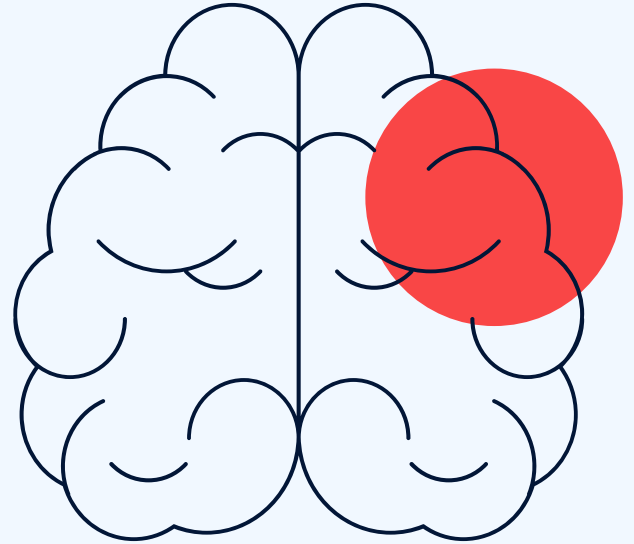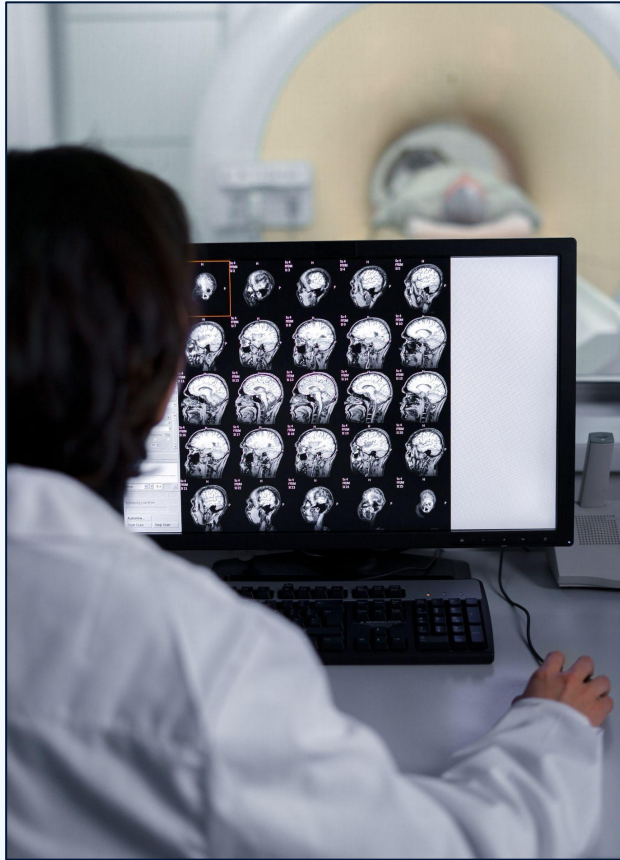# Detection of Alzheimer's Disease
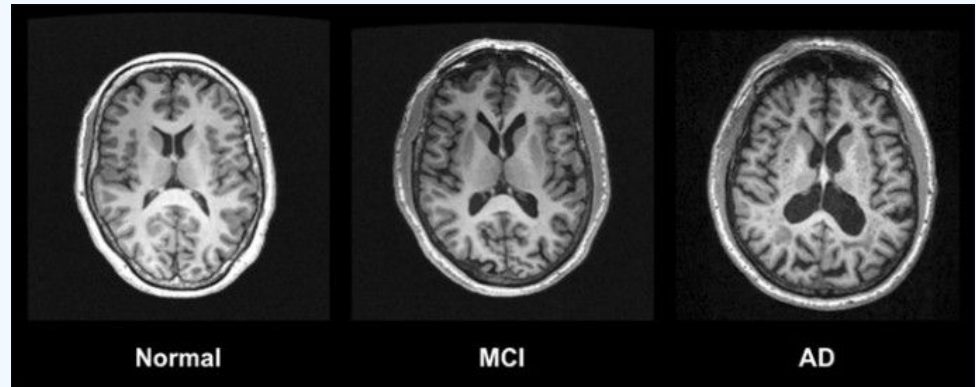
Theresa Sumarta, Adeline Chin
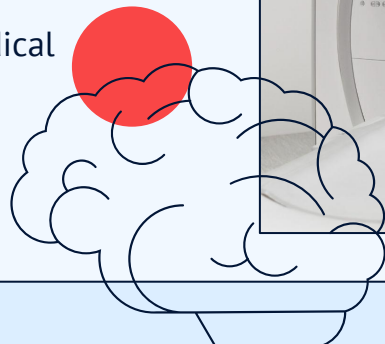
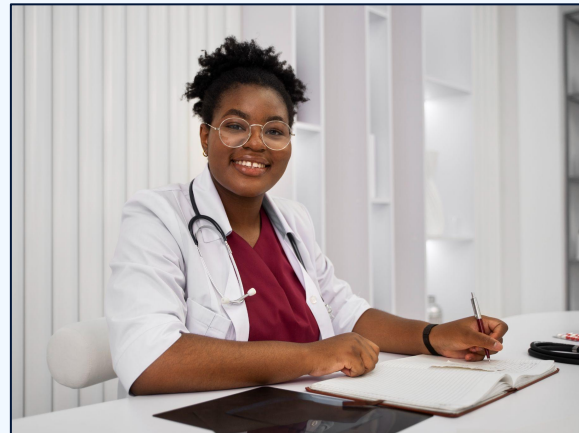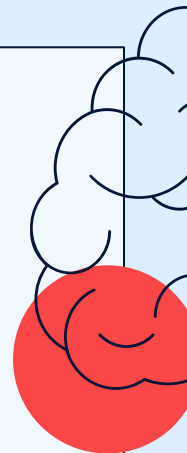# Alzheimer's Disease

