QBS 101.5_ Project Proposal

Topic: Machine learning-based brain MRI analysis for early Alzheimer's disease detection
Dataset: https://www.kaggle.com/datasets/yasserhessein/dataset-alzheimer

Introduction:
Alzheimer's disease (AD), a neurodegenerative disorder potentially caused by the development of beta-amyloid plaques in the brain, is the most common form of dementia and the fifth-leading cause of death among people over the age of 65 (AlSaeed & Omar, 2022). Although there is currently no cure for AD, early diagnosis and disease detection can help to improve quality of life in the elderly and slow down further deterioration in those who have already started to develop cognitive problems (Rasmussen & Langerman, 2019). With brain MRI scans, healthcare providers and researchers have the opportunities to detect brain abnormalities associated with mild cognitive impairment and can use these scanned images to visualize any potential shrinkage of the hippocampus.

This project aims to utilize machine learning to analyze brain MRI scans and identify patterns associated with Alzheimer's disease. The objective of this project is to create an image classification model by extracting meaningful features from MRI images and training a Convolutional Neural Network (CNN) model to differentiate between healthy and disease-affected brains. The designed model will be trained to learn spatial patterns and relevant features from the MRI data, enabling accurate and time-efficient generalization to new, unseen MRI scans. Ultimately, the trained model can be deployed in clinical settings, providing automated diagnosis, aiding in early disease detection, and supporting timely intervention for improved patient outcomes.

Prior work:
In recent years, deep-learning-based approaches have achieved great success across many diseases, including but not limited to Alzheimer's disease analysis. In particular, lots of efforts have been made to develop strategies for early detection with the aid of machine learning and Convolutional Neural Networks. Several studies have proposed using image classification techniques for the diagnosis and detection system in AD (AlSaeed & Omar, 2022).

In early 2015, Liu et al. (2015) proposed an inherent structured-based multi-view learning method (ISLM) for AD classification. It mainly involves 1) multi-view feature extraction using multiple templates and gray matter tissues as tissue-segmented brain image for feature extraction, 2) subclass clustering-based feature selection through using voxel selection, and 3) use a support vector machine (SVM) based ensemble classification. The experiment result shows that proposed ISML method obtains an accuracy of 93.83% and specificity of 95.69%, and sensitivity of 92.78% for AD vs. normal controls (Liu et al., 2015).

More recently, AlSaeed and Omar (2022) proposed to apply transfer learning on a pre-trained AD diagnosis model ResNet50 using Softmax in the classifier layer. They then tested the proposed approach on the ADNI and MIRIAD dataset and concluded that the model with Softmax classifier outperforms SVM and RF in all performance measures (e.g., accuracy, specificity, and sensitivity). AlSaeed and Omar (2022) believe that a transfer learning approach using CNNs could be used to develop a low-cost and non-invasive diagnostic tool for Alzheimer's disease.

Based on these published papers and other prior works, the **most cost-effective solution** for an MRI image classification system is to **use and deploy a previously trained model (e.g., ResNet50)**. This approach will not only save time and money for healthcare providers but also allow them to achieve

higher levels of disease detection accuracy. Ultimately, the goal of this project is to build a model that could be used to assist and achieve better patient outcomes, reduce unnecessary workloads for healthcare providers, and realize a more accessible and affordable healthcare environment.

Data Description:
The dataset is composed of two main sub-datasets: 'Test' and 'Train'.

'Train' has 717 "MildDemented" scans, 52 "ModerateDemented" scans, 2560 "NonDemented" scans, and 1792 "VeryMildDemented" scans. In contrast, 'Test' has 179 "MildDemented" scans, 12 "ModerateDemented" scans, 640 "NonDemented" scans, and 448 "VeryMildDemented" scans. Both 'Test' and 'Train' datasets only have images in JPEG format. The total number of observations are 6400.

Aims:
Aim 1: Training model
1.1. Ensure to resize all MRI scan images to a uniform size
1.2. Train the model on the training data with a specified number of epochs

Aim 2: Testing/classification model
2.1. Classify brain MRI scans into different stages of Alzheimer's disease progression with the use of TensorFlow
2.2. Evaluate on testing data using an accuracy metric
2.3. Provide a guideline for this AD classification model (e.g., model accuracy, potential caveats)

Expected outcome:
This project has the potential to develop a Python-based machine-learning project to analyze brain MRI scans for AD diagnosis. It will use the Kaggle dataset and a convolutional neural network (CNN) to extract features from the imported MRI scans.

Potential limitations:
- Limitation No.1: The dataset may not be large enough to train a CNN that is accurate enough for clinical use
- Limitation No.2: Potential issue of imbalanced dataset (i.e., different numbers of images for each level of AD) might cause bias and error in precision

Potential mitigation strategies:
- Mitigation strategies for Limitation No.1:
    - Deploys and try other previously trained models (with a greater number of dense layers) to increase the size of trained dataset for brain MRI images
- Mitigation strategies for Limitation No.2:
    - Resampling (oversampling the minority class or under-sampling the majority class)
    - Assigns different costs to misclassifications in different classes

References:

AlSaeed D, Omar SF. Brain MRI Analysis for Alzheimer's Disease Diagnosis Using CNN-Based Feature Extraction and Machine Learning. Sensors (Basel). 2022 Apr 11;22(8):2911. doi: 10.3390/s22082911. PMID: 35458896; PMCID: PMC9025443.

Liu M., Zhang D., Adeli E., Shen D. Inherent Structure-Based Multiview Learning With Multitemplate Feature Representation for Alzheimer's Disease Diagnosis. *IEEE Trans. Biomed. Eng.* 2016;63:1473–1482. doi: 10.1109/TBME.2015.2496233

Rasmussen J, Langerman H. Alzheimer's Disease - Why We Need Early Diagnosis. Degener Neurol Neuromuscul Dis. 2019 Dec 24;9:123-130. doi: 10.2147/DNND.S228939. PMID: 31920420; PMCID: PMC6935598.

Kavitha C, Mani V, Srividhya SR, Khalaf OI, Tavera Romero CA. Early-Stage Alzheimer's Disease Prediction Using Machine Learning Models. Front Public Health. 2022 Mar 3;10:853294. doi: 10.3389/fpubh.2022.853294. PMID: 35309200; PMCID: PMC8927715.

2022 Alzheimer's disease facts and figures. Alzheimers Dement. 2022 Apr;18(4):700-789. doi: 10.1002/alz.12638. Epub 2022 Mar 14. PMID: 35289055.