ISyE 6740 – Fall 2024

Project Proposal

Brain Tumor MRI Dataset Analysis

Team Members

- Ng, Wee Ding (wng47@gatech.edu)
- Liang, Yawen (yliang375@gatech.edu)
- Deng, Kun (kdeng66@gatech.edu)

Problem Statement

Brain tumors are among the most severe and life-threatening medical conditions, requiring accurate and timely diagnosis for effective treatment. Magnetic Resonance Imaging (MRI) is one of the most commonly used imaging techniques for detecting and diagnosing brain tumors. However, the manual interpretation of MRI scans is expensive, time-consuming, prone to human error, and highly dependent on the expertise of radiologists. With the increasing volume of medical imaging data, there is a pressing need for automated techniques to assist healthcare professionals in diagnosing tumors more efficiently and accurately.

The objective of this project is to develop a model to automatically detect and classify brain tumors from MRI images. The goal of this project is to develop a robust pipeline for analyzing brain tumor MRI images using both unsupervised and supervised learning methods. The project will involve applying dimensionality reduction (PCA, MDS), clustering (Spectral Clustering), and multiple classifiers (Bayes' Theorem, K-Nearest Neighbors, Logistic Regression, and SVM) to explore and predict the presence of tumors. Additionally, we aim to improve diagnostic accuracy and reduce human error through automated image analysis.

Data Source

- Brain Tumor MRI Dataset: https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset/data

Data Description

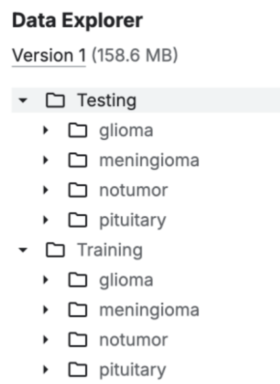

Figure 1: Folder structure of the downloaded dataset from [1]

This dataset contains 7023 images of human brain MRI images which are classified into 4 groups: glioma, meningioma, no tumor and pituitary, where ~18% are testing data, and ~82% are training data. The sizes of the images are not uniform; therefore, images preprocessing must be conducted before we use the images.

We further explored the images. 10 images are picked randomly in training set, for each category, and show the images below: -

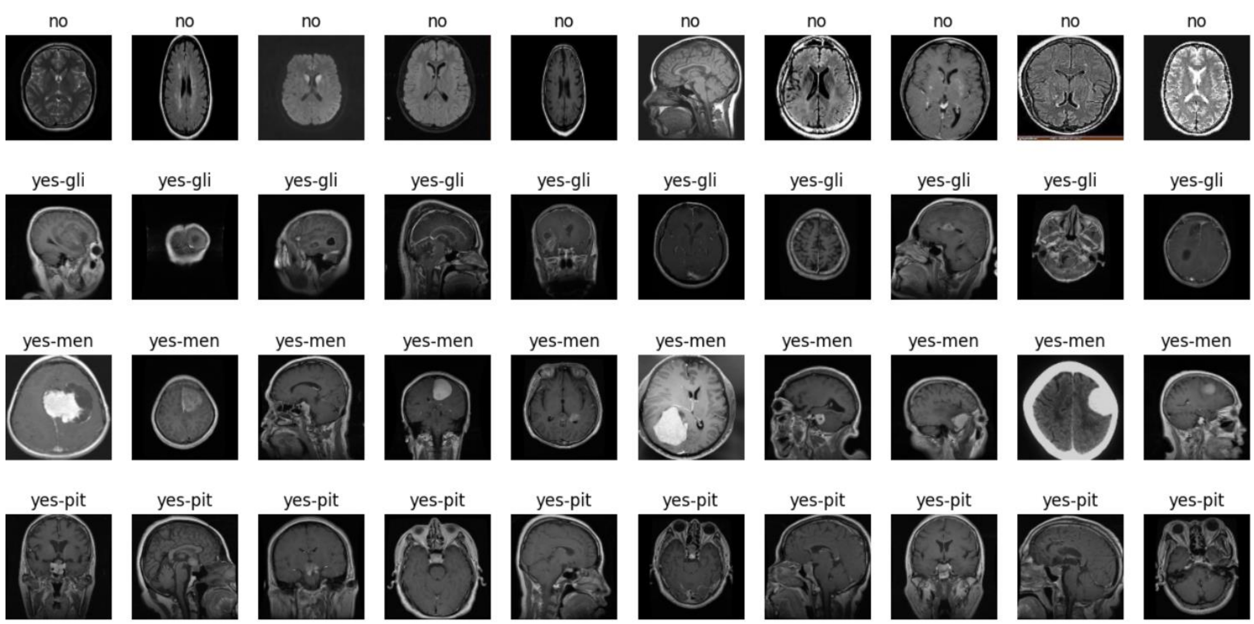

Figure 2: 10 MRI images of no tumor, and 10 images of each of the 3 types of brain tumors (glioma, meningioma, pituitary) [1]

