

Deep Learning-Based Tumor Detection in MRI Images

Nadeem Ahmad

Department Of Science, Wilfrid Laurier University

CP322: Machine Learning

Dr. Sukhjit Sehra

April 6, 2023

Abstract

The final project was centered around applying machine learning to real world problems. The submission I made involved the development of deep learning models to detect and classify brain tumors using MRI scans. The aim was to aid in the early diagnosis and treatment planning of cancerous tumors. Deep learning models were chosen due to their efficacy in image analysis, particularly in the medical industry. Several studies have shown the effectiveness of these models in segmenting MRI images and classifying brain tumors, aiding medical professionals in the process of diagnosis and treatment.

To create a robust model and prevent overfitting, different data sets consisting of multiple types of brain scans were combined and extensively preprocessed before being used to train the models. Binary classification was used to distinguish between healthy and tumor-affected brains, and a separate model was used for categorical classification to determine the type of tumor present.

The tumor detection and identification models had accuracies of 91% and 94% respectively, with an 88% accuracy rate when used together on unseen data. The project demonstrated the potential of machine learning in the medical industry and its ability to improve patient outcomes and quality of life.

Introduction

In recent years medical imaging such as Magnetic Resonance Imaging (MRI) has become a core pillar in the detection, diagnosis, and treatment of many diseases such as brain tumors. Even though MRI is widely used, it has flaws which can cause issues when it comes to patient care. The main concern is time, as image interpretation by medical professionals can be an

arduous and time-consuming affair, which can lead to worse health outcomes for patients. This though does not need to be a permanent fixture in the medical industry as machine learning potentially offers some solutions.

For the final project I developed deep learning models which handled the problem of brain tumor detection and classification using MRI scans. The objective of the project was to help in early diagnosis and treatment planning of cancerous tumors by building models which can accurately detect and classify multiple different types of brain tumors.

Part of the reason that I decided to go down this route was because of several studies which clearly highlighted the efficacy of deep learning models in image analysis, especially pertaining to the medical industry. An example would be one such study which demonstrated the effectiveness of these models in segmenting MRI images of the brain with a greater accuracy than traditional methods used by the medical industry (Havaei et al., 2017). Another example is a study in which deep learning helped in the classification of brain tumors with a 90% - 95% accuracy rate (Gab Allah et al., 2021). These studies clearly show that there is a place in the world of medicine for modern deep learning methods to take hold.

Multiple different data sets consisting of many different types of brain scans were combined to create a robust model and to prevent overfitting from occurring. This dataset was extensively preprocessed and only then were used to for training the models.

A Binary classification model built on a convolutional neural network (CNN) was used to distinguish between brains which were healthy versus those which had tumors in them. The output of this model was then taken and processed accordingly depending on what was returned. In the case that a brain was flagged as unhealthy, the MRI image would be fed into another model which conducted categorical classification, returning an integer which represented which

tumor type classification the image fell into. This second model was also built on top of a CNN. The tumor detection model and the tumor identification model had accuracies of 91% and 94% respectively and had an 88% accuracy rate when used together on sets of unseen data.

This work demonstrated the potential machine learning can have on many aspects of the medical industry. It allows for the improvement of patient outcomes and opens the door to future innovations which can improve a person's quality of life.

Data Collection

The dataset which was used consists of a large set of MRI scans of the brain. These MRI scans were taken from thousands of different patients in multiple different studies. The data was obtained from data set sites such as Kaggle, OpenNeuro, The Cancer Imaging Archive and the Brain Imaging Data Structure (BIDS). After applying data augmentation techniques to prevent overfitting, the result was a data set of over 22,000 images of unhealthy brains and 300 images of healthy ones. Due to the skew of unhealthy images to healthy images the healthy data was enhanced with data augmentation techniques, such as translations, rotations, and flips. This allowed for a set of 600 healthy images. The data is split up into two directories, healthy and unhealthy with the unhealthy folder having further subdirectories each representing a different type of tumor that was possible.

