**Objective**

Our goal is to predict progressions in Alzheimer’s Disease (AD) through only clinical features using machine learning techniques. AD being an irreversible neurodegenerative disease, affecting an estimated 5.3 million Americans with no discovered preventive cure, the study would help doctors subduing its effects with better information.

**Dataset**

- ADNI dataset: 800 subjects from 50 sites from the US and Canada.
- Some clinical features: MOCA, MODACH, NEUROBAT etc.
- Visits: sc (Baseline), m06 (6 months), m12 (12 months) and so on.

**Descriptive Statistics**

- 348 common observations
- Projections on m24 and m36 month visits
  - Principal Component Analysis: 2D, 3D
  - Independent Component Analysis: 2D, 3D
  - Non negative matrix factorization: 2D, 3D

**Preprocessing**

1. Data Cleaning:
   - Reduced to 21 relevant datasets
   - Visits: Up to 12 months
   - Eliminated features and entries with considerable missing values
   - Filled missing values using linear interpolation
2. Normalization
   - Z-score: \( z = \frac{x - \mu}{\sigma} \)
   - Min-Max: \( X_{\text{norm}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \)

**Dimensionality Reduction**

3D NMF

**Clustering**

1. K-means
2. Gaussian Mixture Models
3. Agglomerative Clustering

**Observations and Results**

- 3D NMF best projects the stages of progression in Alzheimer’s namely Healthy Cohort, Mild Cognitive Impairment (MCI) and Dementia for m24 and m36 visits using the data until 12 months.
- Best Clusters representing stages of Progression are obtained using K-means for m24 and Agglomerative Clustering for m36.
- The axis of most relevant and informative features represent memory, cognition and baseline symptoms.

**Conclusions**

- Prediction of future Alzheimer’s progressive stage using the past
- Extracted features responsible for different progression rates

**Impact**

- Additional information about Alzheimer’s progression velocity
- Predictive tests allow early detection and characterization of distinct disease subgroups based on clinical heterogeneity
- Helps doctors subduing effects with better stage information
- 9.2 million fewer cases of the disease in 2050 (Refer [3])
- Improved clinical trial design
- Individualized clinical care

**Future Directions**

- Predicting onset, progression, and clinical subtypes of Parkinson’s disease using machine learning, Faraz Faghri et. al., Brain

**References**

- Explore features causing progressions in both Parkinson’s and Alzheimer’s disease.
- Correlate CSF biospecimen with Alzheimer’s progression space.
- Identify biomarkers predicting progressions to customize treatments.
- Replicate study on other datasets say AddNeuroMed and ItalianADNI.