Interpretable Machine Learning for Dementia Risk Prediction

Suryansh Manocha

MInf Project (Part 1) Report
Master of Informatics
School of Informatics
University of Edinburgh

2023

Abstract

This paper presents a novel approach to investigating the onset of dementia using interpretable machine learning techniques. Dementia is a debilitating syndrome that imposes significant challenges to the elderly population, ranking as a leading cause of death in the UK. While dementia remains a disease with no known cure or prevention, early detection and intervention can effectively mitigate the risk of its development. For this study, we use a portion of the longitudinal ELSA dataset, utilising 10 years of follow-up data from 2002-2012, with 7379 participants.

The study aims to identify the most influential covariates associated with the onset of dementia utilising survival analysis and machine learning models, including Cox Proportional Hazards model and Random Survival Forest, to predict the risk of adverse outcomes. Notably, this study's distinctive contribution lies in its meticulous consideration of the temporal aspect of the disease and competing risks, which is a crucial aspect often overlooked in other studies.

Moreover, the study employs several sophisticated interpretability techniques, such as Permutation Feature Importance, Cox Score Ranking, Local Interpretable Model-agnostic Explanations (LIME), and Counterfactual Explanations (DICE), to gain valuable insights into the primary variables contributing to the onset of dementia. This methodology has the potential to uncover new insights and perspectives on the complex etiology of dementia and provide valuable insights for future research.

We propose a Random Survival Forest model that exhibits superior performance compared to prior work in dementia research conducted on the ELSA dataset, as cited in [101]. Furthermore, we employ interpretable machine learning methods to examine the model's decision-making process and deduce the top 5 modifiable features that contribute the most to its predictions. We also investigate local explanations at patient level, and propose that these explanations may be more meaningful for individuals, rather than looking holistically at the top 5 global feature we identified.

Research Ethics Approval

This research did not involve the gathering of new data, but rather utilized an existing dataset (ELSA) which contains extensive information on a diverse group of human participants. The ELSA dataset has been previously collected and is readily available for use in scientific research. Specific details are provided in appendix D.

This project was planned in accordance with the Informatics Research Ethics policy. It did not involve any aspects that required approval from the Informatics Research Ethics committee.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Suryansh Manocha*)

Acknowledgements

I would like to express my sincere appreciation and gratitude to those who contributed to the successful completion of this research project.

First and foremost, I owe an immense debt of gratitude to my supervisor, Dr. Sohan Seth, for his unwavering patience, support, and valuable feedback throughout this journey. Without his guidance and expertise, this project would not have been possible.

I would also like to extend my thanks to Dr. Alan Marshall for his generous support and interest in the project, helping set a clear direction for the research. His contributions were truly instrumental in shaping this project.

I am deeply grateful to Dr. Chima Eke for his assistance in navigating the dataset and obtaining the necessary approvals for its use. His guidance and support were invaluable in bringing this research to fruition.

Finally, I would like to express my gratitude to the members of the SHaPE research group for their support and encouragement throughout this project. I am particularly grateful for the opportunity to contribute to the group's research endeavors and to be able to provide value to the group through this project. I am honored to have been a part of such a dynamic and supportive community.

Once again, I extend my sincere thanks to each of these individuals for their invaluable contributions to this research project.

Table of Contents

Chapter 1

Introduction

1.1 Motivation

Predicting the risk of adverse outcomes at patient level is crucial in healthcare; more often than not the earlier a patient can be diagnosed, the higher the likelihood of success for treatment [5], as it reduces the time to treatment initiation. The need for computationally determining risk of adverse outcomes is emphasised by the increasing life expectancy globally; notably, a projected 19% of female newborns in the UK in 2020 are anticipated to attain the age of 100 or beyond, a figure expected to increase to 27% by 2045 [7]. The increasing life expectancy amplifies the obligation of healthcare providers to diagnose patients in a cost-effective manner [29]. However, this phenomenon also intensifies the occurrence of diagnostic errors in primary care, which are relatively frequent and harmful [100]. With the passage of time, it is becoming progressively more feasible to employ computational methods for patient risk prediction, thereby yielding risk assessments that are more cost-effective, consistent, and accurate.

Dementia is widely acknowledged to have a strong correlation with old age, with the latter being commonly regarded as a primary etiological factor of the condition [8]. The leading cause of death in the UK in 2018 was dementia, accounting for 13% of all deaths registered [88], a condition that remains neither preventable nor curable at present [8]. The majority of dementia costs per year are due to social care, costing £12.5bn (50%) per year, with the total cost nearing £25bn [4].

The present study aims to take a novel approach to investigating the factors contributing to dementia by interpreting machine learning models. This methodology differs from traditional studies and has the potential to uncover new insights and perspectives on the complex etiology of the condition. Furthermore, the study seeks to explore both individual and societal measures that could be taken to reduce the prevalence and impact of dementia. An insight into the types of modifiable risk factors for dementia are shown in figure A.4, which is a study on the ELSI-Brazil cohort.

To derive value from risk prediction models, it is imperative to prioritize the interpretability of the model. This refers to the extent to which a human can comprehend the rationale behind a decision rendered by the model [84]. Presently, a plethora of

sophisticated machine learning models, such as Deep Neural Networks, exist that can outperform traditional models (Cox PH, section 2.1.1) in healthcare scenarios [102]. However, the lack of transparency in the hidden layers of these models renders them difficult to interpret. The European General Data Protection Regulation (GDPR) policy mandates that researchers must be capable of elucidating algorithmic decisions made by a machine learning model [102]. Thus, the present study aims to surmount this challenge by leveraging methods that are interpretable while simultaneously aiming to deliver the high performance of a complex model.

1.2 Literature Review

In the field of healthcare, existing risk prediction models are typically linear in nature owing to their high level of interpretability and ease of use in practical settings [102], considering only select informative variables to predict risk. Over time, researchers have adopted two primary strategies to enhance the performance of healthcare risk prediction models. The first approach involves utilizing a highly accurate, albeit less interpretable, complex model [119]. The second approach involves emphasizing interpretability through the use of a simple linear model [118]. In the present study, we pursue a blended approach that seeks to capitalize on the benefits of interpretability that are inherent in simpler models while avoiding the assumptions made by a linear model [116].

There exists a substantial amount of research conducted in disease-specific scenarios using machine learning techniques, such as [30, 118], however they generally fail to address the question in a time to event model context, whereby the models are explored with the use of survival analysis (section 2.1), allowing us to estimate a timescale of the likelihood of event occurrence. According to research in the field of healthcare risk prediction, the ability to monitor changes in risk potential over time provides an opportunity to forecast the stage at which a patient is most susceptible to experiencing an adverse outcome [42]. This, in turn, enables proactive measures to be implemented in a timely manner to mitigate such outcomes.

Recent studies have also endeavored to investigate the interpretability aspect of such models. For instance, in the paper by Jansen et al. [58], a multitude of machine learning models were trained to classify the survivalability of breast cancer patients; model explanations were generated utilizing techniques similar to those employed in our own work, such as LIME (section 2.2.3). However, the study only explored the associated risks of six covariates, the selection of which was determined by unspecified feature selection techniques. Furthermore, the classification models employed in the study did not account for the temporal dimension, as opposed to survival models.

Similarly, the authors in [73] propose an explanation-driven HCI model using machine learning algorithms to segregate patients into demented and non-demented groups, and evaluate the accuracy of their model compared to state-of-the-art approaches. They also use SHAP and LIME explanation algorithms to provide more interpretability for their learning models. The study presents a rudimentary explanation of the interpretability aspect, with only a cursory reference to the outcomes of SHAP and LIME. It is notable

that the study does not explain the extent of alignment or divergence of their findings with existing literature, nor does it offer a complete evaluation of all models used. It is important to note that, once again, the study pertains to a classification model that does not factor in time-to-event considerations.

Research studies that have followed the classification-based methodology have approached the temporal aspect by including time as a predictor in the model, defined as discrete time intervals from the start of the study, rather than incorporating time as the outcome of interest [86]. Nevertheless, this methodology diminishes its significance by not permitting the prognostic capability of the temporal dimension, a crucial component in the diagnosis of medical conditions in healthcare environments.

In the context of dementia, recent research has investigated innovative strategies for risk prediction. For example, [101] reports preliminary findings suggesting that time-to-event models utilizing nonlinear models, such as decision trees, can be applied to the same dataset (ELSA) used in our study. In particular, [101] shows how the use of nonlinear time to event models outperforms traditional models used in healthcare research today, such as the Cox proportional hazards model (2.1.1). However, their experiments have limitations. Firstly, they do not account for competing risks (2.1.3), which, if not correctly accounted for, can result in overestimation of the probability of the occurrence of the event, and also mis-estimation of the magnitude of relative effects of the covariates on the incidence of the outcome [17]. Secondly, they do not emphasize the interpretability of the model, which is a crucial factor in healthcare decision-making. For instance, the authors could have elucidated the reasoning behind the model's decisions, such as discussing the decision tree generated by the model. The main objective of this paper was to achieve improved metric performance with the use of machine learning in a survival analysis context; the paper claims to be the first to propose such a model for dementia using the ELSA dataset. The current literature on this topic suggests a noticeable gap in the exploration of the mechanisms behind the superior performance of the model, the potential inferences that can be drawn about the individual under consideration, and the significant covariate differences between the top-performing models. Addressing these issues could significantly enhance the practical relevance of this research in a clinical setting.

Other research that exists exploring dementia within the ELSA dataset typically concentrate on specific variables and their contributions to the onset or progression of dementia, for example paper [115] explores the impact of air pollution on the incidence of dementia, [16] explores the relationship between personality and dementia, and [62] explores how a social network could reduce the risk of developing dementia. Interestingly, all three of these studies use the same approach in their analyses, that is, to use as single model, that being Cox Proportional Hazards model, without any of these studies attempting to adjust for competing risks, nor mentioning the lack of competing risks. Moreover, none of these papers aim to explore more complex time-to-event models, such as Random Survival Forest (2.1.2), which has previously shown to provide greater confidence toward relative importance of model covariates [32].

1.3 Objectives

The novelty of this research paper stems from its distinctive approach to investigating the onset of dementia in a time to event perspective. Specifically, the study aims to identify the most influential covariates from a well-established set of variables associated with the onset of dementia. To our knowledge, there has been no prior exploration of interpretability techniques in survival machine learning using the ELSA dataset. Therefore, in this research paper, we will investigate interpretability techniques in survival machine learning (section 2.2), and compare this approach with the more conventional methods employed in healthcare research, which involves the use of univariate and regression survival models (section 2.1.1).

The aim of this research is to provide insight into the primary variables that contribute to the onset of dementia. It is hoped that the outcome of this study can shed further light on this issue, and subsequent research can be conducted on a case-by-case basis to investigate precisely why these variables might be contributing to the onset of dementia, and what can be done to mitigate their impact.

To make explicit the contributions of this project *(also reflected in figure 4.1)*:

1. Extensive research for academic literature pertaining to casual relationships with dementia; the purpose of which was to determine the most significant variables to be examined in this study (section 3.1).
2. The process of extracting pertinent information from the raw data involved extensive data wrangling, including imputation procedures to replace missing values; *this adds to the novelty of the paper, as no such pre-processed data is publicly available* (sections 3.2 and 3.3).
3. Training and running several survival models as described in (1), investigating the onset of dementia (chapter 4), including models that account for temporality (time-varying covariates) and competing risk, which are frequently overlooked in studies. Both of these were omitted from our baseline paper [101].
4. Analyse the model outputs, comparing performance between the different models, in addition to the results obtained from existing literature (chapter 5).
5. Interpreting the reasoning of the models using techniques described in section 2.2, to better understand the decisions that lead to the regression predictions made by the model, in order to extract actionable insights (chapter 6). *This adds novelty to the paper, as to our knowledge, no such interpretation techniques have been used for survival models for dementia*.

Chapter 2

Background

2.1 Survival Analysis

Survival analysis is a statistical method used to analyze data for which the binary outcome variable of interest is time until an event occurs [64], where the event of interest is typically the occurrence of a certain outcome, in our case, dementia or death. The goal of survival analysis is to estimate the probability of the event occurring over time and to identify the factors that influence this probability.

Survival analysis is characterized by its ability to consider the possibility that not all individuals may encounter the event of interest during the study period (censoring); a key feature which distinguishes it from other statistical methods. This is a crucial consideration for our research, given that we are utilizing a longitudinal study design (3), where participants may join or leave the study at any point in its duration. In this context, it is important to note that even if a participant is censored and does not experience the event of interest, their contribution to the study duration is still relevant and considered.

The hazard function 2.1 gives us the instantaneous potential for failing (event occurring) at time t per unit time, given survival up to time t. Notably, the hazard function is an unbounded whole number, and the survival model is usually written in terms of the hazard functions.

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \tag{2.1}$$

The hazard function can be estimated using a variety of methods, including non-parametric approaches such as the Kaplan-Meier estimator or parametric approaches such as the Cox proportional hazards model.

The survivor function 2.2 gives the probability that a person survives (does not encounter the event of interest) longer than some specified time t. It can be expressed as a function of the hazard function in 2.1.

$$S(t) = \exp\left(-\int_0^1 h(u)\,du\right) \tag{2.2}$$

2.1.1 Cox Proportional Hazards Model

2.1.1.1 Standard

In order to assess the relationship of explanatory variables to survival time, mathematical modelling is required. The Cox model is the most general and widely used regression model, as it is not based on any underlying assumptions of the survival distribution. We can express the Cox model as such 2.3 [64].

$$h(t,\boldsymbol{X}) = h_0(t)\exp(\sum_{i=1}^{p}\beta_i X_i) \tag{2.3}$$

where $\boldsymbol{X} = (X_1, X_2, ..., X_p)$, represents a feature vector of the model covariates, and $\sum_{i=1}^{p}\beta_i$ represents the effect parameters. Note that the baseline hazard $h_0(t)$ is left unspecified, making the Cox PH a semiparametric model.

The hazard ratio 2.4 can be defined as the ratio of the hazard functions for two groups or levels of a covariate at any given time, it quantifies the relative risk of an event occurring between two groups [64]. The hazard ratio has a similar interpretation to the odds ratio; a hazard ratio of 2 is interpreted as the exposed group (i.e group with dementia) as having twice the hazard of the unexposed group (i.e group without dementia) [64].

$$\hat{HR} = \frac{\hat{h_0}(t)\exp(\sum_{i=1}^{p}\hat{\beta}_i X_i^*)}{\hat{h_0}(t)\exp(\sum_{i=1}^{p}\hat{\beta}_i X_i)} = \exp(\sum_{i=1}^{p}\hat{\beta}_i(X_i^* - X_i)) \tag{2.4}$$

The Cox model cannot be optimised using the standard likelihood function (2.5), since the baseline hazard function is unspecified. Therefore, the Cox model uses a partial likelihood function, allowing it to depend only on the parameter of interest [67].

$$L(\beta) = \prod_{i=1}^{K}\frac{\exp(\beta x_i)}{\sum_{j\in R(t_i)}\exp(\beta x_j)} \tag{2.5}$$

where K is the set of chronologically ordered event times, $R(t_j)$ is the set of individuals at risk at time t_j.

A key assumption made with the Cox Proportional Hazards (PH) Model is that the features are time-independent, meaning that the features are assumed not to change once they are measured for each individual [67]. This has the implication that we are unable to account for change, for example, an individual might reduce their BMI over time, but the model will fail to notice this. A solution to this problem is proposed, using time-dependent Cox model. Another key assumption is that any 2 covariates have hazard functions whose ratio is a constant proportion over time - the hazard ratio is constant. This assumption may not hold true for complex relationships between several covariates and the risk of an event occurring over time, where the effect of a covariate on the hazard may vary over time.

2.1.1.2 Penalized

The standard Cox proportional hazard's model fails to provide valuable insight in the presence of a large number of covariates, because it internally tries to invert a matrix that

becomes non-singular due to correlations present amongst the covariates [6]. When the number of predictor variables is large, the model can become overfit and less accurate in predicting the event of interest on new data, occurring when the model captures noise and spurious relationships in the data, which leads to poor generalization performance.

Regularization techniques [47], such as L1 regularization (Lasso) and L2 regularization (Ridge), can be applied to the Cox model to address the issue of overfitting. These regularization techniques introduce a penalty term to the loss function, which encourages the model to select a subset of the most important predictors and to reduce the magnitude of the coefficients of the remaining predictors. Unlike other regularization techniques that shrink coefficients towards zero, L1 regularization uses a form of continuous subset selection, where a subset of coefficients is set to zero and effectively excluded. This leads to a reduction in the number of features required for prediction, making the model more parsimonious and interpretable [6].

The equation below shows how the standard loss function (2.5) would be modified to accommodate for the regularization term (2.6), where $\beta_1, \ldots, \beta_2$ are the coefficients for p features, $\lambda \geq 0$ is a hyperparameter controlling the shrinkage, α controls the weighting to L1 and L2 penalties (elastic net penalty) [99].

$$L(\beta) = \log\left(\prod_{i=1}^{N} \frac{\exp(\beta^T x_i)}{\sum_{j \in R(t_i)} \exp(\beta^T x_j)}\right) - \lambda\left(\alpha \sum_{j=1}^{p} |\beta_j| + (1-\alpha)\sum_{j=1}^{p} \beta_j^2\right) \tag{2.6}$$

2.1.1.3 Time-varying

The time-varying Cox model permits the covariate effects to exhibit temporal variability, thus capturing the dynamic nature of the association between the predictor variables and the outcome. This implies that the model assumes the impact of a specific covariate on the hazard function may change over time, and therefore it has the capacity to accommodate fluctuations in the hazard ratio as a function of time [20].

The equation 2.7 provides a modification to the original equation 2.3 to consider time varying covariates $X_i(t)$.

$$h(t, \boldsymbol{X}) = h_0(t) \exp\left(\sum_{i=1}^{p} \beta_i (X_i(t) - \bar{X}_i)\right) \tag{2.7}$$

2.1.2 Random Survival Forest

Survival analysis alone provides us a means to perform statistical inference holistically on the data that we have [2]; however, the aim of our research is to be able to generalise this to make predictions about future data, whilst also maintaining a balance between interpretability and predictive power. Typically, the application of machine learning algorithms prioritizes the absence of interpretability as a trade-off for enhanced clarity regarding the relationships present within the data [2]. However, despite the fact that random forest models are not inherently interpretable, it is conceivable to relinquish

some performance to scrutinize the underlying decision trees that constitute the random forest model, thereby enabling greater interpretability [120].

There are a variety of survival tree approaches [25] that could be used, most notably, a bagging approach could have been alternatively used, such as [78], which focuses on dataset with the presence of nonsusceptible patients. The key difference between the two approaches is that bagging aims to retain all features from the input dataset, whereas random forest considers a subset of features [13]; for the purpose of our investigation, we aim to find only the top contributing features, and so we use random survival forest.

Algorithm 1 shows a high level overview of the procedure used for creating the random survival forest, based on the original paper [56].

