Comparative Analysis of Discrimination and Calibration Accuracy of Discrete Survival, Random Forests, and Neural Networks in Health-Related Survival Prediction Models

Audrey Tshepho Ramachela

(Student No: 14014807)

Research project for the Master of Science in Statistics

in the Department of Mathematical and Computational Sciences

Faculty of Science, Engineering and Agriculture

University of Venda, Thohoyandou, Limpopo

South Africa

Supervisor: **Dr. Alphonce Bere**

Co-Supervisor: **Dr. Tshilidzi B Mulaudzi**

Co-Supervisor: **Ms. Lactatia Motsuku**

Date Submitted: July 11, 2025

© University of Venda

University of Venda
Creating Future Leaders

Abstract

Prediction models for survival analysis are commonly used in biomedical sciences to understand the onset of certain diseases. Traditional statistical models have been employed for the previous years, however, their limitations and inability to handle big data sets has made a way for the introduction of machine learning methods which gained recognition due to their ability to learn complex algorithms. However, existing literature indicates that the predictive accuracy of machine learning and statistical models for survival analysis varies significantly across different data sets. This variability underscores the need for further research utilizing data sets with diverse characteristics. Such research is essential to develop generalizable insights into the conditions under which each method performs best. In this research project, we compared the predictive performance of traditional statistical method and machine learning algorithms in discrete survival analysis. The machine learning methods include discrete-time survival trees, discrete-time random survival forests, and discrete-time neural networks. The study uses calibration (measured by the prediction error curves) to assess model fit and discrimination (measured by the Concordance index and area under curve) to evaluate predictive accuracy. These methods were applied to data sets: Breast cancer, age at first alcohol intake and CRASH-2. The discrete-time neural network had the best prediction performance as compared to the rest of the models for survival of breast cancer. The discrete-time random forest with hellinger distance had the overall prediction performance on the age at first alcohol intake. The

© University of Venda

University of Venda
Creating Future Leaders

discrete-time survival model outperformed the rest of the models in predicting survival of bleeding trauma patients from the CRASH-2 data .

Keywords:Discrete-time survival analysis, Statistical methods, Machine learning, Calibration, Discrimination.

© University of Venda

Declaration

I, Ramachela Audrey Tshepho [Student Number: 14014807] declare that this research project titled ”Comparative Analysis of Discrimination and Calibration Accuracy of Discrete Survival, Random Forests, and Neural Networks in Health-Related Survival Prediction Models” for the Master of Science degree in Statistics at the University Of Venda has been written by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. I further declare that it is my own work in design and in execution and that all referenced material contained therein have been fully acknowledged.

Signature:.....Date: 11 July 2025.....

© University of Venda

University of Venda
Creating Future Leaders

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisors Dr Alphonce Bere, Dr Tshilidzi B Mulaudzi and Ms Lactatia Motsuku for the support, guidance and patience throughout my masters journey. I hank my mother Maria Mapitsi Ramachela for the support to keep going even when it seemed impossible.

I would also like to acknowledge the South Africa Medical Research Council for the sponsorship for my MSc in Statistics degree.

© University of Venda

Dedication

This research project is dedicated to my family, specifically my mother, Maria Mapitsi Ramachela and my siblings.

© University of Venda

University of Venda
Creating Future Leaders

Contents

Abstract ii

Declaration iv

Acknowledgement v

Dedication vi

1 Introduction 1

1.1 Background 1

1.2 Problem statement 7

1.3 Research Aim and objectives 7

1.4 Proposed methodology 8

1.5 Layout of the research 8

2 Literature review and Justification 10

2.1 Introduction 10

2.2 Survival analysis 10

2.3 Discrete-time survival models 11

2.4 Machine learning 13

© University of Venda

University of Venda
Creating Future Leaders

2.4.1 Early applications of Machine learning methods in continuous time survival analysis . . . 13
2.4.2 Advancement in deep learning and general ML frameworks for continuous survival data . . . 14
2.4.3 Early Discrete-Time ML Methods . . . 14
2.4.4 Deep Learning Ensemble Methods for Discrete Time . . . 14
2.5 Summary . . . 15

3 Methodology 16

3.1 Introduction . . . 16
3.2 Discrete Survival Analysis . . . 16
3.2.1 Introduction . . . 16
3.2.2 The Discrete Hazard and survival function . . . 17
3.2.3 Discrete-time survival model . . . 17
3.2.4 Parameter estimation . . . 18
3.3 Machine learning methods in survival analysis . . . 21
3.3.1 Discrete-time survival Trees . . . 21
3.3.2 Discrete-time Random Survival Forest(RSF) . . . 26
3.3.3 Neural Networks . . . 31
3.4 Model evaluation measures . . . 34
3.4.1 Introduction . . . 34
3.4.2 Calibration measures . . . 34
3.5 Analysis of Variance (ANOVA) . . . 38
3.5.1 Introduction . . . 38
3.5.2 ANOVA procedure . . . 38

© University of Venda

University of Venda
Creating Future Leaders

3.5.3 The Kolmogorov-Smirnov (K-S) test . . . 39
3.5.4 Bartlett test . . . 40
3.5.5 Krustall-wallis (K-W) rank sum test . . . 40
3.5.6 Dunn-test . . . 41
3.6 Data . . . 41
3.6.1 Breast cancer dataset . . . 41
3.6.2 Time to first alcohol intake dataset . . . 43
3.6.3 CRASH-2 dataset . . . 43
3.7 Software of the study . . . 44
3.8 Flow chart . . . 45

4 Empirical results and Discussion 47

4.1 Breast-cancer results . . . 47
4.1.1 Descriptive statistics for numeric data . . . 48
4.1.2 Descriptive statistics for categorical data . . . 49
4.1.3 Hazard function . . . 49
4.1.4 Discrete-time Survival Models . . . 50
4.1.5 Discrete-time survival tree . . . 53
4.1.6 Accuracy metrics . . . 55
4.1.7 Prediction error curves . . . 55
4.2 Alcohol dataset . . . 59
4.2.1 Hazard function . . . 59
4.2.2 Descriptive statistics . . . 60
4.2.3 Discrete-time Survival Models . . . 61

© University of Venda

University of Venda
Creating Future Leaders

4.2.4 Discrete-time survival tree 63
4.2.5 Accuracy metrics 64
4.2.6 Prediction error curves 65
4.3 Crash-2 68
4.3.1 Hazard function 69
4.4 Descriptive statistics 70
4.4.1 Discrete-time Survival Models 70
4.4.2 Discrete-time survival tree 72
4.4.3 Accuracy Metrics 73
4.5 Discussion 77

5 Conclusion and recommendations **78**
5.1 Introduction 78
5.2 Conclusion 78
5.3 Recommendations 79

References **90**

© University of Venda

University of Venda
Creating Future Leaders

List of Figures

3.1 Visualisation of model building and testing process. 45

4.1 Hazard of breast cancer patients 49
4.2 Hazard plot for Discrete-time survival Model 52
4.3 Discrete-time survival tree 53
4.4 Breastcancer: Prediction error curves 55
4.5 Boxplot for Breast-cancer cindex 58
4.6 Alcohol Hazard function 59
4.7 Alcohol: classification tree 63
4.8 Alcohol: Prediction error curves 65
4.9 Boxplot for alcohol C-Index 68
4.10 Crash2 Hazard function 69
4.11 Discrete-time survival tree 72
4.12 crash-2: Prediction error curves 73
4.13 Boxplot for crash2 cindex 76

© University of Venda

University of Venda
Creating Future Leaders

List of Tables

3.1 Breast cancer patients variables 42
3.2 CRASH-2 indepedent variables 44

4.1 Summary statistics 48
4.2 Summary statistics for the discrete-time survival model 50
4.3 ANOVA for breast cancer cindex 56
4.4 K-S, Bartlett test and krustal wallies test for breast cancer cindex . . . 57
4.5 Comparison of C-Index Values 57
4.6 Independent Variables considered for the analysis of time to first alcohol intake 60
4.7 Discrete-time survival model 61
4.8 ANOVA for alcohol C-Index 66
4.9 K-S, Bartlett test and krustal wallies test for alcohol cindex 66
4.10 Comparison of C-Index Values 67
4.11 Patient Characteristics by Time Since Injury 70
4.12 Summary statistics for the discrete-time survival model 71
4.13 ANOVA: CRASH-2 cindex 74
4.14 K-S, Bartlett and krustal-Wallis test: crash2 c-Index 75

© University of Venda

University of Venda
Creating Future Leaders

4.15 Comparison of C-index Values 75

© University of Venda

University of Venda
Creating Future Leaders

Chapter 1

Introduction

This chapter discusses the basic concepts in discrete-time survival analysis, discrete-time machine learning algorithms, problem statement, aim, research questions, objectives, and layout of the study~

1.1 Background

Survival analysis is part of statistics that studies time until an event of interest occurs. For example in biomedical, survival models are normally utilised at certain baseline time, such as diagnosis or treatment time, to respond to questions such as "What is the probability that this patient will not be readmitted in 10 years, given their baseline covariate information?" This forecasted probabilities can be utilised by doctors for decisions making concerning patient care, such as increasing monitoring frequency or implementing particular therapies (Suresh et al., 2020). Traditional statistical models have been employed for the past; however, their limitations and inability to handle big data have made way for the introduction of machine learning methods which gained recognition based on their ability to learn complex algorithms. With a significant increase of a number of complications within

© University of Venda

University of Venda
Creating Future Leaders

the health care setup it's important to have predictive models that will learn complex relationships between covariates and survival time, and can lead to more accurate predictions making it more easier for doctors to make crucial decisions concerning patient care, such as adjusting monitoring frequency or developing particular therapies (Suresh et al., 2020).

Survival data are special type of data which requires a specific set of statistical methods due of several features that are normally discovered within the data set. Firstly the response covariate which is time until the event of interest occurs cannot be negative $(T0)$ (Tutz and Schmid, 2016). Other special features for survival data are censoring, truncation and time-varying covariates which will be discussed within the section. Survival data are typically analyzed using two key probabilities: survival and hazardClark et al. (2003). The survival probability, also known as the survivor function $S(t)$, represents the likelihood that an individual will survive from the starting point up until a specific time t. This concept is central to survival analysis, as the survival probabilities at various time points offer essential insights into the survival experiences of a group being studied. These values provide a clear picture of the cohort's survival patterns.

The hazard probability refers to the likelihood of an individual experiencing the event (e.g., death) at a specific time t, given that they have survived up to that point (Tutz and Schmid, 2016). Unlike the survival function, which emphasizes the absence of the event, the hazard function focuses on the occurrence of the event. This is important because it sheds light on the rate at which failures happen, and it helps in developing survival models. In essence, while the hazard function represents the rate of the event happening, the survival function captures the overall probability of the event not happening. Clark et al. (2003) provides a helpful distinction between these two concepts. The probable conditions of survival data

© University of Venda

University of Venda
Creating Future Leaders

discussed in this study are censoring, truncation and time-varying covariates.

Censoring is an essential aspect of survival analysis. It happens when the event of interest has not taken place for some individuals by the time the study concludes. For instance, in a study tracking the time to death, if some participants are still alive when the study concludes, their survival times are considered censored. There are several types of censoring. Left censoring happens when the event occurs before data collection begins, such as when a participant dies before a treatment trial starts. Right censoring occurs when the event isn't fully observed, for example, if a participant leaves the study early or the study ends before the event happens. Interval censoring refers to situations where the exact time of the event is unknown, but it falls between two known time points (Kartsonaki, 2016).

Truncation is a statistical occurrence where we fail to observe individuals whose event times fall outside a certain range. For example, in a study on the survival of retirement home residents, only those who survive to a certain age are included in the sample. This means that individuals who pass away earlier are not part of the analysis, resulting in left truncation of the data. Left truncation occurs when data is only recorded for individuals whose survival time exceeds a certain threshold, while right truncation happens when data is recorded only for those whose survival time surpasses a different threshold. In cases like mortality studies, longer lifespans may be excluded due to truncation. If both left and right truncation occur simultaneously, it is referred to as bi-truncation.

A time-varying covariate is a term used in survival analysis to describe a covariate whose value changes over the course of the follow-up period. These covariates can be classified into two types: internal and external. Internal covariates are those that are influenced by the survival status of the subject. These

© University of Venda

University of Venda
Creating Future Leaders

covariates typically result from a stochastic process and are only observed as long as the subject survives and is not censored. On the other hand, external covariates are fixed and do not change based on the subject's survival status. Although external covariates can impact the failure rate over time, their path up to a specific time point is not affected by the occurrence of the failure event. External covariates are usually predetermined or derived, such as the age of a participant in a long-term study or environmental factors like air pollution, which might be linked to the frequency of asthma attacks Zhang et al. (2018).

In survival analysis, time can be measured in either discrete or continuous intervals. When the exact time of an event is available, continuous-time survival analysis is generally preferred. One widely used continuous model is the Cox Proportional Hazards (PH) model, introduced by Cox (1972). This model explores the relationship between predictors and the survival time of individuals, providing an estimate of the hazard ratio while accounting for censored data. As a semi-parametric model, the Cox PH model does not make assumptions about the baseline hazard function. However, it does not directly estimate survival time since it focuses on predicting the hazard rather than the actual time to event (Baek et al., 2021).

Discrete-time survival analysis is used for analysis of the risk of an event occurring when the time is measured discretely. According to Suresh et al. (2020), discrete-time survival models are used to analyze time-to-event data, specifically to estimate the likelihood of experiencing an event within a defined time interval. This approach allows the use of various binary classification methods to calculate conditional survival probabilities. In discrete-time analysis, time is treated as discrete, where observations are either represented by whole numbers or as grouped versions of continuous time Maposa et al. (2022). The primary goal of

© University of Venda

University of Venda
Creating Future Leaders

discrete-time survival analysis is to create models that predict the chances of survival within these time intervals. This formulation is versatile, enabling the use of a wide range of binary methods and models for prediction (Suresh et al., 2020).