Normal   MCI   AD

Objectives

1. Classification of those with Alzheimer's and those without using only brain MRIs

2. Prediction of cognitively normal brain MRIs and MRIs with mild cognitive impairment that will later develop Alzheimer's

3. Classification of those with Alzheimer's and those without using only symptom and medical history data
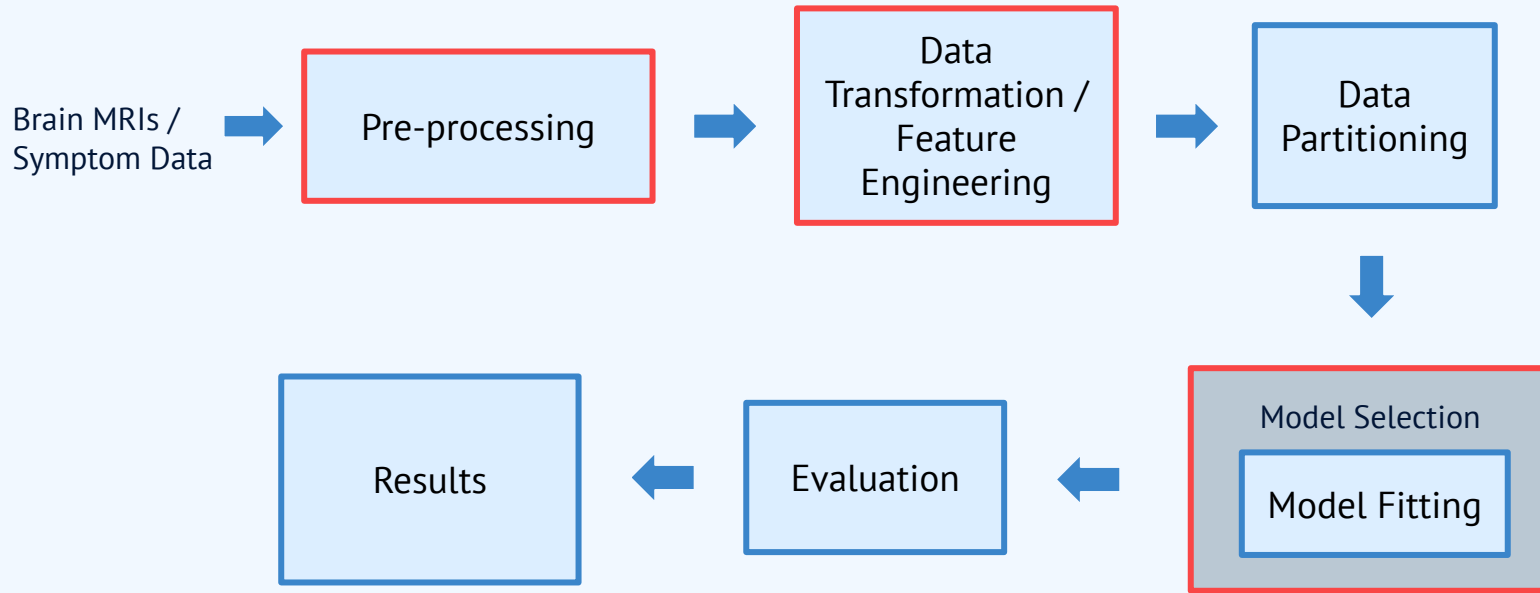
Datasets

