B.A.Sc. Thesis

Division of Engineering Science
UNIVERSITY OF TORONTO

this page is intentionally left empty

Identifying Longitudinal Trajectories of Dementia Patients: A Population-based Study

by

Yiqing (Louis) Luo

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Bachelor of Applied Engineering and Science

in the
Division of Engineering Science
Faculty of Applied Science & Engineering

Supervised by Dr. Frank Rudzicz

April 20th, 2020
UNIVERSITY OF TORONTO

Abstract

Population-based studies on dementia and Alzheimer’s Disease are few in numbers due to a lack of longitudinal data collected on the same individual over many years. Recently, the Canadian Institution of Health Information released longitudinal data on clinical and functional characteristics of residents in Canadian hospitals and residential care facilities collected since 2003 and this provides an excellent opportunity for analysis. First, latent states were determined via latent transition analysis, and several characteristics, including cognitive performance and functional dependence, were identified to associate with both dementia and Alzheimer’s. In addition, patient-specific representations of trajectories were constructed via variational recurrent auto-encoders. This thesis has shown that these vectors can be used to classify trajectories that will lead to dementia and Alzheimer’s. Subgroups of patient trajectories for developing dementia were identified and possible interpretations were discussed.

Keywords: population-based, dementia, Alzheimer’s Disease, latent transition analysis, variational recurrent autoencoder, Continuing Care Reporting System

Acknowledgements

First, I would like to thank my supervisor, Dr. Frank Rudzicz, for his expert guidance and unwavering support throughout the thesis. He has been a role model not just academically, but in life in general.

My gratitude extends to Dr. Geoffrey Anderson, who has provided many insights into the subject field and empowered me with opportunities to succeed. My appreciation also goes to the entire Indoc Research team, for supporting me with their technical infrastructure.

In addition, I am indebted to my family for all the sacrifices they made for me over the years, and their continued belief in me.

At last, I am grateful for my grandmother, who has been an inspiration for me throughout this thesis, as, despite experiencing dementia related memory loss, her love for the family has not withered.

Table of Contents

List of Tables

List of Figures

this page is intentionally left empty

Chapter 1.

Introduction

Dementia is a debilitating disease that places a huge burden on Ontarian, Canadian, and international healthcare systems. Affecting over 45 million people worldwide in 2015, this disease is considered by both the WHO and the G8 nations to be a public health priority, with the global population with this disease forecasted to be over 130 million by 2050 [1]. With no immediate cure in sight, various studies have shown that early detection of the disease allows for patients to not only have access to treatments to slow down the progression of dementia, but also plan and prepare for their future [2] [3]. In recent years, many investigators have conducted longitudinal studies analyzing various risk factors to the disease using population-based studies, sampling from sources such as Twitter, and national databases [4] [5]. However, while many of these studies identify important factors related to the disease, most do not account for the patient-specific individuality. Hence, this thesis aims to fulfill this gap by describing hidden clusters of individuals with different trajectories pertaining to their physical and cognitive functions using latent variables. Specifically, this thesis will address the following objectives:

- Derive descriptive insights from some statistical model that identifies underlying subgroups based on trajectories of developing dementia and Alzheimer’s Disease, a type of dementia.
- Predict the future possibility of Alzheimer’s Disease and other forms of dementia for patients prior to clinical diagnosis using questionnaire data.

Chapter 2.

Background

2.1. Dementia and Alzheimer's Disease

Dementia is an umbrella term covering diseases and conditions that pertains to the loss of cognitive functioning and behavioral abilities to such an extent that the person's daily life and activities are disturbed [6]. Alzheimer's Disease typically refers to an abnormal brain cell death rate caused by plaques and neurofibrillary in the brain, and is the most common cause of dementia, accounting for 60-80% of cases [7]. Other medical conditions that classify under dementia include, Vascular dementia, Lewy body dementia, and frontotemporal disorders [6]. It is also possible to have more than one type of dementia [7]. Signs and symptoms of dementia occur when healthy neurons lose connection with other brain cells and die at an abnormal rate, and usually progress slowly as one ages [7].

2.2. Canadian Institution of Health Information

The Canadian Institution of Heath Information (CIHI), in coordination with InterRAI (an international research network), recently released its data on clinical assessments from the Canadian Care Reporting System (CCRS) implemented across Ontario around 2003 [8]. The CCRS contains demographic, administrative, clinical, and resource utilization information on individuals who received continuing care services in hospitals or long-term homes in Canada. These data, or often called the Minimum Data Set (MDS), are completed by nurses every 3 months. Furthermore, these nurses must complete a training program with a minimum passing score of 90% within the program to be eligible to assess patient situations, thereby ensuring a level of data consistency [9]. The data construct takes the form of structured questionnaires with multiple-choice assessments and each individual is followed up with subsequent assessment with the same questionnaire questions. This provides a sequential record of psychological and cognitive markers for each patient, which is perfect for applying longitudinal analysis.

2.3. Literature Review

Longitudinal analysis is a well-studied field in the health science space. In contrast with cross-sectional studies that examine subject information and compare population groups at a single point in time, longitudinal studies focus on the sequential observation of the same subjects over a period of time. Longitudinal analysis detects changes or developments in the characteristics of a population at a group level, and sometimes at an individual level, whereas cross-sectional analysis focuses on a comparison of groups at a single point in time [10].

Within the domain of dementia, several longitudinal approaches have previously been used to study dementia-related symptoms. McAdams-DeMarco et al [11] modeled the relationship between cognitive decline and mortality with the Cox proportional hazards model, and statistically confirmed the existence of a high-declining subgroup of patients. The Cox proportional hazards model models time as a survival (hazard) function, which accounts for any individuals who died during the period of assessment. In fact, for these studies, survival models are excellent as either the individual dies or contracts the disease [11].

Mixed effect models have also shown promise in classifying trajectories by introducing more than one source of randomness to a generalized linear model. Specifically, by prescribing subject-level random effects, mixed effect models can have random slopes and intercepts, so that any subject-level variance is inherent in the model. In other words, mixed effect models allow inferences to be made on the individual variability of each subject. Wattmo et al [12] found that one subgroup with different Alzheimer's progression can be differentiated by the presence of ChEI agents and other socio-economic status. Samtani et al [13] modeled progression of patients using structured models supplemented by random effects. Both Wilson et al [14] and Yu et al

[15] also used mixed effects regressors to characterize the path of change in cognitive functions but introduced a change point with respect to time in order to indicate a shift in trajectory. Burnham et al [16] incorporated the Cox proportion hazard model with mixed-effect model with pre-defined groups. However, despite various applications, mixed effect models do not differentiate subgroups unless explicitly parameterized to. These models have constant covariate effects across the population, which is inconvenient in identifying unknown subgroups with differing trajectories.

Over the past decade, latent variable models gained significant traction in dementia analysis. These models especially gained popularity dealing with discrete data by being unbiased in introducing underlying cofactors and subgroups within a population. Proust-Lima et al extended generalized mixed models by introducing latent classes to the linear mixture model [17]. Specifically, the latent statuses dictate class-specific fix effects on top of the common fixed effects over all classes and the random component is also conditioned on the individual classes. Carriere et al [18] applied this latent class mixture model to model trajectories of various factors and found that in addition to cognitive functions, men, recent stressful events, ischemic pathologies, are a few among others that increases symptomatology over time. Proust-Lima et al [19] also introduced a joint model that combines survival functions with the latent class mixture models. This joint latent class model assumes that each homogeneous latent subgroup shares the same marker trajectory and the same risk of some event, the latter in most cases being death. This method applied to dementia data showed that certain markers, such as education level, and ApoE4 carrier status (a gene usually targeted for Dementia treatment) significantly affects the patient degradation process towards dementia. However, no explicit subgroups to the populations were defined.

Latent transition analysis (LTA), also known as Latent Markov models, are Hidden Markov Model approaches to the longitudinal problem. While latent class mixture models cluster homogeneous subjects based on their entire trajectory, Latent Markov models allow subjects to move between classes across timesteps [20] [21]. The timesteps are usually discrete and can be simplified to be a latent class model when the transition matrix is the identity matrix. LTA has previously been applied to forecasting household debt [22] and identifying patterns of cyberbullying [23]. In both of these analyses, individual time-invariant subgroups were identified, which can be very useful in dementia analysis as a dynamic approach to grouping individuals.

Recently, deep models have proven to be state-of-the-art in a variety of subjects, from computer vision to language translation [24] [25] [26]. In fact, latent variable models have a natural extension to deep neural networks and can be derived from the expectation-maximization algorithm [27]. In computer vision, variational autoencoders often encode the multi-dimension image into a latent code, which is subsequently decoded to reconstruct the image. Often, the latent code is a vector that condenses input information into lower dimensions and can be viewed as a continuous variant to the discrete subgrouping previously discussed in mixture models and LTA. For sequential records, variational recurrent autoencoders have been developed to cluster different trajectories together [28]. In 2019, Antelmni et al [29] applied variational autoencoders to longitudinal imaging and clinical data to understand relationships between data heterogeneity. They summarized the trajectories into normal aging and dementia aging. However, the number of clinical covariates is much smaller compared to CIHI data. In addition, latent representation learning has been applied to the health records space with promising results [30].