Algorithm 1 Random Survival Forest

Require: p: features of original dataset, B: the number of trees to grow, m: the number of predictors to consider at each split.
$forest \leftarrow \emptyset$
for $b = 1$ to B **do**
Draw a bootstrap sample from original data (on average, 37% out-of-bag data)
Grow a survival tree for the bootstrap sample:
At each node, randomly select p variables
Split the node using the candidate variable that maximizes survival difference between daughter nodes
Grow the tree to full size under the constraint that a terminal node should have no less than d_0 unique deaths
Calculate a cumulative hazard function (CHF) for the survival tree
Save the survival tree in the forest
end for
Calculate the ensemble CHF by averaging the CHFs of all trees in the forest
Calculate the prediction error for the ensemble CHF using the out-of-bag data
return The RSF, consisting of B survival trees and the ensemble CHF

2.1.3 Competing Risk Models

Competing risks analysis in our context enables the distinction between various pathways of failure, such as different causes of death that a person may experience. Our research focuses on identifying factors that are associated with the specific mode of failure, whether a person dies due to dementia or some other cause. Theoretically, the use of competing risks should provide more accurate estimates when analysing the marginal probability for cause-specific events [3], allowing us to accommodate for the competing nature of multiple causes for the same event of interest.

In the context of our research, a competing risk is defined as an event that obstructs the occurrence of the event of interest and simultaneously alters the likelihood of the event of interest. There are two main approaches to competing risk [43], the cause-specific hazard function [52], and the sub-distribution hazard function [41].

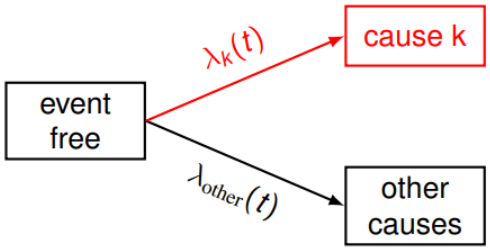


Figure 2.1: Illustration of cause-specific competing risk [45]

In order to understand the former, let us revisit equation 2.1 showing the hazard function, the new cause-specific hazard function would then be as shown in equation 2.8, meaning the probability of the death event occurring at time t for event type c, given that the death event has not occurred yet from event type c [17].

The cause-specific model can be better explicated by reference to Figure 2.1, which delineates the two conceivable states of transition from an individual free of the event. The diagram explicitly illustrates that such an individual can either transition into the outcome of interest (depicted in red, as cause k), or transition into an outcome that does not pertain to our interest (i.e., other causes). Specifically, $\lambda_k(t)$ denotes the transition rate to cause k (equivalent to c in eq 2.8), where $\lambda_k(t) = P(t \leq T_k < t + \Delta t | T_k \geq t)$ [45]. This in turn means that $\lambda_{other}(t) = P(t \leq T_{k'} < t + \Delta t | T_{k'} \geq t | k \neq k')$.

$$h_c(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T_c < t + \Delta t | T_c \geq t)}{\Delta t} \tag{2.8}$$

One key difference between the two models is how they handle individuals who have experienced a competing event. In the sub-distribution hazard model, individuals who have experienced a competing event continue to be included in the risk set for the event in question [41]. In contrast, in the cause-specific hazards model, individuals who have experienced a competing event are treated as censored observations [15]. The sub-distribution hazard model is shown in equation 2.9.

$$h_{c,subdist}(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T_c < t + \Delta t | T_c > t \text{ or } T_{c'} \leq t, c' \neq c)}{\Delta t} \tag{2.9}$$

Specifically, we intend to use the cause-specific model in this paper, as it is better suited for studying etiological questions [69]. One important assumption with the cause-specific model is that the competing risks are independent, this is because the model assumes that all other events, apart from the event of interest, are censored, and censoring has the underlying assumption of being independent.

2.2 Interpretable Machine Learning

2.2.1 Permutation Feature Importance

Permutation feature importance is a method that measures the impact of a feature on a model's score by randomly shuffling the values of the feature and calculating the decrease in the model's performance [26]. The random shuffling disrupts the relationship between the feature and the target, and the drop in model performance indicates the level of dependence on the feature. This approach is model agnostic and allows for repeated calculations using different permutations of the feature [1].

An overview of the algorithm (2) is provided based on the implementation we use in this paper [1]. The greater the decrease in performance, the more important that feature is to the model.

Algorithm 2 Permutation Feature Importance

Compute the baseline score s of the model on the dataset D
for each feature $f \in F$ **do**
  for each repetition k in $1,\ldots,K$ **do**
    Randomly shuffle f to get corrupted dataset $\hat{D}_{k,f}$
    Compute the score $s_{k,f}$ of the model on the permuted data $\hat{D}_{k,f}$
  end for
  Compute the importance score $i_f = s - \frac{1}{K}\sum_{k=1}^{K} s_{k,f}$
end for
  return Importance scores $\forall f \in F$

2.2.2 Cox Score Ranking

This is a novel technique that allows us to individually analyse each of the features from our model in a case-by-case manner [38]. This involves fitting a Cox Proportional Hazards model (2.3) to each variable separately and recording the c-index (2.3.1) on the training set [6], allowing us to determine the best risk predictor among the features. Algorithm 3 below shows the general procedure. Note however, that this ranking is based on a univariate Cox model.

Algorithm 3 Individual Cox Ranking

Require: s_f to denote the score for a feature
for each feature $f \in F$ **do**
  Fit a Cox Proportional Hazards model using only values of f
  Compute the score s_f of the model
end for
  return Importance scores $\forall f \in F$

2.2.3 Local Interpretable Model-agnostic Explanations (LIME)

LIME [95] works by approximating the predictions of any classifier or regressor locally with an interpretable model [35]. It generates a new dataset consisting of perturbed samples and the corresponding predictions of the black box model. Then, it trains an interpretable model, such as a linear model on this data; this model is trained to approximate the behavior of the black-box model in the local region around the instance of interest. Lastly, use the trained interpretable model to compute the importance of each feature for the prediction of the black box model. Equation 2.10 shows how explanations produced by LIME are obtained [95]; the goal being to find the best interpretable model g that minimizes this loss function while being as simple as possible, as measured by the regularization term $\Omega(g)$.

$$L_{lime}(x) = \arg\min_{g \in G} L(f, g, \pi_x) + \Omega(g) \tag{2.10}$$

Where f is the black-box classifier being explained, g is the interpretable model being trained to explain f, x is the instance to be explained, G is the set of interpretable models, π_x is a weighting function that assigns weights to the perturbed instances based on their similarity to x, and $\Omega(g)$ is a regularization term that penalizes complex models.

There are many similar methods of feature selection, such as mean decrease in accuracy (MDA) [48], and SHAP [74]; however, a recent study [76] suggests that LIME is more stable than both MDA and SHAP on features with high importance scores, and is better suited for human interpretability.

2.2.4 Counterfactual Explanations (DICE)

A significant challenge of utilizing LIME (2.2.3) or SHAP is that due to their reliance on simpler substitutes of the model to derive explanations, there is no assurance that these explanations accurately represent the original model [85]. Counterfactual explanations, on the other hand, provide factual insights into the model and extend beyond the confines of the training dataset. This makes counterfactual explanations a promising alternative for explaining machine learning models.

Counterfactual explanations represent another form of model-agnostic interpretability method that elucidates local interpretations, focusing on explainability of individual instances, as opposed to the overall model decisions. A counterfactual explanation characterizes the smallest alteration to the feature values that would result in a change in the prediction to a predefined output [84]. This enables us to describe situations in the form of a causal relationship, e.g ”If an individual decreases their body mass index to 25, the prediction of developing dementia would reduce by 20%”. This adds more depth as to how we are able to interpret the model, as it allows us to not only see one of the most prominent causal features in a specific case, but to also show the smallest change to that feature that would help mitigate the risk of dementia. This practical and applicable approach provides insights for preventative measures by identifying the ”easiest” change that could be made to an individual to help them specifically reduce the risk of developing dementia.

Various approaches are available to generate counterfactual explanations, such as the naive approach of randomly altering feature values of the instance of interest and stopping when the desired output is predicted. Other methods follow a white-box approach, such as the foil trees approach [108], which relies on the knowledge of the model. Alternatively, black-box approaches, such as DiCE [85], are model-agnostic.

In this paper, we will be using DiCE [85], which is a method that generates sets of diverse counterfactual examples for any differentiable machine learning classifier using determinantal point processes - probabilistic models of configurations that favour diversity [66]. Diversity in this manner relates to being able to give a wide range of suggested counterfactual examples, whilst also being close in proximity to the original input (ease of the change suggested). This method allows us to generate any number of counterfactual examples for a given input, and the implementation allows us to implement domain specific constraints on the counterfactual examples, such as weights and constraints; for example, in our case, we would wish to constraint counterfactual explanations relating to uncontrollable variables, such as age.

A loss function is defined 2.13, that takes as input the counterfactual and the desired outcome. This measures how far the predicted outcome of the counterfactual is from the predefined outcome; we also measure how far the counterfactual is from the instance of interest 2.12. Finally, we find counterfactual explanations by minimising this loss using an optimisation algorithm, as defined in 2.11.

$$C(\boldsymbol{x}) = \underset{\boldsymbol{c}_1,\ldots,\boldsymbol{c}_k}{\arg\min} \frac{1}{k}\sum_{i=1}^{k} \text{yloss}\left(f\left(\boldsymbol{c}_i\right),y\right) + \frac{\lambda_1}{k}\sum_{i=1}^{k} \text{dist}\left(\boldsymbol{c}_i,\boldsymbol{x}\right) - \lambda_2 \text{dpp\_diversity}\left(\boldsymbol{c}_1,\ldots,\boldsymbol{c}_k\right) \tag{2.11}$$

where $\boldsymbol{c}_i$ is a counterfactual example, k is the total number of counterfactual examples to be generated, f is the blackbox ML model, yloss is a metric that minimizes the distance between f's prediction for c_is and the desired outcome y (usually 1), d is the total number of input features, x is the original input, and dpp_ diversity is the diversity metric [85]. λ_1 and λ_2 are hyperparameters that balance the three parts of the loss function.

$$\text{dist}\left(\boldsymbol{c},\boldsymbol{x}\right) = \begin{cases} \frac{1}{d}\sum_{p=1}^{d} \frac{|\boldsymbol{c}^p - \boldsymbol{x}^p|}{MAD_p} & if continuous \\ \\ \frac{1}{d}\sum_{p=1}^{d} I\left(\boldsymbol{c}^p \neq \boldsymbol{x}^p\right) & if discrete \end{cases} \tag{2.12}$$

where, in each case, d denotes the number of continuous or discrete variables accordingly. For continuous variables, the total distance function is defined as the sum of the Manhattan distance weighted with the inverse median absolute deviation MAD_p. For discrete variables, we assign a distance of 1 if the counterfactual example's value for any discrete variable differs from the original input, otherwise it assigns zero.

$$\text{yloss } = \max(0, 1 - z * \text{logit}(f(\boldsymbol{c}))) \tag{2.13}$$

Hinge loss function, where z is -1 when $y = 0$ and 1 when $y = 1$, and $\text{logit}(f(c))$ is the unscaled output from the ML model.

2.3 Evaluation Metrics

Given that survival analysis accounts for censoring, standard regression evaluation metrics such as root mean squared error are not appropriate for measuring performance in survival analysis, and so we use specialised evaluation metrics [109].

2.3.1 Concordance Index (C-index)

The C-index is a measure of how well a model ranks individuals in terms of their risk of experiencing the event of interest, measured by comparing predicted outcomes to actual outcomes for pairs of individuals; two individuals are concordant if the observation with the higher predicted survival time also has a higher observed survival time. The C-index is the number of concordant pairs of observations divided by the number of comparable pairs; a value of 0.5 indicates that the model performs no better than random chance, while a value of 1.0 indicates perfect predictive accuracy. The higher the C-index, the better the model's predictive performance [49].

Equation 2.14 below shows how the concordance index is calculated, where the numerator counts the number of concordant pairs, and the denominator counts the number of comparable pairs, where a pair is considered comparable if the observation with the smaller observed survival time is not censored [97].

$$C = \frac{\sum_{i,j} I(\tilde{T}_i > \tilde{T}\, j) \cdot I(\eta_j > \eta_i) \cdot \delta_j}{\sum i, j I(\tilde{T}_i > \tilde{T}_j) \cdot \delta_j} \tag{2.14}$$

where i and j refer to pairs of observations in the sample, $\tilde{T}_i$ and $\tilde{T}_j$ represent the observed survival times respectively, η_i and η_j represent the predicted survival times. I is an indicator function takes on the value 1 if the condition inside the parentheses is true and 0 otherwise, δ_j is an indicator variable that takes on the value 1 if observation j is not censored (i.e., the event of interest has occurred), and 0 otherwise [97].

It is worth noting that the C-index is a measure of discrimination [31], which is the ability of a model to distinguish between individuals who experience the event of interest and those who do not. However, discrimination alone may not be sufficient to evaluate the overall performance of a survival model, as it does not take into account the calibration of the model (how closely the predicted probabilities match the observed probabilities). Therefore, it is often recommended to use multiple performance metrics when evaluating survival models [104].

2.3.2 Brier Score

In the context of survival analysis, the Brier score is used to evaluate the accuracy of a predicted survival function at a given time. It measures the mean squared difference between predicted and observed event times, with lower scores indicating better model

performance [105]. While the C-index measures how well a model discriminates between individuals with different survival times, the Brier score measures how well a model is calibrated (how close the predicted probabilities are to the actual outcomes). Using both measures together can provide a more comprehensive evaluation of a survival model's performance [50].

The general form of the Brier score is shown in equation 2.15, showing the mean squared difference between predictions f_t and observations o_t.

$$BS = \frac{1}{N}\sum_{t=1}^{N}(f_t - o_t)^2 \tag{2.15}$$

A more tailored equation for our purposes in the context of survival analysis is shown in 2.16, which shows the average squared distances between the observed survival status and the predicted survival probability.

$$BS(t) = \frac{1}{N}\sum_{i=1}^{N}(1_{T_i>t} - \hat{S}(t,x_i))^2 \tag{2.16}$$

where $1_{T_i>t}$ is the true status of a new test subject, and $\hat{S}(t,x_i)$ is the predicted survival probability for all $t \in \mathbb{R}^+$.

2.3.3 Brier Skill Score (BSS)

The Brier Score is an inadequate metric in situations of class imbalance, such as the scenario at hand where only a small proportion of patients are expected to develop dementia. The Brier Score can be modified using a reference score, as demonstrated in previous research [111]. In our specific case, the reference score $BS_{ref}(t)$ will be derived from a model that predicts the absence of dementia in all individuals, resulting in a notably low Brier Score due to the imbalanced class distribution; this serves .

$$BSS(t) = 1 - \frac{BS(t)}{BS_{ref}(t)} \tag{2.17}$$

2.4 MICE Imputation

Multiple Imputation by Chained Equations (MICE) is often used instead of other methods of imputation because it can reduce bias in subsequent analyses on datasets [81]; it has been shown to have a positive impact on feature selection compared to datasets imputed by basic techniques or non-imputed incomplete datasets [81].

The algorithm for MICE is presented in Algorithm 4, which is based on the approach detailed in the literature [18]. MICE runs a series of regression models, where each variable with missing data is modeled conditional upon the other variables in the data. MICE is a flexible approach, allowing for modeling of different types of variables, and has been used in datasets with thousands of observations and variables [51, 103].

Regarding the imputation model, we adopt the methodology depicted in Figure 2 of [117], in which the authors argue that the complexity of the relationships present in the data makes it challenging to specify an appropriate imputation model, and instead a random forest imputation model should be used. Therefore, we employ MICE with random forest imputation as a means of addressing this issue, utilizing a random forest as our imputation model.

Algorithm 4 Multiple Imputation by Chained Equations (MICE)

Require: Dataset with missing values: D, variables with missing data V
Ensure: Imputed dataset: $\hat{D}$
  Perform simple imputation (e.g. mean imputation) for every missing value
  Set maximum number of iterations and convergence threshold
  while *converged* $\leftarrow$ False **do**
    for each variable $v \in V$ **do**
      Remove the mean imputations for that variable
      Construct imputation model f using other variables as predictors, $v' \in V, v' \neq v$
      Estimate missing values for that variable using imputation model f
      Update $\hat{D}$ with new imputed values
    end for
    Check for convergence by comparing parameter estimates from current and previous iterations
  end while
  return $\hat{D}$

Chapter 3

Dataset

This paper draws data from the English Longitudinal Study of Ageing (ELSA) [37], which is a longitudinal study designed to track the health and well-being of a representative sample of the English population aged 50 years and older. It is a nationally representative study, meaning that participants are chosen randomly from the population and their characteristics reflect the population as a whole.

ELSA was first conducted in 2002 and has since been repeated every two years (1 wave). The study collects data on a wide range of variables related to health, economic status, social participation, and well-being; the exact number of variables tracked in ELSA can vary slightly from wave to wave of data collection, but typically the study measures thousands of variables for each participant [37].

This paper specifically utilises 10 years of follow-up data, dating from 2002-2012; the reasoning behind this relates to the lack of ease of availability for mortality data beyond this time-frame, which forms a key part of our analysis. Overall, after all pre-processing steps, we consider 7379 participants, 46.33% of whom are male, with each participant having 226 features. *It should be noted that the dataset was not readily available, and a significant portion of time was dedicated to pre-processing the data; the process for which are described in this chapter.*

3.1 Feature Selection

Due to the large size of the dataset and the diversity of its features, it was deemed impractical to run the models with all covariates, since high-dimensional data presents statistical challenges for machine learning models [59, 22], thus it is empirical to reduce this dimensionality before exploring the effects of the covariates on the outcome variable. This however requires domain expertise of the topic at hand, therefore, we take inspiration from similar work relating to dementia conducted on the ELSA dataset by domain experts in order to extract the key topics for the features that we will select. Table 3.1 shows the general topics of the features, along with the associated research study that shows a relationship between those features and dementia.

A common theme among the existing studies is the exploration of a causal link between

Variable	Reasoning
Level of education Leisure activities Complexity of occupation Cognitive abilities	Markers of cognitive reserve and dementia incidence, study on the ELSA dataset [12]
Income, Debt, Net wealth Housing tenure Property value	Inequalities in health at older ages, onset of ilness and survival effects, study on the ELSA dataset [80]
Other financial assets and wealth metrics	
Age, BMI	Predicting risk of dementia with survival machine learning - top features from feature importance analysis [101]
Hours of sleep / sleep measures Presence of APOE4 gene	
Measure of memory / Processing speed Social isolation Diet / fruit consumption	
Dementia diagnosis	Dementia metrics
Intensity/frequency of physical activity	Physical activity attenuates the risk of dementia, study on ELSA [40]
Occupational physical work	
Sex	Dementia diagnosis rates in the UK ethnic groups [90]
Ethnicity	
Diagnosis of heart diseases Diagnosis of diabetes Diagnosis of cancer	Associations between dementia and other disease, comorbidities [21, 44, 28]
Smoking	Associations between dementia and behavioural traits [89, 94]
Alcohol consumption	

Table 3.1: Table showing the variables of interest needing to be extracted from the ELSA dataset, the reasoning column links relevant paper(s). The variables highlighted are those that were only available on special request, with a cost associated with them [37] - therefore these variables are omitted from this study.

specific aspects of an individual's life and the outcome variable of dementia onset. For instance, [89] examines the specific association of smoking and dementia, revealing an increased risk of Alzheimer's disease with smoking. Such studies can inform the selection of relevant features for our analysis, assuming that they have sufficient explanatory power. Table 3.1 illustrates the diverse variable categories that we select, covering behavioural traits and financial measures among others.

