Brain Tumor Classification of MRIs with Machine Learning; A Study of Early Detection and Classification in Medical Imaging

Rohan Sangal

9/14/2022

Abstract

Early classification and diagnosis of Brain Tumors are essential for providing the right treatment to a patient. It is crucial to get treatment as soon as possible because the survival rate for someone with an untreated brain tumor can range from as low as 3 months to as high as 5 years. In this project, we classified brain tumor images into 4 categories: glioma, meningioma, pituitary, and no tumor. With the use of baseline and deep learning models, the deep learning models demonstrated a significantly higher performance due to their ability to analyze images. The model with the highest accuracy was the MobileNet, a pre-trained transfer learning model trained on 5,608 images. This model yielded a validation accuracy of 98.24%. Using metrics including Kappa cohen score, precision, and recall, we validated the machine learning model's performance. We deployed the MobileNet model to a web app using Streamlit, where users submit MRI images and receive diagnoses of tumor class. We found that the model performed very well while utilizing the web app, indicating that it is safe to be used. However, since we only have 4 classes and there are over 150 total types of brain tumors, it could easily get a diagnosis wrong if it is not in one of these 4 classes.

1. Introduction

In the United States, there are over 200,000 cases of brain tumors yearly, putting it in the Common category, and worldwide, over 251,329 people died from them in 2020. When Brain Tumors are left untreated, they grow more, which also leads to more aggressive symptoms such as seizures, dramatic changes in blood pressure, and extreme brain damage, eventually leading to death which can occur within a time span as low as 3-4 months. The process of diagnosing a Brain Tumor involves getting an MRI first to examine if there is actually a tumor, followed by a biopsy to actually determine what type of tumor it is. However, with the use of Machine Learning, we could use the MRI to classify the tumor based on only its appearance, which could replace or add on to the use of a biopsy. With technology continuing to get better every day, a new approach with Machine Learning can help improve medical imaging greatly and help detect brain tumors early on to help a patient get the treatment needed to survive.

2. Background

With other studies conducted on this topic, one of the most common approaches was deep learning since the performance can be extremely good. Some other methods that were used included the bounding box method, and combining gaussian and canny edges. We also noticed a frequent problem, which was a desire for very large amounts of data to yield a good accuracy and it is expensive in time. To overcome these problems, we used a good amount of data that would not be too much or too little, and we also made sure that we could overcome the problem of time, by making sure the models would train as fast as possible by reducing image size. With very large images, models tend to take much longer to train because there are many more pixels the network needs to learn from, therefore increasing the training time. With the use of AI in medical imaging, it will result in better outcomes for patients as well as doctors.

Tumor	**Glioma**	**Meningioma**	**Pituitary**
Region	Glial cells	Brain and spinal cord	Pituitary gland
Severity	malignant	benign	benign
Symptoms	● Headaches ● Seizures ● Personality changes ● weakness in arms, face, and/or legs. ● numbness ● speech problems ● nausea and vomiting ● vision loss ● dizziness	● Changes in vision, blurriness or seeing double. ● Headaches ● Hearing loss or ear ringing. ● memory loss ● loss of smell ● Seizures ● Weakness in arms and/or legs.	● Headache ● Vision loss ● Nausea and vomiting ● Weakness ● Feeling cold ● Less frequent or no menstrual periods. ● Sexual dysfunction ● Increased amount of urine ● Unintended weight loss or weight gain

3. Dataset

For this research, a dataset from Kaggle was used that contained 7022 Brain Tumor MRI images. Each image from the dataset we downloaded can be classified into one of four classes; Glioma, Notumor (healthy), Meningioma, and Pituitary. To examine this dataset further, we created a *count_files()* function that would loop through the training and testing files and return a count of how many images were in each class. Each class was well represented across the testing and training set, except the notumor class; in both the training and testing set, there were about 100 more images than in the other classes. One of the problems that occur when we have an imbalanced dataset can be bias and low accuracy, as well as incorrect predictions. To visualize the data, we displayed some of the images within the testing and training set and tried different methods of displaying such as blurring, resizing, grayscale, and zooming in with the package cv2. We resized all the images from 512 to 150. The reasoning behind reducing the image size is to reduce the number of parameters in the model, increasing speed. We split the training and testing data 80/20 with the *train_test_split()* function from the Python machine learning package sci-kit-learn, setting a total of 5618 training samples and 1405 testing samples. 80/20 is considered a good split as it allows the model to train on a wide array of images without overfitting. This means that models can generalize to unseen data well.