2.4. Main Technical Models

2.4.1. Latent Transition Analysis

Latent transition analysis (LTA) allows subjects to transition between different latent statuses across timesteps. More formally, LTA is a Markov model that estimates transitions between each timepoint given latent statuses. consider a vector of response variables for individual i’s responses $\mathbf{Y}_i = \left(\mathrm{Y}_{I,11} \ldots \mathrm{Y}_{I,1M} \ldots \mathrm{Y}_{I,T1} \ldots \mathrm{Y}_{I,TM}\right)$, where each response can take on values 0, 1, …, r_m, and there are M categorical response items measured at T different timepoints. Let s_{ti} = 1, 2, … n_s, be the individual I’s latent status membership at time t. Let I(y=k) be an indicator function which equals 1 if y equals to k, else 0. Let G_i denote individual i’s group memberships determined a priori. Note for multi-group LCA, the parameters estimated will be conditioned on each group, allowing for tests for group invariance. Let X_i be the variates for i^{th} individual. Then the latent transition model can be expressed as:

$$P(\mathbf{Y}_i = \mathbf{y} \mid X_i = x, G_i = g)$$
$$= \sum_{s_1=1}^{n_s} \cdots \sum_{s_T=1}^{n_s} \delta_{s_1|g}(x) \tau_{s_2|s_1,g}(x) \cdots \tau_{s_T|s_{T-1,g}}(x) \prod_{m=1}^{M} \prod_{k=1}^{r_m} \prod_{t=1}^{T} \rho_{mk|s_t,g}^{I(y_m=k)}$$

where

1. ρ is the item-response probabilities conditional on latent status and time

2. $\delta_{(s|g)(\boldsymbol{x})}$ is defined as the probability of being in latent membership s given some grouping g a priori. Furthermore, it is determined by a logistic regressor with weightings for each covariate vector **x**:

$$\delta_{(s|g)(\boldsymbol{x})} = P(\mathrm{S}_{1\mathrm{i}} = s|X_i = \boldsymbol{x}, G_i = g\ ) = \frac{\exp\left(\beta_{(0,\,s|g)} + \boldsymbol{x'\beta}_{(\mathbf{1},\,\boldsymbol{s}|\boldsymbol{g})}\right)}{1\ +\ \sum_{j=1}^{n_s-1} \exp\left(\beta_{(0,\,s|j)} + \boldsymbol{x'\beta}_{(\mathbf{1},\,\boldsymbol{s}|\boldsymbol{j})}\right)}$$

3. $\tau_{(s2|s1,\,g)(\boldsymbol{x})}$ = the transition probability of an individual i moving from the current latent category, at time 1 for instance, and the next latent category, at time 2 for instance, given some grouping g a priori. Furthermore, it is similarly determined by a logistic regressor with weights conditioned upon the previous latent state:

$$\tau_{(s2|s1,\,g)(\boldsymbol{x})} = P(\mathrm{S}_{2\mathrm{i}} = s_2|\mathrm{S}_{1\mathrm{i}} = s_1, X_i = \boldsymbol{x}, G_i = g\ )$$
$$= \frac{\exp\left(\beta_{(0,\,s2|s1,\,g)} + \boldsymbol{x\beta}_{(\mathbf{1},\,\boldsymbol{s2}|\boldsymbol{s1},\,\boldsymbol{g})}\right)}{1\ + \sum_{j=1}^{(n_s-1)} (\beta_{(0,\,s2|s1,\,j)} + \boldsymbol{x\beta}_{(\mathbf{1},\,\boldsymbol{s2}|\boldsymbol{s1},\,\boldsymbol{j})})}$$

Overall, five sets of parameters are estimated: $\delta, \tau, \rho, \beta_{(s|g)}, and\ \beta_{(s_2|s_1,\,g)}$. Specifically, δ, τ will depend on the estimated $\rho, \beta_{(s|g)}, and\ \beta_{(s_2|s_1,\,g)}$. Note that the notations used in this section are borrowed from [31].

2.4.2. Variational Auto-encoders

Variational autoencoders (VAE) is comprised of a probabilistic encoder model $q(t_i|x_i,\phi)$, which generates a mean μ and Σ of the latent variables. Then, the latent variable vectors z for individual i is constructed via the following equation:

$$z = \mu(X) + \Sigma(X) * \epsilon, \qquad where\ \epsilon \sim Gaussian(0, I)$$

The loss functions that we want to minimize in VAE is as follows:

$$loss = E_{X\sim D}\ [E_{Z\sim Q}\ [log\ \ P(X|z)] - D[Q(z|X)\ ||\ P(z)]$$

where

D(A||B) is the KL divergence between distribution A and B,

X is the input variable matrix,

Z is the latent variable matrix,

Q is the Encoder function (usually implemented as neural networks),

P is the Decoder function (usually implemented as neural networks).

The overall architecture can be seen in more detail in Figure iv [24].

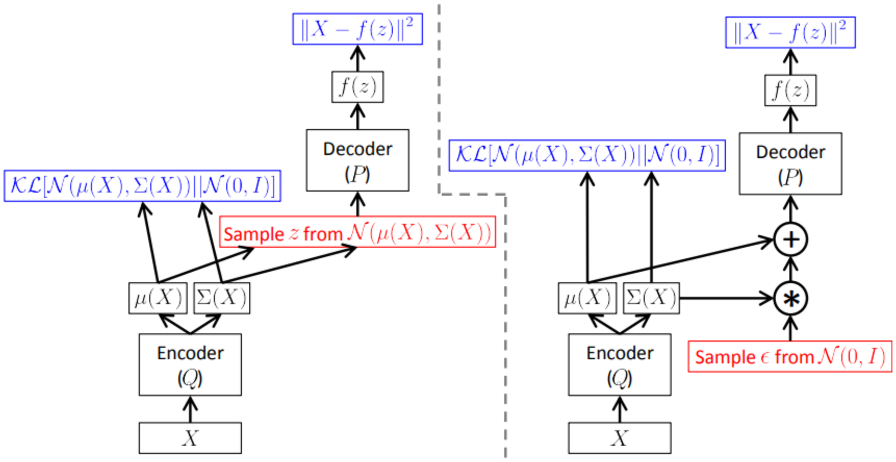
$\|X - f(z)\|^2$
$f(z)$
Decoder (P)
$\mathcal{KL}[\mathcal{N}(\mu(X),\Sigma(X))||\mathcal{N}(0,I)]$
Sample z from $\mathcal{N}(\mu(X),\Sigma(X))$
$\mu(X)$
$\Sigma(X)$
Encoder (Q)
X
$\|X - f(z)\|^2$
$f(z)$
$\mathcal{KL}[\mathcal{N}(\mu(X),\Sigma(X))||\mathcal{N}(0,I)]$
Decoder (P)
$+$
$*$
$\mu(X)$
$\Sigma(X)$
Encoder (Q)
Sample ϵ from $\mathcal{N}(0,I)$
X

Figure i VAE structure. The red box illustrate the sampling process for the latent variable z. The blue boxes refer to loss functions to backpropagate.

2.4.3. Variational Recurrent Auto-encoders

Variational Recurrent Auto-Encoders (VRAE) is a sequence-to-sequence extension to the VAE model. Specifically, in lieu of independent observations, X refers to a sequence of clinical records for an individual. While ordinary VAE usually use a feed-forward neural network as the encoder and decoder function, the VRAE uses a recurrent neural network for the encoder and a feed-forward neural network as the decoder.

The VRAE model can be described using the following set of equations. The encoder can be described as the following:

$$\begin{aligned} h_{t+1} &= \tanh(W_{enc}^T h_t + W_{in}^T x_{t+1} + b_{enc}) \\ \mu_z &= W_\mu^T h_{end} + b_\mu \\ log(\sigma_z) &= W_\sigma^T h_{end} + b_\sigma \end{aligned}$$

where h_0 is initialized as a vector of 0's. Z is sampled in the same approach from the previous section. The decoder can be described as the following:

$$\begin{aligned} h_0 &= \tanh(W_z^T z + b_z) \\ h_{t+1} &= \tanh(W_{dec}^T h_t + W_x^T x_t + b_{dec}) \\ x_t &= \text{sigm}(W_{out}^T h_t + b_{out}) \end{aligned}$$

The above notations in the equations is borrowed from [28].

Chapter 3.

CCRS Data and Preprocessing

3.1.1. Data Overview

The CCRS data is divided into three connected sub-datasets: episode, assessment, and medication. The episode dataset contains administrative information for each client, including gender, location, date of assessments, etc. The assessment dataset contains discrete multi-choice questionnaire form answers. The answer textual expression has already been converted to numerical categorical variables. The medication dataset contains medication names and dosage information. Variable "client_episode_id" connects episode data with assessment data; Variable "assessment_id" connects assessment data with medication data. In addition, each individual is assigned a unique client identifier. This relationship is summarized as a flowchart in Figure ii.

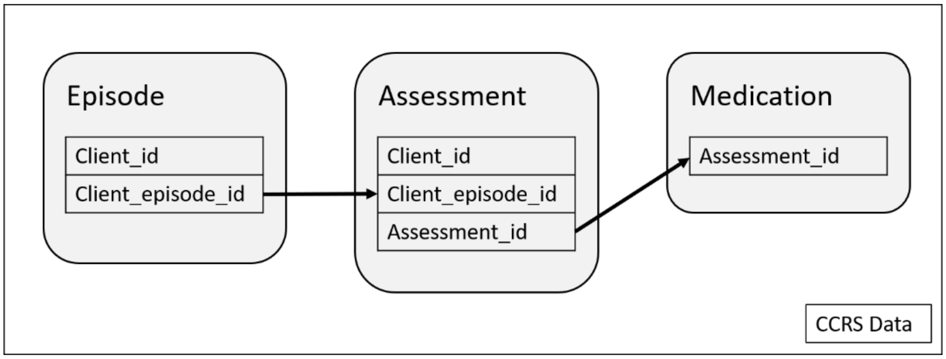

Figure ii Relations between CCRS data. The tables are connected by common variables.

The CCRS questionnaire elements are recorded as encoded characters and numbers. A description of all encoded characters mentioned in this thesis is provided in Appendix A. Please also note that all mentioning of dementia hereinafter excludes

Alzheimer’s Disease, and refers to the questionnaire item “Dementia excluding Alzheimer’s Disease” (J1H).

Both Alzheimer’s Disease (J1G) and dementia (J1H) have three response entries, seen in Table 1.

Response entry	**Definition**
0	Not present
1	Present; Not subject to forces treatment or monitoring by home care professionals
2	Present; Monitored or treated by home care professionals

Table 1 Dependent Variable’s categories and definitions

All data is delivered in Sas7bdat format within a secure RDEN environment from Indoc Research via the remote desktop protocol.

3.1.2. General Preprocessing

Data resulting from joining tables via previously mentioned columns are preprocessed before feeding to different models and analysis. Records with missing targets and columns with missing data were dropped. Several variables were timestamps and they were normalized based on each individual’s birthday. Individuals with only one assessment were dropped, as no longitudinal information can be drawn from them.

