Executive Summary:

Alzheimer's Disease Prediction from Structural Brain MRI

Authors: Gokcen Buyukbas, Selman Ipek, Cisil Karaguzel, Tushar Pandey

Overview

This project aims to develop a machine learning-based approach for predicting Alzheimer's disease from structural brain MRI (Magnetic Resonance Imaging) scans. Alzheimer's disease is a progressive neurodegenerative disorder that affects millions of people worldwide, and early detection is crucial for timely intervention and management. By leveraging advanced deep learning and machine learning techniques, we aim to classify early Alzheimer's disease into two classes: cognitively normal (CN) and mild cognitive impairment (MCI).

The data for this project comes from the ADNI3 cohort of the Alzheimer's Disease Neuroimaging Initiative (ADNI) study. We used "Accelerated Sagittal MPRage MRI" scans of 561 subjects - 183 with Mild Cognitive Impairment (MCI) and 378 with Cognitively Normal (CN). Data preprocessing involved selecting a single MRI session per subject, ordering the slices, choosing the middle 65 slices, and converting them into pixel arrays. The preprocessed dataset was split into training (70%), validation (15%), and test (15%) sets for robust model development and generalizability assessment.

Methodology

We developed three models: a Convolutional Neural Network (CNN) and a hybrid CNN + XGBoost model in two versions. A CNN is a deep learning model that consists of multiple building blocks, such as convolution layers, pooling layers, and fully connected layers, and is designed to automatically and adaptively learn spatial hierarchies of features(pixels) in images through a backpropagation algorithm. XGBoost is a gradient boosting algorithm that iteratively trains an ensemble of decision trees, where each new tree is trained to correct the errors made by the previous trees, and the final prediction is a weighted sum of the predictions from all the trees. We trained a CNN model to classify MRI scans, achieving an AUC of 0.58 on the test set. The hybrid model leverages the CNN's feature extraction capabilities and combines them with the predictive power of the XGBoost algorithm. We developed two versions of the hybrid model. The first one uses the CNN to find a lower, 65-dimensional, representation of the MRI scans, down from 61440, and these features, along with patients' age and sex, serve as inputs for XGBoost. This hybrid model improved the performance, achieving 0.61 AUC. The second version used the probability output of the CNN classifier as a new feature along with age and sex for the XGBoost Classifier; however, this model performed similarly to the initial CNN classifier, with 0.59 AUC.

Conclusion, Potential Impact, and Future Directions

In conclusion, this project contributes to the growing body of research by demonstrating the potential of deep learning and hybrid machine learning approaches in predicting Alzheimer's disease from structural brain MRI scans. The successful development of such a highly accurate model for early detection would immensely improve the management of the disease, patient outcomes, cost reduction, and advancements in neuroimaging research. To further enhance the model's performance and robustness, future directions include expanding the dataset with more examples, preprocessing improvements, model optimization, and incorporating a multi-modal approach.