Alzheimer's Detectives - Project Proposal Report

CMPT 310 - D200: Introduction to Artificial Intelligence and Machine Learning

Group Information

Manjari Prasad - mps12@sfu.ca

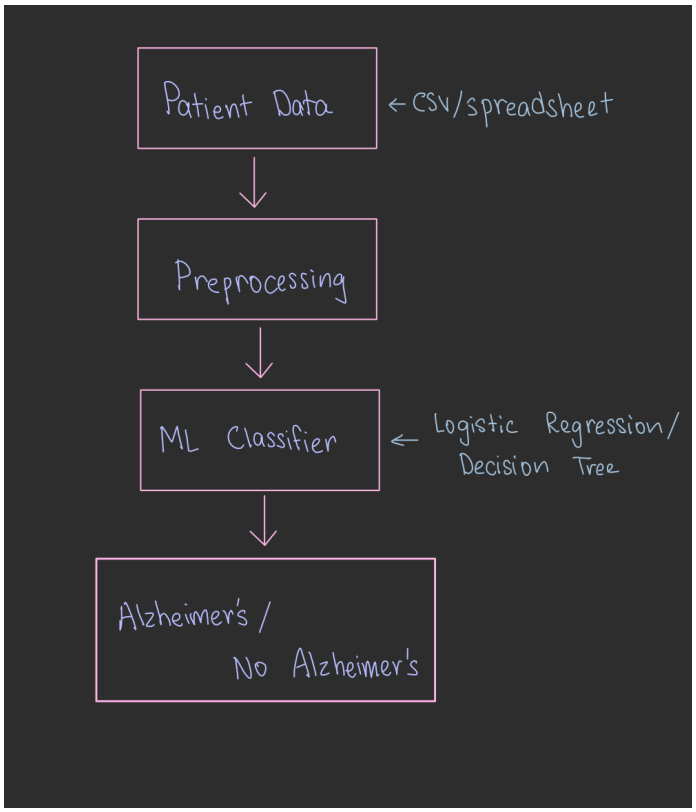
Wan Yu Wendy Wong - wyw9@sfu.ca

Beyzanur Kuyuk - bka81@sfu.ca

Project Idea

Our project aims to build an AI-based system to assist in the early detection of Alzheimer's disease using patient health records. The primary goal is to classify individuals as either at-risk or not at-risk for Alzheimer's based on clinical and cognitive features such as age, memory test scores, family history, and MRI-based indicators (if available). This system will serve as a low-cost, accessible decision-support tool that can flag patients who may benefit from further medical evaluation.

The system will take structured patient data as input. Each row will correspond to a patient, and each column will represent a clinical or demographic feature (such as MMSE scores, age, gender, etc.). The output will be a binary classification: "Alzheimer's" or "No Alzheimer's." In the first iteration, we plan to use supervised learning with classification techniques like logistic regression and decision trees. If time permits, we will explore more advanced methods such as Random Forests or neural networks. We may also explore adding basic interpretability features such as showing which input features contributed most to a given prediction.



System Diagram

Tools and Resources

Programming Language:
Our project will be implemented in Python. We will leverage widely-used libraries such as scikit-learn, NumPy, and Pandas for data processing and model evaluation. For model training and experimentation, we will use PyTorch.

Datasets:
We plan to use publicly available Alzheimer's datasets from platforms such as UCI and Kaggle. These datasets contain clinical and cognitive data relevant to early diagnosis.

Dataset Generation:
We do not intend to generate our own dataset, as high-quality medical data is typically large-scale and requires careful curation. However, we may perform data cleaning and preprocessing (e.g., handling missing values) as needed to prepare the dataset for training.

Simulator vs. Real-world Data:
We will work exclusively with real-world data sourced from public repositories. No simulators (e.g., PyBullet or OpenAI Gym) will be used in this project.

Project Timeline

Milestone 1 (July 2):
Download and explore the dataset
Handle missing values and select key features (e.g., age, MMSE, education)
Train a baseline model using logistic regression
Evaluate the model using accuracy and confusion matrix
Create plots to visualize early performance

Milestone 2 (July 30):
Train and compare additional models (e.g., decision tree)
Perform error analysis (e.g., which inputs are often misclassified)
Visualize decision trees or model coefficients
Evaluate model performance across different features (e.g., by age group)
Begin writing report and how-to guide

Minimal Viable System

What is the simplest working version of your system?
The simplest version of our system will read a spreadsheet of patient records or surveys (such as age, memory test scores, and clinical ratings). It will use that information to make predictions about whether a patient may be showing early signs of Alzheimer's disease. We will use a basic machine learning algorithm like logistic regression to learn the patterns from the data and apply them to make predictions on new cases.

What is the core functionality you'll prioritize first?
The core functionality we would like to prioritize is to have the system take in a patient's data and predict whether they have Alzheimer's or not. After being trained, our system should be able to classify new patient records as "Alzheimer's" or "No Alzheimer's". Getting this prediction working efficiently and reliably will be our primary goal.

How could your system be useful or insightful even in a basic form?
Even in its basic form, our system will help identify characteristics that are most strongly linked to Alzheimer's risk. This can help screen people who may need further clinical testing or early action. It can also assist in diagnosis and in exploring patterns in the data that lead to different diagnosis results. In addition, it can open doors for more research on factors that play a role in increasing Alzheimer's risk and even support discovering treatments.