3.2 Data Wrangling

The dataset poses several challenges for data wrangling; firstly, variables may vary across waves, which necessitates careful attention to ensure that the appropriate variables are chosen for analysis, and that their corresponding values are matched accordingly. Secondly, variables are not always readily accessible, which can make it time-consuming when trying to ascertain their definitions, especially because these definitions can change

between waves. Table C.1 depicts the substantial number of variables involved, which further complicates the process of variable selection and definition.

One of the goals of data wrangling was to establish both time-independent and time-dependent variables for analysis; this in turn enables us to analyse models such as the time varying cox proportional hazards model (section 2.1.1), which facilitates the investigation of models that take into account the changing variables over time for an individual, such as the time-varying Cox proportional hazards model (section 2.1.1). This, in turn, allows for the derivation of a concrete baseline when comparing our machine learning model to simpler, more traditional models, such as the cox model, as we account for the different forms of the baseline model, as well as providing an opportunity to compare the results of the variations. To obtain these two distinct datasets (as depicted in 4.1), specific preprocessing steps were undertaken separately for each dataset, including the creation of binary columns for each categorical variable and conversion to time-series.

3.3 Handling Missing Data

Handling missing data is an essential component of any statistical analysis, incomplete data can lead to biased estimates, reduced statistical power, and decreased precision of results [71]. Some variables and some individuals had a large number of missing values, which would need to be accounted for in order to perform our analysis. There are several techniques to deal with missing data, which can be categorised into either deletion or imputation [96]. There are a handful of missingness mechanisms which describe the nature of the data, and attempt to explain why the data might be missing; we are specifically interested in MCAR (Missing Completely at Random) [53], which refers to data which was collected randomly, and is missing at random, where missingness is not related to any other variables in the dataset [96]. Under this assumption, we can use the technique of 'case deletion', in which we can selectively delete missing data column-wise, or row-wise [96].

To handle incomplete data more effectively, we can combine the technique of deletion and imputation, like the authors do here [113]. This first involves selectively deleting missing data, and then performing an imputation technique to handle the remaining portion of missing data. Therefore, we first perform deletion based on two criterion: participants who have more than 40% missing features, and features which have more than 60% missing values. The percentages are chosen somewhat arbitrarily based on visual inspection of the data, shown in Figure 3.1, which shows the the distribution of the data based on the two criteria before and after deletion. A lower threshold was chosen for participants with missing features, since it is more difficult to impute for specific individuals of the study accurately, especially since our models rely on the features from individuals, rather than the feature values more holistically.

The deletion of data prior to imputation is a crucial step in ensuring that an excessive amount of data is not imputed, as this can have negative consequences, such as decreased reliability of the resulting data, as highlighted by Jakobsen [57]. There are several imputation methods available to handle missing data, and they can be broadly

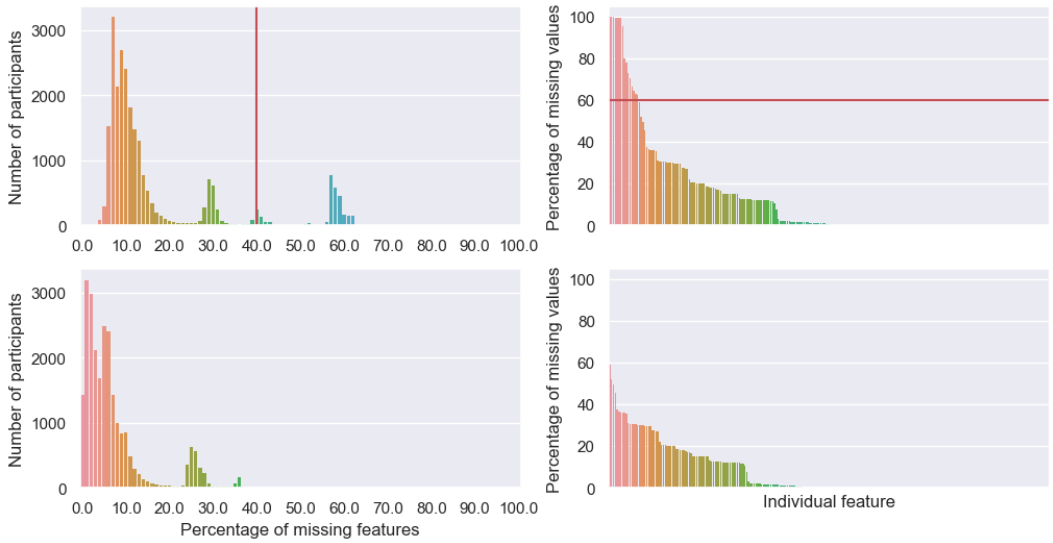


Figure 3.1: Figure showing the deletion criteria for missingness of data. The left side shows the percentage of missing values for each feature before (top) and after (bottom) cut-off; likewise the right side shows percentage of missing features for participants.

categorized into two types: single imputation methods, and multiple imputation methods. In our study, we adopted a strategy of harmonization through multiple imputation, owing to its potential advantages when implemented correctly. As noted by Siddique [98], these advantages include the preservation of variables on their original scale, the elimination of the need for specialized analytical methods after the imputation process, the retention of relationships among variables, and the accommodation of between-trial variability. The choice of imputation method depends on various factors such as the type of missing data, the sample size, the distribution of the data, and the analysis goals [107]. MICE (Multiple Imputation by Chained Equations) imputation is often preferred because it provides a flexible, robust, and efficient approach to handle missing data, particularly when the missingness is non-ignorable and related to other variables in the dataset.

Figure 3.2 shows the effect on the distribution of a selection of variables before and after applying imputation. We can observe that the general distribution of the variables post imputation remains similar to pre imputation, with the exception of the 'loneliness' feature; a reason for this could be that the missing data mechanism is not missing completely at random (MCAR), meaning that the missingness is related to the values of the variable itself or to other variables in the dataset.

3.4 Exploratory Data Analysis

The objective of this section is to provide a comprehensive understanding of the data, along with providing a notion of the quality of the data after performing the various preprocessing techniques described in the previous subsections.

Figure 3.3 shows the distribution of the study participants according to the number of

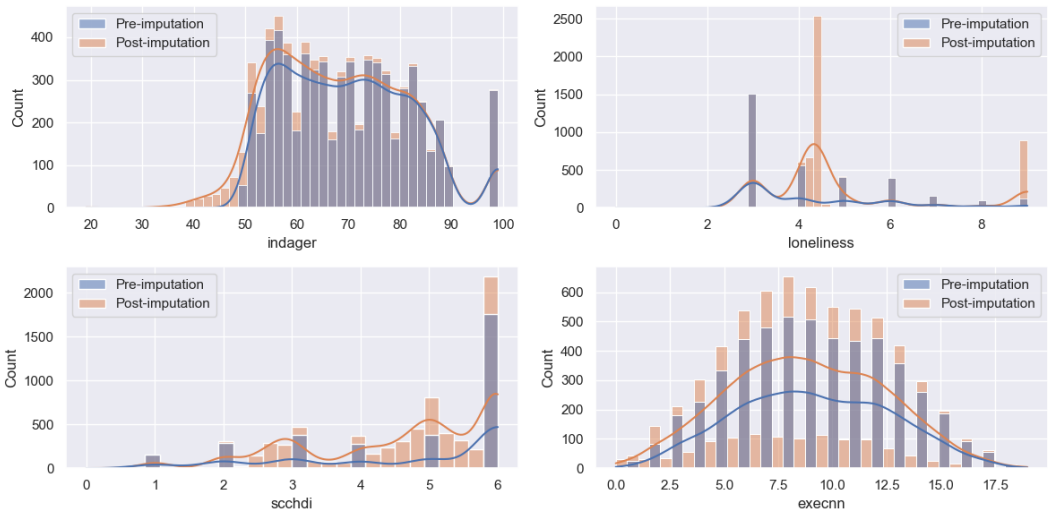


Figure 3.2: Figure showing the effect of imputation on the distribution of a sample of covariates (time independent).

months that they participated in the study. From this figure, it is clear that a large portion of participants were only in the study for a single recorded period (0 months), which hinders the time-to-event modelling capability that we strive to achieve. Nonetheless, we have a good distribution of participants that remain in the study for different durations of time, with the frequency spiking at around the time of each wave (2 year period). Although not visible in the log frequency graph in Figure 3.3, the frequency graph below makes clear the difference between participants with the event of interest of dementia, compared to the total observations. This class imbalance is something however that is to be expected when dealing with healthcare data [60], and the evaluation metrics (section 2.3) that we use must account for this imbalance.

Figure 3.4 explores the univariate Kaplan-Meier model, stratified across different feature values. The Kaplan-Meier plot displays the probability of survival (being dementia free) on the y-axis and time on the x-axis; the curve starts at 1.0, which represents the probability of survival at the beginning of the study. As time passes, the curve drops to reflect the proportion of individuals who have experienced the event of interest, dementia. The results that we obtain generally match with other literature, for example Figure 2 in [68], which shows a similar graph for the probability of being dementia-free over time, stratified by age, where the older age groups have a magnitude of rate of change of the probability of being dementia-free over time greater than younger groups. The same observation can be made with our results in Figure 3.4, where the older age groups are increasingly more at risk of dementia over time. Moreover, Figure 2 in [68] shows another example stratified by gender, showing females having a greater magnitude of rate of change, matching that of our results in Figure 3.4.

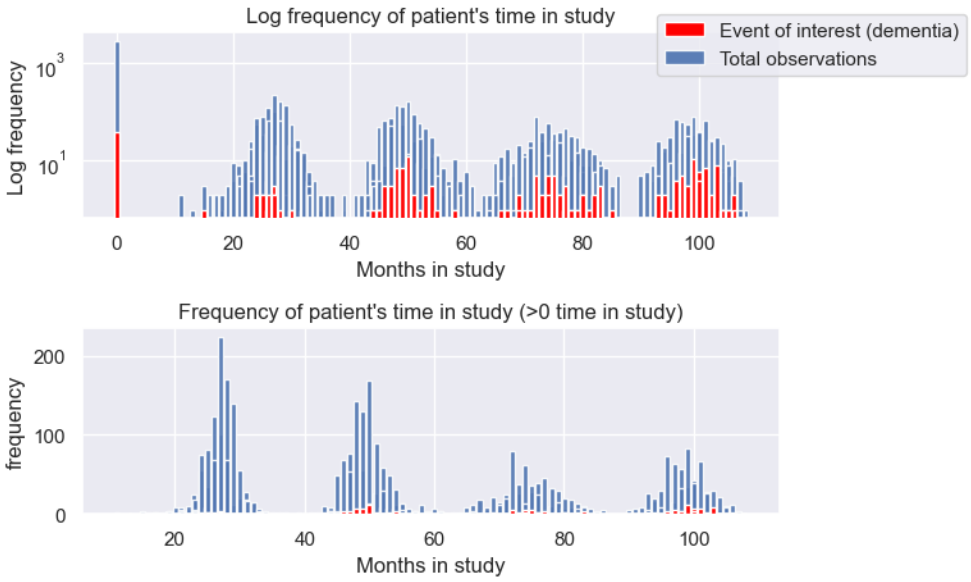


Figure 3.3: Figure showing the survival time distribution of our data (time independent).

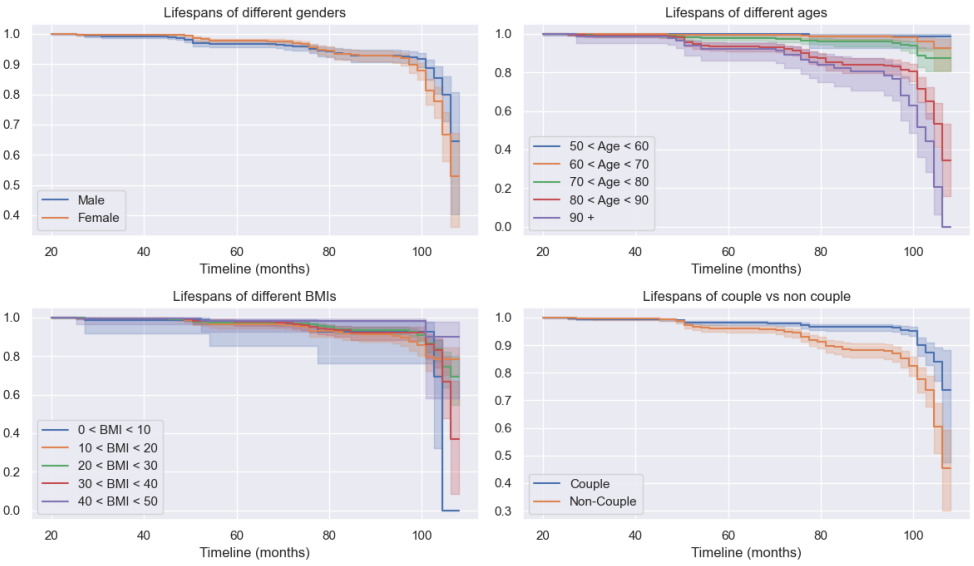


Figure 3.4: Figure showing the univariate Kaplan-Meier distribution, stratified amongst different covariate values (time independent).

Chapter 4

Experiments

We aim to investigate the following core questions in our investigation:

1. **Question:** Is it possible to improve upon existing research on predictive models for anticipating the likelihood of dementia onset over time for this dataset?
   The event of interest is dementia. The competing risk is death.
2. **Question:** Can these predictions be explained on an individual level - identifying the factors that contribute to an increased or decreased risk of dementia onset for a particular person?
   Such that can help individuals make informed decisions to take preventative measures.
3. **Question:** Can these explanations be generalised to the broader population - identifying the factors that contribute most commonly to dementia onset?
   Such that can aid in influencing health policies.

This chapter aims to explore **Question 1**, discussing the design choices, experimental details, and the results obtained by our models.

4.1 Experimental Setup

The objective of this section is to consolidate all the various methodologies that were previously discussed in chapter 2, with the intention of demonstrating their collective integration and providing justification for their selection, prior to delving into the outcomes produced by these methodologies (section 4.2).

Figure 4.1 shows a pipeline of the project's investigation, illustrating the sequential steps involved in the analyses. The first step begins with performing preprocessing on the raw ELSA dataset, which is described further in chapter 3, the result of which provides us with two cleaned datasets: time dependent data, and time independent data.

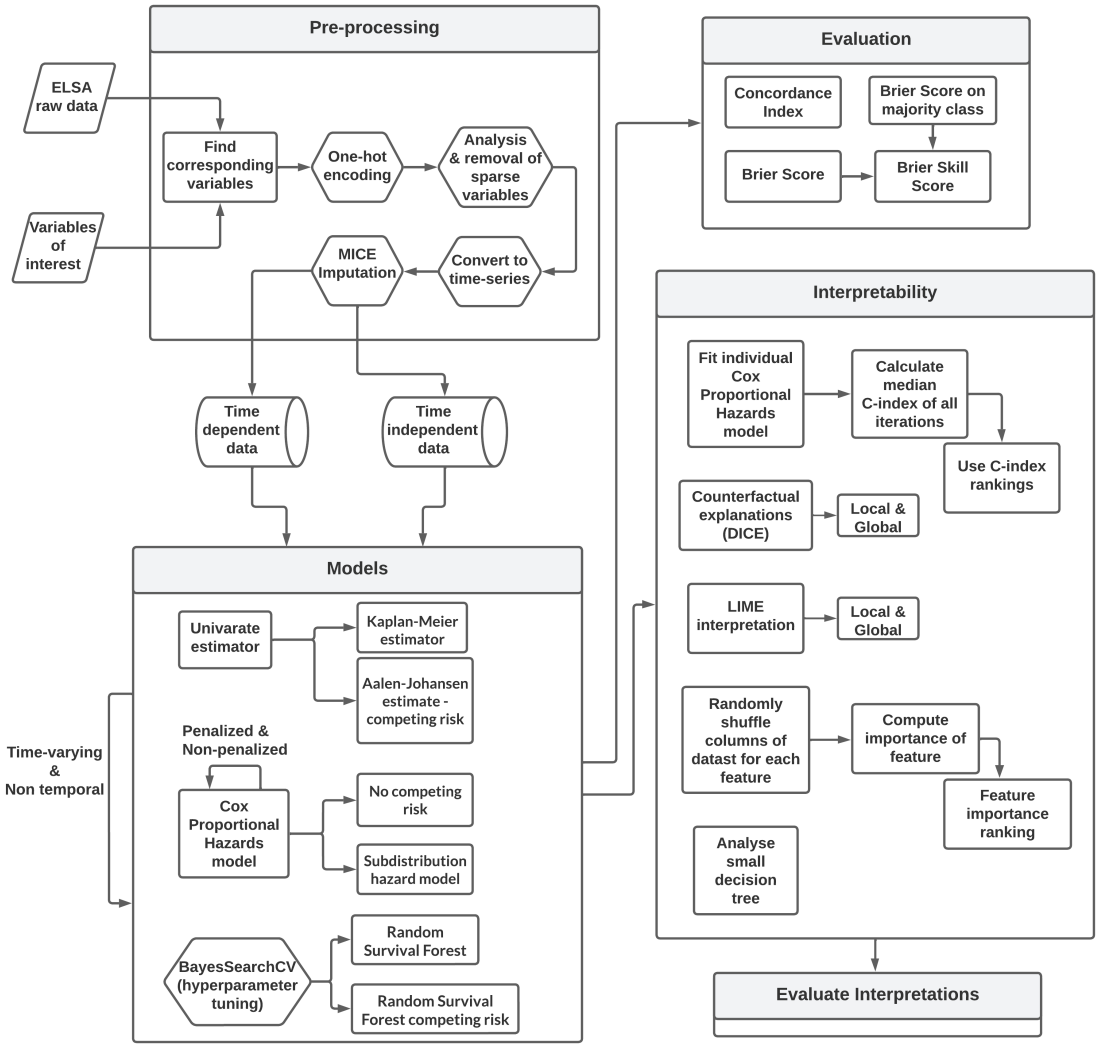


Figure 4.1: An abstract overview of the pipeline of the project.

4.1.1 Temporal Aspect

Investigating time-varying effects is imperative to the conduct of Cox survival analyses [20]. The identification and consideration of time-varying effects provides critical information on specific temporal patterns that may not be apparent through alternative methods. For instance, the conventional Cox model may not display a statistically significant effect of a particular variable; however, tests on the same variable may reveal compelling evidence that its effect is non-constant over time. This change in the variable's effect may explicate the non-significant hazard ratios reported in previous Cox analyses, such as that of [101]. Therefore, an examination of time-varying effects is crucial in preventing erroneous deductions and obtaining crucial insights that would have otherwise been overlooked [20].