In addition to the discrete nature of time measurements, there are other reasons to choose discrete survival models. One key advantage is that hazards in these models can be expressed as conditional probabilities, making them more intuitive to interpret. Discrete survival models also avoid issues with tied event times and can be integrated into the broader class of generalized linear models, simplifying the estimation process. Moreover, certain discrete survival models offer flexibility by relaxing the proportional hazards assumption, which is a strict condition in continuous-time models formulation (Suresh et al., 2020) .

Traditional statistical models typically require defining the relationships between covariates and outcomes, which demands a deep understanding of how covariates interact and behave (Suresh et al., 2020). In contrast, machine learning (ML) algorithms focus on identifying the relationships between predictors through numerical computation and mathematical procedures. The primary goal of these models is prediction. ML algorithms iteratively optimize to capture complex, nonlinear associations between predictors and minimize prediction errors. However, this flexibility can make it challenging to extract and interpret the effects and interactions of individual covariates. This differs from traditional statistical methods, which are designed for inference and interpretability, although they can also be used for prediction. Several ML techniques, such as decision trees, random forests, and neural networks, have been adapted to handle the censoring present in survival data (Suresh et al., 2020).

© University of Venda

University of Venda
Creating Future Leaders

Decision trees can either be classification and regression tailored to handle censoring within the data. The main reason behind these trees is to continually divide the data using a specific splitting criterion, ensuring that similar objects, as determined by the event of interest, are grouped together in the same node. The key distinction between a decision tree and a standard decision tree lies in the splitting criterion employed (Wang et al., 2017).

According to Mogensen et al. (2012) Bagging is an ensemble technique that trains several models on different subsets of the training data and then aggregates the predictions made by each model. Bagging typically reduces the variances of the base models that are used (Wang et al., 2017).

The random forest is an advanced version of bagging, where, in addition to selecting subsets of data, it also randomly chooses subsets of features for each data sample. This non-parametric machine learning method is useful for constructing risk prediction models in survival analysis (Mogensen et al., 2012). According to Wang et al. (2017), the key distinction between random forests and bagging is that, at each decision node, random forests use only a random selection of available features, instead of considering all features, to determine the best split.

As noted by Chi et al. (2007), artificial neural networks (ANNs) are particularly effective at handling complex interactions between variables and can create non-linear prediction models. Compared to traditional methods, they offer greater flexibility in predicting survival time. ANNs are capable of modeling censored survival data with flexibility, accommodating both discrete and continuous time models through their generalized approach.

© University of Venda

University of Venda
Creating Future Leaders

1.2 Problem statement

Survival analysis finds applications in numerous areas. Within the health-related setup it is used for predicting patient outcomes, estimate the effectiveness of treatments and identify hazardous factors for oncologies. It's useful in shaping health-care methods to individual patient needs and improving overall decision-making. The significance of the proposed project lies in its effort to address the variation in the relative performance of traditional statistical methods and machine learning approaches across different datasets. Previous studies have shown that the predictive accuracy of discrete-time survival models, random forests, and neural networks can vary significantly depending on the characteristics of the data used (Suresh et al., 2020). This variation has made it challenging to draw broad generalizations about the effectiveness of these methods. Therefore, this project aims to diversify the datasets used for comparison, ensuring that a wider range of health-related survival data sets are considered. By comparing the accuracy, strengths, and limitations of these methods across different contexts, the project seeks to provide a more comprehensive understanding of how each approach performs under various conditions. This will contribute valuable insights to the field, potentially influencing future choices of survival modeling techniques in healthcare settings.

1.3 Research Aim and objectives

The aim of the study is to compare the discrimination and calibration performance of traditional statistical model and machine learning algorithms in discrete survival analysis for prediction of health-related survival outcomes. The objectives are as follows:

- To assess and compare the discrimination accuracy of discrete survival

© University of Venda

University of Venda
Creating Future Leaders

models, decision tree, discrete survival random forests, and neural networks in predicting health-related survival outcomes, such as age at first alcohol intake and breast cancer survival.

- To evaluate the calibration accuracy of discrete survival models, discrete survival random forests, and neural networks in health-related contexts, focusing on their ability to provide reliable risk predictions.
- To identify the strengths and limitations of each method (discrete survival, random forests, and neural networks) in terms of performance, interpretability, and application to real-world health data for survival prediction.

1.4 Proposed methodology

The first step will be preprocessing of the data which is simply data cleaning. The preprocessed data will be split into two sets, 80% training set and 20% testing set. The training set will be utilized to train the proposed models and the testing set will be utilized to asses the predictive accuracy of the models. The models to be compared are discrete-time survival model, discrete-time survival trees, discrete-time random forests and discrete-time neural networks. The models will be evaluated using two measures, discrimination and calibration measure. The evaluation matrices to be used are the prediction error curves and Concordance Index. The analysis will be performed on the following datasets (Breast cancer, time to first alcohol intake and CRASH-2) using R-statistical software.

1.5 Layout of the research

The chapters of this research are arranged as follows: **Chapter 2** discusses the literature review of the methods. **Chapter 3** discusses the methodology, **Chapter 4**

© University of Venda

University of Venda
Creating Future Leaders

presents the empirical results on four health-related data-sets and **Chapter 5** concludes the findings in **Chapter 4** .

© University of Venda

University of Venda
Creating Future Leaders

Chapter 2

Literature review and Justification

2.1 Introduction

This section presents related literature on the application and comparison of traditional statistical model and machine learning models in discrete-time survival analysis.

2.2 Survival analysis

In the late 1950s, Kaplan and Meier (1958) introduced a groundbreaking statistical method to estimate survival curves while accounting for incomplete observations. Their approach, now widely known as the Kaplan-Meier (KM) estimator, became the gold standard for reporting patient survival in medical research. Later, Cox (1972) developed the proportional hazards (PH) model, which enabled the analysis of censored survival data while incorporating covariates. The Cox model is classified as semi-parametric since it does not impose assumptions on the baseline hazard function. However, because it is based on hazard ratios, it does not directly predict survival time (Baek et al., 2021).

© University of Venda

University of Venda
Creating Future Leaders

To estimate the model's regression parameters, Cox (1972) introduced the concept of conditional likelihood, which eliminates the need to explicitly model the baseline hazard. Building on this work, his 1975 paper refined this concept by introducing partial likelihood, clarifying that the "conditional likelihood" from his earlier work was actually a special case of partial likelihood Cox (1975). Additionally, Cox (1972) proposed a discrete-time survival model, a type of logistic regression suited for discrete-time data. This approach is particularly useful when dealing with a large number of tied events, where multiple individuals experience the event at the same time (Allison, 1982)

2.3 Discrete-time survival models

The idea of discrete-time survival analysis was firstly established by Andersen (1970). Discrete-time survival analysis Cox (1972) is intended for instances where there are few time points measured and the specific time that the event occurred may not be known. According to Suresh et al. (2020) discrete time-to-event methods are a way to predict the likelihood of an event taking place within a particular time interval. They are utilized in medical research and other fields to analyze data that is collected in discrete chunks. Thompson (1977) and Prentice and Gloeckler (1978) are part of the first authors to introduce discrete time-to-event data in biometrics. Guo (2010) gave a review of using the discrete-time models, the binary logistic regression which on the analysis of a single event occurring, and the multinomial logit regression that focuses on more on event occurring. Allison (1982) compares the discrete-time and continuous models, the authors states that it is simple to establish explanatory variables that varies with time in discrete analysis than in continuous, and that using these models one can analyse large data-sets at a low cost.

© University of Venda

University of Venda
Creating Future Leaders

Mantel and Hankey (1978) explored the application of logistic regression analysis for modeling response time data, specifically when the hazard function varies over time. In this approach, the author substituted the baseline hazard parameters with a polynomial function to better capture the time-dependent nature of the hazard. Building on this idea, Efron (1988) introduced the use of regression splines as a flexible method for modeling a large number of baseline hazard parameters. Additionally, he provided an in-depth discussion on how the binary response model can be used to represent discrete hazard models effectively. More recently, Bere et al. (2021) proposed employing both polynomials and spline functions as a means of reducing the number of parameters required to estimate the baseline hazard function, thereby improving model efficiency.

The concept of representing the baseline hazard as a smooth function of time can also be applied to model the effects of continuous variables, allowing for a more flexible and dynamic estimation of their influence over time. In most cases, continuous covariates enter survival models in a linear fashion, assuming a constant effect across time. However, this assumption may be too restrictive, as association between predictors and the hazard function can be inherently non-linear. Generalized Additive Models (GAMs) provide a flexible alternative by allowing continuous covariates to enter the model through smooth functions rather than imposing strict parametric forms. These models enhance interpretability and predictive accuracy by capturing complex, data-driven relationships. Berger and Schmid (2018) explores semi-parametric extensions, including nonlinear and tree-based methods, in the context of discrete time-to-event data, demonstrating their application to unemployment duration analysis in the US. Similarly, Kearns et al. (2019) emphasize the importance of flexible survival models, illustrating how GAMs and other extensions of Generalized Linear Models (GLMs), such as fractional polynomials and spline models, can improve model fit and hazard estimation.

© University of Venda

University of Venda
Creating Future Leaders

2.4 Machine learning

Machine learning approaches records the association among predictors using numerical methods to perform mathematical calculations. Back in 1943, Walter Pitts and Warren McCulloch laid the groundwork for neural networks by developing a mathematical model that aimed to mimic human thought processes (Piccinini, 2004). More than a decade later, Rosenblatt (1957) took this concept further by creating the perceptron, the first neural network designed for computers, which attempted to simulate how the human brain processes information. Although the idea of machine learning had been around since the $1940s$, it wasn't until 1959 that Samuel (1959) officially introduced the term "machine learning."

2.4.1 Early applications of Machine learning methods in continuous time survival analysis

The foundational cox PH model Cox (1972) inspired many early efforts to incorporate computational methods for survival analysis. Gordon and Olshen (1985) regression trees for survival analysis, by extending CART to handle time-to-event data. LeBlanc and Crowley (1993) improved the survival tree by developing relative risk trees, which improves node splitting criteria for censored data. Faraggi and Simon (1995) expanded on the Cox proportional hazards model by incorporating neural networks, allowing for more flexible modeling of survival data. Later, Ishwaran et al. (2008) introduced the Random Survival Forest (RSF), a nonparametric ensemble method designed to handle right-censored survival data using decision trees.

© University of Venda

University of Venda
Creating Future Leaders

2.4.2 Advancement in deep learning and general ML frameworks for continuous survival data

Katzman et al. (2018) introduced DeepSurv, a deep neural network version of the cox model for handling nonlinear covariate effects. Bender et al. (2020) proposed machine learning methodology for survival analysis, transforming survival tasks into standard Poisson regression problems. Wiegrebe et al. (2022) reviewed deep learning methods for continuous-time survival analysis, focusing on handling high-dimensional data such as images, text, and omics data.

2.4.3 Early Discrete-Time ML Methods

Bou-Hamad et al. (2011a) Introduced discrete time-to-event trees and forests with time-varying predictors. Bou-Hamad et al. (2011a) Published a review of survival trees, summarizing their development and applications.

2.4.4 Deep Learning Ensemble Methods for Discrete Time

Lee et al. (2018) introduced deep learning model for discrete time-to-event analysis that estimates probability mass functions for competing risks. Kvamme and Borgan (2019) developed a neural network-based discrete-time survival models, parameterizing hazard rates and probability mass functions. Schmid et al. (2022) introduced discrete time-to-event forests using the hellinger Distance decision trees, improving tree-splitting techniques for survival data.

© University of Venda

University of Venda
Creating Future Leaders

2.5 Summary

The models explained and applied in the literature review have been effectively used to various survival data-sets. The performance of these models in making predictions largely depends on the specific characteristics of the dataset being used (Suresh et al., 2020). Hence, in this project we compare the traditional statistical method (discrete-time survival model) with the machine learning methods (random forests, and neural networks) with regards to their accuracy, strength, and constraints for predicting survival of patients of different health-related survival data sets, where such comparison were not applied.

© University of Venda

Chapter 3

Methodology

3.1 Introduction

This chapter outlines the proposed methodology. Additionally a brief discussion of the datasets used for the analyis in this project.

3.2 Discrete Survival Analysis

3.2.1 Introduction

Discrete time-to-event analysis is a statistical approach used to examine time-to-event data recorded in discrete time intervals, such as days, months, or years. This method focuses on estimating the likelihood of an event occurring within each interval and identifying factors that may influence this probability.

© University of Venda

University of Venda
Creating Future Leaders

3.2.2 The Discrete Hazard and survival function

According to Tutz and Schmid (2016), given the times $t = \{1, 2, \ldots, k\}$ at which the event can occur, the main quantity of interest in discrete survival analysis is the discrete hazard function given by

$$\lambda(t|x) = P(T = t|T \geq t, x), t = 1, ..., q$$

.

This gives the conditional probability of failure at time t given that time t has been reached. The discrete survival function is given by

$$S(t|x) = P(T > t|x) = \prod_{j=1}^{t}(1 - \lambda(j|x))$$

3.2.3 Discrete-time survival model

The discrete-time survival model expresses the hazard function through an equation of the form

$$P(T = 1|x) = h(\gamma_{ot} + x^T\gamma). \tag{3.3}$$

In equation 3.3(3.3) h(.) is the link function. Since probabilities are restricted to the interval [0,1] natural candidates for the link function h(.) are distribution functions and γ_{ot} is hazard function which measures the effect of time on the hazard.

© University of Venda

University of Venda
Creating Future Leaders

3.2.4 Parameter estimation

According to Tutz and Schmid (2016) we let the discrete-time survival model $\lambda(S(t|x_i) = h(\gamma_{ot} + x_i^T\gamma)$ be given in the form

$$\lambda(S(t|x_i) = H(x_{it}^T\beta), \tag{3.4}$$

where t $x_{it}^T = (0, ..., 0, 1, 0, ..., 0, x_i^T)$ are vector covariates with 1 being the t_th digit and all parameters store in $\beta^T = (\gamma_{01}, ..., \gamma_{0q}, \gamma^T)$.

