Seminar Report

On

Assessment of Cognitive Impairments and Dementia using AI/ML Methods

Aayush Prasad (18CS30002)

Guided By: Prof. Partha Pratim Chakrabarti

Jointly Guided by: Prof. Aritra Hazra, Prof. Rajlakshmi Guha

Department of Computer Science and Engineering

Indian Institute of Technology, Kharagpur

Spring Semester, 2020-21

April 9, 2021

Contents

Abbreviations

AD	**A**lzheimer's **D**isease
MCI	**M**ild **C**ognitive **I**mpairment
LBD	**L**ewy **B**ody **D**ementia
FTD	**F**ronto **T**emporal **D**ementia
MRI	**M**agnetic **R**esonance **I**maging
PET	**P**ositron **E**mission **T**omography
ADNI	**A**lzheimer's **D**isease **N**euroimaging
AIBL	**A**ustralian **I**maging, **B**iomarker & **L**ifestyle Study of Ageing
MMSE	**M**ini-**M**ental **S**tate **E**xam
ADAS-13	**A**lzheimer's **D**isease **A**ssessment **S**cale
LSN	**L**ongitudinal **S**iamese **N**etwork
GMM	**G**aussian **M**ixture **M**odel
BIC	**B**ayesian **I**nformation **C**riterion
RF	**R**andom **F**orest
AUC	**A**rea **U**nder **C**urve
EBM	**E**vent **B**ased **M**odel
DEBM	**D**iscriminative **E**vent **B**ased **M**odel
CSF	**C**erebro**s**pinal **F**luid
CRBM	**C**onditional **R**estricted **B**oltzmann **M**achine
SVM	**S**upport **V**ector **M**achine
ANN	**A**rtificial **N**eural **N**etwork
EST	**E**xample **S**imilarity **T**able

Chapter 1

Cognitive Impairment and Dementia

1.1 Dementia and its types

Cognitive Impairment

- Cognitive impairment is when a person has trouble **remembering, learning new things, concentrating, or making decisions** that affect their everyday life.
- Cognitive impairment ranges from mild to severe. With mild impairment, people may begin to notice changes in cognitive functions, but still be able to do their everyday activities.
- Severe levels of impairment can lead to losing the ability to understand the meaning or importance of something and the ability to talk or write, resulting in the inability to live independently.

Dementia Dementia is a general term for loss of memory, language, problem-solving and other thinking abilities that are severe enough to **interfere with daily life**.

