Identifying Brain Tumors

Authors: Bradley Igercich, Sahana Sankar, Laura Lubben, Chloe McGlynn
UC Berkeley School of Information (MIDS) - w207

Table of contents

1

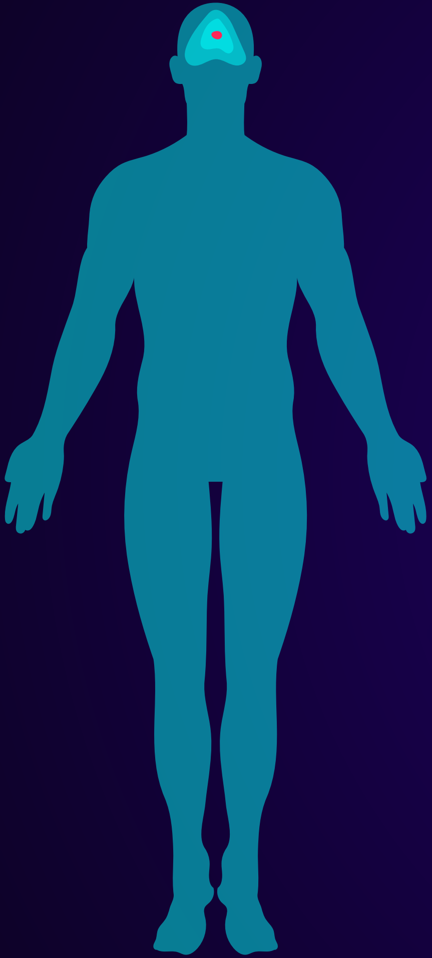
Introduction to the problem

Introduction

- According to the American Brain Tumor Association, approximately 700,000 people in the United States are living with a primary brain tumor, and nearly 80,000 new cases are diagnosed each year .
- Magnetic Resonance Imaging (MRI) is one of the most effective and non-invasive tools for brain tumor detection.
- However, the manual analysis of MRI images by radiologists can be time-consuming.
- Despite advancements in medical technology, the five-year survival rate for patients with malignant brain tumors remains low, at around 36%.
  This underscores the importance of early detection, which can lead to more effective treatment options, such as surgery, radiation, and chemotherapy.

2

Research Question



Research Question

Can we train a model to use MRI brain scans to reliably predict the presence of no brain tumor vs. presence of a glioma, meningioma or pituitary tumor?

Implications for ML Tool:

- Earlier tumor detection for patients undergoing MRI scans
- Validation support for physicians writing a diagnosis/prognosis
- Improved surgical planning
- Patient-tailored brain tumor management
- Opportunities for researchers to build on tool for more advanced models

3

About the Data

Brain Tumor MRI Dataset:

- **Data Sources:**
  - Combined image datasets from figshare.com & other similar kaggle projects
- **Data Shape:**
  - **7023** human brain MRI images total (including no tumor, meningioma, glioma, pituitary)
    - **1621** images featuring a glioma
    - **1645** images featuring a meningioma
    - **1757** images featuring a pituitary tumor
    - **2000** images featuring no tumor
- **Main Features:**
  - Greyscale color // color values
  - 224 x 224 pixels // pixel labels
  - Scan orientation
  - Ensure no individual patient's images are separated across training and test sets

Three types of tumors

Glioma

A type of brain tumor that starts in the glial cells, which are the supportive cells that help protect and nourish nerve cells in the brain.

Pituitary

A growth in the pituitary gland, which can affect hormone levels and overall body functions.

Meningioma

A tumor that forms on the membranes covering the brain and spinal cord, usually benign and slow-growing.

