

QBS 101.5_ Project Proposal

Research Topic: Machine learning-based brain MRI analysis for early Alzheimer's disease detection

Dataset: <https://www.kaggle.com/datasets/yasserhessein/dataset-alzheimer>

Related paper:

https://www.researchgate.net/publication/348486602_Deep_learning_based_prediction_of_Alzheimer's_disease_from_magnetic_resonance_images

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 that are associated with Alzheimer's disease. Models will be trained on a dataset of different classes of MRI images from people with and without Alzheimer's disease, and will then be used to classify new brain MRI images and help to improve the accuracy of AD diagnosis and early detection.

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 (which could potentially be transferred to *NumPyarray* and act like array variables in Python). 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 (i.e., Very Mild, Mild, Moderate, None) with the use of scikit-learn, PyTorch, and TensorFlow (*subject to change*)
- 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:

- The dataset may not be large enough to train a CNN that is accurate enough for clinical use
- Potential issue of imbalanced dataset (i.e., different numbers of images for each level of AD) might cause bias and error in precision

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.

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.