Additionally, a cognitive performance scale (CPS) is appended to the dataset using information already present in the data. The CPS describes the cognitive status of an individual. Validated against the Mini-Mental State Examination (MMSE) and the Test for Severe Impairment (TSI), the CPS score ranges from 0 to 6, with a higher number indicating more severe impairment. The CPS score is a standard method of measuring clinical cognitive performance and is a metric that is inherent to the CCRS dataset.

The calculation of CPS can be performed via a predefined decision tree [32] shown in Figure iii.

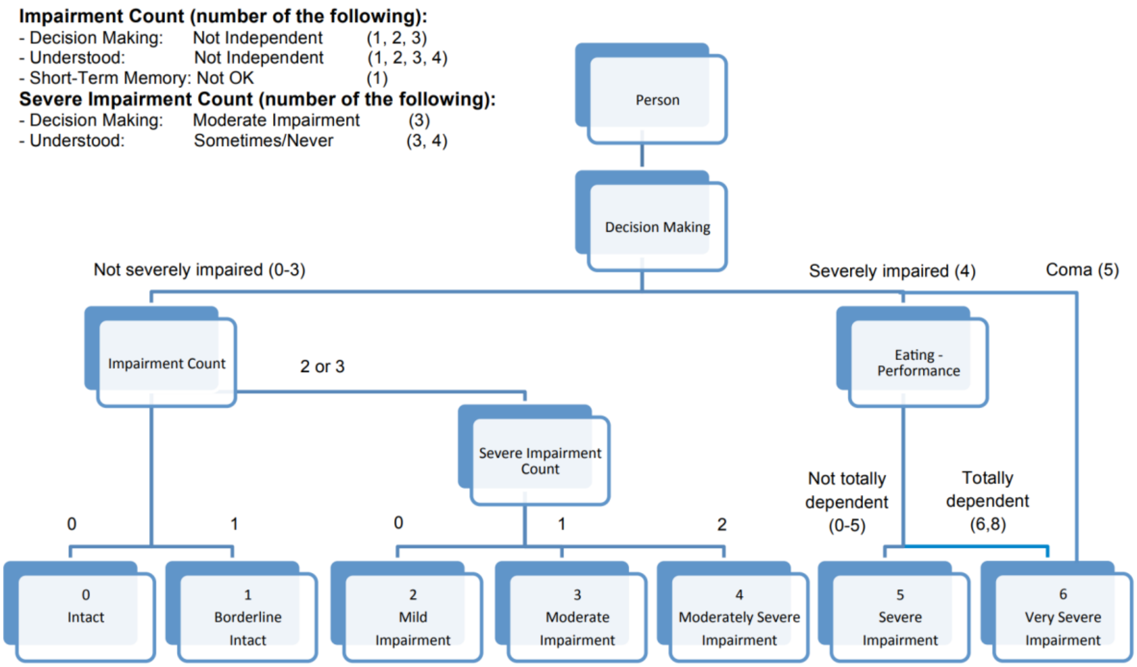
Impairment Count (number of the following):
- Decision Making: Not Independent (1, 2, 3)
- Understood: Not Independent (1, 2, 3, 4)
- Short-Term Memory: Not OK (1)
Severe Impairment Count (number of the following):
- Decision Making: Moderate Impairment (3)
- Understood: Sometimes/Never (3, 4)
Person
Decision Making
Not severely impaired (0-3)
Severely impaired (4)
Coma (5)
Impairment Count
2 or 3
Eating - Performance
Severe Impairment Count
Not totally dependent (0-5)
Totally dependent (6,8)
0
1
0
1
2
0 Intact
1 Borderline Intact
2 Mild Impairment
3 Moderate Impairment
4 Moderately Severe Impairment
5 Severe Impairment
6 Very Severe Impairment

Figure iii CPS decision tree.

All preprocessing is implemented in R and Python 3.

3.1.3. Model-specific Preprocessing

Each model has additional stages of preprocessing but shares the preprocessing steps illustrated in section 3.1.2. The details of the model-specific preprocessing will be described in their respective sections.

Chapter 4.

Latent Transition Analysis

4.1. Introduction

Latent Transition Analysis, or LTA, is a stochastic Markovian Process that allows for latent statuses to be assigned to each individual. In the following section, LTA will be used to provide a descriptive analysis and determine latent statuses on both dementia and Alzheimer's Disease.

4.2. Methods

A population-wide LTA without any prior grouping or feature selection is applied. All ordinal variables with more than 2 levels of responses are collapsed to only 2 levels: presence or no presence. While LTA is capable of handling responses with more than 2 options, Festl et al [23] discussed that collapsing the responses into binary variables has previously shown adequate results. Categorical variables dummy encoded into additional variables for each category. Any administrative variables were dropped. As the dataset contains a varying number of records for each individual at varying time points, only the first and the last records of an individual is considered. In the future, potential smoothing functions could be considered.

The number of latent status to use for LTA was determined by running the latent class model on a single timestep. The number of latent statuses from the best model in terms of performance metrics was selected. The model performance metrics are in Appendix B, and the number of latent statuses is chosen to be 4. The group membership prior is assumed to be uniform.

The code for LTA is available at [33] in R. This implementation is a direct translation from the PROC LTA in MPlus software.

In selecting the covariates used to determine latent statuses at each timestep, two approaches were considered. First, the continuous variable age was considered as the sole covariate. Being the only non-categorical variable in the questionnaire data, age seems to be a logical choice for the covariate. Then, as the computation capacity allows, all features, except the target columns, are treated as membership covariates. All features are used to populate the state space of the LTA. Note that a separate result and discussion will be provided for each covariate set.

There is a total of 384143 clients ids in the data post-preprocessing. After dummying categorical variables, there are a total of 206 features used in training.

4.3. Results

4.3.1. Covariate Set: Age

First, the covariates for determining the latent membership is limited to Age. The item-response probabilities (ρ) for all items with large variance (>0.15) across statuses are shown in Figure iv; the item-response probability is summarized in Table 2; the probability of being in a latent state is summarized in Table 3.

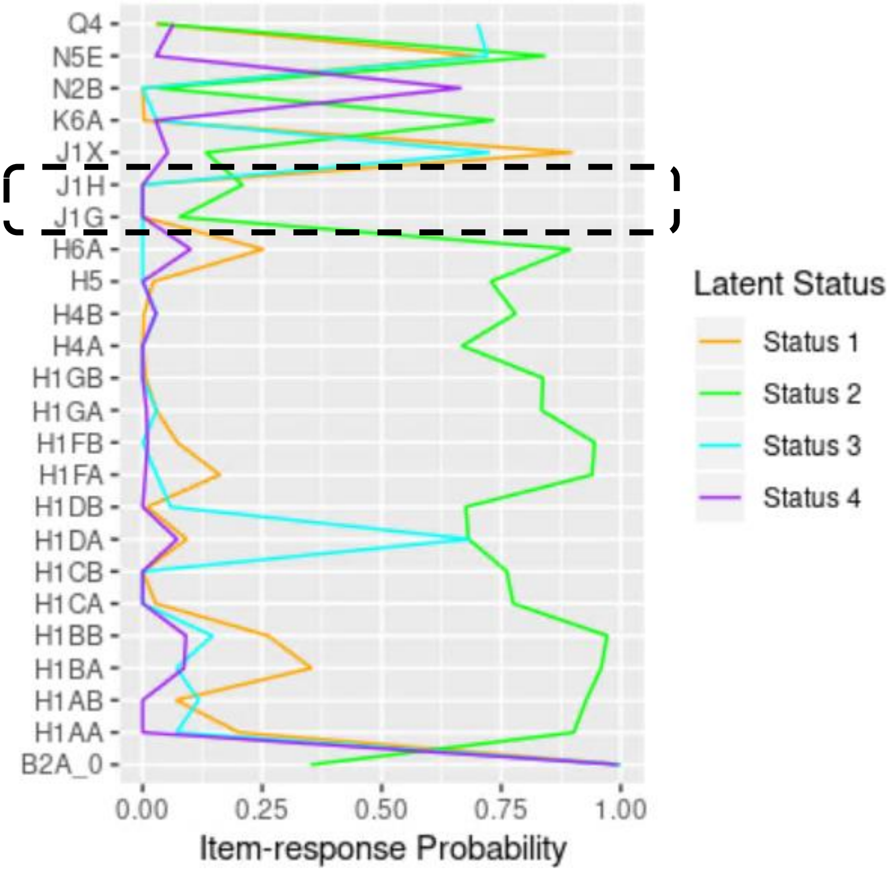

Figure iv The item response probabilities (ρ) of being in each latent status. The descriptions corresponding to each key variable are given in Appendix A. Different colors correspond to different statuses, and interpretations of the statuses are given in the bottom. The two elements boxed, J1G and J1H, refer respectively to Alzheimer's and Dementia's excluding Alzheimer's.

	Status 1	**Status 2**	**Status 3**	**Status 4**
Status 1	0.62	0.38	0.00	0.00
Status 2	0.00	0.99	0.00	0.00
Status 3	0.00	0.35	0.65	0.00
Status 4	0.00	0.26	0.00	0.74

Table 2 Transition matrix (τ) between different statuses across time 1 and time 2. Values rounded to 2 significant figures.

	Status 1	Status 2	Status 3	Status 4
Time 1	0.002	0.997	0.00	0.00
Time 2	0.001	0.997	0.00	0.00

Table 3 Latent Membership Probability (δ) across time 1 and time 2. Values rounded to 2 significant figures

4.3.2. Covariate Set: All Features

In order to incorporate more information in determining latent statuses, all numerical variables are considered as covariates. This is the maximum amount of information available, except for possible interactions between terms. However, as the questionnaire data is categorical with mostly binary and tertiary range, interactions were not considered as part of the analysis.

Again, the item-response probability is shown in Figure v; the transition matrix in Table 4; and the latent membership probability across timesteps in Table 5. Note that the item-response probabilities in Figure iv and Figure v are identical, as covariates only affects how individuals transition between states.