Let T_i be the response time for the ith subject, C_i be the censoring time i^{th} subject and δ_i be the event indicator, then the observed is written like $t_i = min(T_i, C_i)$, that is the minimum of the response T_i and the censoring time C_i. The event indicator is given by

$$\delta_i = \{$$

1ifT$_i \geq C_i$, 0 if $T_i > C_i$ (3.5)

We assume that censoring occurs at the end of the interval. The entire dataset, including any censored observations, is denoted as (t_i, δ_i, x_i) for $i = 1, ..., n$, where n represents the sample size. The likelihood of observing the exact survival time (t_i, δ_i, x_i) is expressed as:

$$P(T_i = t_i, \delta_i = 1) = P(T_i = t_i)P(C_i \geq t_i). \tag{3.6}$$

The likelihood of experiencing censoring at time t is expressed as:

$$P(T_i = t_i, \delta_i = 0) = P(T_i > t_i)P(C_i = t_i). \tag{3.7}$$

If an event occurs within the interval $[a_{ti-1}, a_{ti})$ implies $P(C_i \geq t_i$ and censoring in interval $[a_{ti-1}, a_{ti})$ implies survival beyond a_{ti}. The likelihood function for the i_{th} observation is represented as:

$$L_i = P(T_i = t_i)_i^{\delta}P(T_i > t_i)^{1-\delta_i}P(C_i \geq t_i)_i^{\delta}P(C_i t_i)^{1-\delta_i}. \tag{3.8}$$

© University of Venda

University of Venda
Creating Future Leaders

We assume that the contribution from censoring, represented by C_i, does not depend on the parameters that influence the survival time, and we treat this factor separately $c_i := P(C_i \geq t_i)_i^{\delta} P(C_i t_i)^{1-\delta_i}$ to obtain the simpler form

$$L_i = c_i P(T_i = t_i)_i^{\delta} P(T_i > t_i)^{1-\delta_i}. \tag{3.9}$$

By using the discrete hazard function and the covariates, we can obtain

$$L_i = c_i \lambda(t_i|x_i)_i^{\delta} \prod_{j=1}^{ti-1} (1 - \lambda(t_i|x_i)). \tag{3.10}$$

The likelihood function for the discrete survival model is given by

$$l = \sum_{i=1}^{n} \sum_{s=1}^{t_i} y_{is} log\lambda(s|x_i) + (1 - y_{is}) log(1 - \lambda(s|x_i)). \tag{3.11}$$

The expression $s_t = 1, \ldots, t_i$ represents the log-likelihood function of a binary regression model. This implies that standard techniques for parameter estimation in binary regression models can be applied in this case. To estimate the parameters, we begin by differentiating the log-likelihood with respect to each unknown parameter and then setting each derivative equal to zero.

Before fitting the binary model for transitions, we first generate the binary observations. In this case, we use a discrete hazard model with a linear predictor, denoted as $\eta_{ir} = \gamma_{or} + x_i^T$. The linear predictor is expressed as $x_{ir}^T \beta$, where the parameter vector $\beta^T = (\gamma_{0,1}, \ldots, \gamma_{0q}, \gamma^T)$ contains the model parameters. When calculating the linear predictors for the binary model, it is important to adjust the covariate vector according to the specific observations being considered. While constructing the design matrix, we need to account for the distinction between censored and non-censored observations. Specifically, if $T_i = t_i$ and $\delta_i = 1$, then the

© University of Venda

University of Venda
Creating Future Leaders

binary observations and design variables for t_i are given by.

Binary observations		Design variables
0	$1\ 0\ 0\ \cdots\ 0$	x_i^T
0	$0\ 1\ 0\ \cdots\ 0$	x_i^T
0	$0\ 0\ 1\ \cdots\ 0$	x_i^T
$\vdots$		
1	$0\ 0\ 0\ \cdots\ 1$	x_i^T

if $T = t_i$ and $\delta_i = 0$, the t_i binary observations and design variables are given by the augmented data matrix

Binary observations		Design variables
0	$1\ 0\ 0\ \cdots\ 0$	x_i^T
0	$0\ 1\ 0\ \cdots\ 0$	x_i^T
0	$0\ 0\ 1\ \cdots\ 0$	x_i^T
$\vdots$		
0	$0\ 0\ 0\ \cdots\ 1$	x_i^T

we also construct the observation with running time t, which is considered to be a factor. For $T_i = t_i$ and $\delta_i = 1$ we use the coding given by

Binary observations	Time	Design variables
0	1	x_i^T
0	2	x_i^T
0	3	x_i^T
$\vdots$		
1	t_i	x_i^T

© University of Venda

University of Venda
Creating Future Leaders

For $T_i = t_i$ and $\delta_i = 0$ we use the coding given by

Binary observations	Time	Design variables
0	1	x_i^T
0	2	x_i^T
0	3	x_i^T
$\vdots$		
0	t_i	x_i^T

Using the long format training set we fit the models in r-studio using the following packages, generalized linear model using the $glm()$ function (stats package).

3.3 Machine learning methods in survival analysis

Numerous machine learning approaches have been established to model censored survival data. These algorithms can develop survival prediction models for high-dimensional, sophisticated data that can have better accuracy than traditional statistical methods. Commonly used machine learning approaches for survival data include survival trees and random survival forests, and neural networks.

3.3.1 Discrete-time survival Trees

According to LeBlanc and Crowley (1993) survival trees are like discrete-time survival tree which is developed by iteratively splitting tree nodes. Recursive partitioning with binary outcome has the following procedure:

© University of Venda

University of Venda
Creating Future Leaders

3.3.1.1 Step 1: Specification of the data structure

The log-likelihood function (3.11) shows that the binary outcome variable is represented by the first columns of the expanded data matrices.

Binary observations						Design variables
0	1	0	0	$\cdots$	0	x_i^T
0	0	1	0	$\cdots$	0	x_i^T
0	0	0	1	$\cdots$	0	x_i^T
$\vdots$						
1	0	0	0	$\cdots$	1	x_i^T

and

Binary observations						Design variables
0	1	0	0	$\cdots$	0	x_i^T
0	0	1	0	$\cdots$	0	x_i^T
0	0	0	1	$\cdots$	0	x_i^T
$\vdots$						
0	0	0	0	$\cdots$	1	x_i^T

For both censored and uncensored observations, the same matrices have been used earlier to estimate the parametric discrete-time hazard models in (3.2). The hazard function is then defined as follows.

$$\lambda(t|x) = P(T = t|T \geq t, x), t = 1, ...$$

.

This suggests that conditioning can be done based on the predictor values x and the event "$T \geq t$". To account for the ordinal nature of T, the baseline hazard's dummy variables in the design matrices are replaced with the column vector $(1, 2, ...,)^T$ Tutz and Schmid (2016). As a result, the input data for each individual

© University of Venda

University of Venda
Creating Future Leaders

observation is

Binary observations	Time	Design variables
0	1	x_i^T
0	2	x_i^T
0	3	x_i^T
$\vdots$		
1	t_i	x_i^T

and

Binary observations	Time	Design variables
0	1	x_i^T
0	2	x_i^T
0	3	x_i^T
$\vdots$		
0	t_i	x_i^T

For both censored and uncensored observations, the augmented dataset used in tree construction is created by combining the matrices previously discussed. By definition, the design matrix consists of $n := \sum_{t=1}^{n} t_i$ rows. This approach, where T is treated as an ordinal covariate during tree construction, provides more flexibility than the conventional use of the time trend paramete γ_{0t} in a discrete hazard model (Tutz and Schmid, 2016).

3.3.1.2 Step 2: Tree Construction and Estimation of the Hazard Function

The likelihood function's binomial structure, as shown in equation (3.11), allows recursive partitioning techniques for binary outcomes to be applied to the augmented dataset described in Step 1. Once a proper splitting criterion is established, this method generates a set of terminal nodes, each characterized by binary outcome values $y_j \in \{0, 1\}$ for $j \in \{1, \ldots, n\}$. These binary values can be used to estimate the hazard function as follows: T serves as a potential splitting

© University of Venda

University of Venda
Creating Future Leaders

variable during tree construction, resulting in terminal nodes linked to values of x and time intervals

$$T_{(1)}.\bigcup T_{(1)}\bigcup\ldots\bigcup T_{(Q)}=\{1,\ldots,q\}\text{ (each being a subset of }\{1,\ldots,q\})$$

. Therefore, for each combination of the predictor variable x, the hazard function , estimates of $\lambda(t\epsilon T_{(.)}|.)$ can be obtained from the proportions of ones in the corresponding terminal nodes Tutz and Schmid (2016). Let $\hat{\lambda_1},...,\hat{\lambda_q}$ represent these estimates. The conditional survival function $S(t|x)$ can be derived from these estimates as follows:

$$S(t|x)=P(T>t|x)=\prod_{i=1}^{t}(1-\lambda(i|x))$$

to $\hat{\lambda_1},\ldots,\hat{\lambda_q}$. The selection of an appropriate splitting criterion in discrete-time survival trees emphasizes the accurate estimation of probabilities $\hat{\lambda_1},...,\hat{\lambda_q}$ rather than the correct classification of individual values $y_j\epsilon 0,1$. The objective is to reduce the discrepancies between the true node probabilities and the estimated probabilities at each node during the construction of the tree. Let N represent the set of observations at a given node. A natural metric for assessing node impurity is the Brier score.

$$\begin{aligned}BS_N &:= \frac{1}{|N|}\sum_{j\in N}\Big[\big(\hat{P}(y_j=1\mid j\in N)-I(y_j=1)\big)^2\\ &\qquad +\big(\hat{P}(y_j=0\mid j\in N)-I(y_j=0)\big)^2\Big]\\ &=\frac{1}{|N|}\sum_{j\in N}\Big[\big(\hat{\lambda}_N-I(y_j=1)\big)^2+\big((1-\hat{\lambda}_N)-I(y_j=0)\big)^2\Big]\end{aligned}\tag{3.14}$$

which quantifies the squared distance between the observed hazards $y_j\in\{0,1\}$, where $j\in N$, and the estimated hazards $\lambda_N:=\frac{y_j}{|N|}$ The Brier score is a proper measure, as it attains its minimum when the true probabilities $P(y_j=1|j\in N)$ are utilized in place of λ_N. The binary nature of the outcome values y_j renders the Brier score, as a metric of node impurity, equivalent to the Gini impurity measure.

© University of Venda

University of Venda
Creating Future Leaders

3.3.1.3 Step 3: Selection of Tuning Parameters

To optimize tree construction for discrete survival times, one can use the minimum number of observations in terminal nodes as the key tuning parameter, a technique known as ”cardinality pruning.” This approach means that tree growth stops when further splitting of existing nodes would result in child nodes containing fewer observations than the predetermined minimum node size (Tutz and Schmid, 2016). The ideal minimum node size can be determined through the use of information criteria, which are defined by

$$-2l + \varepsilon.(Q-1) \tag{3.15}$$

The letter l represents the binomial log-likelihood (3.11), which is evaluated at the estimated values $\hat{\lambda_1}, ..., \hat{\lambda_q}$, and where $(Q-1)$ refers to the estimated complexity of the tree, calculated based on the number of binary splits. By setting $\varepsilon = 2$ and $\varepsilon = \log(n)$ leads to the AIC and BIC criteria, respectively. The complexity term $(Q1)$ does not account for the exploration of the covariate space during each split, resulting in an underestimation of the tree's true complexity. Alternatively, cross-validation of the log-likelihood or the integrated prediction error curve can be used. The optimal minimum node size is determined by the tree that yields the highest performance based on the chosen criterion.

3.3.1.4 Step 4: Estimation of Conditional Survival Functions

The last step involves calculating an estimate of the conditional survival function $S(t|x_j^T)$ for a new observation from the test sample.

$(\tilde{\mathrm{t}}_j^T, \delta_j^T, \tilde{x}_j^T), j = 1, ..., n^T$ (3.16)

© University of Venda

University of Venda
Creating Future Leaders

This is achieved by combining the new covariate values x_j^t with each possible time point t, resulting in a set of vectors $((x_j^t)^T, 1), ..., (x_j^t)^T, t_{max}), j = 1, ..., n^T$. The vectors are propagated through the tree, and the estimates $\hat{\lambda}_q$ in the corresponding terminal nodes constitute the estimated hazard function $\hat{\lambda}(t|x_j^T)$.

3.3.1.5 Discrete-time survival tree in r

To compute and visualize the discrete-time survival tree we use the $rpart()$ and $rpart.plot()$ function found in Rpart package in r, the model uses gini-impurity as the splitting criterion throughout the study. It begins by defining a model formula with various predictors, then converts the dataset to a long format for survival analysis. The data is balanced using oversampling and cross-validation folds are created. The model is optimized by testing different minsplit values, and evaluating each tree's performance using AIC, BIC, and cross-validation log-likelihood. These metrics penalize overly complex models to avoid overfitting. The best minsplit value is chosen based on BIC, and the final tree model is fitted using this optimal parameter. The approach ensures a well-generalized and efficient tree model for predicting survival outcomes. We will only use the discrete-time survival tree from $rpart()$ package to visualize the tree results using $rpart.plot()$ because there's no available function to visualise the trees in ranger package.

3.3.2 Discrete-time Random Survival Forest(RSF)

Bagging is a machine learning technique that involves training multiple models on different subsets of the training data, and then aggregating the predictions from each model. Random Forest builds on this approach by also randomly selecting subsets of features for each data sample in addition to the data subsets. As stated in Schmid et al. (2022) The discrete-time RSF algorithm aims to provide a nonparametric estimate of the discrete hazard function $\lambda(t|x)$. The RSF algorithm is

© University of Venda

University of Venda
Creating Future Leaders

applied to a dataset comprising i.i.d. learning data with n independent observations and covariates $(\tilde{T}_i, \delta_i, X_i)$, where $i = 1, \ldots, n$. Here, $\tilde{T}_i$ represents the sample variables of the (potentially right-censored) observed survival time $\tilde{T}_i := (T_i, C_i)$, and the event indicator $\delta_i := T_i \geq C_i$. The procedure for acquiring a discrete-time random survival forest is outlined as follows:

3.3.2.1 Initialization

Prior to the initiation of the algorithm, it is necessary to define the number of trees in the Random Survival Forest (RSF), denoted as $(ntree)$, and the number of covariates permitted for splitting at each node, referred to as $(mtry)$. Typical selections include $mtry = \lfloor \sqrt{p} \rceil$ and $ntree = 500$. The default values in the R package ranger are outlined Wright and Ziegler (2017).