Common Types of Dementia

- **Alzheimer's Disease**

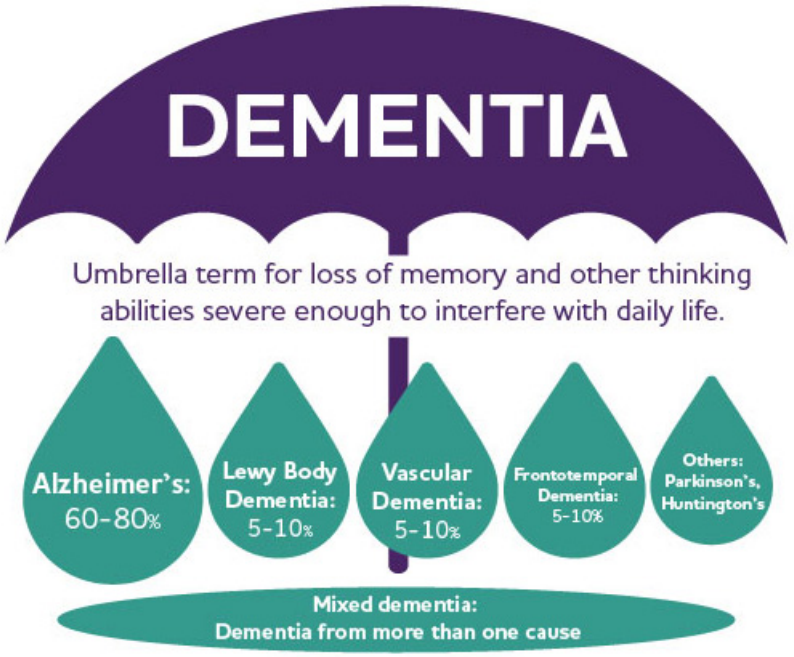

FIGURE 1.1: Types of dementia

  - Alzheimer's is the **most common cause of dementia**. Alzheimer's disease accounts for 60-80% of dementia cases.
  - Alzheimer's is a **progressive disease**, where dementia symptoms gradually worsen over a number of years. In its early stages, memory loss is mild, but with late-stage Alzheimer's, individuals lose the ability to carry on a conversation and respond to their environment.
  - Symptoms:
    * Early Symptom: Difficulty remembering newly learned information
    * Disorientation
    * Mood and behavior changes
    * Deepening confusion about events, time and place
  - People with Alzheimer's tend to develop **plagues** and **tangles** in a predictable manner beginning in areas important for memory before spreading to other regions

- **Lewy Body Dementia**
  - Lewy body dementia (LBD) is a type of **progressive dementia** that leads to a decline in thinking, reasoning and independent function because of **abnormal microscopic deposits** that damage brain cells over time.

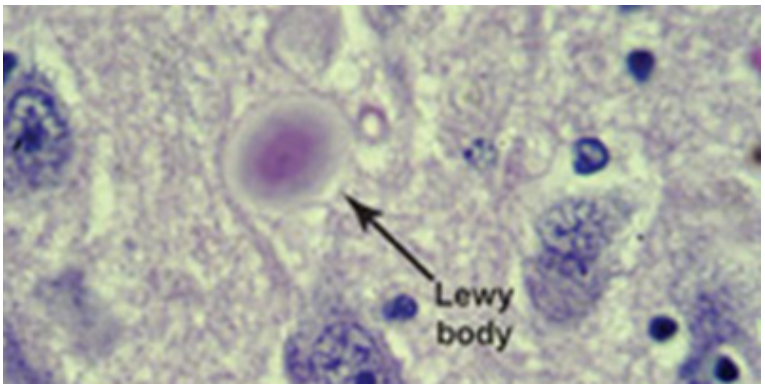

FIGURE 1.2: Lewy Body

  - Many people with both Lewy body dementia and Parkinson's dementia also have plaques and tangles.
  - Symptoms:
    * Changes in thinking and reasoning
    * Confusion and alertness that varies significantly from one time of day to another or from one day to the next.
    * Slowness, gait imbalance and other parkinsonian movement features.
    * Well-formed visual hallucinations.

- **Frontotemporal Dementia**
  - Frontotemporal dementia (FTD) refers to a group of disorders caused by progressive nerve cell loss in the brain's **frontal lobes or its temporal lobes**.
  - The nerve cell damage caused by frontotemporal dementia leads to loss of function in these brain regions, which variably cause deterioration in behavior, personality and/or difficulty with producing or comprehending language.
  - Types of frontotemporal dementia:
    * Frontal variant: affects behavior and personality.
    * Primary progressive aphasia: difficulty communicating
  - Symptoms:
    * Behavior and/or dramatic personality changes
    * Socially inappropriate, impulsive, or repetitive behaviors
    * Impaired judgment
    * Apathy

FIGURE 1.3: Temporal and frontal lobes

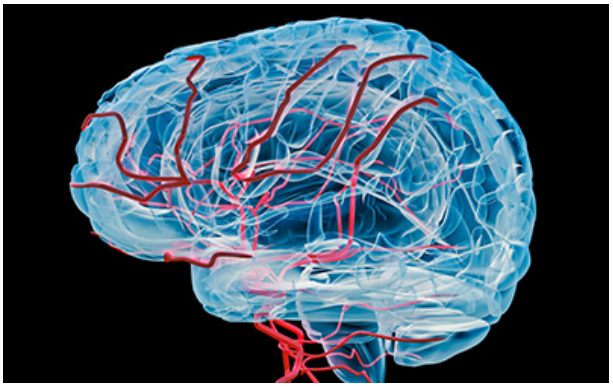

FIGURE 1.4: Blood Flow in Brain in Vascular Dementia

- **Vascular Dementia**
  - Vascular dementia is a decline in thinking skills caused by conditions that **block or reduce blood flow** to various regions of the brain, depriving them of oxygen and nutrients.
  - Vascular dementia is widely considered the second most common cause of dementia, accounting for 5% to 10% of cases.
  - Symptoms:
    * Confusion
    * Disorientation
    * Trouble speaking or understanding speech
    * Physical stroke symptoms, such as a sudden headache
- **Mixed Dementia**
  - Mixed dementia is a condition in which brain changes of **more than one cause** of dementia occur simultaneously.