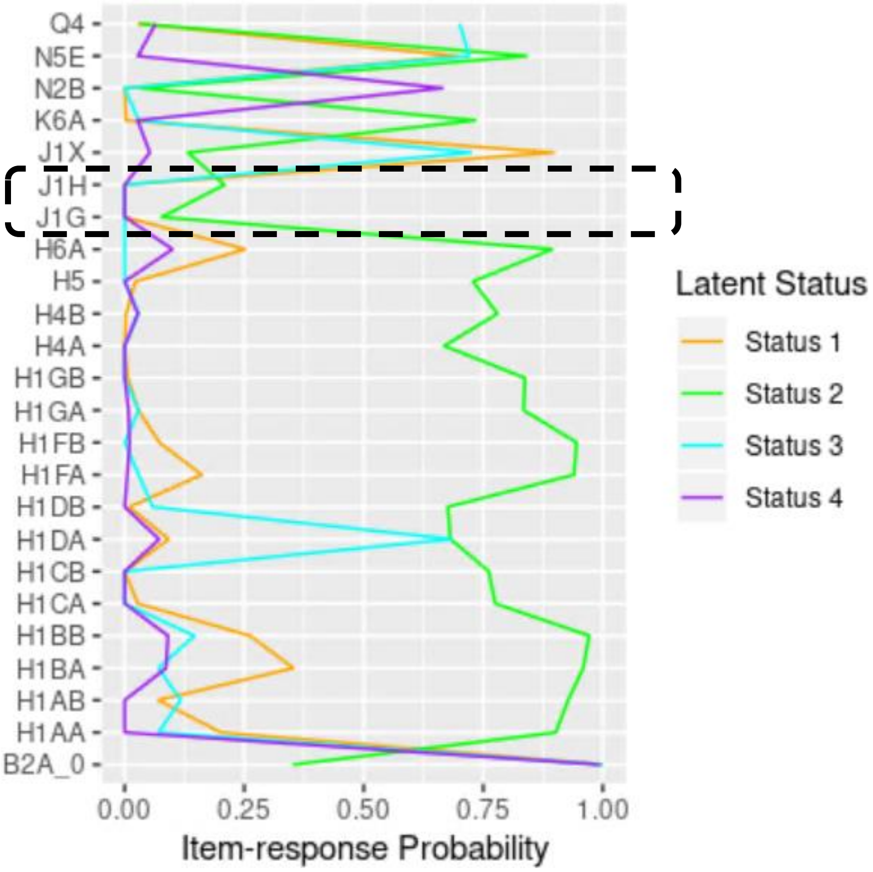

Figure v The item response probabilities (ρ) of being in each latent status. The descriptions corresponding to each key variable are given in Appendix A. Different colors correspond to different statuses, and interpretations of the statuses are given in the bottom. The two elements boxed, J1G and J1H, refer respectively to Alzheimer's and Dementia's excluding Alzheimer's.

	Status 1	**Status 2**	**Status 3**	**Status 4**
Status 1	0.75	0.00	0.00	0.25
Status 2	0.00	0.74	0.00	0.26
Status 3	0.00	0.00	0.63	0.37
Status 4	0.00	0.10	0.00	0.90

Table 4 Transition matrix (τ) between different statuses across time 1 and time 2. Values rounded to 2 significant figures.

	Status 1	**Status 2**	**Status 3**	**Status 4**
Time 1	0.000	0.168	0.001	0.830
Time 2	0.000	0.205	0.001	0.794

Table 5 Latent Membership Probability (δ) across time 1 and time 2. Values rounded to 3 significant figures.

As most records are concentrated in statuses 2 and 4, the item-response probabilities with large variance (>0.2) for those two statuses are provided in Figure vi.

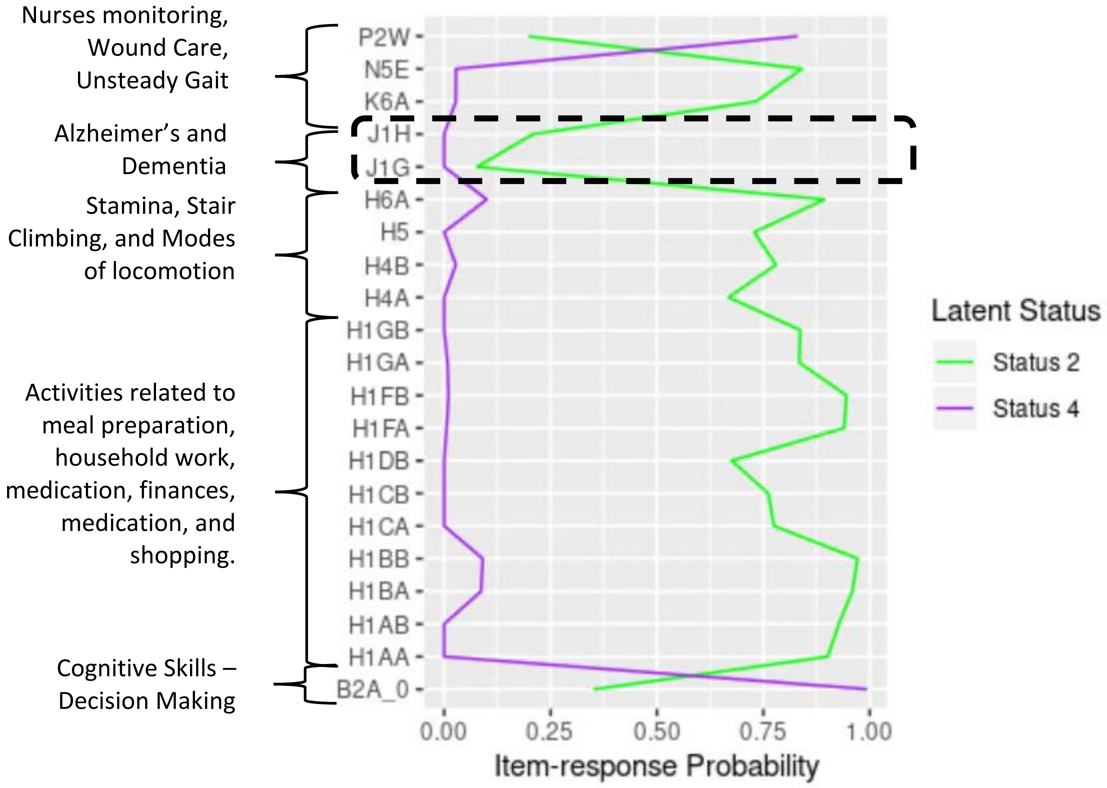

Figure vi The item response probabilities (ρ) of being in latent status 2 and 4 with large variance (>0.2).

4.4. Discussion

4.4.1. Covariate Set: Age

In the converged LTA model, patients with either Alzheimer (J1G) or dementia (J1H) belong only to latent status 2 while the other remaining statuses have a negligible item-response probability of these two indicators being positive. In fact, this status group exhibited a high probability of lack of independence in everyday behaviors (features starting with “H1”). However, it is important to note that in the transition matrix in Table 2, the probability of belonging in Status 2 is almost certain for all clients. Hence, more information in the covariate set is required in determining the sub-groups.

4.4.2. Covariate Set: All Features

As expected with a more diverse covariate set, the model is able to capture more latent patterns in trajectory. As seen in Figure v, the item-response probabilities did not change, and status 2 is still the only latent status that has an appreciable item-response probability in the two entries. The latent membership probability seen in Table 5 suggests that most users have a high probability of belonging in status 2 and 4 and the following discussion will be mostly based from those two statuses.

The results from the item-response probability are indeed consistent with the clinical symptoms of Alzheimer’s and/or Dementia. In Figure v, individuals in status 2 are much more likely than those in status 4 to have Alzheimer’s and/or Dementia. Individuals from status 4 tend to have a higher cognitive ability (B2A) and have the capability to more independently perform various daily tasks (H1**). All of these features are clinical symptoms commonly associated with both diseases. Those in status 4 are more likely to not have a wound that requires care (N5E) and have more overall physical capacity (H4B, H5, and H6A). The former may be a superset of head injuries or be consequences of general loss of cognitive capabilities; the latter may be interpreted as the results of declining functional capabilities. For the feature P2W, i.e. Nurse monitoring daily, individuals in status 4 tend not to have nurses monitoring over them, while those in status 2 do as they require greater assistance in performing ordinary tasks and are more prone to incidents due to the cognitive decline.

To further differentiate the behaviors of those in status 2 and status 4, the variances between item probabilities for status 2 and 4, seen in Figure vi, are computed and they are most significant at two tasks: Ordinary Housework (H1BA and H1BB), and Shopping (H1FA and H1FB). The former requires both physical dexterity and cognitive capability to perform the sequential task of housework tasks; the latter not only requires memorization of items to purchase, but also ability to remember routes to the location of shopping.

In addition, the transition matrix seen in Table 4 suggests that the two statuses mostly transition between themselves. Individuals who start off in status 2, over time, has about 24% probability to transition into status 4. This is significantly more probable than for individuals to start off in status 4 and then end up in status 2, which is about 10%. This means that those who exhibit dependence in behaviors such as household tasks, shopping, etc., and other traits aforementioned, are much more likely to manifest independence and improve in their condition in a later assessment. While seemingly counterintuitive, this can be explained by the probabilistic nature of the LTA model. In fact, being in status 2 does not imply that the item-response probability of J1H and J1G is always positive. As Alzheimer's Disease and Dementia are both progressive diseases, those that transition away from status 2 are likely individuals who do not have the diseases. In other words, the item-response probabilities must be interpreted only as the effect, not the latent cause. Nonetheless, those exhibiting these behavioral cues warrant more attention than those without.

4.5. Future Work

While the current approaches obtained results that coincides with the clinical symptoms of both Alzheimer's Disease and Dementia, other constructions of the model can be considered to attempt to uncover hidden behavioral patterns. One approach is to expand the number of timesteps considered, as the current method only considers the oldest and newest records. Another is to attempt different preprocessing and feature engineering, such as reducing features which induce sparsity and hence the ordinal variables can populate the state space without binarization.