3.3.2.2 Data-reshaping

Instead of directly fitting the discrete-time RSF to the learning data $(\tilde{T}_i, \delta_i, X_i), i = 1, .., n$, as suggested by Tutz and Schmid (2016), we propose defining the input for the RSF algorithm using a modified set of "augmented" data that includes a binary outcome variable. This approach is based on optimizing the log-likelihood derived from the data-generating process described earlier. The main goal is to express the log-likelihood of a discrete-time survival model in terms of the hazard $\lambda(t|X)$, which results in a binomial log-likelihood function.

$$l = \sum_{i=1}^{n} \sum_{t=1}^{\tilde{T}_i} y_{it} log \lambda(t|x_i) + (1 - y_{it}) log(1 - \lambda(t|x_i)) \tag{3.17}$$

Consider the sequence $y_i = (y_{i1}, ..., y_{i\tilde{T}_i})$ of length $\tilde{T}_i$, defined as $y_i := (0, ..., 0, 1)$. If $\delta_i = 1$, then $y_i := (0, \ldots, 0, 0)$. If $\delta_i = 0$ each entry of y_I corresponds to a specific time point $t = 1, 2, \ldots, \tilde{T}_i$, where $y_{it} = 0$ signifies survival beyond t and $y_{it} = 1$ indicates an event occurring at time t. According to the results presented in (2), estimates of

© University of Venda

University of Venda
Creating Future Leaders

$\lambda(t|X)$ can typically be derived by applying a statistical model with a binary outcome to the augmented data set, which is defined for each individual as follows:

$$M_i = \begin{bmatrix} 0 & 1 & x_i^T \\ 0 & 2 & x_i^T \\ 0 & 3 & x_i^T \\ \vdots & & \\ 1 & \tilde{T}_i & x_i^T \end{bmatrix}$$

if $\delta = 1$ and

$$M_i = \begin{bmatrix} 0 & 1 & x_i^T \\ 0 & 2 & x_i^T \\ 0 & 3 & x_i^T \\ \vdots & & \\ 0 & \tilde{T}_i & x_i^T \end{bmatrix}$$

if $\delta = 0$ (Tutz and Schmid, 2016).

3.3.2.3 Tree building

We suggest using the classification and regression trees (CART) technique, as outlined by Breiman et al. (1984), for constructing the tree. The CART process starts with the root node, representing the entire covariate space, and recursively divides this space into terminal nodes. This partitioning is achieved through the optimization of a splitting criterion, based on the learning data and chosen according to the type of outcome variable. At each split, one variable $X_{(j)}$, where $j \in \{1, \ldots, p\}$, is selected, and its values are split into two exclusive subsets known as child nodes. The goal is to partition the data to maximize homogeneity within the child nodes regarding the outcome variable's distribution. The method for selecting the split criterion depends on the outcome variable's scale. When the outcome is

© University of Venda

University of Venda
Creating Future Leaders

categorical, a common approach is to build a classification tree or a probability estimation tree (PET), which minimizes a measure of impurity, such as the Gini index, in the child nodes. Building on the log-likelihood definition in (2), Schmid et al. (2016) developed a discrete-time survival tree algorithm using the CART approach with a binary outcome variables Y applied to the previously defined augmented data matrix. The rows of matrices (3) and (4) are recursively partitioned, implying that (i) each individual corresponds to multiple rows in the augmented learning data, and (ii) each terminal node of the discrete-time survival tree consists of a set of binary values (zeros and ones).

The discrete hazard function, $\lambda(t|X)$, is estimated based on the proportions of ones within the terminal nodes. This hazard function takes into account both the covariates X and the event $T \geq t$. To achieve this, the values X and $\tilde{t}_i := (1, 2, \ldots, \tilde{T}_i)$, from the second columns of the matrices (3) and (4), are utilized for node splitting. The vector $(\tilde{t}_1, \ldots, \tilde{t}_n)^T$ is treated as an ordinal covariate during tree construction, allowing the estimates to capture the interaction between the covariates and time. Moreover, because of the recursive partitioning and the design of the augmented data matrix, each terminal node corresponds to a hazard estimate that remains constant within a specific time interval $T \subset \{1, \ldots, k\}$ Schmid et al. (2016).

For each individual, the discrete hazard function $\lambda(t|X)$ across the entire time span $t = 1, \ldots, k$ is computed by concatenating the hazard estimates (proportions of ones) from the terminal nodes. As an example, if an individual $i \in \{1, \ldots, n\}$ has covariates X_i and an observed survival time $\tilde{T}_i$, this results in $\tilde{T}_i$ rows in the augmented data matrix. These rows are passed through the discrete-time survival tree, which yields hazard estimates for each time interval T. By concatenating these time intervals along with their corresponding hazard estimates, the individual's hazard function is estimated over the full time range $t = 1, \ldots, k$. This methodology

© University of Venda

University of Venda
Creating Future Leaders

produces a flexible, nonparametric estimator for $\lambda(t|X)$, capable of capturing higher-order interactions and time-dependent covariate effects, as demonstrated by Schmid et al. (2016) and Berger and Schmid (2018).

3.3.2.4 Ensemble estimation

According to Tutz and Schmid (2016), the discrete-time Random Survival Forest (RSF) ensemble is formed by combining the binary outcomes from multiple trees. These trees are created using bootstrap samples, which were generated in the initial data preparation step. We suggest building unpruned trees, where the construction continues until the tree reaches a pre-specified minimum size for the terminal nodes. Once the ensemble is built, estimates of the discrete hazard function $\lambda(t|X)$ can be computed as described below: for each new observation with covariate data X_{new}, the augmented data matrix is used.

$$M_{new} = \begin{bmatrix} 1 & X_{new}^T \\ 2 & X_{new}^T \\ 3 & X_{new}^T \\ \vdots & \\ k-1 & X_{new}^T \end{bmatrix}$$

covering the time range $t = 1, ..., k1$. Next, the k 1 data lines are dropped down each of the ntree trees, resulting in sets of estimates $\hat{\lambda}_b(1|X_{new}), \hat{\lambda}_b(2|X_{new}), ..., \hat{\lambda}_b(k-1|X_{new}), b = 1, ..., ntree$, which are given by the proportions of ones(computed from the learning data) in the respective terminal nodes. Note that it isnot necessary to include a k-th row in the augmented matrix, as $\lambda(k|x_{new}) = 1$ by definition Finally, the discrete-time RSF estimate of $\lambda(t|x_{new})$ is computed by averaging over the ntree trees, i.e.,

© University of Venda

University of Venda
Creating Future Leaders

$\lambda_{RSF}(t|x_{new}) := \frac{1}{ntree}\sum_{b=1}^{ntree}$

To evaluate the performance of the discrete-time RSF with a single real-valued score, it is beneficial to aggregate the discrete hazard estimates over time as presented in (9). This can be accomplished by calculating the sum of the cumulative discrete hazard estimates, as expressed by

$_{RSF}(x_{new}) := \sum_{t=1}^{k-1}(k-1)$

The estimate in (10), which can be computed separately for each individual in a set of independent test data, will be used as a predictive marker to evaluate the performance of discrete-time RSF Ishwaran et al. (2008). To obtain the discrete-time tree and discrete-time RFS in r-statistical package we used the $ranger()$ function (in ranger package), using the hellinger distance as a splitting criterion, and setting $mtry = sqrt(noofvariablesinadataset)$.

3.3.3 Neural Networks

Feed-forward neural networks offer a versatile approach to extending linear regression functions. The input units serve to distribute the inputs to the hidden units in the subsequent layer, effectively creating a 'fan out' mechanism. The units aggregate their inputs, incorporate a constant (referred to as the 'bias'), and apply a predetermined function ϕ_h to the outcome. The output units maintain the same structure, yet they incorporate the output function ϕ_h. Therefore

$$y_k = \phi_0(\alpha_k + \sum_h w_{hk}\phi_h(\alpha_h + \sum_i w_{ih}x_i)) \tag{3.20}$$

© University of Venda

University of Venda
Creating Future Leaders

The 'activation function ϕ_h of the hidden layer units is almost always taken to be the logistic function

$$l(z) = \frac{exp(z)}{1 + exp(z)} \tag{3.21}$$

The output units can be categorized as linear, logistic, or threshold units. The latter exhibit $\phi_0(x) = I(x > 0)$. Observe the resemblance to projection pursuit regression (see (3.20)), characterized by linear output units while incorporating general smooth hidden units. (Nevertheless, arbitrary smooth functions can be approximated through the summation of rescaled logistics.) The broad definition permits multiple hidden layers and also accommodates 'skip-layer' connections from input to output when present.

$$y_k = \phi_0(\alpha_k + \sum_{i \to k} w_{ik} x_i + \sum_{j \to k} w_{jk} \phi_h(\alpha_j + \sum_{i \to j} w_{ij} x_i)) \tag{3.22}$$

that enables the non-linear units to influence a linear functional form. By introducing an input unit 0 that is consistently set to +1 and connects to every other unit, we can effectively remove the biases α_i. The regression function f is parametrized by the weights W_{ij}, assigned to each link in the network, with a value of zero for links that do not exist. A seminal model of a neuron as a binary thresholding device in discrete time was published by McCulloch and Pitts (1991), specifically that

$$n_i(t) = H(\sum_{j \to i} w_{ji} n_j(t-1) - \theta_i) \tag{3.23}$$

The total is calculated over the neurons j that are connected to neuron i. In this context, H represents the Heaviside threshold function. $H(x) = I(x > 0), n_i(t)$ represents the output of neuron i at time t, where $0 < W_{ij} < 1$ denotes the attenuation weights. Consequently, the outcome is to establish a threshold for a weighted sum of the inputs at the value θ_i. Real neurons exhibit greater complexity; they demonstrate a graded response instead of the binary thresholding characteristic of the McCulloch-Pitts model, operate in continuous time, and are capable of executing a broader range of non-linear functions based on their inputs,

© University of Venda

University of Venda
Creating Future Leaders

including logical functions. The McCulloch-Pitts model has played a significant role in shaping the evolution of artificial neural networks. Feed-forward neural networks can also be interpreted as a method to parameterize a relatively broad non-linear function. The results of the approximation are non-constructive, necessitating the selection of weights in practice to minimize a fitting criterion, such as least squares.

$$E = \sum_{p} \|t^p - y^p\|^2 \tag{3.24}$$

where t_P represents the target and y_P denotes the output for the p-th example pattern. Alternative measures have been suggested, including for $y \in [0, 1]$ 'maximum likelihood' (specifically the negative logarithm of a conditional likelihood) or, equivalently, the Kullback-Leibler distance, which involves minimization.

$$E = \sum_{P} \sum_{k} [t_k^p \log \frac{t_k^p}{y_k^p}] \tag{3.25}$$

This represents half of the deviance for a binary logistic model with a linear predictor as defined in equations (3.22) or (3.23). A multinomial log-linear model with K classes can be represented using a neural network that has K outputs, utilizing the negative conditional log-likelihood.

$$E = \sum_{P} \sum_{k} -t_k^p \log p_k^P, p_k^P = \frac{e^{y_k^p}}{\sum_{c=1}^{k} e^{y_c^p}} \tag{3.26}$$

Since precisely one of the targets, tt, will be one (for the class that occurred), while all others will be zero. This is frequently referred to as 'softmax' fitting. Due to redundancy in equation (3.26), one output may be set to zero, although this is typically not done. One method to guarantee the smoothness of f is to limit the class of estimates, such as by employing a finite number of spline knots. Another approach involves regularization, wherein the fitting criterion is modified to

$$E + \lambda C(f) \tag{3.27}$$

incorporating a penalty C for the 'roughness' of f. Weight decay, in the context of neural networks, employs the sum of the squares of the weights W_{ij} as a penalty.

© University of Venda

University of Venda
Creating Future Leaders

Rescaling the inputs to a range of approximately [0, 1] is necessary for comparability with the outputs of internal units. Weight decay appears to facilitate the optimization process while mitigating the risk of overfitting.

3.4 Model evaluation measures

3.4.1 Introduction

This section presents the model evaluation measures used in this study. The best performing models will be selected using the calibration and discrimination measures

3.4.2 Calibration measures

Calibration measures how well the predicted probabilities of the model aligns with atual observed outcomes. The following are detailed calibration measures used

3.4.2.1 Prediction error curves

Prediction error curves (PE curves) serve as a time-dependent metric for assessing prediction error, calculated through the squared distance between the predicted survival functions $\hat{S}_j(t) = \prod_{i=1}^{t}(1 - \hat{\lambda}_j(s)), j = 1, ..., n^{\tau}$, and the corresponding observed survival function $S^{\sim} = I(t < t_j\tau)$, with $I(.)$ representing the indicator function. At each time point, the PE curves assess the difference between the observed values (i.e., $S_j^{\sim}(t)$) and the values predicted by a statistical model (i.e., $S_j^{\sim}(t)$) (Tutz and Schmid, 2016). At a specific time t, the potential energy curve is represented by

$$w_j^{\tau}(t) = \frac{\delta_j^{\tau}(1 - S_j^{\sim}(t))}{\hat{G}_j(t_j^{\tau} - 1)} + \frac{S_j^{\sim}(t)}{\hat{G}_j(t)} \tag{3.28}$$

© University of Venda

University of Venda
Creating Future Leaders

$$\hat{PE}(t) = \frac{1}{n^{\tau}} \sum_{j=1}^{n^{\tau}} w_j^{\tau}(t) \, (\hat{S}_j(t) - \tilde{S_j^{\sim}}(t))^2 \tag{3.29}$$

The weights $w_j^{\tau}(t)$ include the expression $\hat{G}_j(t) = \hat{P}(C_j > t|x_j^{\tau})$, representing the estimated survival function of the censoring process derived from the learning data. Their role will be addressed subsequently. The second factor in equation (3.29) denotes the fundamental measure of discrepancy, specifically the squared deviation between observed and predicted survival at a fixed time t. The squared distance presents a limitation as it can solely be calculated using the test data. A robust survival model should exhibit prediction error values below 0.25 for all t. This value of 0.25 corresponds to the prediction error curve derived from a non-informative model, where $\hat{S}(t) = 0.25$ for all t.