Alzheimer's Disease Neuroimaging Initiative (ADNI)

- Longitudinal, multicenter, 4-phase study spanning from 2004 to 2016
- Over 5,000 subjects from North America
- Cognitively Normal (CN), Mild Cognitive Impairment (MCI), or Alzheimer's Disease (AD)
- At least 2 visits per subject, baseline and final
- 5 total datasets used for this project
  - 2D image data for classification (CN, AD)
  - 2D image data for prediction of AD (CN, MCI)
  - 3D image data for classification (CN, MCI, AD)
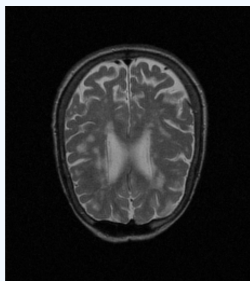 - Symptom dataset and Medical History dataset for classification (approx. 2,000 subjects)
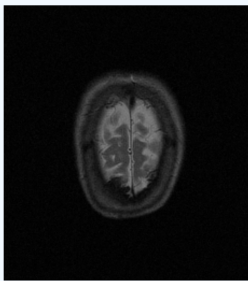
Approach and Methodology

Brain MRIs / Symptom Data → Pre-processing → Data Transformation / Feature Engineering → Data Partitioning ↓

Results ← Evaluation ← Model Selection: Model Fitting

2D Image Classification
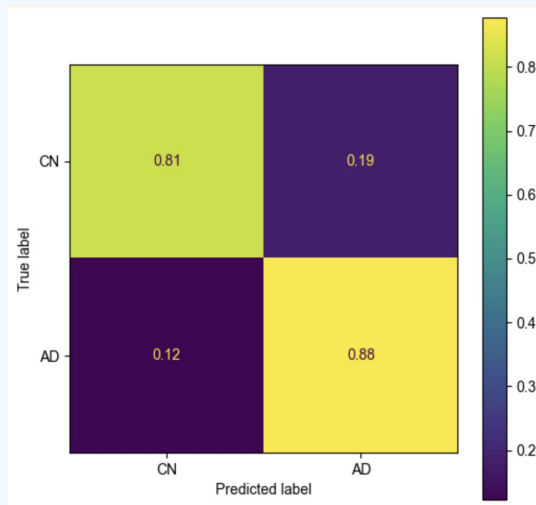
Pre-processing & Data Transformation

Slice #25 example

Slice #48 example

- Scans #22-36 were extracted for each subject
- DICOM to pixel array (224 x 224)
- Resizing
- Brightness adjustments
- Contrast adjustments
- Random flip