To incorporate the temporal aspect of the data in the different models utilized, modified versions were employed, as described in chapter 2. The initial pre-processing steps were undertaken in Python, and it was the intention to continue using Python for the

remaining project; however, some specialised models, such as time varying random survival forest models were only available in R [34], in fact, the referenced paper was only released 2 months prior to the writing of our paper. Therefore, in order to accommodate for this cross-linguality, the processed datasets were stored in forms that could be used by both R and Python, as shown in Figure 4.1.

4.1.2 Competing Risks

In populations with elderly subjects, such as the ELSA dataset, other causes of failure may occur prior to the event of interest (onset of dementia). Competing risk events prevent the occurrence of the event of interest, as well as the potential benefits of an intervention, thus it is imperative that prognostic models incorporate the consideration of competing risk events [114].

We can estimate cause-specific hazard models by censoring patients who experience the corresponding competing event, and subsequently fit standard Cox regression models, as detailed in the tutorial by Putter et al. [92].

4.2 Univariate Models

Whilst univariate models allow for stratification of the population based on covariates, as shown in the right Figure 4.2, these models do not account for multiple covariates, and only provide an insight into how the risk of onset of dementia changes over time.

The paper [36] highlights issues that may arise when using simplified analytic solutions to handle competing events, such as censoring the competing events. The way of handling competing risks can influence how study results are understood and can significantly impact absolute and relative risk estimates, especially if competing events are frequent. The Aalen-Johansen estimator [24] is presented as a way to avoid these issues and obtain more accurate estimates of risk in the presence of competing events. The left plot in Figure 4.2 shows the cumulative density between the two models for the whole population, where the KM model consistently has higher incidence values; a similar evaluation is shown in [87], in which the presence of competing risks, the Kaplan-Meier model is known for overestimating the cumulative incidence.

We can use univariate models to help determine single variable relationships between the onset of dementia and the variable of interest. For the purpose of this section, we focus on gender, widely accepted as disproportionately leading to increased risk of dementia onset for women [19, 112]. The log-rank test (chapter 2) provides a statistical comparison of two groups, which we can use to quantify whether there is a significant difference of the survival curves between the groups. We can test for example, with the null hypothesis H_0: there is no difference between genders in terms of the distribution curve, and H_1 being: the genders have different distribution curves. We can use the chi-squared statistic with one degree of freedom, giving a p value < 0.005, which means we can reject the null hypothesis, and therefore we can show that there is a significant difference between the genders, also shown in 4.2. We can support this claim with other

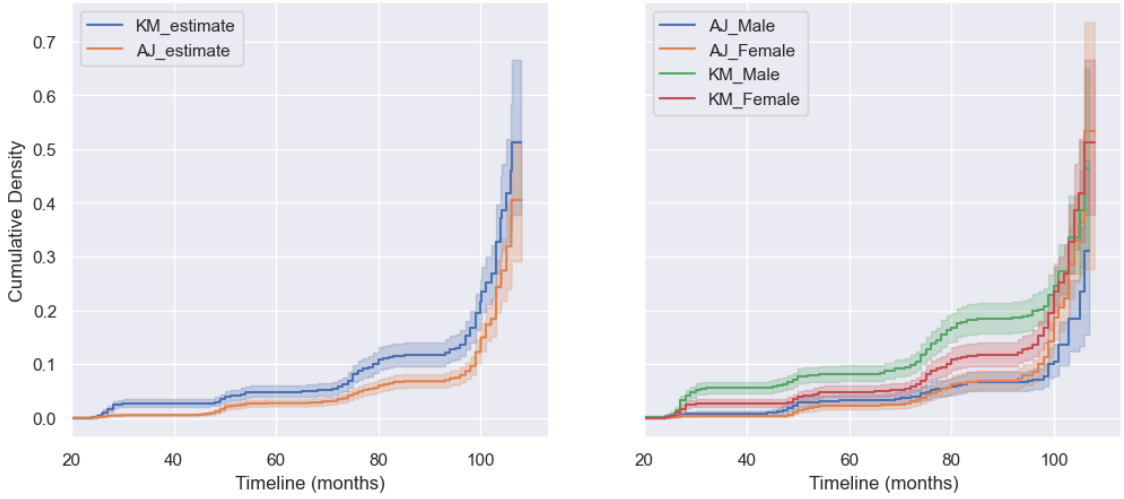


Figure 4.2: Comparison of univariate estimators with competing risk

literature, which states that women have a greater risk of developing dementia during their lifetime [82].

4.3 Cox Proportional Hazards Model

In order to account for the multivariate nature of our data, we can use the Cox model. Given the prevalence of this model in similar literature, and our focus on interpretable machine learning, we use this model as a baseline in our experiments. As denoted in section 2.1.1, there are different variations of the Cox model, and in order to evaluate the best of these, we run experiments on all of these variations, shown in table 4.1, evaluating the performance of the models based on the concordance index, brier score, and brier skill score (section 2.3). The penalised model utilises regularisation with hyperparameter values $\lambda = 0.2$, $\alpha = 0.9$ (from equation 2.6), this value was chosen by performing grid search, defining a grid of hyperparameter values to test, and evaluating the model with each combination of values.

The results in Figure 4.3 illustrate the impact of temporality and competing risks on hazard ratios, specifically for a subset of covariates utilized in the penalized Cox model. The primary objective of this visualization is to provide a rationale for the necessity of accounting for these factors (temporality, and competing risk) in the model, which can ultimately enhance the accuracy and reliability of the model's output. Interestingly, the inclusion of competing risks (4.3b) reveals a similar pattern to the univariate model, whereby the hazard is overestimated before accounting for the competing risk. Moreover, certain variables are more affected by competing risks than others; for instance, the variable 'headlma' exhibits a noticeable reduction in its hazard ratio. On the other hand, the temporal aspect modifies the results to a greater extent than the competing risk alone, as depicted in Figure 4.3c. This modification reduces the range of the confidence interval for all features and skews most features to a hazard ratio of 1. Certain variables, such as 'headlmo,' appear to have a substantial non-constant effect over time, with a hazard ratio increase of 0.2, similar outcomes were observed in [20] when considering

the temporal aspect.

Table 4.1 presents the outcomes of the analyses, with the best model being the penalized Cox model that accounts for both the temporal aspect and the competing risk of death; these results are consistent with those presented in [101]. Notably, our (non-temporal, penalised) model slightly surpasses the performance of the model in [101] in terms of the C-Index metric, unfortunately however it is not possible to compare a discrimination measure (i.e BS), as [101] does not provide such a measure.

Model	Time dependent	Competing Risk	C-Index	Brier Score	BSS
Cox Proportional Hazard	✓	✓	*N/A*	0.0408	0.150
		✗	*N/A*	0.0411	0.144
	✗	✓	0.748	0.0431	0.102
		✗	0.788	0.0438	0.088
Penalized Cox Proportional Hazard	✓	✓	*N/A*	**0.0405**	**0.156**
		✗	*N/A*	0.0407	0.152
	✗	✓	0.779	0.0410	0.146
		✗	**0.801**	0.0415	0.135

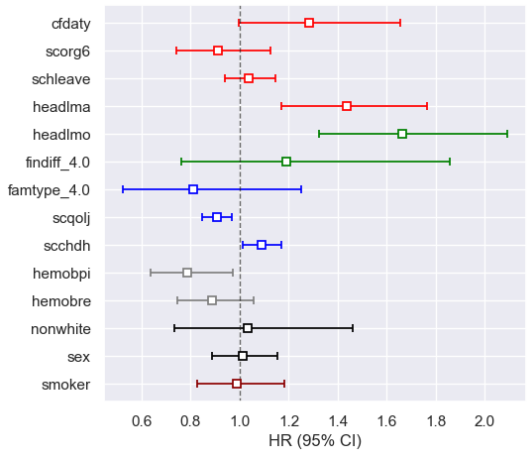
Table 4.1: The model results for the different variations of Cox Proportional Hazard. Note that the C-Index is not applicable for the time varying models, given the definition of the C-Index (chapter 2). 0.0480 is the reference Brier Score used for the calculation of BSS. Values in **bold** indicate the best performing model for each metric. The time used for calculating all Brier Scores was $t = 100$.

Even when using penalized cox proportional hazard, it is difficult to use this model to accurately predict the effect of covariates on the event of interest, this is because of the underlying proportionality assumption (section 2.1.1). The Cox model operates on the premise that all estimated survivor curves possess a consistent underlying structure (proportional hazards), which may not be an accurate depiction of actual outcomes, in fact, we see this assumption break down in the right Figure 4.2, where the male to female risk ratio changes over time. Conversely, the Random Survival Forests method does not make assumptions regarding the proportionality of hazards and is capable of accommodating a wide range of survivor curve shapes that may differ significantly among various subject groups [110].

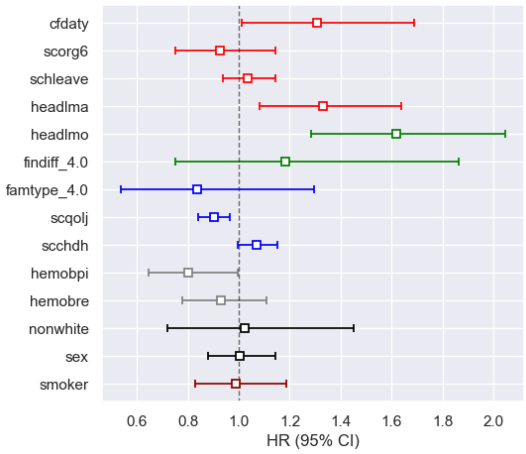
4.4 Random Survival Forest

Random Survival Forests (RSF) methodology is an attractive alternative approach to Cox model, as it is well-suited for handling high-dimensional covariate data and is adaptable to complex associations between predictors and outcomes, such as nonlinear effects or high-order interactions [72]. Moreover, it allows us to investigate techniques in interpretable machine learning to explore the decisions made by our model, which we do in chapter 5.

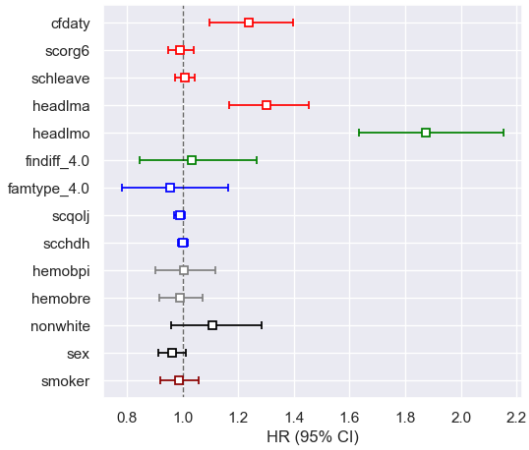
Similar to the previous section on Cox proportional hazards, we conducted four experi-



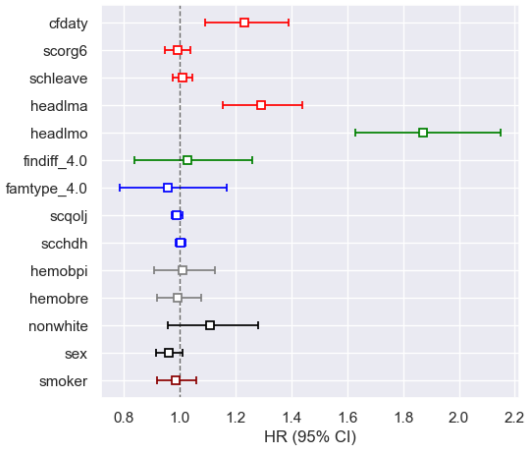
(a) Time invariant; without competing risk.



(b) Time invariant; with competing risk of death.



(c) Time variant; without competing risk.



(d) Time variant; with competing risk of death.

Figure 4.3: The hazard ratios for specific covariates using the penalized cox proportional model. The colours indicate the different types of variables, as discussed in section 3.1, where: red ⇒ cognitive reserve, green ⇒ financial factors, blue ⇒ social isolation and perception, gray ⇒ physical activity, black ⇒ demographic variables, and dark-red ⇒ smoking. *The exact definitions of the variables can be found in table C.1.*

ments using RSF, considering the combination of competing risk and temporality. In order to maximise the model's performance, we first perform hyperparameter tuning on each of the four different RSF models that we run. Specifically, we optimized the number of decision trees in the forest β and the minimum number of samples required to split an internal node m, as in section 2.1.2; one example of this process is shown in Figure 4.4. While tuning the exact number of trees is unnecessary in RSF, a larger number of trees generally leads to better performance. Therefore, we needed to identify the point at which performance plateaus to ensure computational feasibility and optimality. In order to choose these values, we used 5-fold cross validation using Grid Search. Whilst there are potentially other hyperparameters that we can also optimise, such as: the criteria with which to split on (i.e MSE or MAE), the maximum depth of individual trees, and the size of the bootstrapped dataset to train each decision tree with, it was infeasible to account for all of these hyperparameters due to the long training time cost associated with the process.

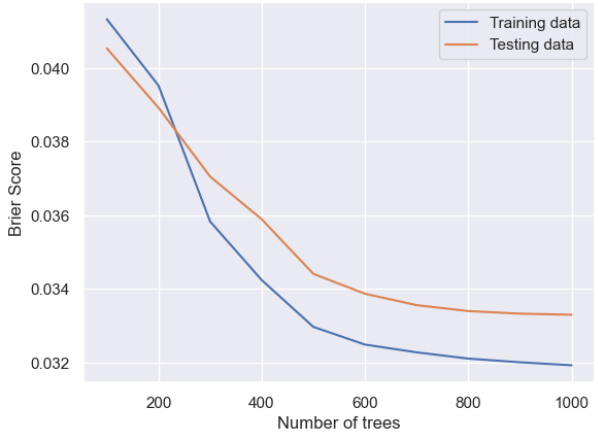


Figure 4.4: Hyperparameter tuning of Random Survival Forest with no competing risk and no temporality; tuning number of trees in relation to the Brier Score metric.

Table 4.2 shows the results obtained from these models. Notably, much like with the Cox model, the best performing model that we obtained is that which considers both the temporal aspect, and competing risk, as it achieves the highest Brier Skill Score. In comparison to existing literature, we achieve comparable results to the models presented in [101] that do not account for the temporal aspect, our models demonstrate improved C-Index scores, which may be attributed to our hyperparameter optimization or the use of different feature selection methods. Similar to [101], we see notable improvements in the performance of our RSF models, as compared to the Cox models, with a 230% increase in the BSS and a 9.5% increase in the C-index when comparing the best models from both approaches. This improvement is expected for the aforementioned reasons, with the lack of underlying statistical assumptions needed for RSF models [121], and the ability for RSF to handle high dimensionality data. We provide a detailed discussion of the results in Section 5, where we also explore the explanations for the decisions made by the model.

We also discuss in chapter 5 the use of a single decision tree for increased interpretability, as it allows us to directly view the decision process made by the model. Decision

trees are however known for overfitting to the training data [83], a common practise to mitigate this is to apply pruning, which is a technique used to reduce its size and complexity by removing certain branches or sub-trees that do not significantly contribute to its accuracy. The pruning process starts by growing the tree to its maximum size, evaluating accuracy, and iteratively removing branches or sub-trees from the leaves if they do not improve accuracy, until no further improvement is possible [83]. After applying this step, our decision tree achieves a C-Index of 0.629 and a BSS of 0.105, although these metrics are significantly lower than those of both RSF and Cox, they still demonstrate predictive ability, as they outperform a simple majority-class allocation of no dementia onset.

Model	**Time dependent**	**Competing Risk**	**C-Index**	**Brier Score**	**BSS**
Random Survival Forest	✓	✓	*N/A*	**0.0311**	**0.353**
		✗	*N/A*	0.0313	0.349
	✗	✓	**0.877**	0.0321	0.332
		✗	0.870	0.0320	0.333

Table 4.2: The model results for the different variations of Random Survival Forest. 0.0480 is the reference Brier Score used for the calculation of BSS. The time used for calculating all Brier Scores was $t = 100$.

Table 4.3 presents a comparison of our model with other related works in the field. It is noteworthy that the two Random Survival Forest (RSF) methods outperform the Deep Neural Network (DNN) approach. This is likely due to the feature selection process, where the DNN method selected only 91 covariates [10], while our model utilized 226 covariates (as shown in Table C.1). The study also mentions that the DNN approach achieves comparable results to a Cox model that uses the same features. The Cox model, as presented in the table, achieves a high C-Index score by utilizing genetic information, as demonstrated in Table 3 of [91]. However, the Brier Score of the Cox model is relatively high, and it is twice that of our model, indicating lower accuracy (similar to Mean Squared Error). Nonetheless, without a baseline score to calculate the Brier Skill Score, it is difficult to ascertain the effectiveness of the model relative to ours.

Model	**Underlying Model**	**C-Index**	**Brier Score**	**BSS**
Ours	RSF	**0.877**	**0.0321**	0.332
[101]	RSF	0.866	-	-
[10]	DNN	0.757	-	-
[91]	Cox*	0.917	0.0640	-

Table 4.3: Results obtained by similar research conducted on longitudinal studies investigating dementia or related diseases. * indicates that the model was exposed to genetic data, blood biomarkers, and therefore cannot directly be used for comparison with the other models that do not use such information.

Chapter 5

Evaluating Explanations

This chapter aims to explore **Question 2** and **Question 3**, using the best model we obtained (RSF) from our investigation in chapter 4, aiming to explore insights that we can obtain from these models at an individual and broader level.

To facilitate the evaluation of local explanations, particularly those pertaining to individual-level explanations, we will conduct two case studies of persons. Specifically, Person ID 6545 will represent an individual with consistently low probability of developing dementia over the duration of the study. On the other hand, Person ID 1692 will represent an individual with progressively increasing likelihood of dementia onset during the study period. Figure 5.1 depicts the survival probabilities of both individuals throughout the study. The use of such divergent case studies will enable us to examine explanations that identify the factors that can heighten or lessen the risk of dementia onset. Additionally, this approach will enable us to investigate the variables that individuals with high and low risk of dementia can modify to decrease their risk, recognizing that these variables may differ between the two groups.