4

Exploratory Data Analysis

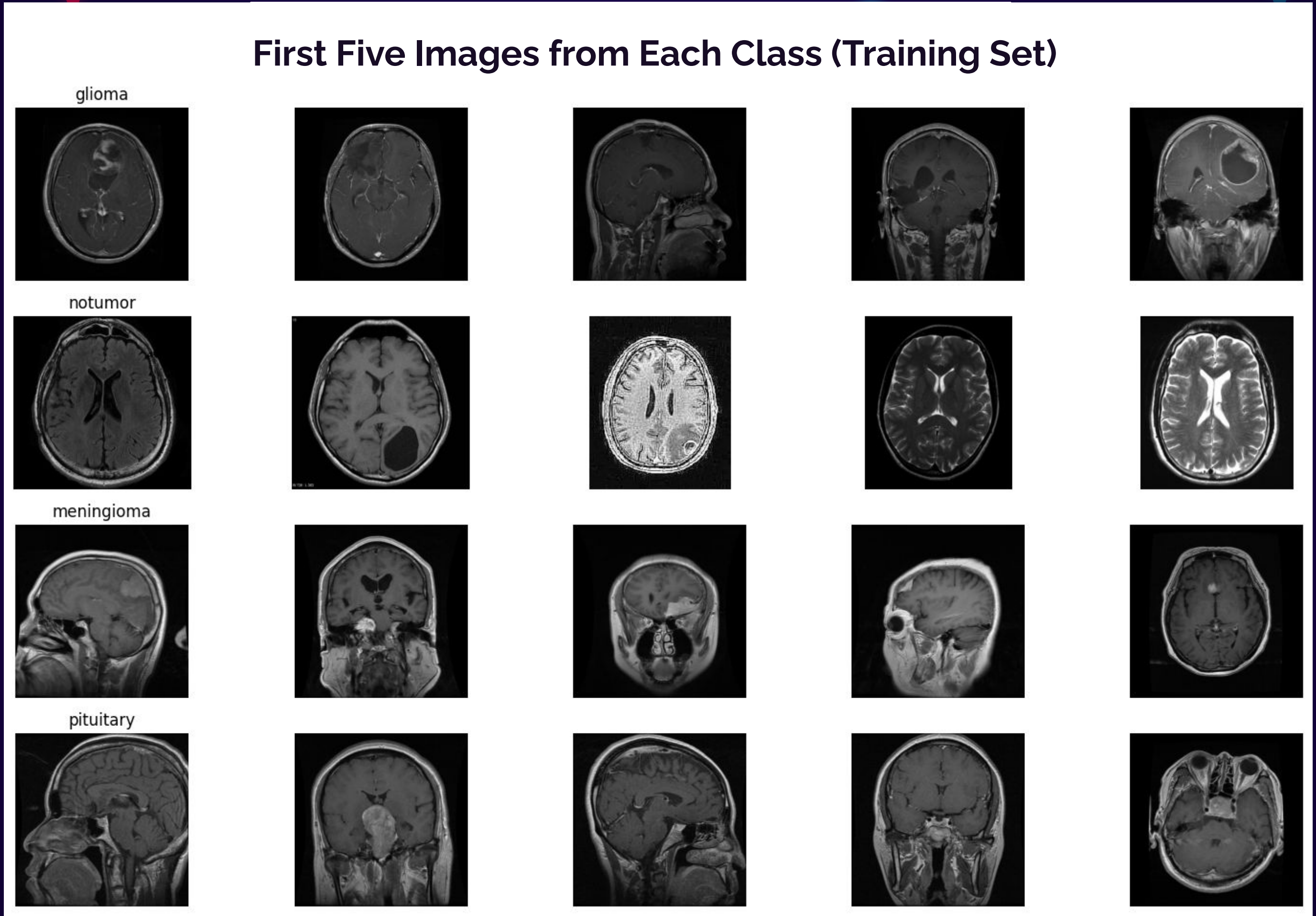
Data Distribution

glioma

no tumor

meningioma

pituitary



Data Preprocessing

Start

Resize Image

- Ensures all images have uniform dimensions (7023, 224, 224, 1)
- Applied to all datasets