CNN Model and Results

Kernel - 3,3
Strides - 1,1
Pool - Max
Pool Size - 2,2
Padding - Same
Learning Rate - 0.001
Optimizer - Adam
Activation - ReLU

Test Accuracy = 87.03%   **Sensitivity = 0.88**
F1 Score = 0.8621   **Specificity = 0.81**

2D Image AD Conversion Prediction
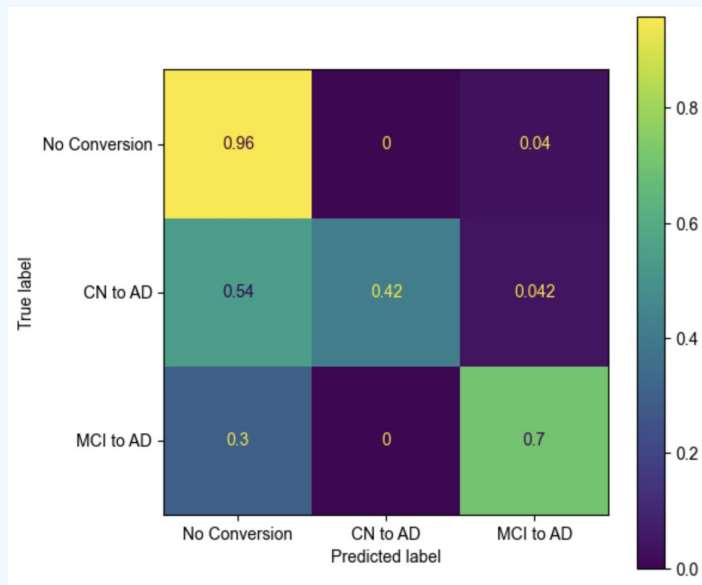
CNN Model

- Used the same pre-processing and data augmentation methods as the previous model

- Used the previous CNN model as a starting point and modified layers and re-tested hyperparameters
  - Dropout layers
  - Batch normalization

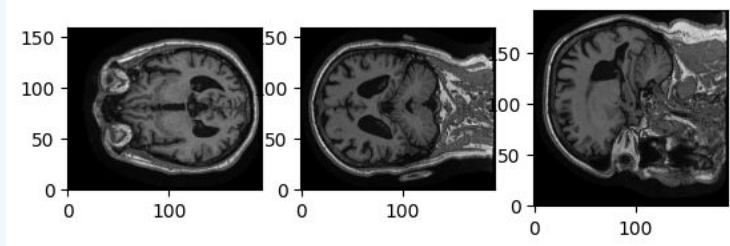**Test Accuracy = 84.28%**
F1 Score = 0.7528
Matthew's Correlation Coefficient = 0.7606

Results

3D Image Classification

Pre-processing & Data Augmentation

- Convert 2000 MRI Images from .NIIfTI files to 3D data arrays of shape 256 x 256 x 166
- Slice 3D data arrays into shape 206 x 72 x 18.
- Z-score normalization
- PCA to reduce dimensionality of large dataset

3D CNN Model

- **Kernel** - 5,5,5
- **Strides** - 2,2,2
- **Pool** - Max
- **Pool Size** - 1,1,1
- **Padding** - Same
- **Learning Rate** - 0.001
- **Optimizer** - Adam
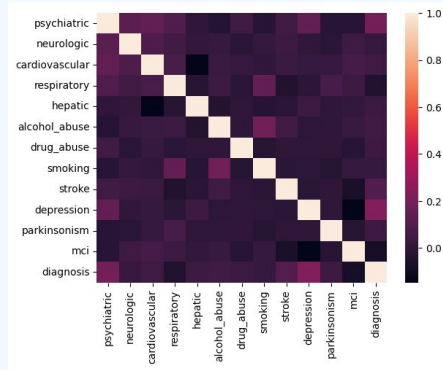- **Activation** - ReLU