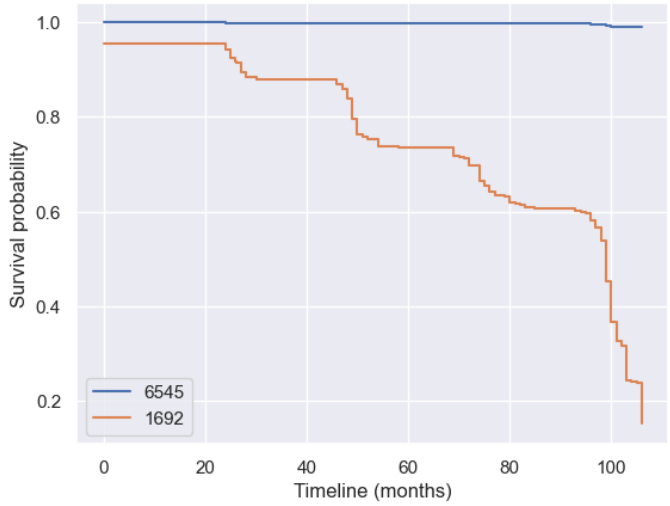


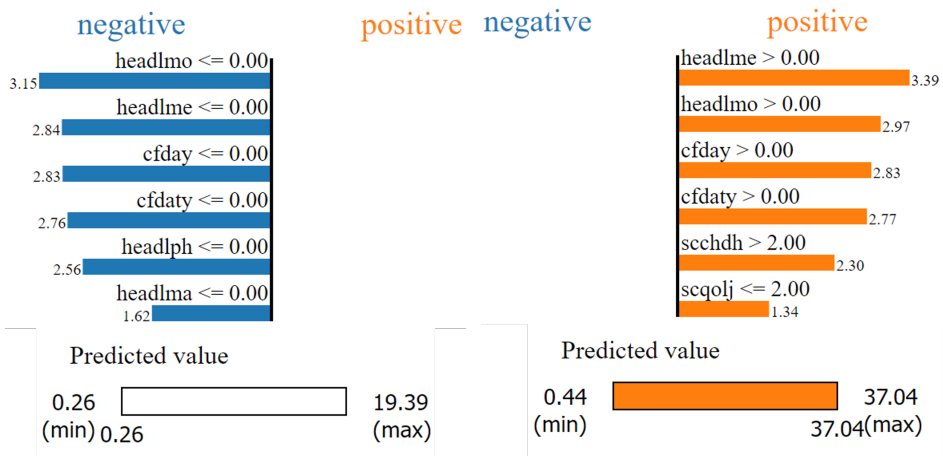
Figure 5.1: Survival curve of chosen individuals. A lower survival probability denotes a higher probability of dementia onset.

5.1 Local Explanations

5.1.1 Interpretations using LIME

LIME produces a list of explanations that shows how much each feature contributes to the prediction of a specific data sample (section 2.2.3). This affords a means of explicability within a restricted context, while also permitting the identification of those attributes whose modification will have the most pronounced effect on the prediction.

The results generated by LIME for our two case studies are presented in Figure 5.2. Each figure displays the feature values that either prevent or lead to the onset of dementia, with the negative side representing the former and the positive side representing the latter; the top features that contribute most to the final outcome are shown. From Figure 5.2a, we can observe the top contributing features that prevented the onset of dementia. It is noteworthy that none of the top contributing features were located on the positive side, which is reflected in Figure 5.1, where the risk of dementia for this particular individual is shown to be particularly low. In contrast, Figure 5.2b illustrates that all top contributing features are located on the positive side. Additionally, not all of the feature explanations are the opposite of those in Figure 5.2a. For instance, we observe the introduction of new feature contributions, such as *scchdh* and *scqolj*. This variation is due to the consideration of two different individuals, and their respective risk factors are influenced by different covariates.



(a) Person ID 6545. (b) Person ID 1692.

Figure 5.2: LIME local explanations. See Table C.1 for variable definitions.

5.1.2 Counterfactual Explanations

Counterfactual explanations allow us to see the minimal changes required to the features to alter the outcome (section 2.2.4). This is an especially powerful tool in a healthcare context like ours, as it provides a clear interpretation of which factors an individual can change to reduce their risk of dementia. Importantly, we can adjust the counterfactual explanation model to disregard unmodifiable factors such as age and BMI, which assures us of obtaining explanations that are genuinely viable, as we have done.

Person ID	Feature Values to Change				Dementia Outcome
1692	*Original*	nright = 0.93	findiff_4.0 = 0.0	scqoli = 2.47	1
	Proposed	nright = 3.1	findiff_4.0 = 0.1	scqoli = 0	0
6545	*Original*	scqold = 2.0	marstat_5.0 = 0.5		0
	Proposed	scqold = 0.6	marstat_5.0 = 0.7		1
6545	*Original*	scqolr = 2.0	scfami_1m = 0.98		0
	Proposed	scqolr = 0.2	scfami_1m = 0.2		1

Table 5.1: Counterfactual explanations on our two case studies, displaying the minimum adjustments necessary to the features to modify the outcome.

Table 5.1 illustrates a subset of possible examples from our two case studies. The table demonstrates, for instance, that if Person 1692 were to enhance their numeracy index (*nright*) by 2.17, and possibly manage their wealth more efficiently (*scqoli*, *findiff-4*), the RSF model would likely classify them as not having dementia. Likewise, if Person 6545 were to perceive life as less full of opportunities (*scqolr*), and correspond less frequently with relatives (*scfami*), our RSF model would predict a higher likelihood of them developing dementia. However, these explanations should not be accepted without scrutiny, as the underlying change in outcome relies on our RSF model, which although accurate, is not infallible. Additionally, we have only selected a subset of possible explanations. We provide another example of a counterfactual explanation for individual 6545, to highlight that multiple counterfactual explanations are feasible, and choosing the most appropriate explanation for a particular individual can be challenging [84].

By taking advantage of this observation, we can execute several explanations for an individual and identify the features that appear most frequently, while considering the distance metric (section 2.2.4). Figure 5.4c displays these features for Person 6545, which, if modified (likely in conjunction with other variables), would lead to the onset of dementia. It is noteworthy that the top features presented in the figure differ entirely from those in Table 5.1, as the table only encompasses a small fraction of the feasible explanations. Interestingly, the variables depicted in the figure incorporate financial aspects, including expenses related to leisure activities (referred to as *leisureu*), which were not identified by the LIME algorithm.

5.2 Global Explanations

5.2.1 Permutation Feature Importance

Permutation importance is a statistical technique for assessing the importance of predictor variables in our RSF [27]. It involves randomly shuffling feature values and observing the impact on model accuracy; by comparing the accuracy before and after permutation of a single predictor variable, the feature's importance can be determined (section 2.2.1).

Figure 5.4a shows the result of this process, which shows that the key indicator of

dementia onset was whether the participant was able to answer correctly the current year (*cfdaty*). This makes sense, since dementia is a condition that leads to memory loss, and an individual not being able to correctly answer this question likely relates to dementia. In fact, almost all of the observations could be said as being related to cognitive abilities, which one would expect, with the exception of *scchdh*, *scqolq*, *scqolj*.

5.2.2 Global Feature Importance (Counterfactual)

Much like in 5.4c in which we gathered several counterfactual explanations for a single individual in order to determine the most likely features that contribute to dementia, we can generalise this idea in order to generate global feature importance. We do this by observing several counterfactual explanations for several individuals in the study population and taking the features that occur most frequently, while considering the distance metric (section 2.2.4).

Figure 5.4d shows the global counterfactual explanations produced. Surprisingly, the model shows us that the top 4 important features are related to socioeconomic variables, such as money spent on leisure and total income; by contrast, permutation feature importance did not include any financial variables in its top features. We observed a similar behaviour for section 5.1.2.

5.2.3 Small Decision Tree Nodes

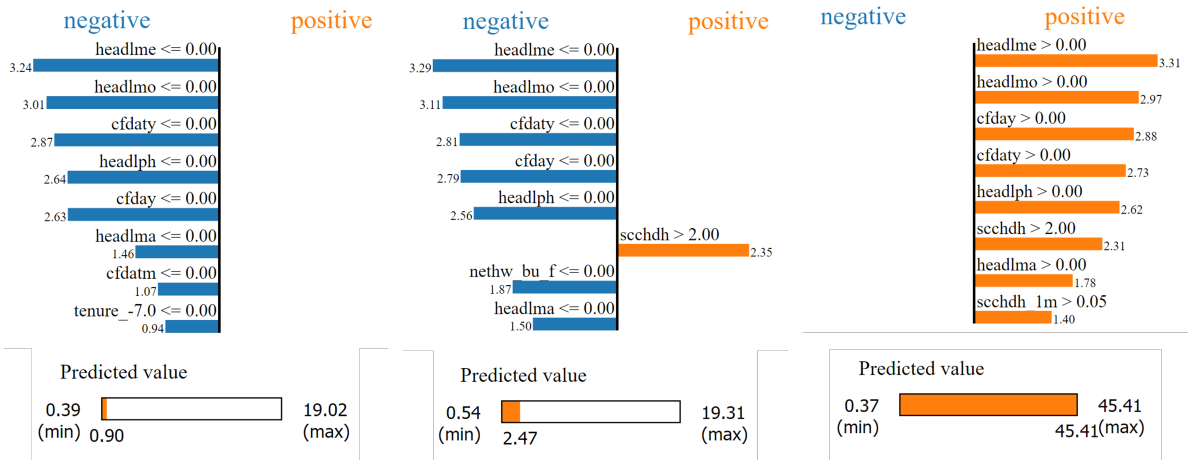
As mentioned in section 4.4, we also trained a small, pruned decision tree in order to have a highly interpretable model. The structure of this model is shown in Figure A.5, which shows the exact decisions made by the model to come to a prediction for dementia onset. Importantly, we can observe that many of the features that the tree considers are similar to our global counterfactual explanations in 5.4d. This may be because both of these models consider covariates in conjunction with one another, they are multivariate explanations of the key model features, whereas cox score ranking (section 5.2.5) provides univariate explanations.

5.2.4 Submodular Pick (LIME)

To provide a global understanding of the model, [95] propose explaining a set of individual instances, called Submodular pick, an extension of LIME. This method selects a few representative instances by identifying diverse and non-redundant explanations. It computes feature importance and defines a coverage function, which measures the total importance of features appearing in at least one instance. The objective is to maximize coverage within a budget constraint, solved through a greedy algorithm adding instances with the highest marginal coverage gain until the budget is met [95]. This results in using a small number of examples that are carefully chosen to provide a comprehensive understanding of the model's behavior.

Figure 5.3 shows these global explanations. We provide three different explanations based on different risk sets, this is to allow diversity in our explanations in order to see

why people with low risk of dementia have low risk, and why people with high risk of dementia have high risk, as these factors are not always the same, as shown in 5.3a and 5.3c. For example, we see cognitive factors (*cfdaty*) in all of the explanations, which we would expect. However, 5.3b and 5.3c both show that speaking to their children less often (*scchdh*) is a significant factor leading to higher risk of dementia. This is something that does not necessarily translate the other way round, for example, we do not see this variable occurring in 5.3a, meaning that it does not mean that speaking to children more frequently than (2, which means at least once or twice a week) leads to lower dementia risk. Interestingly, we also see that difficulty making phone calls (*headlph*) plays an important role in each of the three explanations, which is likely strongly correlated with speaking to children. We explore this further in chapter 6.



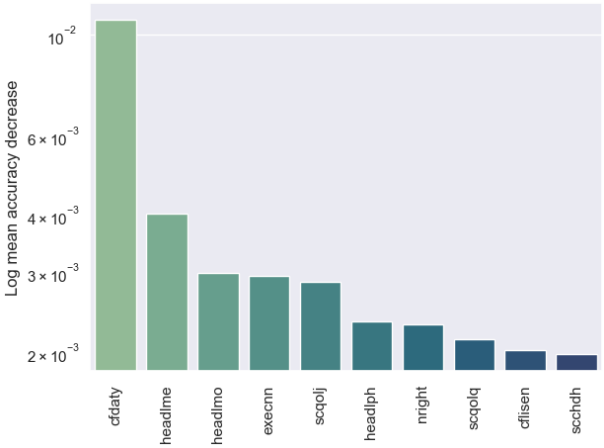
(a) Very low risk of dementia. (b) Low risk of dementia. (c) Very high risk of dementia.

Figure 5.3: LIME global explanations for different risk sets.

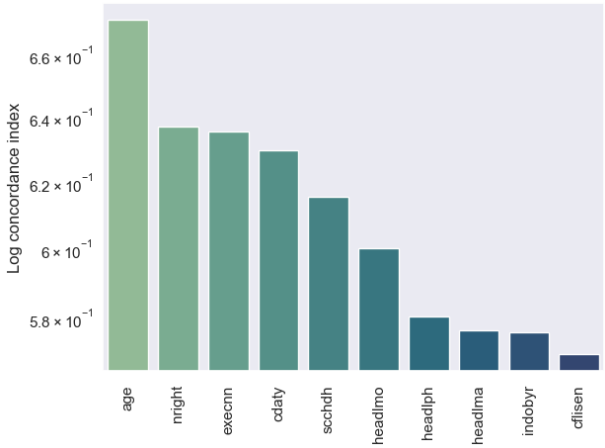
5.2.5 Cox Score Ranking

The Cox score ranking is a univariate approach to assess the association between predictor variables and survival outcomes in a Cox proportional hazards model. For each predictor variable, a separate univariate Cox model is fitted, and the resulting concordance index is used as a ranking criterion (section 2.2.2). This method is used to identify the most important predictors of the outcome and to estimate the magnitude and direction of their effect on the hazard function. Note that unlike all of the other interpretability methods that we used, this method does not investigate our black-box RSF model, but instead simply investigates each feature's individual predictive ability.

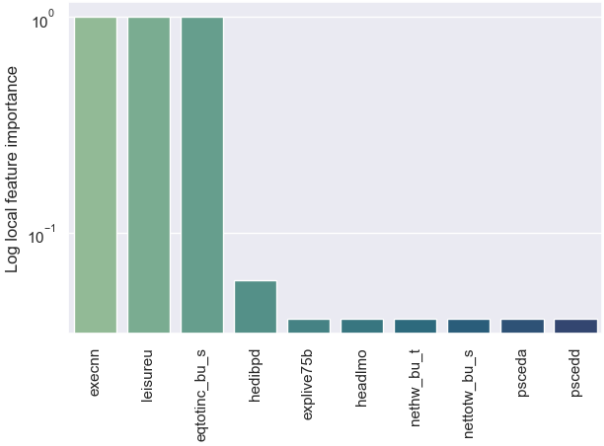
The application of the method to the dataset is depicted in Figure 5.4b. The results demonstrate that age is the most significant univariate predictor of dementia, consistent with prior research [101, 8]. Additionally, several cognitive variables that have previously been identified as important factors for predicting dementia are observed. Notably, the present investigation reveals which covariates, when considered in isolation, are reliable indicators of dementia onset.



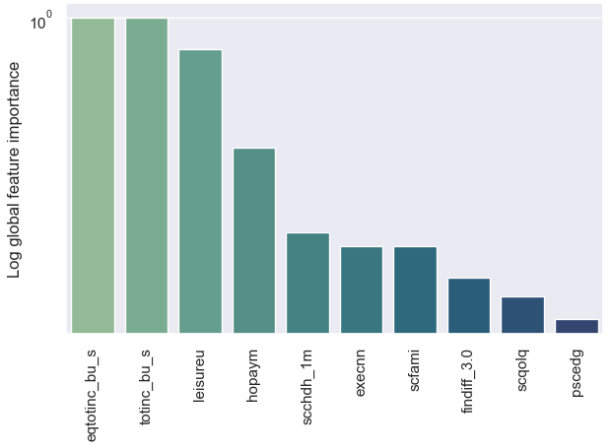
(a) Permutation feature importance.



(b) Cox model applied individually to each variable.



(c) Local counterfactual explanation, conducted on Person 6545 using 50 counterfactual explanations.



(d) Global counterfactual explanations, based on 50 different individuals.

Figure 5.4: Feature ranking based on different explanation techniques.

Chapter 6

Discussion

This chapter aims to discuss the key observations that were found in the analyses conducted in chapter 5, relating these findings to the three research questions posited in chapter 4, while also drawing comparisons to prior research in the field.

6.1 Observations

Observation 1: The most crucial determinants are cognitive and social factors

Upon reviewing Figure 5.4 and Figure 5.3, it becomes evident that the global explanation techniques predominantly emphasize cognitive or social factors. To validate this hypothesis, a simple approach involves summing the occurrences of each covariate across the global model explanations to identify the most frequently featured ones. The following table exhibits the top 5 covariates from the global explanations, wherein social variables are denoted in gray and cognitive variables are represented in red.

Variable	Occurrence
scchdh	6
headlph	5
cfdaty	5
headlmo	5
headlma	4

Observation 2: Interpretation methodology greatly influences explanations

Observing the three global explanation techniques in Figure 5.4, it is apparent that each method yields distinct explanations. Notably, none of the models concur on their top 3 feature contributions, with each model showcasing a unique set of top 3 features. Such discrepancies are likely attributable to underlying dissimilarities in the models. For instance, global counterfactual explanations were executed on a relatively small subset of the population, comprising only 50 individuals, a decision motivated by time constraints; this limited sample size may account for the considerable differences in the covariates chosen by this model compared to the other global models. Furthermore,

Cox Score Ranking operates as a univariate estimator, analyzing variables individually, which diverges from the other explanation methods, which explains its choice of a distinctive subset of covariates. Nevertheless, it is noteworthy that permutation feature importance and global LIME explanations share nearly all of the same covariates.

Observation 3: Interpretations should be conducted at a local level

Upon closely examining the local counterfactual explanations that were investigated in section 5.1.2, it becomes evident that they offer a robust mechanism for determining the factors that an individual can alter to influence their dementia outcome. Notably, the explanations exhibit variations between individuals, as demonstrated by LIME in section 5.1.1, and cannot be adequately captured by a global model, as the feature importance tends to differ between local and global explanations (Fig 5.3). As a result, we suggest that local explanations be employed whenever feasible and that healthcare models employed in practice integrate such explanations, as they are presently underutilized [46].

Our findings align with prior research in the field. For instance, Age UK's report [9] highlights that individuals diagnosed with amnestic mild cognitive impairment (MCI) have an elevated likelihood of developing dementia. Similarly, other studies, including Mayo Clinic [77], Alzheimer's Society [14], and Dementia Australia [33], have arrived at comparable results. Furthermore, there exists a correlation between social isolation and dementia, as demonstrated by studies such as Rafnsson et al. [93] utilizing the ELSA dataset, in addition to other studies [39, 106]. While it is widely recognized that age and gender are significant risk factors for the onset of dementia [79, 9], we specifically concentrate on examining modifiable factors rather than non-modifiable ones such as age.

Figure 6.1: Spearman rank correlation of top 5 selected covariates.

Figure 6.1 indicates that the variable *scchdh* does not demonstrate a significant correlation with the other top covariates. This fact may account for its high rank in the Cox Score Ranking (section 5.2.5), given that it does not fluctuate with other covariates but

serves as an effective standalone predictor. Additionally, we observe that the cognitive variables exhibit low levels of correlation among themselves, contrary to what we might have expected, though is a good indicator that the tests assess distinct cognitive functions.

Finally, we undertook an examination to determine the performance of a non-temporal RSF model trained solely on the five identified covariates, with no consideration given to competing risks. This model resulted in a C-Index of 0.773 and a BSS of 0.143, which is nearly comparable to the original Cox model we trained (section 4.3), the hazard ratios for which are shown in Figure A.3, also see Figure A.2.

6.2 Conclusion

In this study we addressed three core research questions (section 4.1) related to predictive modeling and interpretation of the risk of dementia. First, we proposed a machine learning-based approach for predicting the onset of dementia over time, building upon prior research, demonstrating that our approach outperforms existing methods in the ELSA dataset, which can be attributed to the inclusion of temporality and competing risk (**Question 1**). Second, we proposed several model-agnostic interpretation techniques that allow us to explain the predictions made by our model on an individual level. Our results showed that the identified factors contributing to an individual's risk of developing dementia are heterogeneous and can differ significantly from those that contribute to the risk at the population level (**Question 2**). Finally, we presented our findings on the factors that are most commonly associated with dementia onset, globally. We found that cognitive and social factors, such as memory and social participation, are the most crucial determinants of dementia onset (**Question 3**) when ignoring confounding variables that are non-modifiable, such as age.