Chapter 5.

Variational Recurrent Autoencoders

5.1. Introduction

Variational Recurrent Autoencoders (VRAE) are autoencoders that incorporates a recurrent neural network layer in both the encoder and decoder of an ordinary autoencoder. This recurrent layer allows for individualized representations of the patient's overall trajectory using sequential data, which in this case, are the patient records. While the LTA model was able to describe important features that correlate with the presence of the diseases, VRAE aims to predict the onset of Alzheimer's Disease and dementia before clinical diagnosis.

5.2. Methods

5.2.1. Data Preprocessing

Like the LTA implementation, no feature selection was performed prior to model fitting. Differing from the LTA implementation, the VRAE model is trained on all data whose client has more than 10 questionnaire assessment records. As the count of records differs across individuals, the sequences were padded at the end with 0's. However, this may very well negatively affect the training schemes, as 0's does have an inherent meaning in the data, and hence lead to false training loss.

In terms of feature engineering, while several approaches were attempted, the best approach is presented: the categorical variables were dummy encoded, and ordinal variables were binarized such that only the most severe forms of an indicator are categorized as 1. With this, the most severe forms of the questionnaire indicators are distinguished from the less severe ones. The target column is dropped from the training features to prevent data leakage.

In addition, a new label, called “eventually-positive”, is created to describe whether each client will eventually be clinically diagnosed with Alzheimer’s Disease. This label is not assessment-specific, but patient-specific, as it summarizes the overall trajectory of the patient with respect to either Alzheimer’s Disease or dementia. This new label is not used in training the autoencoder but used in the visualization and, later, a separate discriminator on the latent space. The calculation of this label is as follows:

$$eventually\text{-}positive \;=\; \begin{cases} 1, if\ client\ is\ positive\ in\ any\ of\ his\ or\ her\ records\ for\ a\ given\ condition \\ 0, otherwise \end{cases}$$

After preprocessing, there are 12527 individuals with 788 and 2831 individuals who were not diagnosed with Alzheimer’s Disease and dementia respectively at the first assessment but were eventually diagnosed with the disease. In total, there are 193 features, including the target column, for every record of assessment.

5.2.2. Classifier on the Latent Representation

The latent vector representation from the VRAE contains rich information that allows for the reconstruction of the sequential records. Various classifiers are applied to these latent vectors from the trained model and their performances can be found in Table 6 (top 3 Alzheimer’s Disease), Table 7 (top 3 Dementia) and Appendix C (All). Note that overall training is a 2-step process and the losses from the second stage classifier are not updated into the autoencoder, which is trained unsupervised. Prior to the training of the classifier described below, the training features are oversampled via Synthetic Minority Over-Sampling Technique (SMOTE), a standard approach to oversampling imbalanced data. The overall flow of the model is summarized in Figure viii.

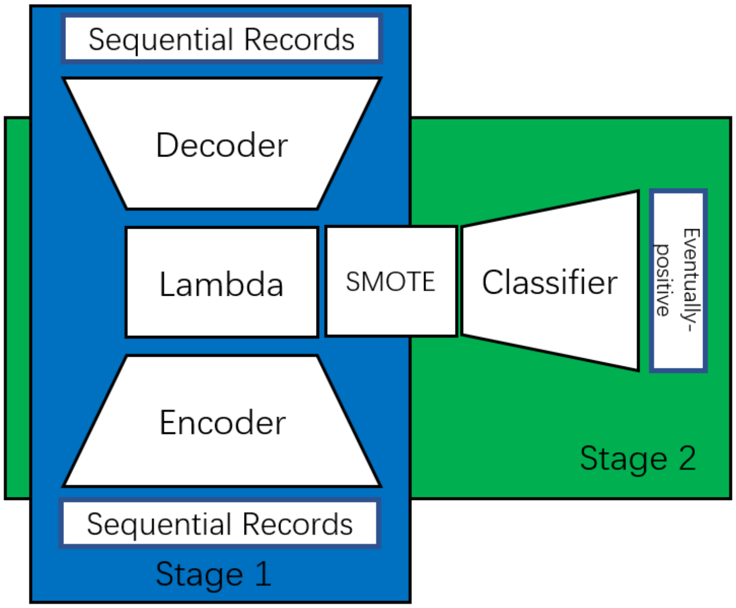

Figure vii VRAE model with a discriminating classifier for eventually-positive labels.

5.2.3. VRAE Training Scheme

The VRAE model is implemented in PyTorch. The RNN vector has a hidden size of 70. The latent variable vector z is of dimension 30. The training scheme was with 0.01 learning rate, 0.02 dropout rate, 50 batch size and 150 epochs. The loss function is optimized with ADAM. The training ran on CPUs. The loss curve for the autoencoder with Alzheimer’s Disease and Dementia as the target variable can be found in Figure viii. Threefold cross-validation was used during training on the training data. Training data comprised of 70% of total data and the remaining 30% were used as testing data.

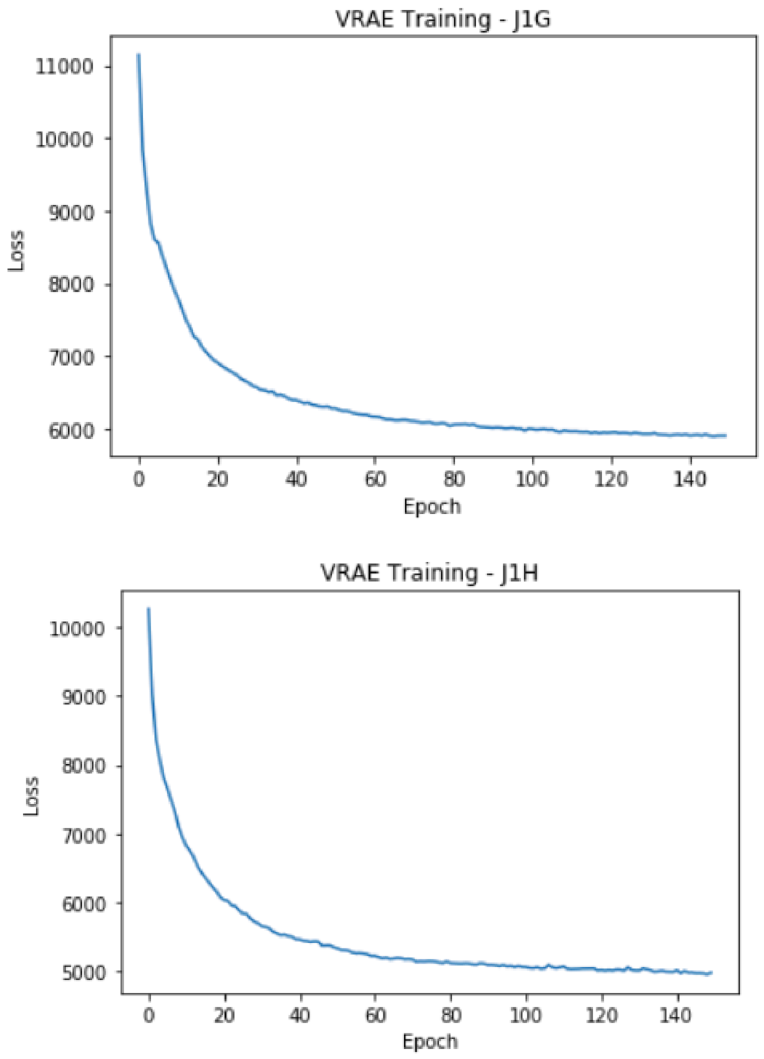

Figure viii VRAE Training Curves with Alzheimer's Disease J1G (top) and dementia J1H (bottom) as target variable

5.3. Results

5.3.1. Target Variable: Alzheimer's Disease

In order to have an intuition of the high dimensional space, T-SNE with perplexity 80, is used to reduce dimensions for the latent vectors from the testing set, which are plotted in Figure ix.

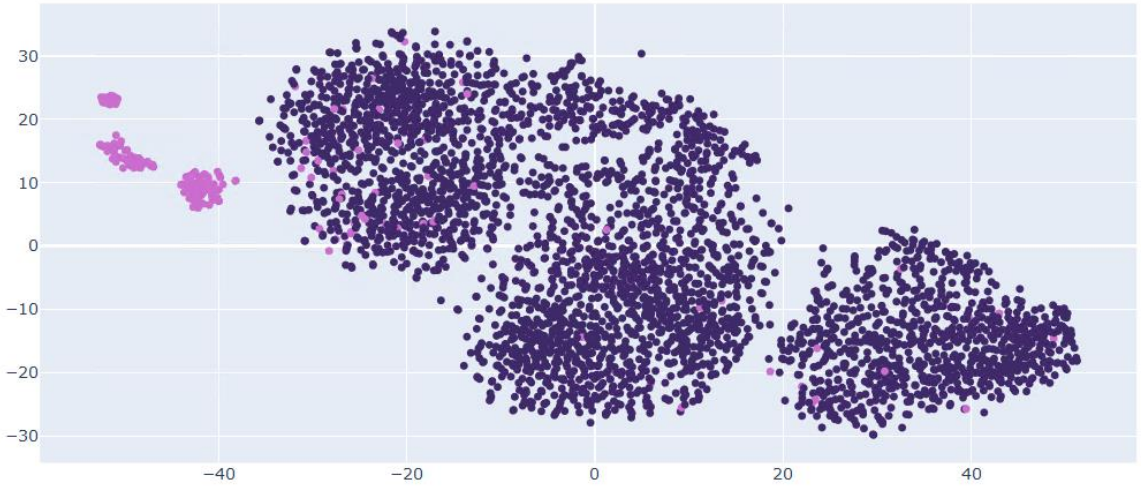

Figure ix T-SNE plot for the high dimensional latent vectors for Alzheimer’s Disease. Each data point represents a patient trajectory. The dark blue dots are patients who are not clinically diagnosed with Alzheimer's Disease in any assessments; the pink ones represent the “eventually-positive” group for Alzheimer’s Disease.