We used the function $prederrcurves()$ found in the discSurv package to obtain the prediction errors of the models which uses the prediction values obtain from the long format testing set, because this function uses the long format dataset. To obtain the long format predictions we used the function $predict()$ found in the discSurv package in r. To visualise the prediction error curves we used the $ggplot()$ function found in the ggplot2 package using r statistical software.

3.4.2.2 AUC and Concordance-Index (C-Index)

The concordance index (C-Index) is a method used to evaluate the prediction accuracy, based on the discrimination measures. The measures uses $\eta = x^T\gamma$ as the predictor. For discrimination measures, one can consider survival outcomes as time-dependent binary outcomes with levels event at t and event after t. Hence, prediction accuracy can be measured by following the established concept for the evaluation of the binary classification rule, that is for each fixed point t survival

© University of Venda

University of Venda
Creating Future Leaders

outcomes have classes, namely, event at t (cases) and event after t (controls). Prediction performance is to be high if the η has a high discriminative power. Discrimination power can be measured using time-dependent sensitivities and specificities . Defined by

sens(c,t) = P($\eta > c|T = t)$)(3.30) and

$$spec(c,t) = (\eta \leq c|T \geq t) \tag{3.31}$$

Where c is the threshold of the predictor η. The concordance index is well-known prediction measure which is linked to the Receiver Operating Characteristics (ROC). The ROC is obatined by summarising the sensitivity and specificity for fixed t, which is defined by

$$ROC(c,t) = 1 - spec(c,t), sens(c,t)c \in \tag{3.32}$$

Where $1 - spec(c,t)$ is the hit rate and $sens(c,t)$ is the positive false rate. The ROC curve plots sensivity vurses false positive rate and usually result in a concave shape. When the ROC curve is strongly concave and has a large area below the curve, then the predictor η has a high discrimination power. Calculating the area under the ROC curve results in the time-dependent Area Under Curve (AUC). The time-dependent AUC measures the discrimination power of the predictor η at each time point. $Sens(c,t)$ and $spec(c,t)$ can be estimated by

$$\text{sens}(c,t) = \frac{\sum_j \delta_j I(\hat{\eta}_j)^\tau \geq t_j^\tau}{\hat{G}_{j(t_j^\tau - 1)}} \Big/ \frac{\sum_j \delta_j^\tau I(t_j^\tau = t)}{\hat{G}_{j(t_j^\tau - 1)}}$$

$$\text{spec}(c,t) = \frac{\sum_j (1-\delta_j) I\left(\hat{\eta_j}^\tau < t_j^\tau\right)}{\hat{G}_{j(t_j^\tau - 1)}} \Big/ \frac{\sum_j (1-\delta_j)^\tau I\left(t_j^\tau \neq t\right)}{\hat{G}_{j(t_j^\tau - 1)}}$$

© University of Venda

University of Venda
Creating Future Leaders

δ_j is the event indicator (1 if the event occurred for subject j, 0 if censored), $\hat{\eta}_j$ is the predicted value or survival probability for subject j, t_j^τ is the observed time for subject j, t is the specific time point at which sensitivity is calculated, $\hat{G}_{j(t_j^\tau - 1)}$ is the estimated survival probability up to the time $t_j^\tau - 1$, $I(\cdot)$ is the indicator function, which equals 1 if the condition is true, otherwise 0 (Tutz and Schmid, 2016).

Using numerical integration of the estimated ROC curve leads to estimates of the Area under curve. Thus, the concordance index can be estimated by

$$C^* = \sum_t \frac{\hat{P}^L(T = t).\hat{P}^L(T > t)}{\sum_t \hat{P}^L(T = t)\hat{P}^L(T > t)} A\hat{U}C(t) \tag{3.33}$$

$\hat{P}^L(T = t)$ is the predicted probability that the event occurs at time t, $\hat{P}^L(T > t)$ is the predicted probability that the event occurs after time t, $A\hat{U}C(t)$ is the Area Under the Curve (AUC) at time t. A well predicting model should have concordance index values above 0.5 across all t. This is because a null model have a constant concordance index value of 0.5. The following section presents the analysis of variance which is used to asses the difference in the models evaluation measures (Tutz and Schmid, 2016).

To obtain the concordance index We used the function $cIndex()$ found in the discSurv package to obtain the concordance index of the models which uses the prediction values obtain from the short format testing set, because function uses the short format dataset. To obtain the short format predictions we used the function $predict()$ found in the discSurv package in r. Further we utilize the $ggplot()$ function found in the ggplot2 package using r to visualise the model's concordance index in a form of boxplots.

© University of Venda

University of Venda
Creating Future Leaders

3.5 Analysis of Variance (ANOVA)

3.5.1 Introduction

This section presents ANOVA used for testing the significant difference between the models evaluation measures. We carry out the ANOVA test and check if the assumptions are met. If the assumptions are met we can rely on the results of the ANOVA test. If not, we will have to use the alternative tests. If the ANOVA test does not show significant differences we conclude accordingly i.e that the mean C-Index/Brier score values are the same across the models. If we reject the null hypothesis our conclusion from ANOVA is that at least two models differ in mean Cindex/brier score values. Hence, we perform post-hoc tests.

3.5.2 ANOVA procedure

Step 1:

- Null hypothesis (H_0) = There's no difference between the means of the models
- Alternative hypothesis (H_1) = There's a significant difference between the means of the models

Step 2: we compute the sum of squares and corresponding degrees of freedom

- $SS_{within} = \sum_{j=1}^{k} \sum_{j=1}^{l} (X - \hat{X}_j)^2$ and $df_{within} = k - 1$
- $SS_{between} = \sum_{j=1}^{k} (\hat{X}_j - \hat{X})^2$ and $df_{between} = n - k$
- $SS_{total} = \sum_{j=1}^{n} (\hat{X}_j - \hat{X})^2 and$ df$_{total} = n - 1$

Step 3 : we compute the mean squares and the F-statistic

- $MS_{within} = \frac{SS_{within}}{df_{within}}$

© University of Venda

University of Venda
Creating Future Leaders

- $MS_{between} = \frac{SS_{between}}{df_{between}}$
- $F = \frac{MS_{between}}{MS_{within}}$

step 4 : we compute F(critical value) from the F-table and make decision. If the F-statistic is greater that F-value , we reject the null hypothesis and conclude that there's significant difference between the means of the models, otherwise we fail to reject the null hypothesis. We use the following tests (Kolmogorov-Smirnov test and Bartlett test) to check if the ANOVA assumptions are met and if the assumptions are met we proceed to perform a post-hoc test for the pairwise comparision of the models, and if not we proceed with the Kruskal-Wallis test then compute the suitable post-hoc test for pairwise comparision of the models. We use the function $aov()$ in r to obtain ANOVA test for the mean difference of the models.

3.5.3 The Kolmogorov-Smirnov (K-S) test

The Kolmogorov-Smirnov test is used to test if the evaluation matrix is normally distributed. The null-hypothesis of the K-S test assumes that the data are normally distributed. The test-statistic D of K-S test is defined by

$$D = max|F_0(x) - F_r(x)| \tag{3.34}$$

where,

$F_0(x)$ is the Observed cumulative frequency distribution of a random sample of n observations and $F_r(x)$ is the theoretical frequency distribution. We use test $ks.test()$ in r to obtain the Kolmogorov-Smirnov test to asses if the ANOVA assumption are satisfied.

© University of Venda

University of Venda
Creating Future Leaders

3.5.4 Bartlett test

The Bartlett test is employed to assess the homogeneity of variation. The null hypothesis of the Bartlett test states that there is no disparity in variances among the models. The test statistic for the Bartlett test is denoted as T and is defined by

$$T = \frac{(N-K)lns_p^2 - \sum_{i=1}^{k}(N_i)lns_i^2}{1+(1/(3(k-1)))((\sum_{i=1}^{k}1/(N_i-1)-1/(N-k)} \tag{3.35}$$

In the above, s_i^2 represents the variance of the ith group, N denotes the total sample size, N_i indicates the sample size of the i_{th} group, k refers to the number of groups, and s_p^2 signifies the pooled variance. The pooled variance represents the weighted average of the variances of the models and is defined as follows:

$$s_p^2 = \sum_{i=1}^{k}(N_i-1)s_i^2/(N-k) \tag{3.36}$$

We use $Bartlett.test()$ in r to obtain the Bartlett test.

3.5.5 Krustall-wallis (K-W) rank sum test

When the sample sizes are small and normality checks showed that the assumption of normality had not been met, that is ANOVA is not suitable for comparing the groups.Krustal wallies test is non-parametric test equivalent to ANOVA that is used to compare the distribution between two or more independent models. The null hypothesis of the K-W test states that all the models have the same median value. The test statistics H for the K-W rank sum test and it is defined by

$$H = \frac{12N(N+1)\sum_{i=1}^{k}R^2}{n_i} - 3(N+1) \tag{3.37}$$

© University of Venda

University of Venda
Creating Future Leaders

3.5.6 Dunn-test

Dunn's test is a post-hoc test that can be performed after a Kruskal-Wallis test to determine which groups differ. It's a non-parametric test that uses the same data and rankings as the Kruskal-Wallis test.

3.6 Data

This section provides variable description of the data sets used in this research project.

3.6.1 Breast cancer dataset

The dataset used in this study comes from the November 2017 update of the SEER Program, provided by the National Cancer Institute (NCI), which offers comprehensive cancer statistics based on population data. The data specifically includes female patients diagnosed with infiltrating duct and lobular carcinoma breast cancer (SEER primary sites recode NOS histology codes $8522/3$) between 2006 and 2010. Patients who had missing data on tumor size, examined regional lymph nodes, or regional positive lymph nodes, as well as those with survival times shorter than one month, were excluded from the analysis. As a result, the final dataset consists of 4024 patients. The primary outcome variable is the time to death for breast cancer patients, with individuals who survived throughout the study period being treated as censored. A detailed description of the variables is provided in Table 3.1.

© University of Venda

University of Venda
Creating Future Leaders

Table 3.1: Breast cancer patients variables

Variable name	**Description**
Age	Age at the time of diagnosis is categorized in three groups
Race	Race of patients
Marital Status	Marital status of the patient
T Stage	The higher the number after the T, the larger the tumor or the more it has grown into nearby tissues.
N Stage	The higher the number after the N, the more lymph node contains cancer
A stage	a simplified version of the stage to check the cancer spread.
Grade	I- Tumor cells and tissue looks most like healthy cells and tissue.
Tumor size	indicating the size of the tumor and categorized into three groups. Measured in mm.
Survival time	survival of patients until an event of interest occurs measured in months.
Regional nodes examined	Records the total number of regional lymph nodes that were removed and examined by the pathologist.
Regional node positive	Records the exact number of regional nodes examined by the pathologist and found to contain metastases
Estrogen status	labeled positive and Negative if cancer has receptors for estrogen or not.
Progesterone status	labeled positive or negative if cancer has progesterone receptors or not.
Status	the event indicator, 1 for patients who deceased before the end of the study, and 0 for patients who were censored.

© University of Venda

University of Venda
Creating Future Leaders

3.6.2 Time to first alcohol intake dataset

The dataset used in this study was generously provided by Hulisani G. Sithuba, who gathered the data as part of his honors project in 2017. The data was collected from two tertiary institutions: a university and a vocational training college (VTC). A total of 745 students participated by completing a self-administered questionnaire designed to collect information on the demographic and socio-economic characteristics of the respondents, as shown in Table 3.1 (Bere et al., 2021). The questions from items 7 to 17 aimed to determine whether the respondents had experienced specific events between the ages of 1 and 15. The primary outcome variable in the study was the age at which individuals first consumed alcohol. Respondents who reported never having consumed alcohol were considered censored at the age they participated in the study.

3.6.3 CRASH-2 dataset

The CRASH-2 (Clinical Randomisation of an Antifibrinolytic in Significant Haemorrhage-2) trial was a randomized controlled study conducted across 274 hospitals in 40 countries. This international study involved $20,207$ trauma patients who were either experiencing or at risk of significant hemorrhage, determined by signs of hemodynamic instability. The dataset includes variables such as patient sex, age, the time between injury and hospital admission, injury type, systolic blood pressure, heart rate, respiratory rate, central capillary refill time, and the overall Glasgow Coma Scale score. The outcome of interest is the time until death for patients with bleeding trauma. Data from the CRASH-2 trial are accessible in the Hmisc package in R. A complete description of the explanatory variables can be found below.

© University of Venda

University of Venda
Creating Future Leaders

Table 3.2: CRASH-2 indepedent variables

Name	Description
sex	Gender of the patient
age	Age of the patient
injurytime	Hours Since Injury
injurytype	Type of injury
sbp	Systolic Blood Pressure mmHg
rr	Respiratory Rate /min
cc	Central Capillary Refille Time
hr	Heart Rate /min
gcs	Glasgow Coma Score Total
ddeath	Date of Death
cause	Main Cause of Death
scauseother	Description of Other Cause of Death
status	Status of Patient at Outcome if Alive

The data represented in Table 3.2 has 20207 observations and 44 variables, maximum number of NAs:17121.

3.7 Software of the study

The R-statistical package will be used to perform statistical analysis for the proposed study.