As you can see, it wouldn't be straight forward to classify the tumors manually. Hence, a automatic systematic approach is needed.

Methodology

We like to perform classification based on the models we learnt in IsYE 6740. The following models are proposed: -

- PCA / MDS with Spectral clustering
    - We will use PCA to reduce the dimensionality of MRI images while retaining key variance components. Identify patterns in the data and detect which principal components are most indicative of tumor-related features.
    - We would also use multi-dimensional scaling methods to preserve the distances between MRI images in the reduced space, capturing the true structure of the data.
    - By comparing the result with PCA to observe differences in how dimensionality reduction affects image clustering.
    - Additionally, we will analyze the analyze the data's structure and group similar images together with spectral clustering, which will help us identify key components that are most representative of tumor-related features.

- Expectation-Maximization (EM) algorithm vs K-means algorithm
    - After PCA is performed and we can project the images on a 2D space, where multiple classification models can be performed further for comparison, such as: -
        - EM clustering, it is considered soft clustering algorithm as compared to K-means clustering. We can try running multiple iterations on the training images until it converges to train the model. After that, we can test the model with the testing set to analyze the accuracy.
        - K-means clustering, this is the unsupervised clustering algorithm where we specify the K with the number of clusters to be clustered into, and iteratively assign the points to the centroids with the centroids repeatedly being updated. We'll perform K-means clustering on the images and compare the predictive results with other models.

- Kernel-Density Estimator (KDE), Gaussian Mixture Models (GMM)
    - We'll perform KDE and GMM for data visualization. We will analyze how the four classes of MRI images (glioma, meningioma, no tumor, and pituitary) are distributed and shaped, gaining deeper insights into the underlying data structure.

- Bayes' Theorem / KNN/ Logistics / SVM classifier

- We use these supervised classification classifiers to evaluate by different matrix. For Bayes's Theorem, we use a probabilistic approach to classify MRI image by evaluating the confusion matrix and accuracy metrics.
- KNN method used in similarity-based classification by assigning labels based on the closest neighbors. We can evaluate the precision, recall, and F1-score for the model accuracy to identify the images.
- The logistics regression model uses a linear model to predict the probability of tumor presence. We can visualize the ROC-AUC curve to evaluate the model performance.
- The SVM model is used to find the optimal decision boundary for separating tumor and non-tumor classes. We will examine the images by evaluating the accuracy, precision, recall, and F1-score for the model performance.

We'll be using scikit-learn library for this project, which will be built on NumPy, SciPy and matplotlib.

If time allows, we'll continue to explore with neural network models, such as CNN, utilizing TensorFlow built-in library for model's implementation.

Results Interpretation

We will evaluate the performance of each model based on classification accuracy and other key metrics such as precision, recall, and F1-score. The final report will include a fully implemented pipeline for brain tumor classification, as well as classification reports, confusion matrices, and visual results like ROC curves and sample classifications. A summary of findings, model performance, and recommendations for future work will be provided.

References

[1] *Brain Tumor MRI Dataset* https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset/data

[2] *Machine Learning Approach to Brain Tumor Detection and Classification*, By Alice Oh, Inyoung Noh, Jian Choo, Jihoo Lee, Justin Park, Kate Hwang, Sanghyeon Kim, Soo Min Oh https://arxiv.org/abs/2410.12692

[3] *Brain Tumor Types Classification using K-means Clustering and ANN Approach,* Angona Biswas; Md. Saiful Islam, and more, IEEE https://ieeexplore.ieee.org/document/9331115

[4] *Pre-trained deep learning models for brain MRI image classification*, Srigiri Krishnapriya, Yepuganti Karuna https://pmc.ncbi.nlm.nih.gov/articles/PMC10157370/