuracy as well as loss. This implements binary-cross-entropy as well.

6 Conclusion

In conclusion, our group was able to create and train a Convolutional Neural Network model with a very high accuracy and precision to detect brain tumors. Our model could assist medical staff quickly detect and recognize brain tumors in MRI scans, which exactly aligns with our group's mission of helping medical patients and healthcare professionals using machine learning. The importance our model brings to the problem of brain tumor detection would be to make the process of reviewing MRI scans more efficient and attempt to accelerate the process of detecting brain tumors using a machine learning model. As medical staff and patients receive quicker results from our model, further necessary actions can be determined for the patient, as tumors could be time sensitive.

Potential further improvements our model could implement is outlining the brain tumor, proper image processing, and acquiring adequate data of MRI scans. Applying a red box outlining the brain tumor in the MRI scan would be the first improvement our group would have been specified more time. Building this into our model could help medical professionals precisely determine where the brain tumor or supposed tumor is located in the MRI scan. Another way our group could improve this model would be to properly pre-process all the images in our data set. From the MRI scans our model misclassified had an overall similarities were the different background colors of the images. Given more time, our group could have done a better job masking and cutting the images to only focus on the brain, leaving no background that could possibly interfere with detecting a pattern in our model. The data set we utilized in this project only consisted of 253 images, which is not the ideal number of data to build and train a machine learning model. If our group could prepare for this project next time, we would apply to gain access to MRI scans straight from the sources to properly build and train our model to be optimized. Overall our project was successful, with these simple improvements with more time could have taken our project to the next level.

7 Acknowledgment

Firstly, we are grateful to have been given this opportunity. This summer internship has been an all around great experience. Academically, we all have learned so much through the process of colabs, projects, and the capstone. The relationships we built will be forever marked within our lives. The atmosphere inside of the classroom was a unique experience and really helped to get to know the professors and teaching assistants. We all felt extremely comfortable with each professor that lectured us throughout the summer and enjoyed getting to know them. We distinctly want to recognize and acknowledge our capstone professor, Dr. Zhang, for guiding us throughout the entire process and providing feedback. She really helped to push for a more perfect finished product. Dr. Zhang's support greatly motivated us to keep moving forward and improve our project. Also, we would like to thank each teaching assistant for providing assistance when running into issues. Kashu, Dat, Duc, and Khoa are phenomenal people and were very easy to work with. On the other hand, we want to thank each professor that provided the necessary skills and resources to even begin this project. Dr. Le, Dr. Rainwater, and Dr. Luu indirectly and directly assisted with our capstone project. Additionally, our fellow peers provided exceptional feedback and comments about various parts of our capstone and we also appreciate their aid. As mentioned, we feel blessed to have this experience and are thankful for the financial, emotional, and educational support that has been given to us. We look forward more opportunities such as this and hold high expectations for anything else to come within our journey.