  - In the most common form of mixed dementia, the abnormal protein deposits associated with Alzheimer's disease coexist with blood vessel problems linked to vascular dementia
  - Mixed dementia symptoms may vary, depending on the types of brain changes involved and the brain regions affected

1.2 Diagnosis of Dementia

Early Diagnosis of Dementia

- A diagnosis may well provide **long-awaited answers** for a failing memory, communication problems and changes in behaviour.
- An early diagnosis opens the door to **future care and treatment**.
- It helps people to plan ahead while they are still able to make important decisions on their care and support needs and on **financial and legal matters**.
- It also helps them and their families to receive **practical information, advice and guidance** as they face new challenges.
- It will help to eliminate the possibility of other, potentially treatable, conditions with dementia-like symptoms being responsible for memory, communication, behaviour and other problems.
- It can help people with dementia to have **access to relevant information**, resources and support, make the most of their abilities and potentially benefit from drug and non-drug treatments available

Medical Assessment for Dementia A medical assessment for dementia generally includes:

- **Medical history**: Typical questions about a person's medical and family history might include asking about whether dementia runs in the family, how and when symptoms began, changes in behavior and personality, and if the person is taking certain medications that might cause or worsen symptoms.
- **Physical exam**: Measuring blood pressure and other vital signs may help physicians detect conditions that might cause or occur with dementia. Some conditions may be treatable.
- **Neurological tests**: Assessing balance, sensory response, reflexes, and other cognitive functions helps identify conditions that may affect the diagnosis or are treatable with drugs.

Tests to Diagnose Dementia The following procedures also may be used to diagnose dementia:

- **Cognitive and neuropsychological tests**: These tests are used to assess memory, problem solving, language skills, math skills, and other abilities related to mental functioning.
- **Laboratory tests**: Testing a person's blood and other fluids, checking levels of various chemicals, hormones, and vitamins, can help find or rule out possible causes of symptoms.
- **Brain scans**: These tests can identify strokes, tumors, and other problems that can cause dementia. Scans also identify changes in the brain's structure and function. The most common scans are:
  - **Computed tomography (CT)**, which uses x rays to produce images of the brain and other organs
  - **Magnetic resonance imaging (MRI)**, which uses magnetic fields and radio waves to produce detailed images of body structures, including tissues, organs, bones, and nerves
  - **Positron emission tomography (PET)**, which uses radiation to provide pictures of brain activity
- **Psychiatric evaluation**: This evaluation will help determine if depression or another mental health condition is causing or contributing to a person's symptoms.
- **Genetic tests**: Some dementias are caused by a known gene defect. In these cases, a genetic test can help people know if they are at risk for dementia.

Chapter 2

Literature

2.1 Problem Statements

Disease Progression Prediction

- **Disease Progression**: The worsening of a disease over time. This concept is most often used for chronic and incurable diseases where the stage of the disease is an important determinant of therapy and prognosis.
- By developing machine learning models where progression information from various time-points is accumulated, we obtain insight into how performance can change over time, which provides us with a certainty measure.
- It also enables the possibility to perform an exploratory study concerning the chronology of important predictors and its influence, which may allow a clinician to understand the patient's disease progression and thus, support its decision concerning the best treatment strategy.

Identification of Clusters

- This is important as clustering algorithms can **find patterns** across people with cognitive impairments that are difficult for medical practitioners to find.
- Clustering can also serve as a useful data-preprocessing step to identify homogeneous groups on which to build supervised models.