Normalize

- Scales pixel values to a standard range [0,1]
- Applied to all datasets

Adjust Brightness and Contrast

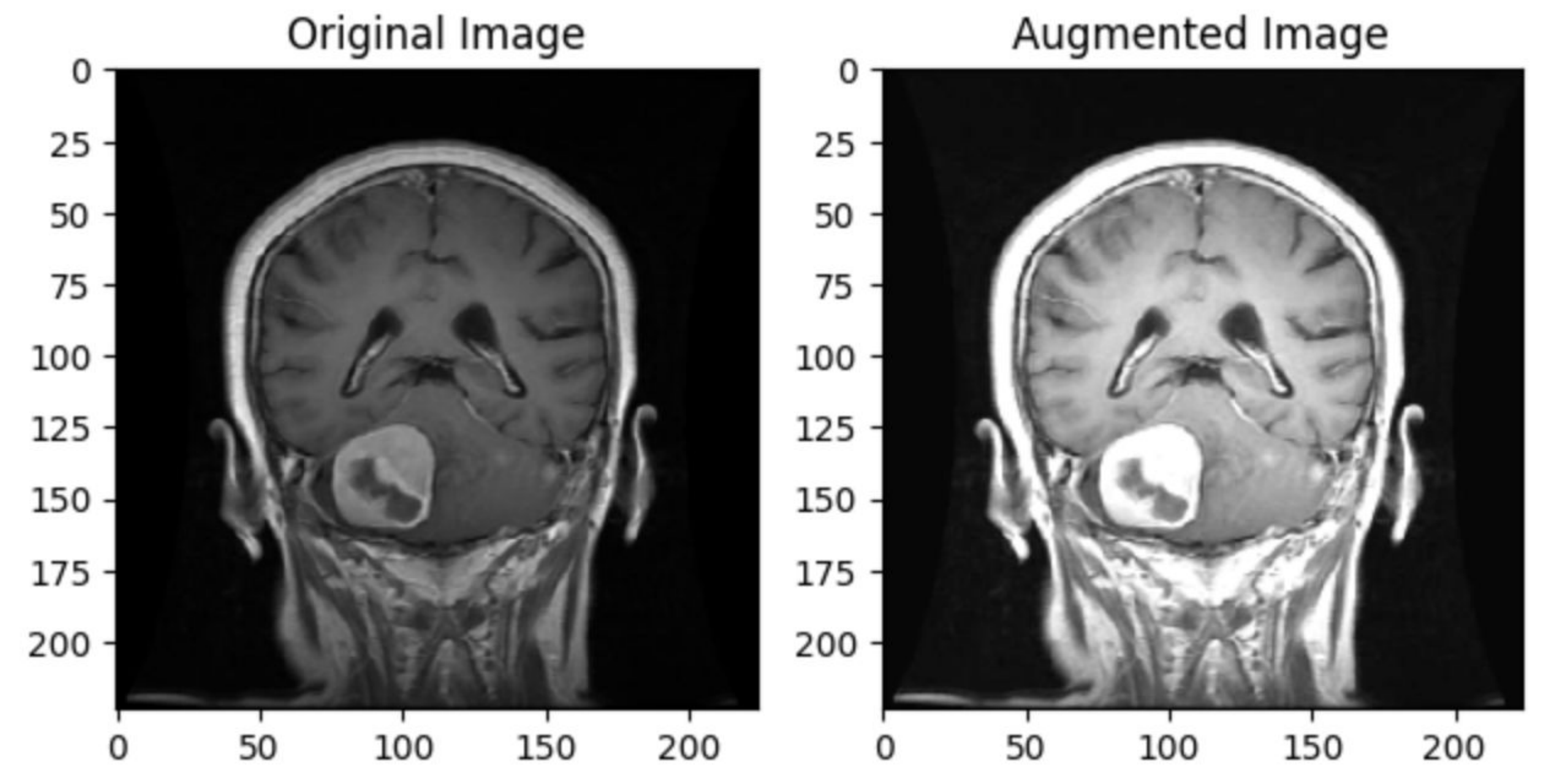
- Improves the visibility of tumors.
- Applied to training set only.

Random Flip

- Helps model generalize better and reduces overfitting by increasing diversity of training dataset.
- Applied to training set only.

End

Data Preprocessing (Example)



5

Prediction Algorithm

Prediction Algorithm:

CNN Model

Feature Generation
- Convolution, Activation and Pooling layers

Pros
- Well established techniques
- Interpretable

+

Transformer Model

Feature Generation
- Self-attention, Feed-forward, multi-head and positional layers

Pros
-Scalability and flexibility
- Recent success with ViTs

Suggested Hybrid Model

Start with a CNN feature Extractor that would include Convolution Layers and Pooling Layers and then layer Transformer Encoding which would include Positional Encoding and Self-Attention Mechanisms. Lastly, we will include a fully connected layer an an output layer.

Pros: Takes strengths of each model - using localized feature extraction from CNNs with contextual understanding of Transformers.

6

Evaluate Results

Baseline Model

- Evaluate performance of our image classification algorithm with a good baseline model appropriate for comparison of performance.
- Use the majority class classifier, which would predict ‘no tumor’ in each case.
- Defaulting our baseline prediction to ‘no tumor’ is beneficial since brain tumor diagnosis is rare.

Class	Percentage (%)	Rate of Prevalence in Reality
No Tumor	28%	
Pituitary	25%	
Glioma	23%	
Meningioma	23%	

Evaluating Results

Minimize False Negatives & Maximize Recall	• Recall = positive cases that were correctly identified by the model • Critical to early & effective treatment. We do not want the model to miss positive cases of cancer.

Multiclass ROC Analysis - Assess Model Discriminatory Ability

- **Approach**: Use One-vs-Rest (OvR) strategy for each tumor class:
  - Class A (no tumor) vs. (Class B + Class C + Class D)
  - Class B (meningioma) vs. (Class A + Class C + Class D)
  - Class C (glioma) vs. (Class A + Class B + Class D)
  - Class D (pituitary) vs. (Class A + Class B + Class C)

- **Metric**: Compute Area Under the Curve (AUC) for each class-specific ROC curve.
- **Micro-averaging**: Average AUC across all classes, weighted by their support.
- **Macro-averaging**: Average AUC independently for each class, then average these values.
- **Benefits**:
  - Each class=specific ROC curve plots the TPR (recall) against FPR at different thresholds
  - Comprehensive evaluation of model's ability to distinguish between tumor types.
  - Visual representation (ROC curves) aids in threshold selection and comparison.
  - Handles class imbalances naturally, crucial for varied tumor prevalence.