4. Methodology / Models

After processing the data, the next step was developing the models. First, we created six different baseline models from scikit-learn(Logistic Regression, RandomForest, Ridge Classifier, Decision Tree, K Neighbors Classifier, and Support Vector Classifier) to see how they would perform with the data. We saw good results, ranging from 80-90% accuracy with the highest accuracy from the K Neighbors Classifier which came out at about 90%. Then, we explored deep learning models from Tensorflow Keras such as the Convolutional Neural Network, AlexNet, and MobileNet. Also known as transfer learning models, the AlexNet and MobileNet models are pre-trained models that are used as the starting point for a model on a new task. Because of the ability of these models to analyze and extract features within images, the accuracy of these models was much higher than the baseline models, ranging from 94-98%. After running all these models, we did some hyperparameter tuning to ensure that the models were yielding the highest accuracy possible. After creating the Convolutional Neural Network model, we also tried out cross-validation, a resampling method that uses different portions of the data to train and test the model on different iterations. With the train_test_split function, we split the previous X_train and y_train variables to replace the inputs with the new cross-validation variables and tested it on the training samples, and the validation accuracy stayed approximately the same as the original Convolutional Neural Network. Out of all the models that were utilized, the best performing model was the MobileNet model which yielded an accuracy of 98.4%, and the worst performing model, the Ridge Classifier had an accuracy of around 82.1%.

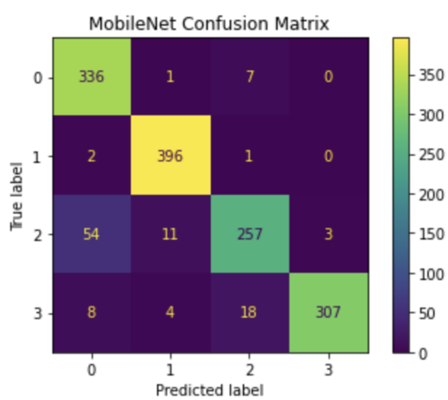
Model	**Description**
Logistic Regression	Logistic Regression is used to make a prediction about a categorical variable versus a continuous one.
Random Forest	The random forest algorithm is made up of a collection of decision trees, and each tree in the ensemble is comprised of a data sample drawn from a training set with replacement, called the bootstrap sample
Ridge Classifier	The Ridge Classifier converts the target values into {-1,

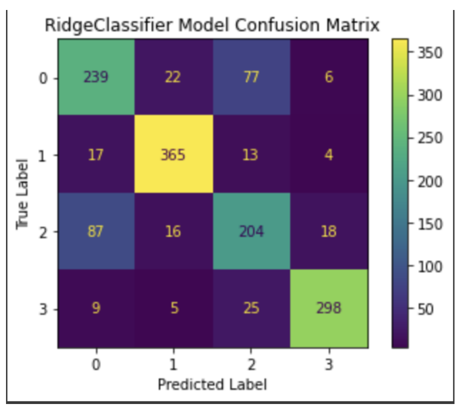
	1} and then treats the problem as a regression task(multi-output regression in the form of multiclass).
Decision Tree	Decision trees use the tree representation to solve a problem in which each tree node corresponds to a class label.
KNeighborsClassifier	Uses proximity to make classifications or predictions about the grouping of a data point.
Support Vector Classifier	Maps data to a high dimensional feature space so data points can be categorized.
Convolutional Neural Network	A type of Artificial Neural Network for image processing and recognition to process pixel data.
AlexNet	A CNN architecture known to be one of the most efficient created.
MobileNet	A CNN designed for mobile and embedded vision applications.

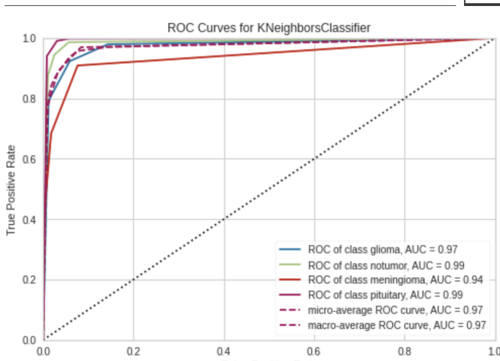
5. Results and Discussion

After creating these models, we used different metrics from the sci-kit-learn library to evaluate the models' all-around performance including accuracy, precision, recall, f1, cohen kappa, and ROC/AUC scores. These metrics all have a different way of evaluating the model's predictions based on different factors such as data imbalance, the type of predictions being made (either true positive, true negative, false positive, or false negative), and the models' ability to differentiate between each class. The goal is to reduce the number of false positives (classifying non-tumors as tumors) and more importantly, false negatives (classifying tumors as non-tumors). We also displayed a confusion matrix to visualize the predictions the models made to show what common errors may be made between certain classes and what the model is predicting correctly. With the confusion matrix, a common error found among each model was meningioma being predicted as notumor and notumor being predicted as meningioma. The possibilities for why this was happening could be directly from the dataset itself, and the training and testing images for each class were very similar, or because of certain hyperparameters. With the different metrics, we found that most of them returned a very similar result, except the cohen kappa score. This is because the cohen kappa score uses data imbalance into account when analyzing the model's predictions. From the MobileNet model, it yielded about a 96% recall, precision, and F1 score while also having a 95% cohen kappa score. For its ROC/AUC score, it came out to roughly 99.7%.