- Types of Clustering Algorithms:
  - **Centroid-based Clustering**: Centroid-based clustering organizes the data into non-hierarchical clusters
  - **Density-based Clustering**: Density-based clustering connects areas of high example density into clusters. This allows for arbitrary-shaped distributions as long as dense areas can be connected.
  - **Distribution-based Clustering**: This clustering approach assumes data is composed of distributions, such as Gaussian distributions. As distance from the distribution's center increases, the probability that a point belongs to the distribution decreases.
  - **Hierarchical Clustering**: Hierarchical clustering creates a tree of clusters. Hierarchical clustering is well suited to hierarchical data, such as taxonomies.
- **Multi-layer Clustering**: In some domains, the set of attributes may be partitioned in two or more disjoint subsets (layers) according to some criteria, such as the physical meaning of the attributes or the way data on specific attributes have been collected.
  - For example, in the Alzheimer's disease domain, the first layer can be the laboratory data, while the second layer can be the clinical data.
- In some other domain, different layers may contain the same attributes but collected in various time periods.

The developed machine learning solutions potentially offer substantial clinical impact by augmenting clinical decision-making for physicians and healthcare specialists.

2.2 Predicting Disease Progression

Literature Survey

1. **Modeling and prediction of clinical symptom trajectories in Alzheimer's disease using longitudinal data [Bhagwat et al., 2018]**
   - A computational framework comprising machine learning techniques for
     - modeling symptom trajectories, and
     - prediction of symptom trajectories

     using multimodal and longitudinal data is presented.

- Primary analyses on three cohorts from Alzheimer's Disease Neuroimaging Initiative (ADNI), and a replication analysis using subjects from Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing (AIBL) is done.
- The prototypical symptom trajectory classes is modelled using clinical assessment scores from mini-mental state exam (MMSE) and Alzheimer's Disease Assessment Scale (ADAS-13) at nine timepoints spanned over six years based on a hierarchical clustering approach.
- Euclidean distance between longitudinal clinical score vectors as a similarity metric and Ward's method as linkage criterion for **clustering** was used.
- For prediction, a **longitudinal Siamese neural-network (LSN)** with novel architectural modules for combining multimodal data from two timepoints is presented.
- The trajectory modeling yields two (stable and decline) and three (stable, slow-decline, fast-decline) trajectory classes for MMSE and ADAS-13 assessments, respectively.

2. **Learning the progression and clinical subtypes of Alzheimer's disease from longitudinal clinical data [Satone et al., 2018]**

   - Unsupervised and supervised machine learning approaches are used for subtype identification and prediction.
   - Machine learning methods are applied to the extensive clinical observations available at the Alzheimer's Disease Neuroimaging Initiative (ADNI) data set to identify patient subtypes and to predict disease progression.
   - Unsupervised clustering method **Gaussian Mixture Model (GMM)** was used to identify clusters. **Bayesian Information Criterion (BIC)** was used to select the optimum number of clusters in GMM.
   - Various supervised learning classifiers were compared to predict an individual patient's progression. **Random forest (RF)** classifier provided the most accurate model.
   - The analysis depicts the progression space for the Alzheimer's disease into low, moderate and high disease progression zones.
   - The work will enable early detection and characterization of distinct disease subtypes based on clinical heterogeneity.

3. **Forecasting the progression of Alzheimer's disease using neural networks and a novel preprocessing algorithm [Albright, 2019]**

   - This study investigated machine learning approaches to use clinical data to predict the progression of AD in future years.

- Data from 1737 patients were processed using the **"All-Pairs" technique**, a novel methodology created for this study involving the comparison of all possible pairs of temporal data points for each patient.
- Machine learning models were trained on these processed data and evaluated using a separate testing data set (110 patients).
- A neural network model was effective (mAUC 5 0.866) at predicting the progression of AD, both in patients who were initially cognitively normal and in patients suffering from mild cognitive impairment.
- Such a model could be used to identify patients at early stages of AD and who are therefore good candidates for clinical trials for AD therapeutics.