The top three performance of the classifiers on the testing dataset is summarized in Table 6. The performance of the other classifiers can be found in Appendix C. The F1 scores for the classifiers are plotted in Figure x.

Classifier	Label	Precision	Recall	F1-Score	Support
Linear SVM	0	1	0.99	0.99	3536
	1	0.85	1	0.92	214
	macro avg	0.92	0.99	0.96	3750
Quadratic Discriminant Analysis	0	1	0.99	0.99	3526
	1	0.85	0.96	0.9	224
	macro avg	0.92	0.99	0.99	3750
Multilayer Perceptron **	0	1	0.99	0.99	3508
	1	0.86	0.89	0.87	242
	macro avg	0.93	99	0.96	3750

Table 6 Performance Report for top 3 classifiers in terms of macro average F1 scores for VRAE for Alzheimer’s Disease.

** Maximum Depth of tree is 5*
***Training parameters: hidden size = 100, and learning rate = 0.5 with early stopping*

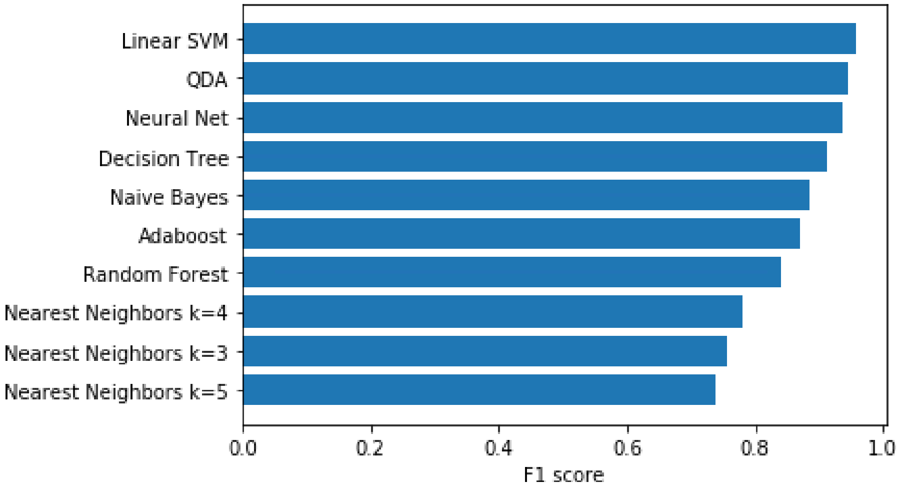

Figure x Macro F1 scores for classifier of eventually-positive Alzheimer's Diseases.

5.3.2. Target Variable: Dementia

The T-SNE plot is again shown in Figure xi, with the target column being Dementia. From empirical observation, two groups were identified. In addition, a 3D scatter plot of the top principle components of same latent space is shown in Figure xii.

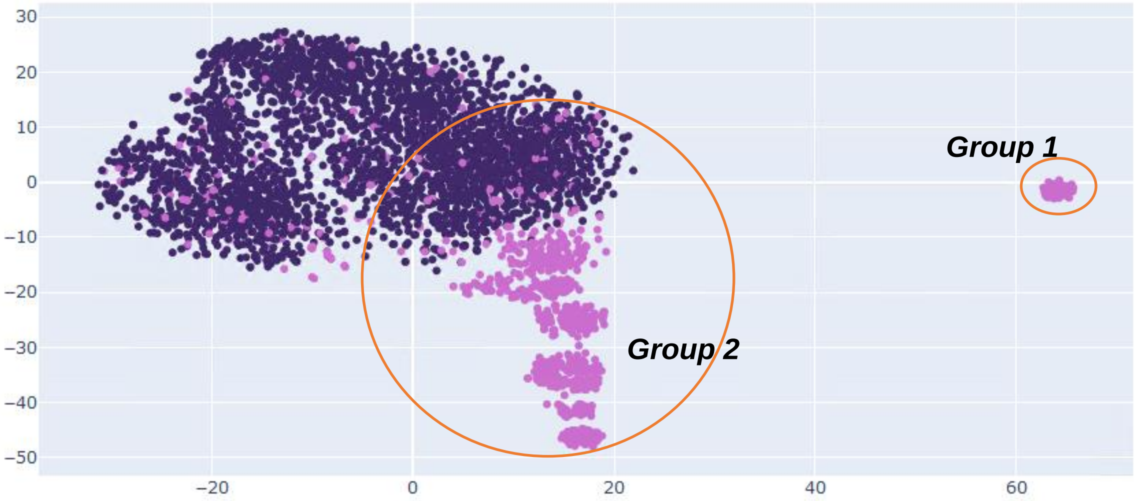

Figure xi T-SNE plot for the high dimensional latent vectors. Each data point represents a patient trajectory. The dark blue dots are patients who are not clinically diagnosed with Alzheimer's Disease in any assessments; the pink ones represent the "eventually-positive" group for Alzheimer's Disease. Two subgroups were empirically identified.

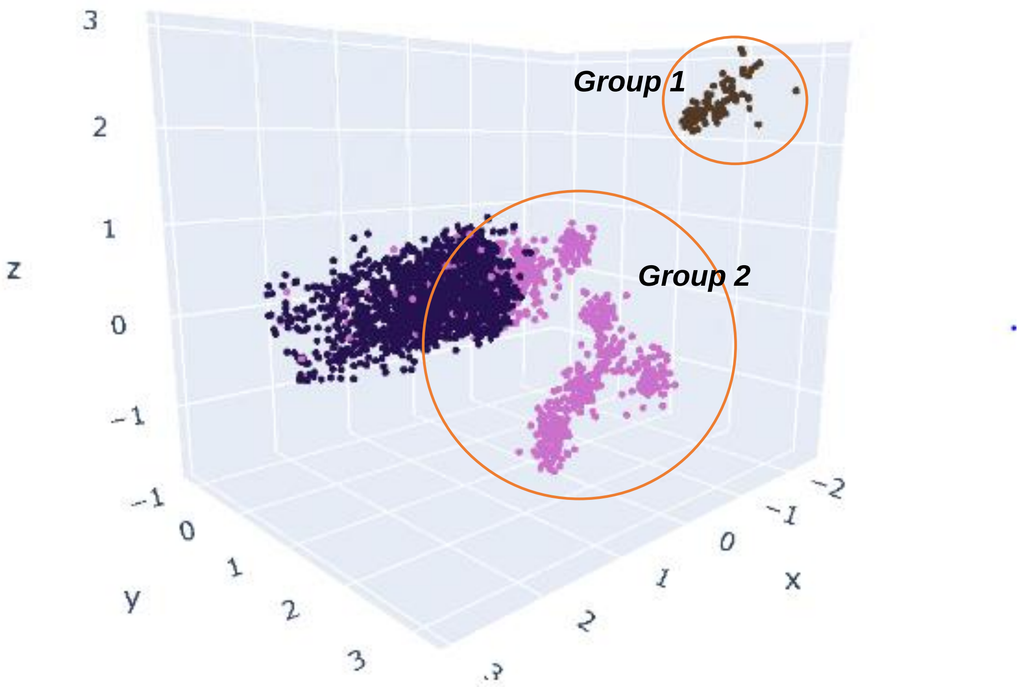

Figure xii 3D scatter plot of the first three principle components for the high dimensional latent vectors. Each data point represents a patient trajectory. The dark blue dots are patients who are not clinically diagnosed with Alzheimer's Disease in any assessments; the brown ones represent group 1of the "eventually-positive" group for Alzheimer's Disease; and the pink ones represent group 2 of the "eventually-positive" group for Alzheimer's Disease.

In order to identify the difference between the two groups, for each feature, the variance between the feature magnitude on average of the original records inputted for each group is calculated and a bar plot of the top features shown in Figure xiii. The top 3 features that differentiate the two groups are K4C, K4B, and P5, which corresponds to "Pain that disrupts usual activities", "Intense pain", "Having Treatment Goals" respectively. In fact, individuals in group 1 do not experience an increase in pain (K4C and K4B) on average, while those in group 2 do not have treatment goals on average. Note that the severity of these characteristics is increasing as one becomes more likely to be clinically diagnosed with the disease.

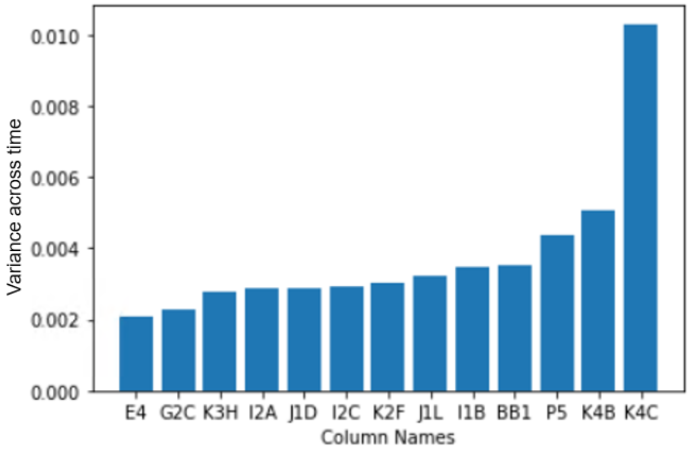

Figure xiii The columns with top variances between the mean input records of the two groups.

The top three performance of the classifiers on the testing dataset is summarized in Table 7. The performance of the other classifiers can again be found in Appendix C. The F1 scores for the classifiers are again plotted in Figure xiv.

Classifier	Label	Precision	Recall	F1-Score	Support
Multilayer Perceptron **	0	1.00	0.95	0.97	2858
	1	0.83	0.99	0.90	742
	macro avg	0.91	0.97	0.94	3600
Decision Tree *	0	1.00	0.95	0.97	2857
	1	0.83	0.98	0.90	743
	macro avg	0.91	0.97	0.93	3600
Adaboost	0	0.98	0.95	0.96	2787
	1	0.85	0.92	0.88	813
	macro avg	0.91	0.94	0.92	3600

Table 7 Performance Report for top 3 classifiers in terms of macro average F1 scores for VRAE for Dementia.