© University of Venda

University of Venda
Creating Future Leaders

3.8 Flow chart

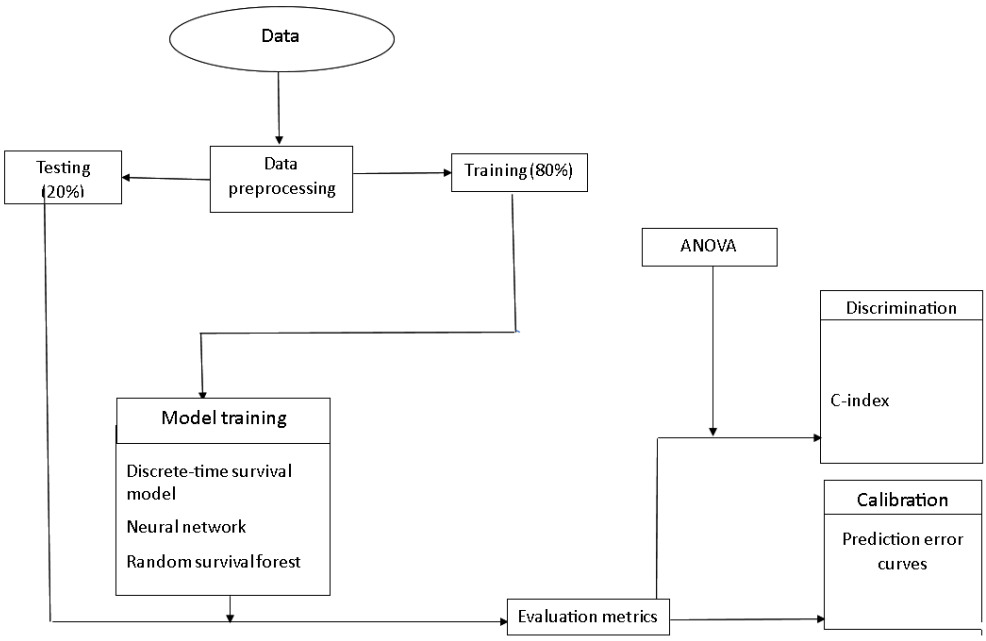

Figure 3.1: Visualisation of model building and testing process.

Figure 3.1 illustrates the flowchart of the data analysis plan for this research project. To obtain the results, we implemented three separate R scripts to analyze three distinct datasets: breast cancer patients, age at first alcohol intake, and CRASH-2 data. All analyses were conducted using the R statistical software (`RStudio`). The scripts followed the same procedure to ensure fair and unbiased comparisons across datasets.

Each dataset was cleaned and preprocessed to support discrete-time survival analysis and to allow fitting of the selected models. We then split each dataset into two subsets: an 80% training set and a 20% testing set, both initially in short format. These were converted into long format using the `dataLong()` function from the `discSurv` package, which restructures the data to enable time-interval-based modeling.

© University of Venda

University of Venda
Creating Future Leaders

Using the long-format training data, we fitted four types of models and repeated the analysis fifty times with different random samples to evaluate their discrimination and calibration performance. First, the discrete-time survival model was implemented as a generalized linear model using the `glm()` function from the `stats` package, with a binomial family and logit link to model the probability of an event occurring in each interval. Second, the discrete-time survival tree was fitted using the `ranger` package, employing the Hellinger distance as the split rule to handle imbalanced outcomes and improve class separation. Third, we used the same `ranger` package to fit the discrete-time random survival forest, also applying the Hellinger split rule to guide ensemble construction across multiple trees. Lastly, the discrete-time neural network was implemented using the `nnet` package, configured with one hidden layer containing 10 nodes (`size = 10`) and a maximum of 100 iterations (`maxit = 100`). For each model, predictions were made on the corresponding long-format and shot-format testing datasets, and performance was evaluated using the following discrimination and calibration metrics:

To asses the model fit we use the prediction error curves as a measure of calibration, and to evaluate predictive accuracy we use the Concordance index (C-Index) as measures of discrimination. To test for the significant difference between the models evaluation measures we use ANOVA. We use Kolmogorov-Smirnov test $ks.test()$ and Bartlett test $Bartlett.test()$ to asses if the ANOVA assumption are satisfied. If any of the assumptions is violated the use a non parametric test equivalent to ANOVA for the median pairwise comparison, the Kruskal-Wallis test which uses the function $kruskal.test()$. We further perform dunn test post hoc using the function $dunnTest()$ found in the FSA package to obtain the median pairwise comparison for the evaluation measures of the models. We also visualizes the evaluation measures using box-plots to asses if they align with the post hoc test results. Hence from the results we can conclude which model has better prediction accuracy.

© University of Venda

Chapter 4

Empirical results and Discussion

This chapter present data analysis using the models and methods discussed in the previous chapter.

4.1 Breast-cancer results

The breast cancer data was exported as microsoft excel file using the $read_excel()$ found in readxl package which is available in r statistical software. We computed a ratio of two variables, regional node examined and regional positive node to obtain Lymph Node Ratio (LNR), because LRN is a better predictor than just using regional positive node. We use a loop in r we discretized the data in half-yearly intervals to obtain the short format of the dataset, as to reduce the number of parameters of the discrete survival model to be estimated. The data is split into 80% training set and 20% testing set. Further we use the function $datalong()$ function in disSurv package to obtain long format training and testing sets. Hence the data was ready for model training and testing.

© University of Venda

University of Venda
Creating Future Leaders

4.1.1 Descriptive statistics for numeric data

This section present descriptive statistics for numeric variables.

Table 4.1: Summary statistics

Variable name	Event	Mean	Minimum	Maximum
Age	Dead(616)	55.15	30	69
	Alive(3408)	53.76	30	69
Tumor size	Dead(616)	37.14	1	140
	Alive(3408)	29.27	1	140
Lymph Node Ratio (LNR)	Dead(616)	-0.373	-0.919	3.75
	Alive(3408)	0.0674	-0.919	6.55
Survival Months	Dead(616)	45.61	2	102
	Alive(3408)	75.94	1	107

From table 4.1, indicates that the age of patients who died during the study period varied from 30 to 69 years, while the age of those who survived (censored) also fell within the same range of 30 to 69 years. The average age of patients who died was slightly above 55 years, whereas the average age of censored patients was just over 53 years. The mean survival duration for deceased individuals was 45 months, while the mean follow-up period for censored patients was 76 months. The mean tumor size for patients who succumbed during the study period is 37, while it is 29 for censored patients.

© University of Venda

University of Venda
Creating Future Leaders

4.1.2 Descriptive statistics for categorical data

Table **??** it is clear that out of 2042, $616(15.31\%)$ died during the period of the study and $3408(84.69\%)$ survived through out the period of the study. The majority of the patients were whites (85%), followed by other (8%), and blacks (7%). (82%) of patients who died during the study period were white, (12%) were black, and (6%) other race, (85%) of patients who survived were whites. Majority of the patients were married (66%).

4.1.3 Hazard function

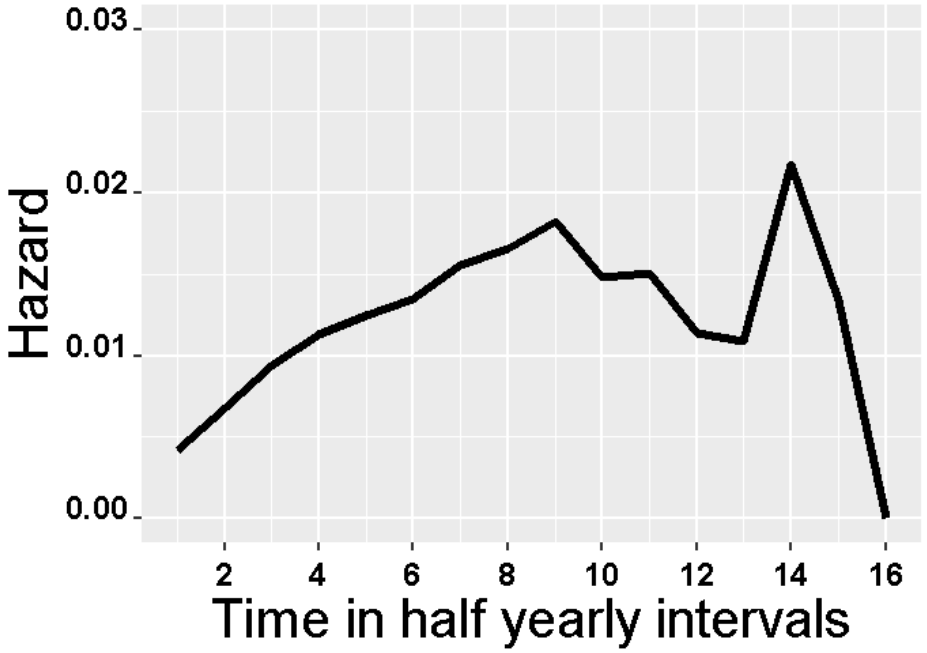

Figure 4.1: Hazard of breast cancer patients

Figure 4.1 visualises the overall hazard for patients with breast cancer in which time is discretized in half yearly intervals. Figure 4.1 shows that the hazard of dying from breast cancer has a peak at interval nine and fourteen.

© University of Venda

University of Venda
Creating Future Leaders

4.1.4 Discrete-time Survival Models

This section present discrete survival model

Table 4.2: Summary statistics for the discrete-time survival model

					Odds ratio CI	
Variable	Estimate	Std Error	Odds ratio	p-value	lower CI	upper CI
RaceOther	-1.25	0.30	0.29	0.00	0.15	0.51
RaceWhite	-0.52	0.15	0.60	0.00	0.45	0.81
N_StageN2	0.63	0.27	1.88	0.02	1.13	3.29
N_StageN3	1.29	0.20	3.65	0.00	2.49	5.39
Estrogen_StatusPositive	-0.70	0.17	0.49	0.00	0.35	0.70
MaritalMarried	-0.09	0.72	0.91	0.90	0.29	5.59
MaritalSingle	-0.25	1.25	0.78	0.84	0.04	8.42
Age	0.14	0.05	1.15	0.01	1.04	1.28
Grade2	0.46	0.21	1.58	0.03	1.06	2.47
Grade3	0.62	0.22	1.86	0.01	1.22	2.94
Gradeanaplastic; Grade IV	1.46	0.45	4.32	0.00	1.67	9.88
Tumor_Size	0.09	0.05	1.09	0.07	0.99	1.21
x6th_StageIIB	0.33	0.18	1.39	0.06	0.99	1.98
x6th_StageIIIA	-0.10	0.30	0.90	0.73	0.49	1.59
x6th_StageIIIB	0.63	0.38	1.89	0.10	0.86	3.83
Progesterone_StatusPositive	-0.58	0.13	0.56	0.00	0.43	0.73
Lymph Node Ratio (LNR)	-0.32	0.09	0.72	0.00	0.61	0.85

Looking at results in Table 4.2 we are 95% confident, that after controlling for the rest of the variables, for every year increase in age the odds of a women dying from breast cancer are 1.15 times higher, age is significant with a p-value (< 0.0001).

© University of Venda

University of Venda
Creating Future Leaders

After controlling for the rest of the variables, the odds of a woman with positive estrogen status dying from breast cancer are between 35 and 70% higher than those for a woman with negative estrogen status, the point estimate of the relative increase in the odds is 95%. After controlling for the rest of the variables, we are 95% confident that the odds of a women with positive progesterone status dying from breast cancer are between 43 and 73% lower than those for a woman with negative progesterone status. The odds of a women from other races dying from breast cancer are between 15% and 51% lower than those for a black woman, and the odds of a white woman dying from breast cancer are between 45% and 81% lower than those for a black women.

For every unit increase in Tumor-size the of a women dying from breast cancer are increased by 1.09, and for every unit increase in Lymph Node Ratio (LNR) the odds of a women dying from breast cancer are 0.72 times larger. The odds of a women with NStage-N2 dying from breast cancer are between 113% and 329%, and between 239% and 549% for a women with NStage-N3 higher than those NStage-N1 . The corresponding odds are 63% and 129% . The odds of a women with x6th-StageIIB dying from breast cancer are between 99% and 198%, and between 86% and 383% for women with x6th-StageIIIB higher than those of x6th-StageIIA. The corresponding odds are 49% and 107% respectively.

© University of Venda

University of Venda
Creating Future Leaders

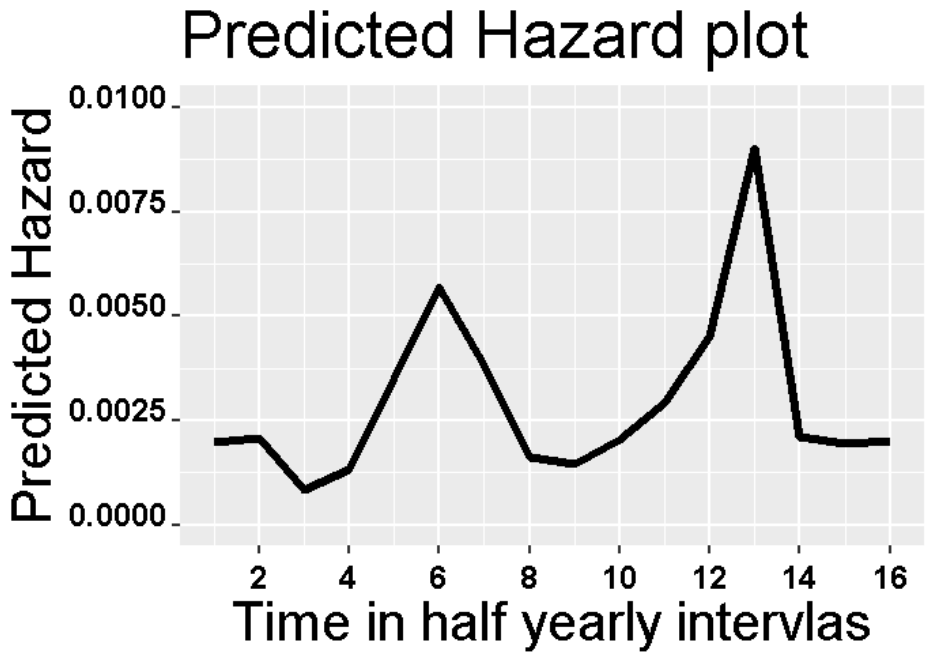

Figure 4.2: Hazard plot for Discrete-time survival Model

Figure 4.11 visualises the overall hazard for patients with breast cancer recorded in half-yearly intervals. The figure shows that the hazard of dying from breast cancer experienced a gradual rise from interval four to interval six, and rises again from interval nine to interval thirteen.