The data was split into separate training, validation, and tests sets, with 64%, 16% and 20% in each respectively. Each time the test set was used to evaluate the model, sets would be created and destroyed accordingly to ensure that models had never seen the MRI's they were being tested on. This ensured accurate scores and helped prevent overfitting.

Methodology

This project involved many steps to complete. I first collected all the brain MRIs into folders with the methods described in the data collection section. Data was then split into two different directories, one for each model: tumor detection and tumor identification.

The tumor detection directory consisted of two subdirectories, healthy and unhealthy. Due to the huge skew in data certain steps had to be taken before the split into training, validation, and test sets. Instead of taking tens of thousands of unhealthy photos, only 600 were selected (the same size as the healthy set). If this was not done the model would most likely have a class imbalance causing inaccurate readings. The subset is made by taking an equal number of photos from each tumor type directory to again prevent any imbalances.

The data was split into separate training, validation, and tests sets, with 64%, 16% and 20% in each set respectively. I preprocessed the images by resizing them to 224x224 pixels and after applied rotation, shifting, shearing, zooming, flipping, and filling techniques to ensure that the model was more robust.

Once image pre-processing is complete two separate models are to be created utilizing the Keras library. The first will be the binary classification model which will detect the presence of a tumor in an MRI image, and the second a multi-class classification model for identifying the type of tumor. The two models utilized sequential model architecture with both consisting of different numbers of convolutional, dense, and max pooling layers along with different activation functions.

After the model's construction, they were trained on the validation set and evaluated on their performance utilizing the test set. Metrics such as accuracy, F1 score and AUC were combined with graphical methods such as confusion matrices and ROC curves to judge the

performances individually. The models were adjusted until both produced benchmarks which deemed satisfactory. This involved modifying the number of convolutional layers, changing activation functions, and adjusting data organization to create a robust model. I also made sure to explore misclassified images as well to get an understanding of the potential weakness of the model and find areas of improvement.

Once satisfactory models were created, I then employed cascading prediction where the first model would be for tumor detection and the second for tumor identification. The detection model would make a prediction on an image it was given, returning a floating-point number between 0 and 1. If this number was found to be greater than 0.5, the MRI was found to be unhealthy and would be passed onto the tumor detection model. The tumor detection model would then take the image and output either 0, 1 or 2, each integer representing a different type of tumor (0: glioma, 1: meningioma, 2 pituitary tumor). This cascading prediction method allowed for two highly accurate models to make good predictions consistently. The outputs were then again scrutinized utilizing metrics such as accuracy, F1 score and AUC along with graphical methods such as confusion matrices and ROC curves to ensure the model was performing at a high level.

Overall, my methodology involved a combination of data preprocessing, model building, training, evaluation and visualization techniques to develop accurate and reliable deep learning models for brain tumor classification.

Results/Experimental Analysis:

The accuracy of the binary tumor detection and multi – categorical tumor identification models were found to be 91% and 94% respectively when evaluated against a test set. This is extremely promising as it indicates the models are both operating at a high level and making good classifications. When combined they return an accuracy of over 88% suggesting that there is a marginal drop when both are used in a cascade prediction. One thing to note is that although this is an acceptable degree of accuracy, it may not be enough for the models to be applied to the medical industry. When it comes to healthcare, devices and technology must have an extremely low margin for error, as there are lives on the line which could easily be lost due to small inconsistencies.

To further analyze the efficacy of the models, ROC curves were used in conjunction with AUC values. This will allow for a further understanding of performance and a visual representation of behaviour over a test set. As can be seen in Figure 1, The ROC curve has an almost ideal shape and has an AUC value of 0.99. This indicates that the binary classification portion of the model is operating nearly perfectly, meaning that often the images passed to the tumor identification model are correct. In Figure 2, the ROC curve for the final output (the average of the AUC of each classification) has a slightly smaller AUC of 0.89 but still nonetheless indicates a very healthy behaviour of the model. This can be due to a combination of factors. The main one which may be causing the overall dip is the fact that the detection architecture may be passing a few false positives, forcing the identification to make decisions on completely healthy brains, thus lowering the overall AUC score.