** Maximum Depth of tree is 5*
***Training parameters: hidden size = 100, and learning rate = 0.5 with early stopping*

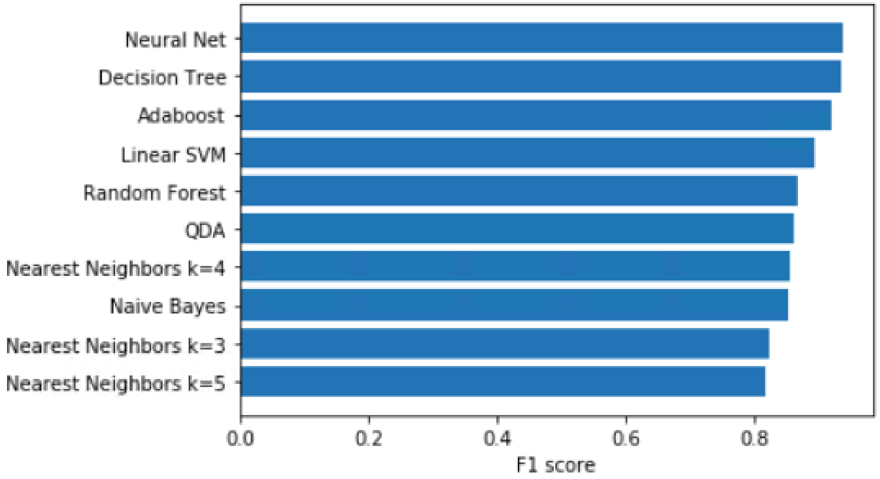

Figure xiv Macro F1 scores for classifier of eventually-positive Alzheimer's Disease

5.4. Discussion

5.4.1. Target Variable: Alzheimer's Disease

A T-SNE visualization of the VRAE model with J1G target variable reveals possible separation between the latent vector of clients who will eventually be diagnosed with Alzheimer's to those without. While there are still overlaps, this assumption is supported by the results of the secondary classifiers. In particular, support vector machine (SVM) with a linear kernel performed the best in F1 score among other classifiers. Linear SVM yields linear decision boundaries and hence this suggests that the eventually Alzheimer's positive individuals can be separated with high probability from the non-Alzheimer's individuals by some hyperplane, even when the training data only contains records from before the onset of Alzheimer's Disease.

The linear SVM performed well with regards to classifying patients who will not be diagnosed with the disease in future. Although the recall for classifying eventually positive individuals is satisfactory, its precision is comparatively lacking. In other words, more individuals will be falsely classified as positive than falsely classified as negative. In fact, the comparatively low precision may be attributed to the lack of eventually positive patients within the training set, as oversampling may have introduced unintended bias.

5.4.2. Target Variable: Dementia

The T-SNE visualization of the VRAE model with J1H target variable again reveals possible separation between the latent vectors of eventually positive clients, and those never diagnosed. Two distinct grouping is observed in the t-SNE plot within the eventually positive population and this hypothesis is validated in the 3-dimensional scatter plot of principle components of the same latent vector. This extraneous group, denoted as group 1 in Figure xi and Figure xii, appears to lack indicators of pain experienced in the questionnaire when compared to the rest of the eventually positive individuals. In addition, individuals belonging to group 1 typically have ongoing treatment goals in place, though the actual treatment is not known from the data.

On the other hand, individuals from group 2 typically experiences pain across many assessments. This is a possible validation of results from a 2017 cohort study by Whitlock et al [34], who described an association between persistent pain, accelerated memory decline, and dementia. It is also worth noting that the distance between the centroid of group 2 to those without onset of dementia is much closer than that of high dimensional representations from group 1, and that on average, group 1 has a much lower response related to pain. As pain is an external factor, it may be that pain may be a causal factor for developing certain types of dementia, though more analysis is needed to affirm this claim.

Nonetheless, this observed separation is again supported by the results of the classifiers. The top classifier for Alzheimer's Disease, linear SVM, again yielded strong results, but the top classifier for the dementia label is the multilayered perceptron, albeit the improvement in F1 score is not significant. In fact, the improvement for F1 score comes from a much higher recall score between the two classifiers, not from the precision score.

5.5. Future Work

While the VRAE autoencoders were able to separate the temporal high-dimensional representation of individuals eventually with or without Alzheimer's Disease and Dementia, the specific trajectories within the eventually positive group should be investigated more thoroughly, as dementia, for instance, can comprise various forms of diseases that may exhibit different symptoms. In addition, the possibility of pain being a significant contributing factor to developing dementia should be investigated further. At last, it would be interesting to see if the latent representations can predict the onset of other features as well.

Chapter 6.

Conclusion

In this thesis, the trajectories of both dementia and Alzheimer's Disease are examined. From the latent transition analysis, four distinct statuses for trajectorial progression were identified, with one of the statuses comprising positive responses from both dementia and Alzheimer's Disease. In particular, the level of independence for ordinary tasks, physical prowess, and current medical attention were determined as probable characteristics of those with either disease. In addition, using a variational recurrent autoencoders, latent space representations of individual-level trajectories were constructed, and by using a discriminator, future occurrences for both conditions were predicted using data prior to onset. Two subgroups for individuals who eventually will develop dementia were determined and the changes in the level of pain, and the presence of current treatment goals were identified as features that differentiate the two subgroups. Several next steps pertaining to each model were introduced in this thesis. Overall, this thesis hopes to be a validator for existing smaller studies and a foundation for detecting dementia prior to clinical diagnosis.

Chapter 7.

References

[1] "Dementia cases set to triple by 2050 but still largely ignored," WHO, 11 APRIL 2012 . [Online]. Available: https://www.who.int/mediacentre/news/releases/2012/dementia_20120411/en/. [Accessed 21 1 2020].

[2] "Benefits of early dementia diagnosis," NHS, 9 July 2017. [Online]. Available: https://www.nhs.uk/conditions/dementia/early-diagnosis-benefits/. [Accessed 21 1 2020].

[3] "World Alzheimer Report 2011 The benefits of early diagnosis and intervention," in *Alzheimer's Disease international*, 2011.

[4] J. M. Robillard,, . T. W. Johnson, C. Hennessey, B. L. Beattie and . J. Illes , "Aging 2.0: Health Information about Dementia on Twitter," *PLoS One,* 2013.

[5] K. M. Park, J. M. Sung, W. J. Kim, S. K. An, K. Namkoong, E. Lee and H.-J. Chang, "Population-based dementia prediction model using Korean public health examination data: A cohort study," *PLOS,* 2019.

[6] "What Is Dementia?," Alzheimer's Association, [Online]. Available: https://www.alz.org/alzheimers-dementia/what-is-dementia. [Accessed 18 4 2020].

[7] "What Causes Alzheimer's Disease?," 24 12 2019. [Online]. Available: https://www.nia.nih.gov/health/what-causes-alzheimers-disease.

[8] "CIHI - Continuing Care," CIHI, [Online]. Available: https://www.cihi.ca/en/continuing-care.

[9] "Continuing Care Reporting," CIHI, 1999. [Online]. Available: https://www.cihi.ca/sites/default/files/document/ccrs-rai-mds-overview-infosheet-en.pdf.

[10] "Cross-sectional vs. longitudinal studies," Institute for Work & Health, Toronto, August 2015. [Online]. Available: https://www.iwh.on.ca/what-researchers-mean-by/cross-sectional-vs-longitudinal-studies. [Accessed 19 01 2020].

[11] M. McAdams-DeMarco, M. Daubresse, S. Bae, A. Gross, M. Carlson and D. Segev, "Dementia, Alzheimer’s Disease, and Mortality after," *American Society of Nephrology,* 2018.

[12] C. Wattmo and Å. Wallin, "Early- versus late-onset Alzheimer's disease in clinical practice: cognitive and global outcomes over 3 years.," *Alzheimer's Research & Therapy,* 2017 .

[13] M. N. Samtani, M. Farnum, V. Lobanov, E. Yang, N. Raghavan, A. DiBernardo and V. Narayan, "An Improved Model for Disease Progression in Patients From the Alzheimer’s Disease Neuroimaging Initiative," *Journal of Clinical Pharmacology,* 2011.

[14] R. S. Wilson, S. E. Leurgans, . P. A. Boyle and . D. A. Bennett, "Cognitive Decline in Prodromal Alzheimer Disease and Mild Cognitive Impairment," *Arch Neurol,* 2011.

[15] L. Yu, . P. Boyle, R. S. Wilson, E. Segawa, S. Leurgans, P. L. De Jager and D. A. Bennett, "A random change point model for cognitive decline in Alzheimer’s disease and mild cognitive impairment," *Neuroepidemiology,* 2012.

[16] S. Burnham , P. Bourgeat, . V. Doré, G. Savage, B. Brown and S. Laws , "Clinical and cognitive trajectories in cognitively healthy elderly individuals with suspected non-Alzheimer's disease pathophysiology (SNAP) or Alzheimer's disease pathology: a longitudinal study.," *The Lancet Neurology,* 2016.

[17] C. Proust-Lima, V. Philipps and B. Liquet, "Estimation of Extended Mixed Models Using Latent," 2016.

[18] I. Carrière, . A. Farré, C. Proust-Lima, . J. Ryan, M.-L. Ancelin and K. Ritchie, "Chronic and remitting trajectories of depressive symptoms in the elderly. Characterization and risk factors," *Epidemiol Psychiatr Sci,* 2017.

[19] A. Rouanet, . P. Joly, J.-F. Dartigues, C. Proust-Lima and . H. Jacqmin-Gadda, "Joint latent class model for longitudinal data and interval-censored semi-competing events: Application to dementia," *Biometrics,* 2015.

[20] . L. Collins and S. Lanza, Latent Class and Latent Transition Analysis, John Wiley & Sons, Inc., 2010.

[21] F. Bartolucci, "Introduction to Latent Variable Models for Cross-Sectional and Longitudinal Data," 2012. [Online]. Available: http://www.econ.upf.edu/~michael/latentvariables/.

[22] P. Bialowolski, "Forecasting household debt with latent transition modelling," *Applied Economics Letters, Taylor & Francis Journals,* 2017.