4. **Multi-study validation of data-driven disease progression models to characterize evolution of biomarkers in Alzheimer's disease [Archetti et al., 2019]**
   - Understanding the sequence of biological and clinical events along the course of Alzheimer's disease provides insights into dementia pathophysiology and can help participant selection in clinical trials.
   - Their objective is to train two data-driven computational models for sequencing these events, the **Event Based Model (EBM)** and **discriminative-EBM (DEBM)**, on the basis of well-characterized research data, then validate the trained models on subjects from clinical cohorts characterized by less-structured data-acquisition protocols.
   - The Alzheimer's Disease Neuroimaging Initiative (ADNI) data set was used as training set for the constriction of disease models while a collection of multi-centric data cohorts was used as test set for validation. Cross-sectional information related to clinical, cognitive, imaging and cerebrospinal fluid (CSF) biomarkers was used.
5. **Predicting Brain Degeneration with a Multimodal Siamese Neural Network [Ostertag et al., 2020]**
   - To study neurodegenerative diseases, longitudinal studies are carried on volunteer patients.
   - During a time span of several months to several years, they go through regular medical visits to acquire data from different modalities, such as biological samples, cognitive tests, structural and functional imaging. These variables are heterogeneous but they all depend on the patient's health condition, meaning that there are possibly unknown relationships between all modalities.

- In this work, **a neural network architecture** for multimodal learning, able to use imaging and clinical data from two time points to predict the evolution of a neurodegenerative disease, and robust to missing values is presented.

6. **Machine learning for comprehensive forecasting of Alzheimer's Disease progression [Fisher and Company, 2019]**

   - The ability to simultaneously simulate dozens of patient characteristics is a crucial step towards personalized medicine for Alzheimer's Disease.
   - Here, an unsupervised machine learning model called a **Conditional Restricted Boltzmann Machine (CRBM)** is used to simulate detailed patient trajectories.
   - Data comprising 18-month trajectories of 44 clinical variables from 1909 patients with Mild Cognitive Impairment or Alzheimer's Disease is used to train a model for personalized forecasting of disease progression.
   - Synthetic patient data including the evolution of each sub-component of cognitive exams, laboratory tests, and their associations with baseline clinical characteristics is simulated.

ML techniques for Diagnosis

- The most widely used classification techniques are **support vector machine (SVM)**, **artificial neural network (ANN)**, and **deep learning and ensemble methods**.
- The primary difference between SVM and ANN is the nature of the optimization problem. SVM gives a globally optimal solution, while ANN gives locally optimal solution.
- SVM based models have been widely used for Alzheimer's disease showing its robustness. This is because techniques like ANN suffers from the drawbacks of local minima, which is not the case with SVM.
- The abundant usage of SVM also stems from the fact that it is easier to interpret as compared to deep neural networks which act as black box models.
- However, ANNs are more versatile and robust when it comes to incremental learning, modelling sequential data, and quantizing high dimensional spaces. Therefore, novel variants of ANN can be used for Alzheimer‘s in future.
- Deep learning and ensemble learning techniques give promising results by modeling highly complex data with high accuracy.

Literature Gaps

- Less focus is given on the **clinical interpretability** of ML models used.
- Researchers have given more importance to the feature extraction phase and not much to the classification phase
- More work is required in formulation of machine learning models which can integrate information from **various modalities** for early diagnosis of Cognitive Disorders
- Identification of **subtle biomarkers** to predict disease progression.
- **Novel ML/AI frameworks** using multi-modal, longitudinal and cross-sectional data to improve prognostic predictions.

2.3 Clusters in Dementia Population

Literature Survey

1. **Homogeneous clusters of Alzheimer's disease patient population [Gamberger et al., 2016a]**
    - A novel clustering tool is used with the goal to identify subpopulations of the AD patients that are homogeneous in respect of available clinical as well as in respect of biological descriptors.
    - Data were obtained from the Alzheimer's disease neuroimaging initiative (ADNI) database.
    - Three clusters of patients with significant problems with dementia was obtained. The evaluation of properties of these clusters demonstrates that brain atrophy is the main driving force of dementia.
    - Positive properties of the **multi-layer clustering approach** are that no explicit distance measure is used, that the algorithm has a well-defined stopping criterion, that examples can include both numerical and categorical attributes with unknown values, and that availability of many attributes is an advantage. The multi-layer concept can be regarded as a way to ensure attribute diversity at least at the level of different layers.
2. **Clusters of male and female Alzheimer's disease patients in the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [Gamberger et al., 2016b]**