To conclude, we propose one of the first machine learning based approaches for survival analysis in a dementia context on the ELSA dataset, building upon prior research in [101]. We demonstrate the effectiveness of our proposed model, while also conducting a thorough analysis and comparison of various models that can account for temporality and competing risk, and present a concrete baseline Cox model along with it. We built a model that achieves state-of-the-art performance for the dataset and task, beating that of prior research, and propose a much simpler model that achieves notable performance on the same task. Moreover, we presented, to our knowledge, the first interpretable machine learning techniques to ELSA in a dementia context, in which we discuss both local and global explanation techniques to facilitate informed individual decision-making and healthcare policy-making.

Chapter 7

Limitations & Future Work

7.1 Limitations

Care needs to be taken when making causal inferences using our results, as altering the method used to account for competing risks may have a significant impact on the interpretation of study findings and can substantially influence both absolute and relative risk estimates [36].

We also use survival analysis under the censoring assumption, which assumes that participants did not drop out of the study due to reasons related to the study, which in practise is unlikely. Despite the presence of measures within the dataset to prevent this, such as conducting follow-ups with relatives and documenting reasons for dropout, there may still be instances where a participant withdrew from the study due to factors such as a deteriorating financial situation, yet provided different reasons for their departure; this could potentially impact the results of our study. Moreover, the ELSA dataset specifically is limited to the UK elderly population, and may not generalise well to other countries or age groups. Further, the ascertainment of dementia and AD involved a combined algorithm based on a physician-made diagnosis and a higher score on the informant reports (IQCODE). Nonetheless, it should be noted that the diagnosis was reliant on self-reporting by the participants or their caregivers, which means that it is likely that only noticable and severe cases of dementia were reported, whereas milder cases were omitted from being reported [101].

7.2 Future Work (MInf Part 2)

Whilst it is true that random forest algorithms exhibit immunity to statistical assumptions [121], they are not optimally suited for handling unbalanced data, and are sensitive to hyperparameter choices, which can significantly affect model performance [121]. Moreover, other methods such as Support Vector Machines (SVM) have been shown to outperform random forests in clinical settings [63]; for the purpose of part 2 of the MInf project, it would be interesting to compare how the model performance of an SVM compares to the models discussed in this paper.

The present study had a central focus on interpretability for feature selection, with the potential application for healthcare policies. However, Part 2 of the MInf project aims to shift the focus to the model itself and enhancing its performance. One reason for not utilizing dimensionality reduction techniques in the current study was to maintain interpretability, as it is not possible to directly interpret the representations of the principal components for example. However, in Part 2 of the project, we intend to employ state-of-the-art latent space representation methods, such as autoencoders, similar to the approach described in a recent study [75], that employs a novel technique using a hybrid autoencoder approach. This approach will allow us to perform dimensionality reduction by representing features in a latent space, prior to feeding this data into our model; using such a technique is likely to give improved model performance, as compared to manually selecting a subset of features like we did in chapter 3, even though our selection criteria was informed based on prior research.

Our goal is to incorporate the latent space representation into a deep neural network, such as the end-to-end deep learning pipeline, DeepHit, proposed by Yoon [70]. DeepHit is a type of deep neural network that learns the survival time distribution directly, thereby avoiding the need to make assumptions about any underlying stochastic processes. As a result, the parameters and structure of the stochastic process utilized by the model are contingent upon the covariates present within the dataset under examination for survival analysis [70]. Moreover, this framework allows us to account for cause-specific competing risks, much like we did throughout this paper, thereby allowing us to make direct comparisons of model performance between this proposed framework and that of standard non-deep-learning based approaches.

Since we have shown in this paper that machine learning (ML) techniques, such as RSFs deliver improved predictive performance as compared to standard methods such as Cox proportional hazards model on our dataset and task, much like [101] shows; we will use ML techniques (RSF, SVM) as our baseline. Our study question involves using a more complex, deep learning based approach with the hypothesis that it will yield improved performance due to the aforementioned reasons. To further our research, we intend to employ similar interpretation techniques such as LIME, SHAP, and counterfactual explanations to comprehend the model's underlying decisions. However, a potential challenge could arise in interpreting these explanations, given that we will potentially be using a latent space representation of our covariates; a simple solution may be to simply feed an individual's raw data through this autoencoder to have it represented in the same latent space, and then evaluate the explanations. Moreover, we aim to use a recently proposed methodology, SurvSHAP which modifies SHAP interpretations, aiming to provide explanations for survival models specifically [11], there exists modifications of this technique to focus on time-dependent effects as well [65]. The use of this will allow for clearer interpretations, as it is specifically designed for survival models, whereas SHAP is designed for point predictions [11].

To the best of our knowledge, there has been no exploration of deep neural networks exploring time-to-event analysis in the context of dementia on the ELSA dataset. Moreover, deep time-to-event models such as that of DeepHit [70] have not explored using interpretability techniques as of yet; incorporating these novel aspects into our MInf Part 2 approach will bring originality to our research.

Bibliography

[1] 4.2. permutation feature importance.

[2] The actual difference between statistics and machine learning. https://towardsdatascience.com/the-actual-difference-between-statistics-and-machine-learning.

[3] Competing risk analysis aims to correctly estimate marginal probability of an event in the presence of competing events. learn how it's done by reading on. https://www.publichealth.columbia.edu/research/population-health-methods/competing-risk-analysis.

[4] Cost by sector in the uk.

[5] Early diagnosis is crucial for cancer patients. https://www.england.nhs.uk/blog/chris-harrison-3/.

[6] Introduction to survival analysis with scikit-survival.

[7] Meeting the challenges of an ageing population. https://www.england.nhs.uk/blog/catherine-thompson/.

[8] Alison Abbott. Dementia: A problem for our age. *Nature*, 475(7355):S2–S4, Jul 2011.

[9] Age UK. Cognitive decline and dementia evidence review. https://www.ageuk.org.uk/globalassets/age-uk/documents/reports-and-publications/reports-and-briefings/health--wellbeing/rb_oct14_cognitive_decline_and_dementia_evidence_review_age_uk.pdf, 2014.

[10] Gloria A Aguayo, Lu Zhang, Michel Vaillant, Moses Ngari, Magali Perquin, Valerie Moran, Laetitia Huiart, Rejko Krüger, Francisco Azuaje, Cyril Ferdynus, et al. Machine learning for predicting neurodegenerative diseases in the general older population: a cohort study. *BMC Medical Research Methodology*, 23(1):1–13, 2023.

[11] Abdallah Alabdallah, Sepideh Pashami, Thorsteinn Rögnvaldsson, and Mattias Ohlsson. Survshap: A proxy-based algorithm for explaining survival models with shap. In *2022 IEEE 9th International Conference on Data Science and Advanced Analytics (DSAA)*, pages 1–10. IEEE, 2022.

[12] Pamela Almeida-Meza, Andrew Steptoe, and Dorina Cadar. Markers of cognitive reserve and dementia incidence in the english longitudinal study of ageing. *The British Journal of Psychiatry*, 218(5):243–251, 2021.

[13] Naomi Altman and Martin Krzywinski. Ensemble methods: bagging and random forests. *Nature Methods*, 14(10):933–935, 2017.

[14] Alzheimer's Society. Mild cognitive impairment (mci). `https://www.alzheimers.org.uk/about-dementia/types-dementia/mild-cognitive-impairment-mci`.

[15] Per Kragh Andersen, Ronald B Geskus, Theo de Witte, and Hein Putter. Competing risks in epidemiology: possibilities and pitfalls. *International journal of epidemiology*, 41(3):861–870, 2012.

[16] Damaris Aschwanden, Angelina R Sutin, Martina Luchetti, Yannick Stephan, and Antonio Terracciano. Personality and dementia risk in england and australia. *GeroPsych*, 2020.

[17] Peter C Austin, Douglas S Lee, and Jason P Fine. Introduction to the analysis of survival data in the presence of competing risks. *Circulation*, 133(6):601–609, 2016.

[18] Melissa J Azur, Elizabeth A Stuart, Constantine Frangakis, and Philip J Leaf. Multiple imputation by chained equations: what is it and how does it work? *International journal of methods in psychiatric research*, 20(1):40–49, 2011.

[19] Christopher R Beam, Cody Kaneshiro, Jung Yun Jang, Chandra A Reynolds, Nancy L Pedersen, and Margaret Gatz. Differences between women and men in incidence rates of dementia and alzheimer's disease. *Journal of Alzheimer's Disease*, 64(4):1077–1083, 2018.

[20] Carine A Bellera, Gaëtan MacGrogan, Marc Debled, Christine Tunon De Lara, Véronique Brouste, and Simone Mathoulin-Pélissier. Variables with time-varying effects and the cox model: some statistical concepts illustrated with a prognostic factor study in breast cancer. *BMC medical research methodology*, 10(1):1–12, 2010.

[21] Geert Jan Biessels, Salka Staekenborg, Eric Brunner, Carol Brayne, and Philip Scheltens. Risk of dementia in diabetes mellitus: a systematic review. *The Lancet Neurology*, 5(1):64–74, 2006.

[22] Verónica Bolón-Canedo, Noelia Sánchez-Maroño, and Amparo Alonso-Betanzos. *Feature selection for high-dimensional data*. Springer, 2015.

[23] Wyllians Vendramini Borelli, Vanessa Bielefeldt Leotti, Matheus Zschornack Strelow, Márcia Lorena Fagundes Chaves, and Raphael Machado Castilhos. Preventable risk factors of dementia: Population attributable fractions in a brazilian population-based study. *The Lancet Regional Health-Americas*, 11:100256, 2022.

[24] Ørnulf Borgan. *Aalen–Johansen Estimator*. 07 2005.

[25] Imad Bou-Hamad, Denis Larocque, and Hatem Ben-Ameur. A review of survival trees. 2011.

[26] Leo Breiman. Random forests. *Machine learning*, 45:5–32, 2001.

[27] Leo Breiman and A Cutler. Random forests manual v4. *UC Berkeley*, 2003.

[28] Frances Bunn, Anne-Marie Burn, Claire Goodman, Greta Rait, Sam Norton, Louise Robinson, Johan Schoeman, and Carol Brayne. Comorbidity and dementia: a scoping review of the literature. *BMC medicine*, 12(1):1–15, 2014.

[29] Julian Buxton. Past and projected period and cohort life tables: 2020-based, uk, 1981 to 2070, Jan 2022.

[30] Mohammad A. Dabbah, Angus B. Reed, Adam T. C. Booth, Arrash Yassaee, Aleksa Despotovic, Benjamin Klasmer, Emily Binning, Mert Aral, David Plans, Davide Morelli, Alain B. Labrique, and Diwakar Mohan. Machine learning approach to dynamic risk modeling of mortality in covid-19: a uk biobank study. *Scientific Reports*, 11(1):16936, Aug 2021.

[31] Ralph B D'Agostino and Byung-Ho Nam. Evaluation of the performance of survival analysis models: discrimination and calibration measures. *Handbook of statistics*, 23:1–25, 2003.

[32] Frank R Datema, Ana Moya, Peter Krause, Thomas Bäck, Lars Willmes, Ton Langeveld, Robert J Baatenburg de Jong, and Henk M Blom. Novel head and neck cancer survival analysis approach: random survival forests versus cox proportional hazards regression. *Head & neck*, 34(1):50–58, 2012.

[33] Dementia Australia. Mild cognitive impairment. `https://www.dementia.org.au/about-dementia-and-memory-loss/about-dementia/memory-loss/mild-cognitive-impairment`.

[34] Anthony Devaux, Cécile Proust-Lima, and Robin Genuer. Random forests for time-fixed and time-dependent predictors: The dynforest r package. *arXiv preprint arXiv:2302.02670*, 2023.

[35] Jürgen Dieber and Sabrina Kirrane. Why model why? assessing the strengths and limitations of lime. *arXiv preprint arXiv:2012.00093*, 2020.

[36] Jessie K Edwards, Laura L Hester, Mugdha Gokhale, and Catherine R Lesko. Methodologic issues when estimating risks in pharmacoepidemiology. *Current epidemiology reports*, 3:285–296, 2016.

[37] ELSA Project. Genetic data access. `https://www.elsa-project.ac.uk/genetic-data-access`, accessed 15 March 2023.

[38] Charles Englebert, Thomas Quinn, and Isabelle Bichindaritz. Feature selection for survival analysis in bioinformatics. In *Proceedings of the Workshop on Advances in Bioinformatics and Artificial Intelligence: Bridging the Gap co-located with 26th International Joint Conference on Artificial Intelligence (IJCAI 2017)*, volume 1942, pages 30–35, 2017.

[39] Jianfeng Feng, Jing Li, Junhui Li, Ye Zhang, Ying Sun, Shuo Chen, and Barbara J. Sahakian. Associations of social isolation and loneliness with later dementia risk: A meta-analysis. *Neurology*, 99(2):e164–e172, 2022.

[40] Natan Feter, Samuel C Dumith, Emily C Smith, Larissa L da Cunha, Júlia Cassuriaga, Jayne S Leite, Ricardo Alt, Jeff S Coombes, and Airton J Rombaldi. Physical activity attenuates the risk for dementia associated with aging in older adults with mild cognitive impairment. findings from a population-based cohort study. *Journal of Psychiatric Research*, 141:1–8, 2021.

[41] Jason P Fine and Robert J Gray. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American statistical association*, 94(446):496–509, 1999.

[42] Robert Flynn. Survival analysis. *Journal of Clinical Nursing*, 21(19pt20):2789–2797, 2012.

[43] Mitchell Gail. A review and critique of some models used in competing risk analysis. *Biometrics*, pages 209–222, 1975.

[44] Mary Ganguli. Cancer and dementia: it's complicated. *Alzheimer disease and associated disorders*, 29(2):177, 2015.

[45] Ronald Geskus. Competing risks, analysis and interpretation.

[46] Arjan M Groen, Rik Kraan, Shahira F Amirkhan, Joost G Daams, and Mario Maas. A systematic review on the use of explainability in deep learning systems for computer aided diagnosis in radiology: Limited use of explainable ai? *European Journal of Radiology*, page 110592, 2022.

[47] Jiang Gui and Hongzhe Li. Penalized cox regression analysis in the high-dimensional and low-sample size settings, with applications to microarray gene expression data. *Bioinformatics*, 21(13):3001–3008, 2005.

[48] Hong Han, Xiaoling Guo, and Hua Yu. Variable selection using mean decrease accuracy and mean decrease gini based on random forest. In *2016 7th ieee international conference on software engineering and service science (icsess)*, pages 219–224. IEEE, 2016.

[49] Frank E Harrell Jr, Kerry L Lee, and Daniel B Mark. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine*, 15(4):361–387, 1996.

[50] Audinga-Dea Hazewinkel, Hans Gelderblom, and Marta Fiocco. Prediction models with survival data: a comparison between machine learning and the cox proportional hazards model. *medRxiv*, pages 2022–03, 2022.

[51] Yulei He, Alan M Zaslavsky, MB Landrum, DP Harrington, and P Catalano. Multiple imputation in a large-scale complex survey: a practical guide. *Statistical methods in medical research*, 19(6):653–670, 2010.

[52] Dennis M Heisey and Brent R Patterson. A review of methods to estimate cause-specific mortality in presence of competing risks. *The Journal of Wildlife Management*, 70(6):1544–1555, 2006.

[53] Daniel F Heitjan and Srabashi Basu. Distinguishing "missing at random" and "missing completely at random". *The American Statistician*, 50(3):207–213, 1996.

[54] Loukas Ilias and Dimitris Askounis. Explainable identification of dementia from transcripts using transformer networks. *IEEE Journal of Biomedical and Health Informatics*, 26(8):4153–4164, 2022.

[55] Hemant Ishwaran, Udaya B. Kogalur, Eugene H. Blackstone, and Michael S. Lauer. Random survival forests. *The Annals of Applied Statistics*, 2(3):841 – 860, 2008.

[56] Hemant Ishwaran, Udaya B Kogalur, Eugene H Blackstone, and Michael S Lauer. Random survival forests. 2008.

[57] Janus Christian Jakobsen, Christian Gluud, Jørn Wetterslev, and Per Winkel. When and how should multiple imputation be used for handling missing data in randomised clinical trials–a practical guide with flowcharts. *BMC medical research methodology*, 17(1):1–10, 2017.

[58] Tom Jansen, Gijs Geleijnse, Marissa Van Maaren, Mathijs P Hendriks, Annette Ten Teije, and Arturo Moncada-Torres. Machine learning explainability in breast cancer survival. In *Digital Personalized Health and Medicine*, pages 307–311. IOS Press, 2020.

[59] Iain M Johnstone and D Michael Titterington. Statistical challenges of high-dimensional data, 2009.

[60] Neesha Jothi, Wahidah Husain, et al. Data mining in healthcare–a review. *Procedia computer science*, 72:306–313, 2015.

[61] Marciane Kessler, Elaine Thumé, Shaun Scholes, Michael Marmot, Luiz Augusto Facchini, Bruno Pereira Nunes, Karla Pereira Machado, Mariangela Uhlmann Soares, and Cesar de Oliveira. Modifiable risk factors for 9-year mortality in older english and brazilian adults: The elsa and siga-bagé ageing cohorts. *Scientific Reports*, 10(1), 2020.

[62] Mizanur Khondoker, Snorri Bjorn Rafnsson, Stephen Morris, Martin Orrell, and Andrew Steptoe. Positive and negative experiences of social support and risk of dementia in later life: An investigation using the english longitudinal study of ageing. *Journal of Alzheimer's Disease*, 58(1):99–108, 2017.

[63] Helmanita Kibtia, Sarini Abdullah, and Alhadi Bustamam. Comparison of random forest and support vector machine for prediction of cognitive impairment in parkinson's disease. In *AIP Conference Proceedings*, volume 2296, page 020093. AIP Publishing LLC, 2020.

[64] David G Kleinbaum and Mitchel Klein. *Survival analysis a self-learning text*. Springer, 1996.

[65] Mateusz Krzyziński, Mikołaj Spytek, Hubert Baniecki, and Przemysław Biecek. Survshap (t): Time-dependent explanations of machine learning survival models. *Knowledge-Based Systems*, 262:110234, 2023.

[66] Alex Kulesza. Determinantal point processes for machine learning. *Foundations and Trends® in Machine Learning*, 5(2-3):123–286, 2012.

[67] Dhananjay Kumar and Bengt Klefsjö. Proportional hazards model: a review. *Reliability Engineering & System Safety*, 44(2):177–188, 1994.

[68] Chih-Cheng Lai, Chung-Han Ho, Chin-Ming Chen, Shyh-Ren Chiang, Chien-Ming Chao, Wei-Lun Liu, Yu-Chieh Lin, Jhi-Joung Wang, and Kuo-Chen Cheng. Long-term risk of dementia after acute respiratory failure requiring intensive care unit admission. *Plos one*, 12(7):e0180914, 2017.

[69] Bryan Lau, Stephen R Cole, and Stephen J Gange. Competing risk regression models for epidemiologic data. *American journal of epidemiology*, 170(2):244–256, 2009.