Figure 1. ROC curve for tumor detection model

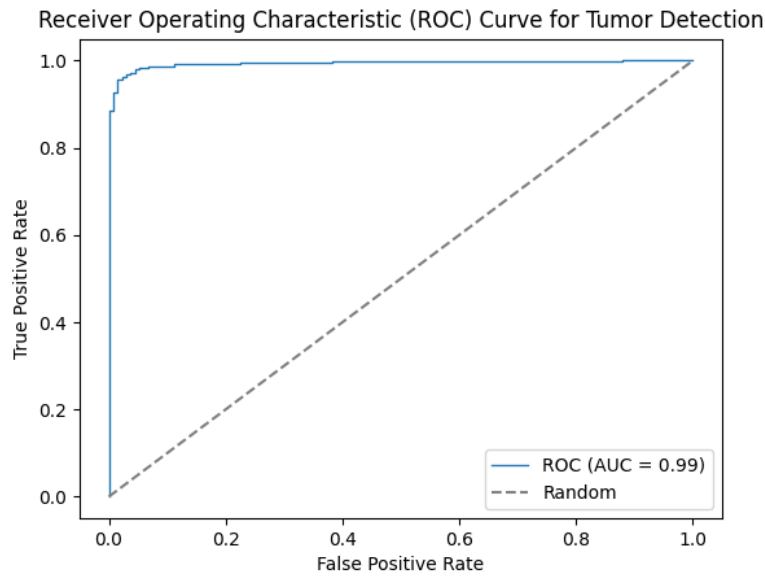
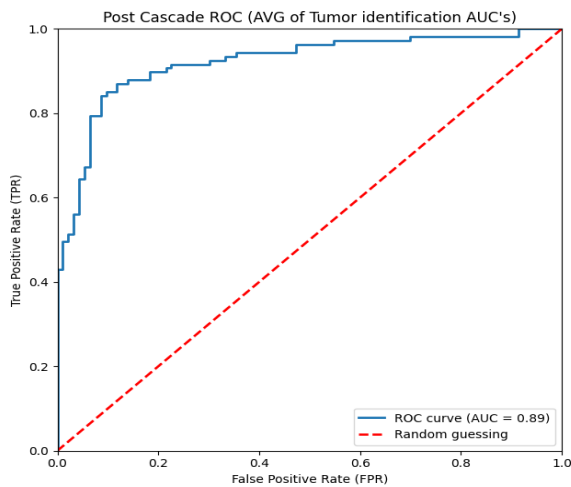


Figure 2. ROC curve after tumor detection and identification models



The F1 - scores tell a very similar story as the ROC curves and the accuracy evaluations with their results shown in Table 1. Both models had strong performances across the board although there was an odd drop in the F1- score of pituitary data. This dip is highlighted later in

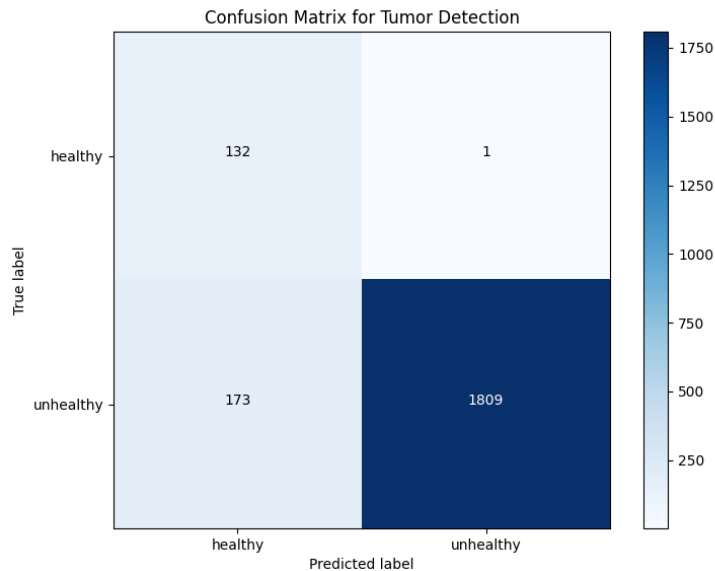
the confusion matrix. The reason it exists may be since the unhealthy data set contains fewer pituitary tumor photos causing the model to have trouble recognizing them because of a lack of training data.