- This paper presents homogeneous clusters of patients, identified in the Alzheimer's Disease Neuroimaging Initiative (ADNI) data population of 317 females and 342 males, described by a total of 243 biological and clinical descriptors.
- **Multi-layer clustering** was performed. The basic lesson learned from redescription mining and multi-view clustering is that the reliability of clustering can be significantly improved by a requirement that the result should be confirmed in two or more attribute layers.
- Properties of the constructed clusters clearly demonstrate the differences between female and male Alzheimer's disease patient groups. The major difference is the existence of two male subpopulations with unexpected values of intracerebral and whole brain volumes.

3. **Identification of clusters of rapid and slow decliners among subjects at risk for Alzheimer's disease [Gamberger et al., 2017]**

   - 5-year longitudinal outcomes and biomarker data from 562 subjects with mild cognitive impairment (MCI) from two national studies (ADNI) was analysed using a novel multi-layer clustering algorithm.
   - The algorithm identified homogeneous clusters of MCI subjects with markedly different prognostic cognitive trajectories.
   - Characterizing subgroups of at risk subjects, with diverse prognostic outcomes, may provide novel mechanistic insights and facilitate clinical trials of drugs to delay the onset of AD.
   - Multi-layer clustering enables the size and the number of clusters to be determined automatically. The algorithm consists of two steps; in the first step an example similarity table (EST) is computed for each data layer and in the second step these tables are used by an agglomerative bottom-up procedure to find an optimal clustering solution.
   - Multi-layer clustering results in improved quality over single layer clustering methods, does not require statistical independence of input data layers, and requires no explicit definition of the distance measure among instances (patients) or the number or size of the resulting clusters.

Literature Gaps

- One problem with the above mentioned studies is that they subtyped AD patients based on very different features varying from neural, cognitive, and clinical variables. Accordingly, it

is thus unclear what the subtypes of AD patients are, given the different features reported in every study.

- None of the existing studies on clustering analysis have used a dataset that includes early-stage vs. late-stage AD patients. Several experimental studies have shown that these two groups differ profoundly in terms of clinical, cognitive, and neural damage [Kauer-Sant'Anna et al., 2008]. Like MCI conversion to AD, clustering analysis can point to several features that underlie the conversion from early-stage AD to advanced AD.

- There are no studies on using **semi-clustering algorithms** in AD. Semi-clustering enhances clustering by using additional information as constraints in the clustering process. Such additional information is often existent in the dataset or provided by neurologists/clinicians to guide the clustering process.

- In **Fuzzy clustering**, the classification function causes the class members to become a relative one and an object can belong to several classes at the same time but with different degrees. Fuzzy clustering has many applications to health sciences, as some individuals may or may not be diagnosed with a certain disorder, depending on different conditions Fuzzy clustering can help us understand the nature of MCI, as some of these individuals may convert to AD, but others may stay healthy.

Importance of Cluster Analysis

- In clinical populations, **substantial heterogeneity** exists in patient characteristics, illness severity and treatment responses. The heterogeneity of Alzheimer's disease contributes to the high failure rate of prior clinical trials.

- Better understanding of such heterogeneity may lead to more **effective and efficient treatment** by personalising care to better suit patient profiles

- Characterizing subgroups of at risk subjects, with diverse prognostic outcomes, may provide novel mechanistic insights and **facilitate clinical trials of drugs** to delay the onset of cognitive disorders

- Segmentation of the AD population may enable comparative evaluation of subpopulations of AD patients, potentially leading to a better understanding of their distinguishing properties.

- Identification of biomarkers for the Alzheimer's disease (AD) is a challenge and a very difficult task both for medical research and data analysis. Clustering helps in identification of biomarkers.

Chapter 3

My Plans

3.1 Plans

Boolean Expression Generator: Build a Boolean Expression Generator for Cognitive Disorder Progression using Longitudinal and Cross-Sectional Data. The Boolean Expression Generator can generate a Boolean expression combining different biomarkers as variables for different cognitive disorders. This will help clinicians in diagnosis.