[70] Changhee Lee, William Zame, Jinsung Yoon, and Mihaela Van Der Schaar. Deephit: A deep learning approach to survival analysis with competing risks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

[71] Roderick JA Little and Donald B Rubin. *Statistical analysis with missing data*, volume 793. John Wiley & Sons, 2019.

[72] Yingxin Liu, Shiyu Zhou, Hongxia Wei, and Shengli An. A comparative study of forest methods for time-to-event data: variable selection and predictive performance. *BMC Medical Research Methodology*, 21:1–16, 2021.

[73] Gaur Loveleen, Bhandari Mohan, Bhadwal Singh Shikhar, Jhanjhi Nz, Mohammad Shorfuzzaman, and Mehedi Masud. Explanation-driven hci model to examine the mini-mental state for alzheimer's disease. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 2022.

[74] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.

[75] Mahdi Mahdavi, Hadi Choubdar, Zahra Rostami, Behnaz Niroomand, Alexandra T Levine, Alireza Fatemi, Ehsan Bolhasani, Abdol-Hossein Vahabie, Stephen G Lomber, and Yaser Merrikhi. Hybrid feature engineering of medical data via variational autoencoders with triplet loss: a covid-19 prognosis study. *Scientific Reports*, 13(1):1–15, 2023.

[76] Xin Man and Ernest P Chan. The best way to select features? comparing mda, lime, and shap. *The Journal of Financial Data Science*, 3(1):127–139, 2021.

[77] Mayo Clinic. Mild cognitive impairment - symptoms and causes. `https://www.mayoclinic.org/diseases-conditions/mild-cognitive-impairment/symptoms-causes/syc-20354578`, 1 2023.

[78] Cyprien Mbogning and Philippe Broët. Bagging survival tree procedure for variable selection and prediction in the presence of nonsusceptible patients. *BMC bioinformatics*, 17:1–21, 2016.

[79] Catriona D McCullagh, David Craig, Stephen P McIlroy, and A Peter Passmore. Risk factors for dementia. *Advances in psychiatric treatment*, 7(1):24–31, 2001.

[80] Anne McMunn, James Nazroo, and Elizabeth Breeze. Inequalities in health at older ages: a longitudinal investigation of the onset of illness and survival effects in england. *Age and ageing*, 38(2):181–187, 2009.

[81] Maritza Mera-Gaona, Ursula Neumann, Rubiel Vargas-Canas, and Diego M López. Evaluating the impact of multivariate imputation by mice in feature selection. *Plos one*, 16(7):e0254720, 2021.

[82] Michelle M Mielke. Sex and gender differences in alzheimer's disease dementia. *The Psychiatric times*, 35(11):14, 2018.

[83] John Mingers. An empirical comparison of pruning methods for decision tree induction. *Machine learning*, 4:227–243, 1989.

[84] Christoph Molnar. Interpretable machine learning. `https://christophm.github.io/interpretable-ml-book/interpretability.html`, Oct 2022.

[85] Ramaravind K Mothilal, Amit Sharma, and Chenhao Tan. Explaining machine learning classifiers through diverse counterfactual explanations. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 607–617, 2020.

[86] Henry Musto, Daniel Stamate, Ida Pu, and Daniel Stahl. A machine learning approach for predicting deterioration in alzheimer's disease. In *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 1443–1448. IEEE, 2021.

[87] Rasmus Oestergaard Nielsen, Michael Lejbach Bertelsen, Daniel Ramskov, Merete Møller, Adam Hulme, Daniel Theisen, Caroline F Finch, Lauren Victoria Fortington, Mohammad Ali Mansournia, and Erik Thorlund Parner. Time-to-event analysis for sports injury research part 2: time-varying outcomes. *British journal of sports medicine*, 53(1):70–78, 2019.

[88] Rhys Owen-Williams. Leading causes of death, uk: 2001 to 2018, Mar 2020.

[89] Ruth Peters, Ruth Poulter, James Warner, Nigel Beckett, Lisa Burch, and Chris Bulpitt. Smoking, dementia and cognitive decline in the elderly, a systematic review. *BMC geriatrics*, 8:1–7, 2008.

[90] Tra My Pham, Irene Petersen, Kate Walters, Rosalind Raine, Jill Manthorpe, Naaheed Mukadam, and Claudia Cooper. Trends in dementia diagnosis rates in uk ethnic groups: analysis of uk primary care data. *Clinical epidemiology*, pages 949–960, 2018.

[91] Vincent Planche, Vincent Bouteloup, Isabelle Pellegrin, Jean-Francois Mangin, Bruno Dubois, Pierre-Jean Ousset, Florence Pasquier, Frederic Blanc, Claire

Paquet, Olivier Hanon, et al. Validity and performance of blood biomarkers for alzheimer disease to predict dementia risk in a large clinic-based cohort. *Neurology*, 100(5):e473–e484, 2023.

[92] Hein Putter, Marta Fiocco, and Ronald B Geskus. Tutorial in biostatistics: competing risks and multi-state models. *Statistics in medicine*, 26(11):2389–2430, 2007.

[93] Snorri Bjorn Rafnsson, Martin Orrell, Eleonora d'Orsi, Eef Hogervorst, and Andrew Steptoe. Loneliness, social integration, and incident dementia over 6 years: Prospective findings from the english longitudinal study of ageing. *The Journals of Gerontology: Series B*, 75(1):114–124, 2020.

[94] Jürgen Rehm, Omer SM Hasan, Sandra E Black, Kevin D Shield, and Michaël Schwarzinger. Alcohol use and dementia: a systematic scoping review. *Alzheimer's research & therapy*, 11:1–11, 2019.

[95] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144, 2016.

[96] Judi Scheffer. Dealing with missing data. 2002.

[97] Matthias Schmid, Marvin N Wright, and Andreas Ziegler. On the use of harrell's c for clinical risk prediction via random survival forests. *Expert Systems with Applications*, 63:450–459, 2016.

[98] Juned Siddique, Peter J de Chavez, George Howe, Gracelyn Cruden, and C Hendricks Brown. Limitations in using multiple imputation to harmonize individual participant data for meta-analysis. *Prevention Science*, 19:95–108, 2018.

[99] Noah Simon, Jerome Friedman, Trevor Hastie, and Rob Tibshirani. Regularization paths for cox's proportional hazards model via coordinate descent. *Journal of statistical software*, 39(5):1, 2011.

[100] Hardeep Singh, Gordon D Schiff, Mark L Graber, Igho Onakpoya, and Matthew J Thompson. The global burden of diagnostic errors in primary care, Jun 2017.

[101] Daniel Stamate, Henry Musto, Olesya Ajnakina, and Daniel Stahl. Predicting risk of dementia with survival machine learning and statistical methods: Results on the english longitudinal study of ageing cohort. In Ilias Maglogiannis, Lazaros Iliadis, John Macintyre, and Paulo Cortez, editors, *Artificial Intelligence Applications and Innovations. AIAI 2022 IFIP WG 12.5 International Workshops*, pages 436–447, Cham, 2022. Springer International Publishing.

[102] Gregor Stiglic, Primoz Kocbek, Nino Fijacko, Marinka Zitnik, Katrien Verbert, and Leona Cilar. Interpretability of machine learning-based prediction models in healthcare. *WIREs Data Mining and Knowledge Discovery*, 10(5), 2020.

[103] Elizabeth A Stuart, Melissa Azur, Constantine Frangakis, and Philip Leaf. Multiple imputation with large data sets: a case study of the children's mental health initiative. *American journal of epidemiology*, 169(9):1133–1139, 2009.

[104] AFG Taktak, A Eleuteri, SP Lake, and AC Fisher. Evaluation of prognostic models: discrimination and calibration performance. In *Proceedings of the 3rd International Conference on Computational Intelligence in Medicine and Healthcare*, 2007.

[105] The PySurvival Team. Brier score and integrated brier score.

[106] University of Cambridge. Social isolation may impact brain volume in regions linked to higher risk of dementia. `https://www.cam.ac.uk/research/news/social-isolation-may-impact-brain-volume-in-regions-linked-to-higher-risk`

[107] Stef Van Buuren. *Flexible imputation of missing data*. CRC press, 2018.

[108] Jasper van der Waa, Marcel Robeer, Jurriaan van Diggelen, Matthieu Brinkhuis, and Mark Neerincx. Contrastive explanations with local foil trees, 2018.

[109] Ping Wang, Yan Li, and Chandan K Reddy. Machine learning for survival analysis: A survey. *ACM Computing Surveys (CSUR)*, 51(6):1–36, 2019.

[110] Brandon Weathers. Comparison of survival curves between cox proportional hazards, random forests, and conditional inference forests in survival analysis. 2017.

[111] Andreas P Weigel, Mark A Liniger, and Christof Appenzeller. The discrete brier and ranked probability skill scores. *Monthly Weather Review*, 135(1):118–124, 2007.

[112] Sue Westwood. Dementia, women and sexuality: How the intersection of ageing, gender and sexuality magnify dementia concerns among lesbian and bisexual women. *Dementia*, 15(6):1494–1514, 2016.

[113] Janine Witte, Ronja Foraita, and Vanessa Didelez. Multiple imputation and test-wise deletion for causal discovery with incomplete cohort data. *Statistics in Medicine*, 41(23):4716–4743, 2022.

[114] Marcel Wolbers, Michael T Koller, Jacqueline CM Witteman, and Ewout W Steyerberg. Prognostic models with competing risks: methods and application to coronary risk prediction. *Epidemiology*, pages 555–561, 2009.

[115] Dylan Wood, Dimitris Evangelopoulos, Sean Beevers, Nutthida Kitwiroon, and Klea Katsouyanni. Exposure to ambient air pollution and the incidence of dementia in the elderly of england: The elsa cohort. *International Journal of Environmental Research and Public Health*, 19(23):15889, 2022.

[116] Songhao Wu. Multi-collinearity in regression. `https://towardsdatascience.com/multi-collinearity-in-regression-fe7a2c1467ea`, Jun 2021.

[117] Jesper N Wulff and Linda Ejlskov Jeppesen. Multiple imputation by chained equations in praxis: guidelines and review. *Electronic Journal of Business Research Methods*, 15(1):41–56, 2017.

[118] Hui Zhang, Ji-Xia Ren, Jin-Xiang Ma, and Lan Ding. Development of an in silico prediction model for chemical-induced urinary tract toxicity by using naïve bayes classifier. *Molecular Diversity*, 23(2):381–392, 2018.

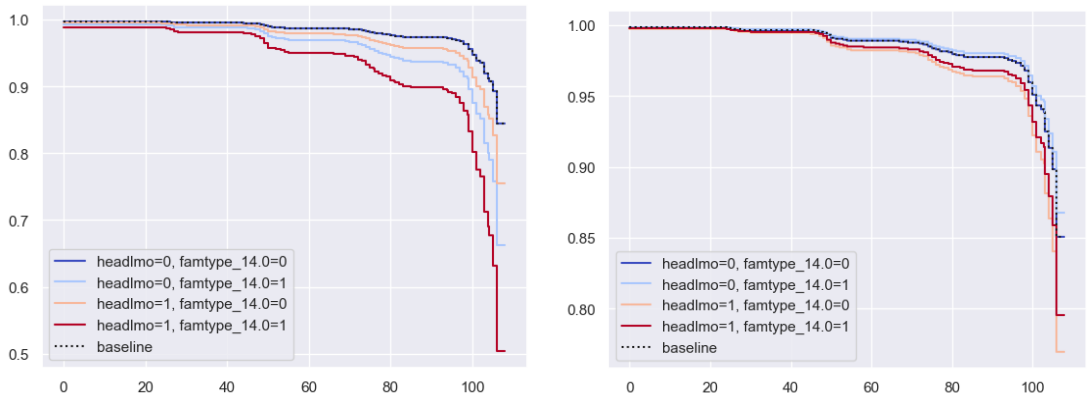
[119] Yanxin Zhang and Ye Ma. Application of supervised machine learning algorithms in the classification of sagittal gait patterns of cerebral palsy children with spastic diplegia. `https://www.sciencedirect.com/science/article/pii/S0010482519300095`, Jan 2019.

[120] Xun Zhao, Yanhong Wu, Dik Lun Lee, and Weiwei Cui. iforest: Interpreting random forests via visual analytics. *IEEE transactions on visualization and computer graphics*, 25(1):407–416, 2018.

[121] Tongtian Zhu. Analysis on the applicability of the random forest. In *Journal of Physics: Conference Series*, volume 1607, page 012123. IOP Publishing, 2020.

Appendix A

Supplementary Figures



(a) Without competing risk.

(b) Accounting for competing risk of death.

Figure A.1: The partial effects on outcome plot, showing a comparison of the model's baseline curve with variations in covariate(s), allowing for variable comparison while keeping other variables constant. *The exact definitions of the variables can be found in table C.1.*

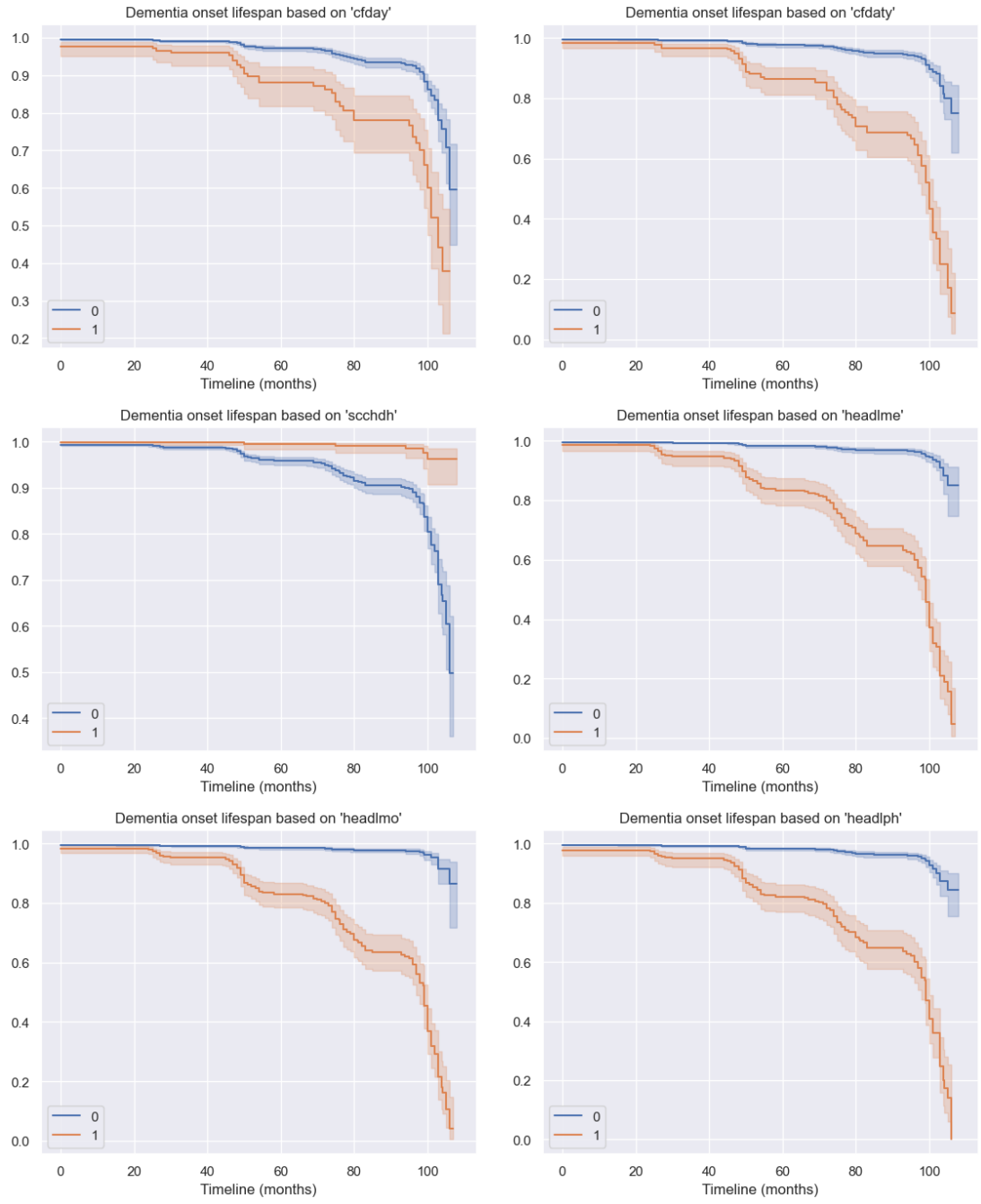


Figure A.2: Univariate Kaplan Miere plot stratified based on the 5 key covariates identified.

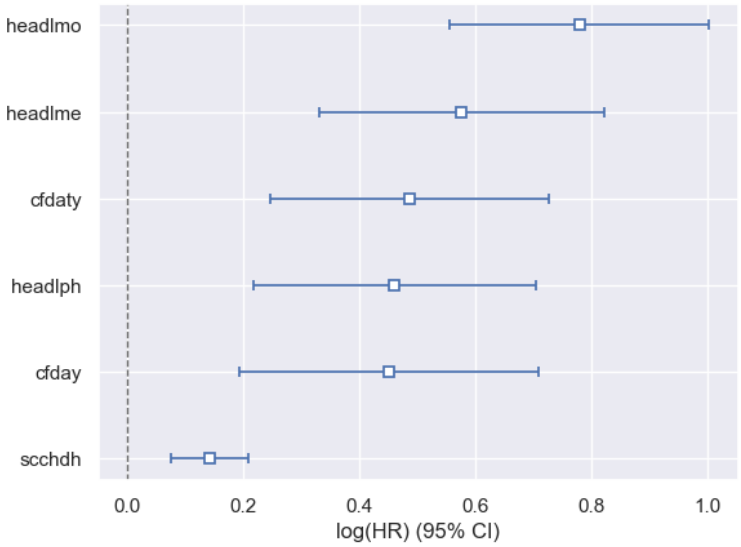


Figure A.3: The hazard ratios attributed to the basic penalized Cox PH model trained solely on the 5 covariates identified.

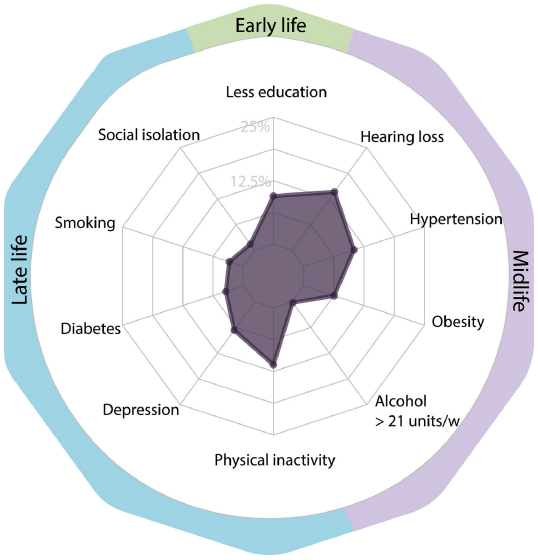


Figure A.4: The Population Attributable Fraction for potentially modifiable risk factors of dementia in Brazil [23].