Table 1. Performance Metrics for MRI model stages

Type	Evaluation Score	F1-Score	AUC
Tumor Detection Model	0.91	0.89	0.99
Tumor Identification Model	0.94	0.87	0.96
Cascading Prediction	0.88	0.84	0.89

To highlight the results in a meaningful manner, the models both had confusion matrices to so that their strengths and weaknesses can be properly understood. An example of this would be, Figure 3 which demonstrates the confusion matrix relating to the tumor binary detection model. It performs extremely well when given unhealthy brain tumors but has some struggles in terms of brain scans which are healthy. This is not due to the training data being lopsided like the pituitary problem since the model is trained on an equal amount of healthy and unhealthy data. It may have something to do with the fact that the healthy data isn't from as many different sources, leading the model to maybe have some overfitting in that area, causing misclassifications.

Figure 3. Confusion Matrix for Binary Classification Model



Overall, the results suggest that binary the binary classification model performs well with identifying brains with or without tumors. On the other hands the multi-class classification model does have quite a bit of room for improvement, with one example being improving its ability to recognize pituitary tumors when given fresh data. A key thing to note in the behaviour of categorical identification model is that some of the misclassifications are not its fault. The binary detection model has a 91% accuracy rate, meaning that there are some healthy images which slip through the to the identification model, clouding the data.

Based on results from the MRI detection and identification models, I've successfully combined the two into a cascading decision structure which would allow an individual to learn weather a brain scan is healthy or unhealthy and what type of tumor it has in the latter case. The structure does this with an overall accuracy rate of about 88%, which indicates that it has an easy time distinguishing between different images. The findings indicate that with more time and investment deep learning techniques do have promise in the field of medicine.

The general methodology involves gathering MRI images of patients with and without brain tumors from multiple different sources. The data is then preprocessed and organized accordingly depending on which model its used for. This allows for two separate models to be created and then combined in a cascading prediction structure to allow for outputs to be given after data is received. Along many of the steps mentioned above, performance was evaluated based on metrics such as accuracy, F1 score and AUC. Additionally, data visualization techniques such as confusion matrices and ROC curves were utilized to all for a physical demonstration of how the models performed at various stages.

The findings I have are consistent with research in the field of different deep learning techniques pertaining to medicine. (Havaei et al., 2017) was also similar as they had an accuracy rate of 92.8%, which is very close to the result I achieved.

In conclusion, the project demonstrated the potential for deep learning to be applied to the field of medicine with promising results. The development of accurate and reliable models for the detection and diagnosis of brain tumors can significantly improve patient outcomes and reduce healthcare costs. My model provides a foundation for future research on medical image analysis and can be applied to other medical fields as well.

References

Gab Allah, A. M., Sarhan, A. M., & Elshennawy, N. M. (2021). Classification of brain MRI

tumor images based on Deep Learning PGGAN augmentation. *Diagnostics*, 11(12), 2343.

<https://doi.org/10.3390/diagnostics11122343>

Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.-

M., & Larochelle, H. (2017). Brain tumor segmentation with deep neural networks. *Medical Image Analysis*, 35, 18–31.

<https://doi.org/10.1016/j.media.2016.05.004>