Logical Explanation of ML/AI based Diagnosis: Work on Logical Explainability of ML/AI Frameworks for Diagnosis of Dementia. Explaining the workings of ML/AL Framework will help in building better Novel Frameworks for diagnosis.

Abductive Learning Framework for Diagnosis: Develop an Abductive Learning Framework for Diagnosis of Dementia using Longitudinal and Cross-Sectional Data.
In abductive learning, a machine learning model is responsible for interpreting sub-symbolic data into primitive logical facts, and a logical model can reason about the interpreted facts based on some first-order logical background knowledge to obtain the final output.

Utilization of Generated Data from New Computer Based Games: Utilize Data Generated from New Computer Based Games which test the Cognitive Abilities of the player

Form a hierarchy of subtypes with the Dementia population: A major challenge for prediction of cognitive disorder and individualized care is the phenotypic heterogeneity that exists within the AD population. Forming a hierarchical characterization of disease subtypes will help clinicians.

Bibliography

Albright, J. (2019). Forecasting the progression of alzheimer's disease using neural networks and a novel preprocessing algorithm. *Alzheimer's Dementia: Translational Research Clinical Interventions*, 5:483–491.

Archetti, D., Ingala, S., Venkatraghavan, V., Wottschel, V., Young, A., Bellio, M., Bron, E., Klein, S., Barkhof, F., Alexander, D., Oxtoby, N., Frisoni, G., and Redolfi, A. (2019). Multi-study validation of data-driven disease progression models to characterize evolution of biomarkers in alzheimer's disease. *NeuroImage: Clinical*, 24:101954.

Bhagwat, N., Viviano, J. D., Voineskos, A. N., Chakravarty, M. M., and Initiative, A. D. N. (2018). Modeling and prediction of clinical symptom trajectories in alzheimer's disease using longitudinal data. *PLOS Computational Biology*, 14(9):1–25.

Fisher, Charles K., S. A. M. W. J. R. S. A. J. E. C. J. C. R. H. D. R. D. H. D. G. D. W. F. V. H. M.-J. B. K. M. K. S. L. M. A. M. W. M. F. N. A. P. C. P. E. P. R. S. T. K. W. D. C. A. M. A. A. f. A. R. A. A. A. F. o. A. A. P. L. B.-M. S. C. C. P. I. C. F. I. E. L. and Company, F. Hoffmann-La Roche Ltd, F. R. I. G. I. G. J. &. J. N. H. C. N. P. C. P. A. N. P. D. F. P. I. s.-a. C. O. C. D. I. S. C. C. E. M. I. (2019). Machine learning for comprehensive forecasting of alzheimer's disease progression. *Scientific Reports*, 9(1):13622.

Gamberger, D., Lavrač, N., Srivatsa, S., Tanzi, R. E., and Doraiswamy, P. M. (2017). Identification of clusters of rapid and slow decliners among subjects at risk for alzheimer's disease. *Scientific Reports*, 7(1):6763.

Gamberger, D., Ženko, B., Mitelpunkt, A., Lavrač, N., and Initiative, T. A. D. N. (2016a). Homogeneous clusters of alzheimer's disease patient population. *BioMedical Engineering OnLine*, 15(1):78.

Gamberger, D., Ženko, B., Mitelpunkt, A., Shachar, N., and Lavrač, N. (2016b). Clusters of male and female alzheimer's disease patients in the alzheimer's disease neuroimaging initiative (adni) database. *Brain Informatics*, 3(3):169–179.

Kauer-Sant'Anna, M., Kapczinski, F., Andreazza, A. C., Bond, D. J., Lam, R. W., Young, L. T., and Yatham, L. N. (2008). Brain-derived neurotrophic factor and inflammatory markers in patients with early- vs. late-stage bipolar disorder. *The International Journal of Neuropsychopharmacology*, 12(04):447.

Ostertag, C., Beurton-Aimar, M., Visani, M., Urruty, T., and Bertet, K. (2020). Predicting brain degeneration with a multimodal siamese neural network.

Satone, V., Kaur, R., Faghri, F., Nalls, M. A., Singleton, A. B., and Campbell, R. H. (2018). Learning the progression and clinical subtypes of alzheimer's disease from longitudinal clinical data.