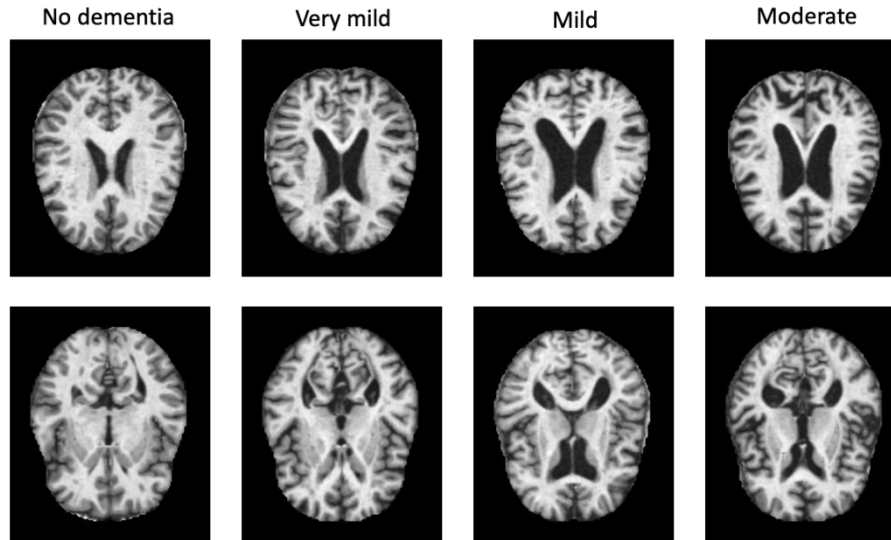# Alzheimer's Image Classification Case Study

A DS 4002 Case Study by Ashley Kim

Alzheimer's is a neurodegenerative condition and the most common form of dementia in the United States, impacting over 6 million Americans as of today. Memory loss is obviously the most notable symptom of Alzheimer's, but it also causes impaired thinking and judgment, changes in personality, and an inability to perform daily tasks.

This chronic nature of dementia development suggests that the patient's brain experiences gradual change over time, and studies have shown that these changes can take place more than a decade before the onset of symptoms. Current medical literature cites tissue atrophy, or shrinkage, as one of the biggest indicators of neuronal death. This atrophy can be seen in medical imaging, and by the use machine learning and data science models, there is hope for early identification and intervention for patients suffering from dementia.

A convolutional neural network (CNN) is a class of artificial neural network that is most commonly used in visual image analysis. As a new data scientist working to help individuals suffering from Alzheimer's, you're tasked with generating an image classification model utilizing a convolutional neural network algorithm that classifies a set of MRI scan images into the correct stage of Alzheimer's with at least an 80% accuracy. With the provided dataset, there are four stages of dementia that each scan can be classified into: no dementia, very mild dementia, mild dementia, and moderate dementia.

Alzheimer's Case Study Rubric

DS 4002 – Spring 2023 – Ashley Kim
Due: May 10th, Noon
Submission format: Upload link to GitHub repository on UVA Canvas

Individual Assignment

Why am I doing this?
This case study allows you to leverage your data science knowledge by producing an image classification model that helps detect Alzheimer stages using MRI scans of the brain. It is the hope that by working through this assignment, it can encourage you to see the impactful ways data can be utilized in a real-world context – especially in the field of medicine and early detection of fatal neurogenerative conditions, data science could very much save a person's life.

What am I going to do?
Use the provided dataset on Kaggle that contains four datasets of MRI images separated by Alzheimer stages to create an image classification model in either R or Python utilizing a convolutional neural network (CNN). In addition, after you build your model, you will need to pass through several images (that I will provide you) through your model to see how well it can predict with new data. A standard workflow would be to first familiarize yourself with the images you're analyzing within the dataset through some exploratory data analysis, then conduct preprocessing, standardizing, and finally building out and evaluating your machine-learning based classification system.

Your final deliverables should include:
- Well documented source code for your image classification model
  - Well commented, well documented and <u>clean</u>!
- A markdown file citing any resources (journal articles, websites, etc.) referenced in helping you create your model
- All these above should be included in a GitHub repository that is submitted to the class Canvas site

Tips for success:
- There are a lot of different ways to go about producing the model. I recommend looking at the initial references provided before doing your own research. These references have been selected to start you off and build your knowledge about convolutional neural network algorithm that can hopefully provoke you into making astute decisions in how to go about building the classification system
- Exploratory data analysis will be of huge help for you to understand the Spotify dataset you're working with. I suggest undergoing a few series of data visualizations in either R or Python to gain valuable insight into all the features of the tracks in the respective playlist.

This can also mean undertaking algorithms like clustering to view the songs by numerical features closely

How will I know I have succeeded? You will meet expectations on this case study when you successfully follow and complete the criteria in the rubric below:

Spec Category	Spec Details
Formatting	- One GitHub repository (submitted via link on Canvas) - Create a new GitHub repository for this assignment titled 'CS2_LastName-FirstName' that contains: - README.md - LICENSE.md - Source Code File - REFERENCES.md
README.md	Brief summary of what you've produced for the case study, this does not have to be super detailed but provide enough information to orient people to your repository
Source Code File	Well documented R Markdown or Jupyter Notebook file that contains the code used to execute your song recommendation system. Suggestions of what to include in the source code are as follows: - Exploratory data analysis visualizations/plots - Preprocessing of data - Feature generation - Song recommendation model
REFERENCES.md	Markdown File titled "REFERENCES.md" with citing any resources (journal articles, websites, etc.) referenced in helping you create your model in IEEE Documentation style. Also include brief annotations under each citation on how each reference informed/helped you for this case study.

Acknowledgements: Thank you to Professor Alonzi for providing the rubric structure!