Test Accuracy = 0.4792
F1 Score = 0.1638
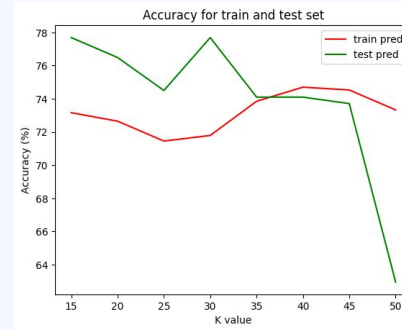**Matthew's Correlation Coefficient = 0.1125**

Symptom Classification
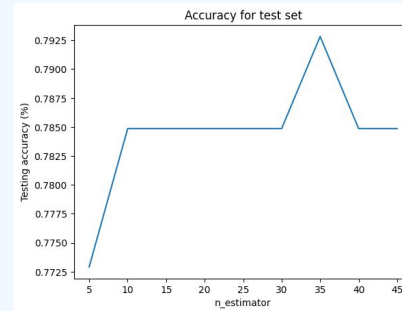
Pre-processing & Data Augmentation

- Medical history: psychiatric, neurologic, cardiovascular, respiratory, hepatic, alcohol abuse, drug abuse and smoking
- Diagnostic summary: depression, parkinsonism, MCI
- One hot encoding
- Drop rows that contained nan values

Model optimization

KNN:
k value = 36

Random Forest:
n_estimators = 35

Symptom Classification results

7 optimized models

Classifier model	Test accuracy (%)
Random Forest	**79.3**
Decision Tree	78.5
Kernel SVM	78.5
Logistic Regression	76.5
SVM	75.3
KNN	73.7
Naive Bayes	62.5

<u>Random Forest Model:</u>

F1 Score = 0.7957

Matthew's Correlation Coefficient = 0.6779

Challenges and Limitations, Bias

- 2D classification: Choosing subset of horizontal slices
- 3D classification: Long computation time, information loss from PCA
- Symptoms and Medical history classification: Many columns containing null values in dataset, small sample size that causes overfitting
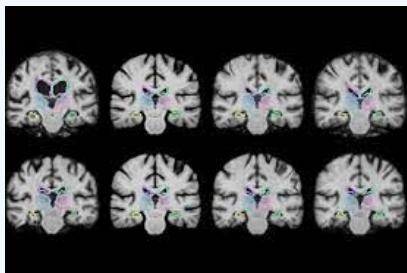
- Imbalanced dataset:
  - Significant number of participants over the age of 70 (80%)
  - Higher proportion of white participants (93%)
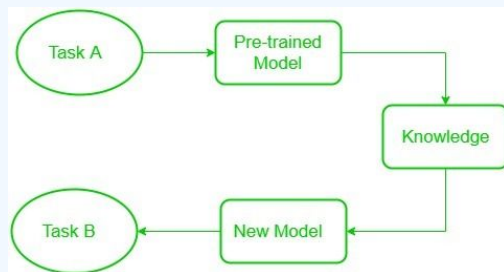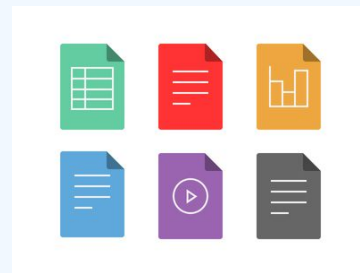- Unintended cognitive biases from individuals who train models
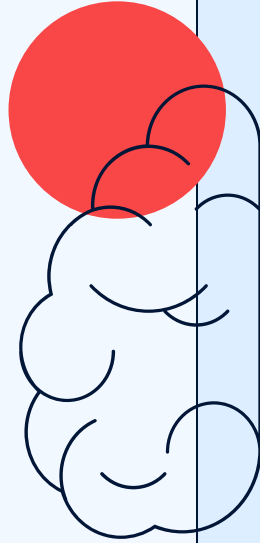
Future Analysis

Models that do not require conversion of medical files

Transfer learning

Models that take multiple types of data

Questions