© University of Venda

University of Venda
Creating Future Leaders

4.1.5 Discrete-time survival tree

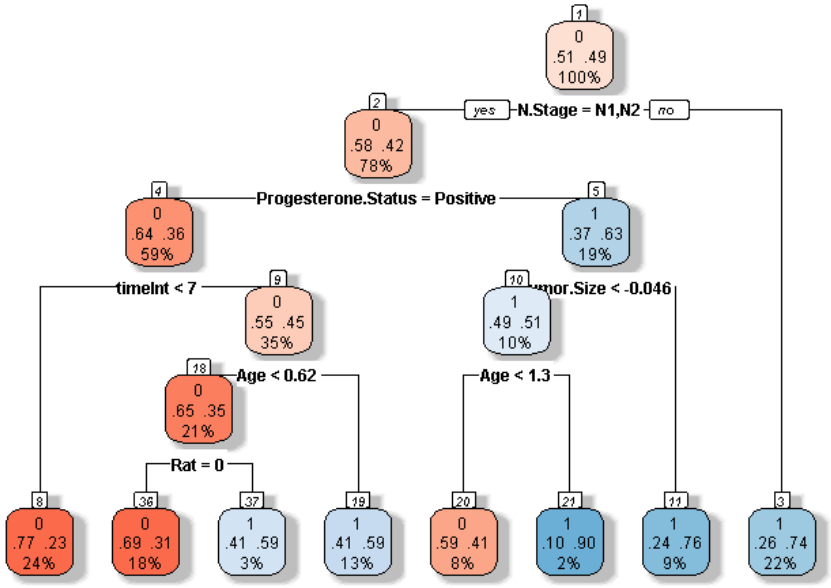

Figure 4.3: Discrete-time survival tree

Figure 4.3 displays the discrete-time survival tree derived from the breast cancer dataset, illustrated using the $rpart()$ and $rpart.plot()$ packages in R, the model uses gini-impurity as the splitting criterion. Each node contains a boolean question (classification rule). If the answer to a question from root node is yes, we move along the arrow to the left child node. Otherwise, if the answer is no, we move along the arrow to the right child node. The discrete-time survival tree begins with $1,000$ observations, with an almost even class distribution at the root node $(50.7\%$ in class 0 and 49.3%in class 1). The first major split is based on N.Stage $(N1, N2 vs. N3)$, indicating that the lymph node stage is a key determinant of classification. For individuals with N.Stage $= N1 or N2 (780 observations)$, the probability of class 0 increases to 57.7%. The next major split occurs based on Progesterone Status. If Progesterone.Status is Positive $(591 cases)$, further division occurs by timeInt (time-related variable) and tumor size.

© University of Venda

University of Venda
Creating Future Leaders

Patients with a Positive Progesterone Status and a time interval less than 6.5 have a significantly higher survival rate, with 76.7% Alive. This suggests that quicker treatment or a shorter time to intervention is associated with better survival outcomes. In this group, the survival rates are more balanced, with 55.5% Alive and 44.5% Dead. This suggests that for patients with Positive Progesterone Status, longer time intervals could be associated with more mixed outcomes, possibly reflecting disease progression. For patients with Positive Progesterone Status, time intervals of 6.5 or more, and younger age , survival rates are higher, with 69.1% Alive. Younger patients in this category have a relatively favorable prognosis. For older patients with Positive Progesterone Status and longer time intervals, the survival rate is lower than for younger patients, with 40.6% Dead. This suggests that age plays a significant role in survival, even within this Positive Progesterone Status group. For patients with Positive Progesterone Status and older age, the survival rate drops, with 40.6% Dead and 59.4% Alive. This again emphasizes that age is a key determinant of survival, even with Positive Progesterone Status.

In the Negative Progesterone Status group, survival outcomes are relatively more favorable than for the N3 stage. 63% of patients are Alive, but 37% are Dead, indicating a poorer prognosis compared to the Positive Progesterone group but not as poor as $N3$. Patients with Negative Progesterone Status and smaller tumor sizes (< -0.0461019) have a better survival rate, with 58.8% Alive. This suggests that smaller tumor sizes are associated with improved outcomes, even in the Negative Progesterone Status group. This small subgroup of older patients with Negative Progesterone Status and smaller tumors has an exceptionally poor survival outcome, with 90% Dead. Age appears to play a crucial role in worsening the prognosis even when the tumor size is smaller. In this branch, survival is nearly equal between Dead and Alive cases, with 51% Alive and 49% Dead. This suggests that the tumor size is not strongly predictive of survival in this case, though it's still important. Larger tumor sizes are associated with worse survival outcomes in

© University of Venda

patients with Negative Progesterone Status. Here, 76.4% are Dead, indicating that tumor size has a stronger negative impact on survival in this group. For patients in the $N3$ stage, survival is significantly poorer, with 74.1% Dead. This shows that the N3 stage is a critical factor in determining prognosis, and most patients in this group are expected to die.

4.1.6 Accuracy metrics

We present the discrimination (C-index) and callibration (Prediction error curves) measures used to asses the performance of four discrete-time Random forest, discrete-time survival tree, discrete-time survival model (glm), and discrete-time neural network models applied to the breast cancer data.

4.1.7 Prediction error curves

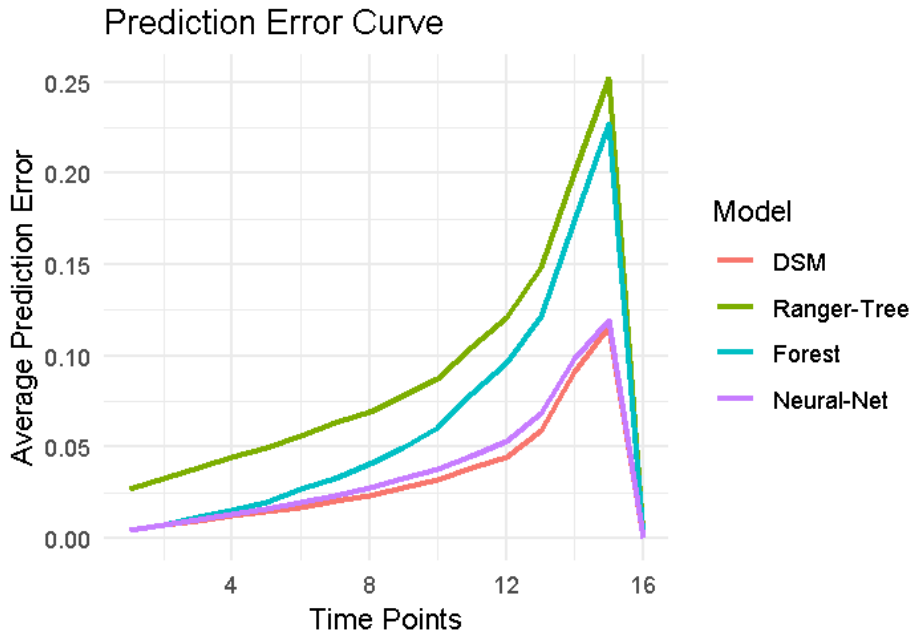

Figure 4.4: Breastcancer: Prediction error curves

© University of Venda

University of Venda
Creating Future Leaders

Figure 4.4 shows the visualised prediction errors for the models applied to the breast cancer data recorded in months nd further group in intervals of 18 . We utilised the $predErrcurve()$ function from the $DiscSurv$ package in r to obtain prediction errors for all models trained over 50 datasets by replacement. Further, we used the ggplot package to visualise the prediction errors. Looking at the Figure 4.4 we see that all the prediction error curves experience a rise between time point 1 and 2, and all have values below 0.25, hence we say the models predict the data accurately . The glm and neural network appears to have lower prediction errors as compared to the other models. The ranger tree prediction error curve shows that the model has higher prediction errors as compared to the other models with a peak at time point 15. We conclude that the neural network outperforms the rest of the models followed by the glm, and that the ranger tree has the worst calibration performance.

4.1.7.1 C-index

Table 4.3: ANOVA for breast cancer cindex

	Df	Sum Sq	Mean Sq	F-vale	Pr(F)
Model	3	0.8135	0.27115	383.3	2e-16
Residuals	196	0.1387	0.00071		

Table 4.3 shows ANOVA for breast cancer I-Index. The The p-value for ANOVA test is $< 2e - 16$, thus we fail to accept the null hypothesis and conclude that there's some difference in the means of the models. Further we perform the post-hoc test for the pairwise comparison, but first we check if the ANOVA assumption are satisfied using the following

© University of Venda

University of Venda
Creating Future Leaders

Table 4.4: K-S, Bartlett test and krustal wallies test for breast cancer cindex

Test	P.adj
Asymptotic Kalmogorov test	0.01198
Bartlett test of homogeneity of variances	0.0004213
Krustal-wallis test	$2.2e-16$

Table 4.4 $p-value < 0.01198$ (Kolmogorov-Smirnov test), we fail to accept the null hypothesis and conclude that the data is normally distribution. Also the Bartlett test of homogeneity of variances results in a $p-value < 0.0004213$, hence we reject the null hypothesis and conclude the variances are not equal across the different models. The $p-value(< 2.2e-16)$ of the Krustal-wallis test, we reject the null hypothesis and conclude that atleast one group has a different distribution.

Table 4.5: Comparison of C-Index Values

Comparison	Z	P.unadj	P.adj
Discrete-time neural network - Discrete-time survival model	7.29	0.00	0.00
Discrete-time neural network - Discrete-time Random forest	3.54	0.00	0.00
Discrete-time survival model - Discrete-time Random forest	-3.76	0.00	0.00
Discrete-time neural network - Discrete-time Random forest	12.24	0.00	0.00
Discrete-time survival model - Discrete-time survival tree	4.94	0.00	0.00
Discrete-time Random forest - Discrete-time survival tree	8.69	0.00	0.00

Table 4.6 shows the pairwise median comparison for breast cancer models C-Index. Figure 4.6 presents strong evidence $(p < 0.00)$ of the difference across all C-Index median pairwise comparison. The results shows that discrete-time neural network model has the highest median throughout the pairwise comparison,

© University of Venda

University of Venda
Creating Future Leaders

followed by discrete-time Random forest model. Discrete-time survival tree has the lowest median across all median comparison, followed by median of discrete-time survival model.

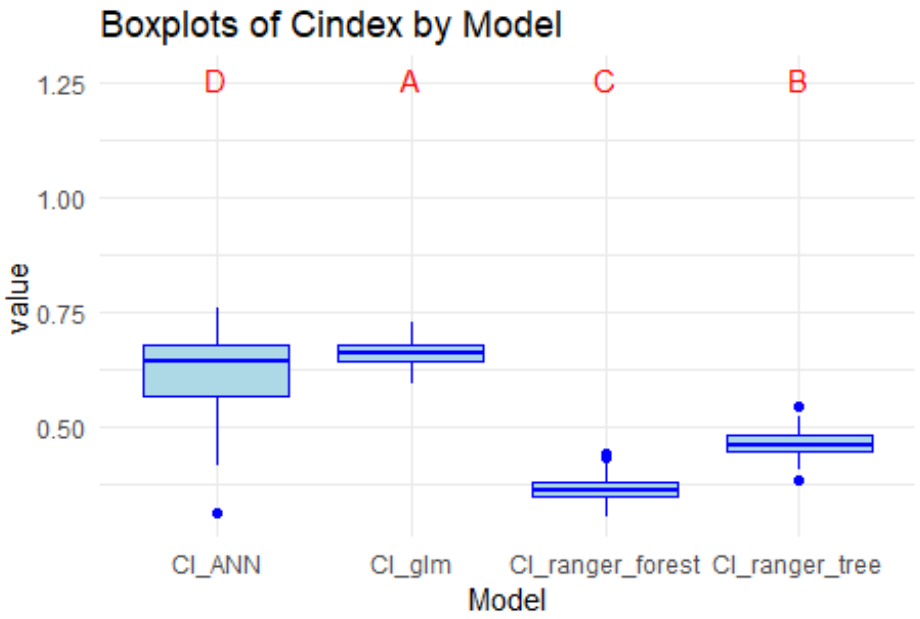

Figure 4.5: Boxplot for Breast-cancer cindex

Figure 4.5 displays the box-plots of the models C-Index obtained from dunn-test. The plot shows that box-plot of discrete-time neural network has the highest median as compared to rest of the models, followed by median boxplot of discrete-time Random forest. Box-plot of discrete-time survival tree has the lowest median box-plot as compared to all models, followed by median box-plot of discrete-time survival model. The results in Table 4.6 and Figure 4.5 show that the discrete-time neural network has highest c-Index median as compared to the of rest of the models.

© University of Venda

University of Venda
Creating Future Leaders

4.2 Alcohol dataset

Age at first alcohol intake data was imported into r-studio using haven package also available in the software. We utilized a loop to discretize the age at first alcohol intake data into intervals of three years, as to reduce the number of parameters of the discrete survival model to be estimated.

4.2.1 Hazard function

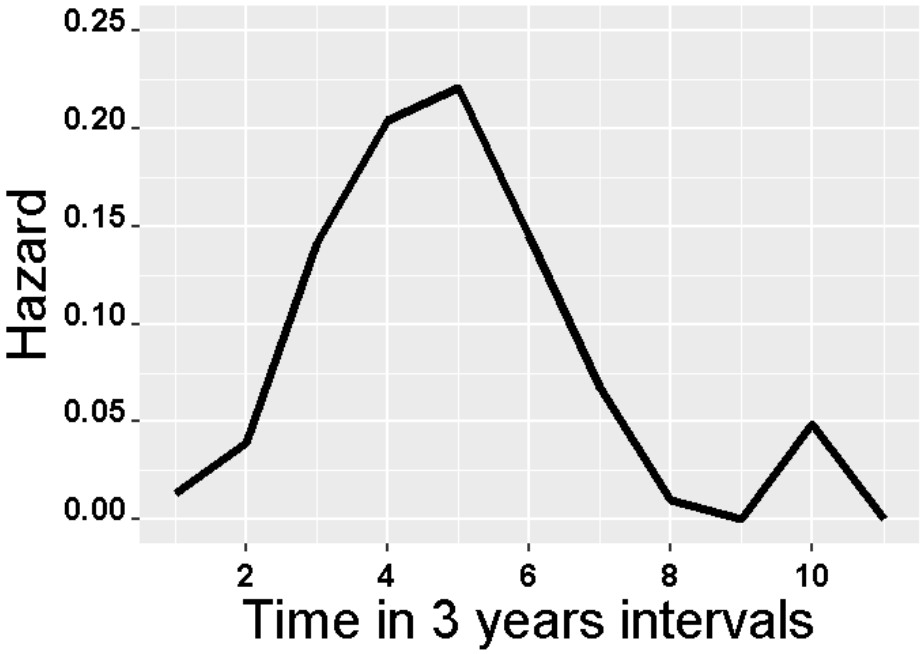

Figure 4.6: Alcohol Hazard function

Figure 4.6 shows the hazard plot for the age at first alcohol intake data-set, we grouped the data into 10 three yearly intervals . Figure 4.6 shows that the chances of experiencing first alcohol intake are predicted to rise from time interval 1 to 5 with a peak at interval 5 then there's a decline. A secondary peak appears at interval 10.