Metric	Formula	MobileNet Model Scores
Accuracy	Correct predictions/Total predictions	98.4%
F1	2(precision*recall)/(precision+recall)	96.3%
Recall	True Positives/(True Positives+False Negatives)	96.3%
Precision	True Positives / (True Positives+False Positives)	96.4%
Cohen kappa	(Total accuracy - Random accuracy)/(1-Random Accuracy)	95%
ROC/AUC	False Positive Rate = False Positives/(False Positives+True Negatives) True Positive Rate = True Positives/(True Positives+False Negatives)	99.7%







6. Conclusions

With the results obtained from this study, we can see great results from the MobileNet model being used to analyze MRI images and return a diagnosis through the web app. With more advancements in medical imaging, the impact can significantly help patients in the future when needing the best treatment in order to survive. However, since the model is only trained on four different classes, it would only work with an MRI from those four classes. A way to expand this study would be to try an immense dataset, with more brain tumor classes, focusing on the most common types of brain tumors first and even uncommon ones. Another way could be compiling more data on symptoms and using an MRI image to predict. With more classes, it can be much harder to develop a model that can yield high accuracy since the data would increase significantly, and training and testing models would be a very difficult process with so many classes. Within this research, some of the limitations we faced within the results itself was a common error the models were making were meningioma and notumor, which falls under a false negative result, which is not good. A reasoning for this could be within the dataset itself depending on how similar the images were for both training and testing data.

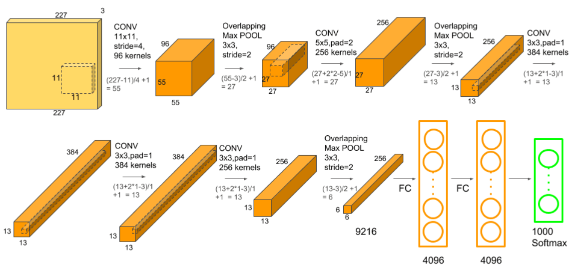
Acknowledgments

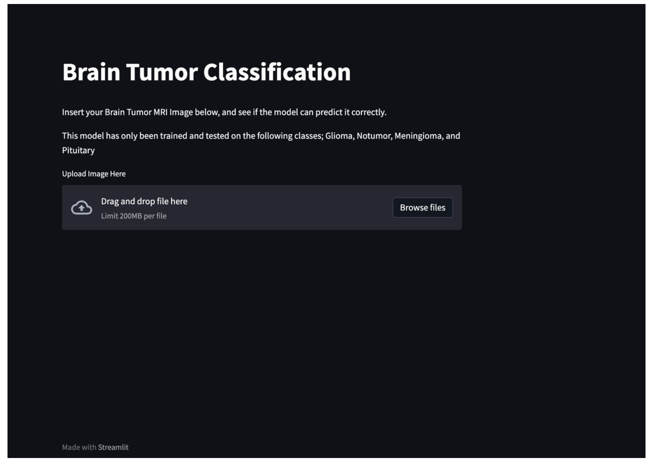
I would like to thank Sriram Hathwar for guiding me throughout this research project as well as the Inspirit AI team for the opportunity to participate in their program.

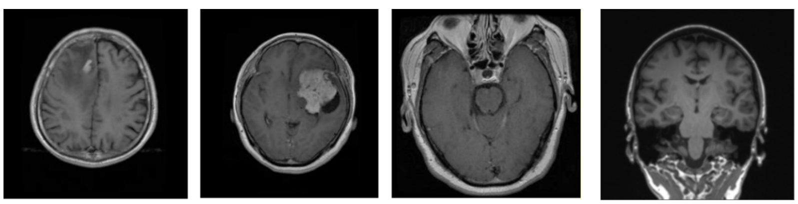
References

1. Msoud Nickparvar. (2021). <i>Brain Tumor MRI Dataset</i> [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/2645886
2. "Meningioma," *Pennmedicine.org*. [Online]. Available: https://www.pennmedicine.org/for-patients-and-visitors/patient-information/conditions-treated-a-to-z/meningioma. [Accessed: 15-Sep-2022].
3. "Gliomas," *Johns Hopkins Medicine*, 11-Feb-2022. [Online]. Available: https://www.hopkinsmedicine.org/health/conditions-and-diseases/gliomas. [Accessed: 15-Sep-2022].
4. R. Mehrotra, M. A. Ansari, R. Agrawal, and R. S. Anand, "A transfer learning approach for AI-based classification of brain tumors," *Machine Learning with Applications*, 03-Oct-2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666827020300037. [Accessed: 17-Sep-2022
5. B. N. Saha, N. Ray, R. Greiner, A. Murtha, and H. Zhang, "Quick detection of brain tumors and Edemas: A bounding box method using symmetry," *Computerized Medical Imaging and Graphics*, 29-Jun-2011. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0895611111000796. [Accessed: 01-Oct-2022].
6. D. A. Dahab, S. Ghoniemy, and G. Selim, *Automated Brain Tumor Detection and Identification Using Image Processing and Probabilistic Neural Network Techniques*, Oct-2012. [Online]. Available:

https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.674.6216&rep=rep1&type=pdf. [Accessed: 01-Oct-2022].







(a) glioma **(b) meningioma** **(c) pituitary** **(d) notumor**