[23] R. Festl, J. Vogelgesang, M. Scharkow and T. Quandt, "Longitudinal patterns of involvement in cyberbullying: Results from a Latent Transition Analysis," *Computers in Human Behavior,* 2017.

[24] K. Diederik P and W. Max, "Auto-Encoding Variational Bayes," 2013.

[25] P. Yunchen, G. Zhe, H. Ricardo, Y. Xin, L. Chunyuan, S. Andrew and C. Lawrence, "Variational Autoencoder for Deep Learning," *NIPS,* 2016.

[26] Y. Zichao, H. Zhiting, S. Ruslan and B.-K. Taylor, "Improved Variational Autoencoders for Text Modeling using Dilated," *Proceedings of the 34th International Conference on Machine Learning,* 2017.

[27] "Tutorial - What is a variational autoencoder?," no. https://jaan.io/what-is-variational-autoencoder-vae-tutorial/.

[28] F. Otto and R. v. A. Joost, "Variational Recurrent Auto-Encoders," *ICLR workshop track,* 2014.

[29] L. Antelmi, N. Ayache, P. Robert and M. Lorenzi, "Sparse Multi-Channel Variational Autoencoder for the Joint Analysis of Heterogeneous Data," *International Conference on Machine Learning,* 2019.

[30] S. Najibesadat, Z. N. Milad, C. Ratna Babu and Z. Dongxiao, "Representation Learning with Autoencoders for Electronic Health Records: A Comparative Study," 2019.

[31] "PROC LCA & PROC LTA Users' Guide," 2015. [Online]. Available: https://www.methodology.psu.edu/files/2019/03/proc_lca_lta_1-3-2-1_users_guide-2ggq4d3.pdf.

[32] J. Morris, B. Fries, D. Mehr, C. Hawes, C. Philips, V. Mor and L. Lipsitz, "MDS Cognitive Performance Scale," *Journal of Gerontology: Medical Sciences,* 1994.

[33] "R package for latent variable models with categorical data," 2009. [Online]. Available: https://msu.edu/~chunghw/downloads.html.

[34] E. Whitlock, L. Diaz-Ramirez, M. Glymour, W. Boscardin, K. Covinsky and A. Smith, "Association Between Persistent Pain and Memory Decline and Dementia in a Longitudinal Cohort of Elders.," *JAMA Intern Med. ,* pp. 1146-1153, 2017.

Chapter 8.

Appendices

8.1. Appendix A – Description of Features

Only a subset of the features which are mentioned in this thesis is described below. Each subsection corresponds to the questionnaire items mentioned in Chapter 4 and Chapter 5 respectively.

8.1.1. Covariate Set: Age only

Name	Description	Code	Meaning
B2A_0 (dummy)	Cognitive Skills - Decision Making	0	Independent
		1	Varying level of impairment
H1AA	Meal Preparation - Self-Performance	0	Independent
		1	Some help, Full help, or By others
H1AB	Meal Preparation - Difficulty	0	No Difficulty
		1	Some difficulty or Great difficulty
H1BA	Ordinary housework - Self-Performance	0	Independent
		1	Some help, Full help, or By others
H1BB	Ordinary housework - Difficulty	0	No Difficulty
		1	Some difficulty or Great difficulty
H1CA	Managing Finances - Self-Performance	0	Independent
		1	Some help, Full help, or By others
H1CB	Managing Finances - Difficulty	0	No Difficulty
		1	Some difficulty or Great difficulty
H1DA	Managing Medications - Self-Performance	0	Independent
		1	Some help, Full help, or By others
H1DB	Managing Medications - Difficulty	0	No Difficulty
		1	Some difficulty or Great difficulty
H1EA	Phone Use - Self-Performance	0	Independent
		1	Some help, Full help, or By others
H1EB	Phone Use - Difficulty	0	No Difficulty
		1	Some difficulty or Great difficulty
H1FA	Shopping - Self-performance	0	Independent
		1	Some help, Full help, or By others

H1FB	Shopping - Difficulty	0	No Difficulty
		1	Some difficulty or Great difficulty
H1GA	Nurse monitoring daily	0	Independent
		1	Some help, Full help, or By others
H1GB	Nurse monitoring daily	0	No Difficulty
		1	Some difficulty or Great difficulty
H4A	Mode of Locomotion - Indoors	0	No Assistive device
		1	Cane, Walker/crutch, scooter, wheelchair
H4B	Mode of Locomotion - Outdoors	0	No Assistive device
		1	Cane, Walker/crutch, scooter, wheelchair
H5	Stair Climbing	0	Up and down stairs without help
		1	Up and down stairs with help or Not go up and down stairs
H6A	Stamina - Days	0	Every day
		1	2-6 days a week, 1 day a week, or no days
J1X	Cancer, Not including Skin Cancer	0	Not Present
		1	Present
K6A	Unsteady Gait	0	No
		1	Yes
N2B	Unsteady Gait	0	No
		1	Yes
N5E	Wound Care - None of the Above (Antibiotics, Dressings, Surgical Wound Care, Other Wound/Ulcer Care)	0	No
		1	Yes
P2W	Nurse monitoring daily	0	Not Applicable
		1	Scheduled, full adherence as prescribed, partial adherence, or not received
Q4	Compliance/Adherence with Medications	0	Always compliant
		1	Not always compliant

Table 8 Description of all questionnaire items mentioned in Chapter 4.

8.1.2. Covariate Set: All Features

Name	**Description**	**Code**	**Meaning**
K4B	Pain - Intensity	0	No pain, mild, moderate pain
		1	Intense pain
K4C	Pain that disrupts usual activities	0	No
		1	Yes

P5	Treatment Goals	0	No
		1	Yes

Table 9 Description of all questionnaire items mentioned in Chapter 5.

8.2. Appendix B – Latent Class Analysis across Timesteps

time	Num of latent groups	AIC	BIC	Max Log likelihood
Earliest record	2	-8867	-8746	4451
	3	-7504	-7322	3779
	4	-7320	**-7078**	**3696**
	5	**-6117**	-7815	4103
Last record	2	-5584	-5463	2810
	3	-4625	-4443	2339
	4	**-4547**	**-4305**	**2309**
	5	-4633	-4331	2361

Table 10 Latent Class Analysis across Timesteps

8.3. Appendix C – Classifiers Performance Summary

8.3.1. Target Variable: Alzheimer's Disease

Classifier	Label	Precision	Recall	F1-Score	Support
K-NN(k=3)	0	0.92	0.99	0.95	3263
	1	0.83	0.43	0.56	487
	macro avg	0.85	0.65	0.7	3750
K-NN(k=4)	0	0.94	0.99	0.96	3320
	1	0.81	0.47	0.6	430
	macro avg	0.87	0.73	0.78	3750
K-NN(k=5)	0	0.91	0.99	0.95	3215
	1	0.83	0.39	0.53	535
	macro avg	0.93	99	0.96	3750
Linear SVM	0	1	0.99	0.99	3536
	1	0.85	1	0.92	214

	macro avg	0.92	0.99	0.96	3750
Decision Tree *	0	0.99	0.99	0.99	3490
	1	0.85	0.83	0.84	260
	macro avg	0.92	0.91	0.91	3750
Random Forest	0	0.99	0.99	0.99	3493
	1	0.83	0.82	0.83	257
	macro avg	0.91	0.9	0.91	3750
Multilayer Perceptron **	0	1	0.99	0.99	3508
	1	0.86	0.89	0.87	242
	macro avg	0.93	99	0.96	3750
Adaboost	0	0.97	0.99	0.98	3434
	1	0.86	0.68	0.76	316
	macro avg	0.93	99	0.96	3750
Naïve Bayes	0	1	0.98	0.99	3563
	1	0.68	0.92	0.78	187
	macro avg	0.84	0.95	0.89	3750
Quadratic Discriminant Analysis	0	1	0.99	0.99	3526
	1	0.85	0.96	0.9	224
	macro avg	0.92	0.99	0.99	3750

Table 11 Performance of VRAE classifiers pertaining to eventually positive with Alzheimer's Disease.

** Maximum Depth of tree is 5*

***Training parameters: hidden size = 100, and learning rate = 0.5 with early stopping*

8.3.2. Target Variable: Dementia

Classifier	Label	Precision	Recall	F1-Score	Support
K-NN(k=3)	0	0.86	0.95	0.90	2471
	1	0.85	0.66	0.75	1129
	macro avg	0.85	0.81	0.82	3600
K-NN(k=4)	0	0.91	0.95	0.93	2609
	1	0.84	0.75	0.79	991
	macro avg	0.87	0.85	0.86	3750
K-NN(k=5)	0	0.85	0.95	0.90	2432
	1	0.86	0.65	0.74	1168
	macro avg	0.85	0.80	0.82	3600
Linear SVM	0	0.95	0.95	0.95	2740
	1	0.83	0.85	0.84	860
	macro avg	0.89	0.90	0.90	3600
Decision Tree *	0	1.00	0.95	0.97	2857
	1	0.83	0.98	0.90	743
	macro avg	0.91	0.97	0.93	3600
Random Forest	0	0.95	0.92	0.97	2858

	1	0.75	0.84	0.80	790
	macro avg	0.85	0.88	0.87	3600
Multilayer Perceptron **	0	1.00	0.95	0.97	2858
	1	0.83	0.99	0.90	742
	macro avg	0.91	0.97	0.94	3600
Adaboost	0	0.98	0.95	0.96	2787
	1	0.85	0.92	0.88	813
	macro avg	0.91	0.94	0.92	3600
Naïve Bayes	0	0.96	0.91	0.93	2859
	1	0.71	0.92	0.78	741
	macro avg	0.83	0.88	0.85	3600
Quadratic Discriminant Analysis	0	0.97	0.91	0.94	2902
	1	0.70	0.89	0.79	698
	macro avg	0.84	0.90	0.86	3600

Table 12 Performance of VRAE classifiers pertaining to eventually positive with dementia.

** Maximum Depth of tree is 5*

***Training parameters: hidden size = 100, and learning rate = 0.5 with early stopping*