# Diagnosing Alzheimer’s Disease with Machine Learning

### By Eugene Fong

### Spring 2023

### Prof Chaojie Wang’s Capstone in Data Science class

### DATA 606 @UMBC

# OVERVIEW

### PHASE 1 – Proposal & Planning

###### Literature search: Alzheimer’s Disease

###### Diagnosing Alzheimer’s is hard!

###### OASIS dataset

### PHASE 2 – Data prep, EDA, & Data Viz

###### Stats

###### Data viz

###### Filtering

### PHASE 3 – Model training & Deployment

###### Data cleansing & transformation

###### Stratified train/test split

###### Model training

###### Prediction evaluation

# PHASE 1 – Proposal & Planning

# Literature search: Alzheimer’s disease (AD)

Alzheimer’s disease (AD) is a progressive neurodegenerative disease which causes a decline in cognitive functions until death. It is the main cause of dementia and rapidly increasing worldwide, roughly doubling every 5 years. The direct cause is still an issue of hot debate, but the literature has many risk factors listed such as: “increasing age, genetic factors, head injuries, vascular diseases, infections, and environmental factors” (Breijyeh & Karaman, 2020). From a clinical standpoint, AD patients exhibit “memory loss… change of personality… progressive loss of cognitive functions” (Breijyeh & Karaman, 2020). From a biomolecular examination, autopsies have revealed neuritic plaques and neurofibrillary tangles, which are predicted to be caused by amyloid-beta (AB plaques). However, whether AB plaques are truly the root cause of Alzheimer’s is now being reexamined after neuroscientist and physician, Matthew Schrag, reexamined the initial research identifying AB plaques as the root cause by Sylvain Lesne. Some of the results from Lesne’s 2006 papers may have been manipulated (*Potential Fabrication in Research Images Threatens Key Theory of Alzheimer’s Disease*, n.d.). Prior to 2021, the AD drugs available could only address some of the individual symptoms of AD, with no effect on the final outcome of the disease (Breijyeh & Karaman, 2020). Beginning in 2021, the FDA approved 3 new AD drugs: Aducanumab in 2021, Lecanemab and Donanemeb in 2023 – all with their own fair share of controversy:

* “Aducanumab does not cure or reverse [AD]… reduced amyloid plaque levels, but that did not translate to any clinical effect… Potentially serious harms are common” (Woloshin & Kesselheim, 2022).
* “Lecanemab slowed clinical decline by 27%” (*Lecanemab, the New Alzheimer’s Treatment*, n.d.).
* “Donanemab slowed mental decline by 35%” (Reardon, 2023).

These 3 new drugs target AB plaques, but none of them can reverse or stop Alzheimer’s completely, which may be an indication that AB plaques may not be the root cause of Alzheimer’s and that as a result of fabricated research results, researchers have been pursuing the wrong target.

The process of diagnosing AD has changed quite a bit over the decades. The Mini-Mental State Exam (MMSE) was one of the early attempts in 1973 to diagnose AD using a simple questionnaire (*Mini-Mental State Exam (MMSE) Test for Alzheimer’s / Dementia*, n.d.). However, the sensitivity ranged from a low 23% up to 76%, while specificity ranged from 40% - 94% (Arevalo-Rodriguez et al., 2021), but was limited to patients who already had mild cognitive impairment (MCI) and works better with repeated follow-ups to mark the pace of mental degradation converting to AD (Arevalo-Rodriguez et al., 2021). The next major iteration was in 1984 with the National Institute on Aging and the Reagan Institute Working Group on Diagnostic Criteria for the Neuropathological Assessment of Alzheimer’s Disease (NIA-Reagan) criteria, which could diagnose patients after death via an autopsy looking for AB plaques and tangle formation (Beach et al., 2012). This method could only identify 3 combinations of dementia with low, medium and high probability ranges, and had a subset of patients that could not be classified at all, leaving a large patient population underserved (Beach et al., 2012). Further refinement led to the creation of the Consortium to Establish a Registry for Alzheimer’s Disease (CERAD) diagnostic in 1991 which has grown to include more methods, such as: autopsy, biopsy, brain scans which can look for more biomarkers of 25 different combinations of AD and be applied in living patients (Beach et al., 2012). Sensitivity improved to 71% - 87%, specificity range of 44% - 70% (Beach et al., 2012). These ranges are certainly an improvement upon the MMSE and NIA-Reagan methods, however, it still leaves a large range of uncertainty.

Recently, there have been attempts to diagnose AD with the help of automated methods and Machine Learning (ML) algorithms. Researchers can now utilize a group of “atlas normalization” techniques to compare brain scans across different patients, while accounting for individual differences in head shape, developmental and internal brain matter.

TODO

* Convert into MD format