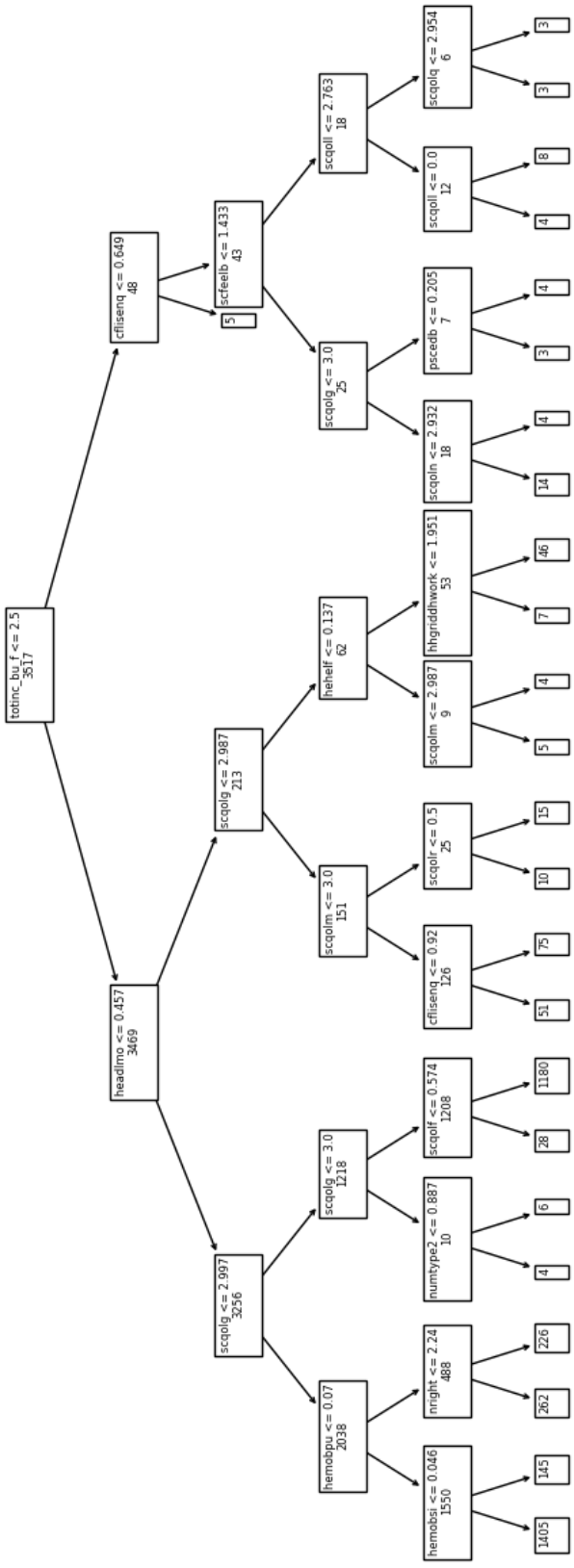


Figure A.5: Pruned decision tree visualised. Terminal nodes indicate the 'estimated rate' - the higher the value, the higher the risk of dementia onset.

Appendix B

Model Replication

Owing to the nature of the data, which requires academic use agreements through UK Data Service, it is not possible to make it publicly accessible. Nevertheless, there exists a possibility of coming to an agreement to share the dataset, subject to further discussion with my supervisor.

Kindly take note that certain modifications were implemented on the underlying packages to enable the acquisition of the figures and data presented. The environment files containing these modifications are available upon request and can be shared accordingly.

Appendix C

Variable Definitions

Variable	Variable label
edend	Age Finished Cont. Full Time Education, Merged From Current And Previous Waves
edqual	Educational Qualification - Info Merged From Current And Previous Waves
schleave	Type Of School Leaver - Info Merged From Current And Previous Waves
nonwhite	Ethic Origin (White/Non-White)
sex	Male, Female
died_p	Respondent Had A Spouse Who Died Since Last Nterview
couple	Is In A Couple (Not Including If Partner In An Institution)
ngrandch	Number Of Grandchildren Or Great-Grandchildren Inside Or Outside The Household
ngrandchinhh	Number Of Grandchildren Living In The Household
nsibs	Number Of Living Siblings
age	Age: Copy Of Indage/Dhager
hhgriddhwork	Whether Was In Paid Employment Last Week - Taken From Hhgrid
hhgriddhwork_p	Whether Partner Was In Paid Employment Last Week - Taken From Hhgrid
wpactive	Current Working Status
hours	Hours Of Work Main Job (Employed Or Self Employed)
everwork	Ever Worked
pp_mem	Currently A Member Of Private Pension Scheme (Cont/Rec/Ret Rights)
pp_rec	Currently Receiving Income From A Private Pension
pp_nrec	Number Of Private Pensions From Which Receiving Income
inhergt100	Chances Of Inheritance Greater Than £100K (Inhergt10 + Exinhe)
expliveb	Chances Of Living To Age X (Banded Version)
explive75b	Chances Of Living To Age 75 (Banded Version)

srh3_hse	Self-Reported Health, Hse Form, 3-Way
llsill	Has (Limiting) Long-Standing Illness (Heill/Helim)
hlimwrk	Has Health Problem That Limits Kind Or Amount Of Work (Helwk/Hetemp)
hemobwa	Mobility: Difficulty Walking 100 Yards
hemobsi	Mobility: Difficulty Sitting 2 Hours
hemobst	Mobility: Difficulty Stooping, Kneeling Or Crouching
hemobre	Mobility: Difficulty Reaching Or Extending Arms Above Shoulder Level
hemobli	Mobility: Difficulty Lifting Or Carrying Weights Over 10 Pounds
headlea	Adl: Difficulty Eating, Such As Cutting Up Food
headlma	Iadl: Difficulty Using Map To Figure Out How To Get Around Strange Place
headlsh	Iadl: Difficulty Shopping For Groceries
headlph	Iadl: Difficulty Making Telephone Calls
headlme	Iadl: Difficulty Taking Medications
headlmo	Iadl: Difficulty Managing Money, Eg Paying Bills,Keeping Track Expenses
smoker	Whether Current Smoker
nright	Number Of Numeracy Questions Correct
numtype2	Numeracy Index - 2 Way Split
execnn	Index Of Executive (Non-Numeracy) Function (0-20)
nrooms	Number Of Rooms In House - Merged Information From Current And Previous Waves
hopaym	How Much Paid For This Property? Info Merged From Current And Previous Waves
leisureu	Money Spent On Leisure Wkly Upper Bound (Holeis/Holeisb)
foodinu	Money Spent On Food In Wkly Upper Bound (Hofood/Hofoofb)
casp_19	Self-Derived
cesd	Self-Derived
cfani	Number Of Animals Mentioned
cfaniq	Refers To Cfani (1 If Cfani = 1, 0.75 If Cfani = 2, 0.5 If Cfani = 3, 0.25 If Cfani = 4, 0 If Cfani = 5)
cfdatd	Whether Correct Day Of Month Given
cfdatm	Whether Correct Month Given
cfdaty	Whether Correct Year Given
cfday	Whether Correct Day Given
cflisd	Number Of Words Recalled After Delay
cflisdq	Refers To Cflisd (1 If Cflisd = 1, 0.75 If Cflisd = 2, 0.5 If Cflisd = 3, 0.25 If Cflisd = 4, 0 If Cflisd = 5)
cflisen	Number Of Words Recalled Immediately

cflisenq	Refers To Cflisen (1 If Cflisen = 1, 0.66 If Cflisen = 2, 0.33 If Cflisen = 3, 0 If Cflisen = 4)
cfmem	Whether Prompt Given For Prospective Memory Test (Remembering To Write Initials)
depression	Self-Derived
eqtotinc_bu_f	Bu Equivalised Total Income - Validation Flag
eqtotinc_bu_s	Bu Equivalised Total Income - Summary Var
eqtotinc_bu_t	Bu Equivalised Total Income - Imputation Flag
headlba	Difficulty Bathing Or Showering
headlbe	Difficulty Getting In And Out Of Bed
headldr	Difficulty Dressing, Including Putting On Shoes And Socks
headlhg	Difficulty Doing Work Around The House Or Garden
headlpr	Difficulty Preparing A Hot Meal
headlwa	Difficulty Walking Across A Room
headlwc	Difficulty Using The Toilet Including Getting Up Or Down
hedibar	Arthritis Dx
hedibas	Asthma Dx
hedibca	Cancer Dx
hediblu	Lung Disease Dx
hedibos	Osteoporosis
hedibpd	Parkinson'S Dx
hedibps	Psychiatric Condition
hediman	Angina Dx
hedimar	Abnormal Heart Rhythm
hedimbp	High Bp Dx
hedimdi	Diabetes Or High Blood Sugar
hedimhf	Congestive Heart Failure
hedimmi	Heart Attack
hedimst	Stroke Dx
heeye	Self-Reported Eyesight (While Using Lenses If Appropriate)
hefla	Whether Fallen Down Since Last Interniew
hefrac	Whether Has Fractured Hip
hehear	Self-Reported Hearing (While Using Hearing Aid If Appropriate)
hehelf	Self-Reported General Health
heji	Whether Had Joint Replacement
hemobch	Difficulty Getting Up From Chair After Sitting Long Periods
hemobcl	Difficulty Climbing One Flight Of Stairs Without Resting
hemobcs	Difficulty Climbing Several Flights Of Stairs Without Resting
hemobpi	Difficulty Picking Up A 5P Coin From A Table
hemobpu	Difficulty Pulling Or Pushing Large Objects
indager	Definitive Age Variable Collapsed At 90 Plus.
indobyr	Year Of Birth Collapsed At 90 Plus

loneliness	Self-Derived
mmpain	Timed Walk: Whether Had Pain Whilst Walking
nethw_bu_f	Bu Total Net Primary Housing Wealth - Validation Flag
nethw_bu_s	Bu Total Net Primary Housing Wealth - Summary Var
nethw_bu_t	Bu Total Net Primary Housing Wealth - Imputation Flag
nettotnhw_bu_f	Bu Total Net Non-Housing Wealth - Validation Flag
nettotnhw_bu_s	Bu Total Net Non-Housing Wealth - Summary Var
nettotnhw_bu_t	Bu Total Net Non-Housing Wealth - Imputation Flag
nettotw_bu_f	Bu Total Net (Non-Pension) Wealth - Validation Flag
nettotw_bu_s	Bu Total Net (Non-Pension) Wealth - Summary Var
nettotw_bu_t	Bu Total Net (Non-Pension) Wealth - Imputation Flag
no_contact_in_1m_chd	Self-Derived
no_contact_in_1m_fam	Self-Derived
no_contact_in_1m_frd	Self-Derived
no_partner	Self-Derived
psceda	Whether Felt Depressed Much Of The Time During The Past Week
pscedb	Whether Felt Everything They Did During The Past Week Was An Effort
pscedc	Whether Felt Their Sleep Was Restless During The Past Week
pscedd	Whether Was Happy Much Of The Time During The Past Week /R
pscede	Whether Felt Lonely Much Of The Time During The Past Week
pscedf	Whether Enjoyed Life Much Of The Time During The Past Week /R
pscedg	Whether Felt Sad Much Of The Time Duing The Past Week
pscedh	Whether Could Not Get Going Much Of The Time During The Past Week
rockwood_frailty	Self-Derived
scchdg	How Often The Respondent Meets Up With Their Children On Average
scchdg_1m	Self-Derived
scchdh	How Often The Respondent Speaks On The Phone To Their Children
scchdh_1m	Self-Derived
scchdi	How Often The Respondent Writes To Or Emails Their Children
scchdi_1m	Self-Derived
scfamg	How Often The Respondent Meets Up With Other Relatives
scfamg_1m	Self-Derived
scfamh	How Often The Respondent Speaks With Other Relatives On The Phone

scfamh_1m	Self-Derived
scfami	How Often The Respondent Writes To Or Emails Other Relatives
scfami_1m	Self-Derived
scfeela	How Often Respondent Feels They Lack Companionship
scfeelb	How Often Respondent Feels Left Out
scfeelc	How Often Respondent Feels Isolated From Others
scfrdg	How Often The Respondent Meets Up With Their Friends
scfrdg_1m	Self-Derived
scfrdh	How Often The Respondent Speaks With Their Friends On The Phone
scfrdh_1m	Self-Derived
scfrdi	How Often The Respondent Writes To Or Emails Their Friends
scfrdi_1m	Self-Derived
scorg1	Organisational Membership: Political Party, Trade Union Or Environmental Group
scorg2	Organisational Membership: Tenants Or Resident Group Or Neighbourhood Watch
scorg3	Organisational Membership: Member Of A Church Or Other Religious Group
scorg4	Organisational Membership: Member Of A Charitable Association
scorg5	Organisational Membership: An Education, Arts Or Music Group Or Evening Class
scorg6	Organisational Membership: Member Of A Social Club
scorg7	Organisational Membership: Member Of A Sports Clubs, Gym, Or Exercise Class
scorg8	Organisational Membership: Member Of Any Other Organisations, Clubs Or Societies
scorg9	Organisational Membership: Not A Member Of Any Organisation, Club Or Society
scqola	Casp19 Scale: How Often Feels Age Prevents Them From Doing Things They Like
scqolb	Casp19 Scale: How Often Feels What Happens To Them Is Out Of Their Control
scqolc	Casp19 Scale: How Often Feels Free To Plan For The Future
scqold	Casp19 Scale: How Often Feels Left Out Of Things
scqole	Casp19 Scale: How Often Can Do The Things They Want To Do
scqolf	Casp19 Scale: How Often Family Responsibilities Prevents Them From Doing Things
scqolg	Casp19 Scale: How Often Feels They Can Please Themselves What They Do

scqolh	Casp19 Scale: How Often Feels Their Health Stops Them Doing What They Want To Do
scqoli	Casp19 Scale: How Often Shortage Of Money Stops Them Doing Things
scqolj	Casp19 Scale: How Often Look Forward To Each Day
scqolk	Casp19 Scale: How Often Feels That Their Life Has Meaning
scqoll	Casp19 Scale: How Often Enjoys The Things They Do
scqolm	Casp19 Scale: How Often Enjoys Being In The Company Of Others
scqoln	Casp19 Scale: How Often Looks Back On Their Life With A Sense Of Happiness
scqolo	Casp19 Scale: How Often Feels Full Of Energy These Days
scqolp	Casp19 Scale: How Often Chooses To Do Things They Have Never Done Before
scqolq	Casp19 Scale: How Often Feels Satisfied With The Way Their Life Has Turned Out
scqolr	Casp19 Scale: How Often Feels That Life Is Full Of Opportunities
scqols	Casp19 Scale: How Often Feels The Future Looks Good To Them
totinc_bu_f	Bu Total Net Income - Validation Flag
totinc_bu_s	Bu Total Net Income - Summary Var
totinc_bu_t	Bu Total Net Income - Imputation Flag
wxwgt	Cross-Sectional Weight
bmivg6	Bmi
famtype_1.0	Household Type (Single)
famtype_2.0	Household Type (Lone Plus Dependent Children)
famtype_3.0	Household Type (Lone Plus Non-Dep Children Aged<30)
famtype_4.0	Household Type (Lone Plus Non-Dep Children Aged>=30)
famtype_5.0	Household Type (Lone Plus Both)
famtype_6.0	Household Type (Couple)
famtype_7.0	Household Type (Couple Plus Dependent Children)
famtype_8.0	Household Type (Couple Plus Non-Dep Children Aged <30)
famtype_9.0	Household Type (Couple Plus Non-Dep Children Aged >=30)
famtype_10.0	Household Type (Couple Plus Both)
famtype_11.0	Household Type (Extended Family)
famtype_12.0	Household Type (Extended Family Plus Children)
famtype_13.0	Household Type (Other Multiple Tax Unit)
famtype_14.0	Household Type (Other Multiple Tax Unit Plus Children)
tenure_-7.0	Tenure (Institution)
tenure_-1.0	Tenure (Not Applicable)
tenure_1.0	Tenure (Own It Outright)

tenure_2.0	Tenure (Buying It With The Help Of A Mortgage Or Loan)
tenure_3.0	Tenure (Pay Part Rent And Part Mortgage (Shared Ownership)
tenure_4.0	Tenure (Rent It)
tenure_5.0	Tenure (Live Here Rent Free (Including Rent Free In Relative/Friends))
socrent_-1.0	Social Renter (Not Applicable)
socrent_0.0	Social Renter (Private Renter)
socrent_1.0	Social Renter (Social Renter)
marstat_1.0	Marital Status (Married)
marstat_2.0	Marital Status (Cohabiting)
marstat_3.0	Marital Status (Single, Never Married)
marstat_4.0	Marital Status (Widowed)
marstat_5.0	Marital Status (Divorced)
marstat_6.0	Marital Status (Separated)
ecpos_2.0	Economic Activity (Employee)
ecpos_3.0	Economic Activity (Self-Employed)
ecpos_4.0	Economic Activity (Seeking Work)
ecpos_6.0	Economic Activity (Sick And Not Seeking)
ecpos_7.0	Economic Activity (Retired)
ecpos_8.0	Economic Activity (Unoccupied)
findiff_-1.0	Financial Stability (Inapplicable)
findiff_1.0	Financial Stability (Manage Very Well)
findiff_2.0	Financial Stability (Manage Quite Well)
findiff_3.0	Financial Stability (Get By Alright)
findiff_4.0	Financial Stability (Don'T Manage Very Well)
findiff_5.0	Financial Stability (Have Some Financial Difficulties)
findiff_6.0	Financial Stability (Have Severe Financial Difficulties)
smokerstat_-1.0	Smoker Status (Refused/Not Asked/Don'T Know)
smokerstat_0.0	Smoker Status (Never Smoked)
smokerstat_1.0	Smoker Status (Ex Smoker - Occasional)
smokerstat_2.0	Smoker Status (Ex Smoker - Regular)
smokerstat_3.0	Smoker Status (Ex Smoker - Dk Freq)
smokerstat_4.0	Smoker Status (Current Smoker)
finstat_C1CM	Final Status Of Respondent (C1Cm)
finstat_C3CM	Final Status Of Respondent (C3Cm)
finstat_C4CM	Final Status Of Respondent (C4Cm)

Table C.1: Table showing the list of variables/features used in the models. This should serve the purpose of being able to cross-reference the interpretations of features mentioned in the paper. There are 226 features listed in the table above, excluding the model outputs (*survival-months, dementia*), which are not shown in the table above.

Appendix D

Participants' Information Sheet

ELSA is funded by the National Institute on Aging (R01AG017644), and by UK Government Departments coordinated by the National Institute for Health and Care Research (NIHR).

Banks, J., Batty, G. David, Breedvelt, J., Coughlin, K., Crawford, R., Marmot, M., Nazroo, J., Oldfield, Z., Steel, N., Steptoe, A., Wood, M., Zaninotto, P. (2021). English Longitudinal Study of Ageing: Waves 0-9, 1998-2019. [data collection]. 37th Edition. UK Data Service. SN: 5050, DOI: 10.5255/UKDA-SN-5050-24