© University of Venda

University of Venda
Creating Future Leaders

4.2.2 Descriptive statistics

Table 4.6: Independent Variables considered for the analysis of time to first alcohol intake

Variable name	Category	Frequency(%)
Gender	Male	396(53.15)
	Female	349(46.85)
Ethnicity	Black	729(97.90)
	Otherace	16(2.09)
Drinking Peers	Yes	318(42.68)
	No	427(57.32)
Physical abuse	Yes	60 (8.05)
	No	685(91.95)
Parental drug/alcohol abuse	Yes	344(46.17)
	No	401(53.83)

Table 4.6 indicates that the sample was evenly distributed regarding gender and childhood place of residence, yet predominantly comprised students from the university. A significant proportion of the respondents had not experienced marriage. Based on the low family incomes, the presence of numerous siblings, and a significant proportion of parents lacking formal education, it can be concluded that the socio-economic status of most respondents is not high. The prevalent experiences among individuals aged one to 15 years included having peers who engage in drinking (43%), maintaining a confiding relationship with a single adult (43%), experiencing stress (89%), having parents who abused alcohol or other drugs (47%), and being subjected to parental discipline (59%). The prevalence of

© University of Venda

University of Venda
Creating Future Leaders

other drugs and cigarettes was relatively low, at approximately 15%. (Bere et al., 2021)

4.2.3 Discrete-time Survival Models

This section present discrete survival model. Due to the large number of variables, we have opted to use variables selected through the use of variable selection by (Bere et al., 2021).

Table 4.7: Discrete-time survival model

Variable	Estimate	Std Error	p-value	Odds ratio	lower CI	upper CI
(Intercept)	-2.87	0.34	0.00	0.06	0.03	0.11
as.factor(timeInt)2	0.86	0.35	0.02	2.36	1.21	4.87
as.factor(timeInt)3	2.59	0.31	0.00	13.32	7.56	25.74
as.factor(timeInt)4	2.81	0.31	0.00	16.66	9.42	32.29
as.factor(timeInt)5	3.17	0.31	0.00	23.92	13.46	46.54
as.factor(timeInt)6	2.52	0.33	0.00	12.44	6.68	24.99
as.factor(timeInt)7	1.77	0.39	0.00	5.86	2.72	12.94
as.factor(timeInt)8	0.32	0.77	0.68	1.37	0.21	5.15
as.factor(timeInt)9	-12.08	320.42	0.97	0.00	0.00	0.00
as.factor(timeInt)10	-12.07	433.09	0.98	0.00	0.00	0.00
SexFemale	-0.51	0.11	0.00	0.60	0.48	0.75
OtherDrNo	-0.71	0.16	0.00	0.49	0.36	0.67
DrPeersNo	-0.20	0.12	0.09	0.82	0.65	1.03
NegetiveLENo	-0.05	0.13	0.71	0.95	0.75	1.23
PhysicAbNo	-0.27	0.19	0.16	0.77	0.53	1.12
PDrIntakeNo	-0.24	0.11	0.03	0.78	0.63	0.98

© University of Venda

University of Venda
Creating Future Leaders

Table 4.7 provides the summary results for the discrete-time survival model. The results show that time intervals 2 through 7 show a strong increase in the odds of the outcome occurring, with odds ratios ranging from $2.36 to 23.92$, all statistically significant with p-value < 0.05. In particular, time intervals $3, 4, and 5$ have the largest odds ratios, indicating that the likelihood of the event occurring increases substantially in these time periods. In contrast, time intervals 8 through 11 have odds ratios near 1 or close to 0, with non-significant p-values p-value > 0.05, suggesting no meaningful effect in these time intervals. Gender also plays a significant role, with females having 0.60 times the odds of the event occurring compared to males, which is statistically significant (p-value < 0.05). Similarly, individuals who did not visit a doctor (odds ratio = 0.49) or who did not follow doctor intake recommendations (odds ratio $= 0.78$) were found to have lower odds of the outcome. Both of these factors are statistically significant (p-value < 0.05). However, other factors such as the presence of peers or negative life events do not show significant effects, with p-values indicating no meaningful influence on the outcome. Overall, time intervals $2 - 7$, female sex, and doctor-related behaviors (doctor visit and intake) were the most significant predictors of the outcome in this model.

© University of Venda

University of Venda
Creating Future Leaders

4.2.4 Discrete-time survival tree

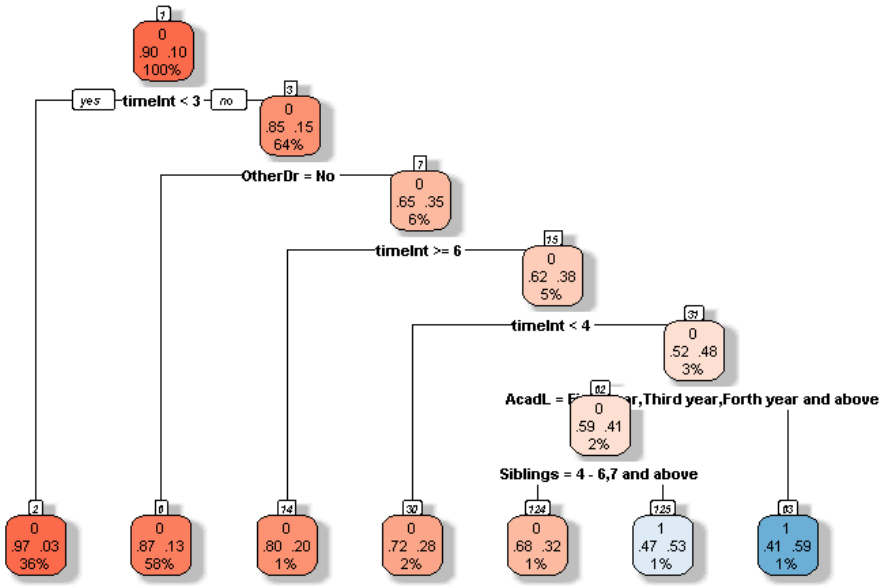

Figure 4.7: Alcohol: classification tree

Figure 4.7 shows the discrete-time survival tree derived from the age at first alcohol intake is illustrated using the $rpart()$ and $rpart.plot()$ packages in R, the model uses gini-impurity as the splitting criterion. Each node contains a boolean question (classification rule). If the answer to a question from root node is yes, we move along the arrow to the left child node. Otherwise, if the answer is no, we move along the arrow to the right child node. The decision tree starts with a total of $4,147$ observations at the root node, where the majority (89.5%) belong to class 0. The first major splitting factor is timeInt, representing a time-related variable. If timeInt < 2.5, the probability of class 0 increases to 97.4%, making this a terminal node with very low risk. However, if timeInt $>= 2.5$, the classification process continues, with $2,667$ remaining observations and a slightly higher proportion (14.8%) classified as class 1. For individuals with timeInt $>= 2.5$, the next important factor is OtherDr, indicating whether other drugs was involved. If no other drug was involved, the

© University of Venda

University of Venda
Creating Future Leaders

probability of class 0 remains high (87.3%), and this branch ends as a terminal node. However, if another drug was involved, the risk of class 1 increases, with 35.3% of cases falling into this category. Among these individuals, those with timeInt $>= 5.5$ still belong to class $0(80\%)$, while those with timeInt < 5.5 show a higher probability (38.2%) of class 1.

Further down, timeInt < 3.5 is associated with a greater likelihood of class $0(72.3\%)$, while those with timeInt $>= 3.5$ experience nearly an equal split between the two classes $(52.3\% vs. 47.7\%)$. Within this latter group, academic level (AcadL) becomes a key determinant. Students in their first, third, or fourth year have a 41.4% probability of class 1, but this probability varies depending on the number of siblings. Those with four or more siblings are more likely to belong to class $0(67.5\%)$, while those with fewer siblings $(0 - 3)$ show a reversal, with a 53.3% likelihood of class 1. Meanwhile, students in their second year have the highest probability of class $1(58.5\%)$, marking this as a terminal node. Overall, the tree highlights timeInt as the most influential factor, followed by doctor involvement, academic level, and sibling count. The model suggests that early time points are strongly associated with class 0, whereas later time points, combined with external factors, increase the likelihood of class 1.

4.2.5 Accuracy metrics

We present the discrimination (C-Index) and callibration (Prediction error curves) measures used to asses the performance of four discrete-time Random forest, discrete-time survival tree, discrete-time survival model, and discrete-time neural network models applied to the age at first alcohol intake data.

© University of Venda

4.2.6 Prediction error curves

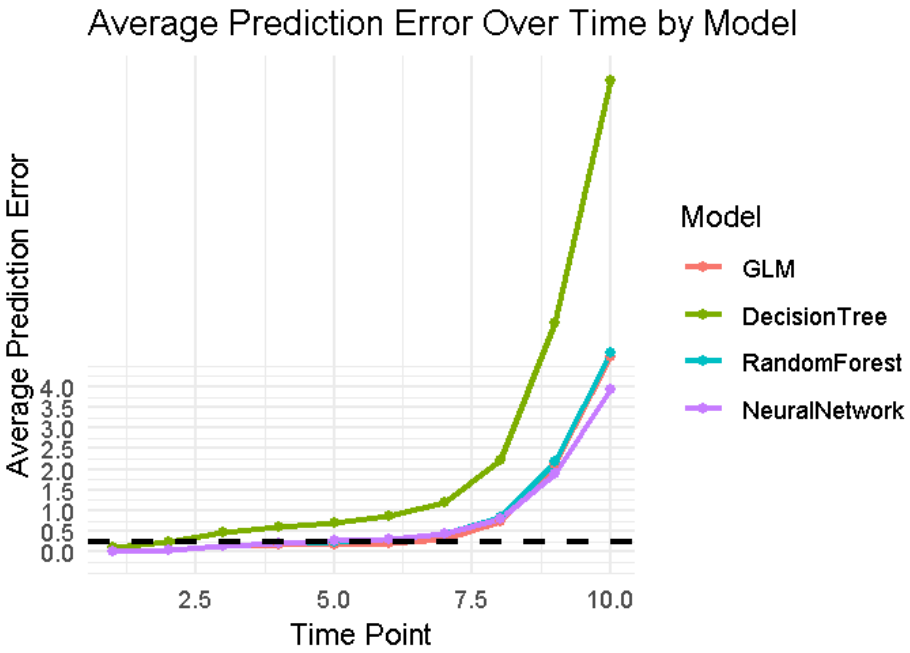

Figure 4.8: Alcohol: Prediction error curves

Figure 4.8 shows the prediction error curves for the models applied to alcohol data which is grouped into 3 years interval (time points). Figure 4.8 shows that three models (discrete-time random survival forest, discrete-time neural network, and discrete-time survival model) seem to predict the data well with values below 0.25 between time point 1 and 6 and rise above 0.25 for the rest of the time points. The decision tree have prediction errors rising above 0.25 from time point 2.5 and throughout.The discrete-time random forests, discrete-time neural network, and discrete-time survival model appear to have consistent lower prediction over each time point. The discrete-time survival tree have higher prediction error at each time point compared to the rest of the models. The discrete-time random survival forest prediction error curve appears to have lower prediction errors compared curves of other models, that is the discrete-time random survival forest performs better as compared to other models, and the discrete-time survival tree appear to perform

© University of Venda

University of Venda
Creating Future Leaders

worse compared to other models.

4.2.6.1 C-Index

Table 4.8: ANOVA for alcohol C-Index

	Df	Sum Sq	Mean Sq	F-vale	Pr(F)
Model	3	3.219	1.0730	809.5	2e-16
Residuals	196	0.260	0.0013		

Table 4.8 shows ANOVA for alcohol C-Index. The $p - value < 2e - 16$, thus we fail to accept the null hypothesis and conclude that there's some difference in the means of the models. Further we perform the post-hoc test for the pairwise comparison, but first we check if the ANOVA assumption are satisfied using the following

Table 4.9: K-S, Bartlett test and krustal wallies test for alcohol cindex

Test	P.adj
Asymptotic Kalmogorov test	$3.689e - 07$
Bartlett test of homogeneity of variances	$4.8e - 06$
Krustal-wallis test	2.2e-16

Table 4.9 shows that $p - value < 3.689e - 07$ (Kolmogorov-Smirnov test), we fail to accept the null hypothesis and conclude that the data is normally distribution. Also the Bartlett test of homogeneity of variances results in a $p - value < 4.8e - 06$, hence we reject the null hypothesis and conclude the variances are not equal across the different models. The $p - value < 2.2e - 16$ of the Krustal-wallis test, we reject the null hypothesis and conclude that atleast one group has a different distribution.

© University of Venda

University of Venda
Creating Future Leaders

Table 4.10: Comparison of C-Index Values

Comparison	Z	P.unadj	P.adj
Discrete-time neural network - Discrete-time survival model	6.95	0.00	0.00
Discrete-time neural network - Discrete-time random survival forest	-0.50	0.62	0.62
Discrete-time survival model - Discrete-time random survival forest	-7.46	0.00	0.00
Discrete-time neural network - Discrete-time survival tree	9.82	0.00	0.00
Discrete-time survival model - Discrete-time survival tree	2.87	0.00	0.00
Discrete-time random survival forest- Discrete-time survival tree	10.32	0.00	0.00

Table 4.10 shows the pairwise median comparison for breast cancer models C-Index. Figure 4.6 presents strong evidence $(p < 0.00)$ of the difference across all C-Index median pairwise comparison, except for the median difference between discrete-time neural network and discrete-time random survival forest with p-value (0.62). The results shows that discrete-time random survival forest has the highest median throughout the pairwise comparison, followed by discrete-time neural network. discrete-time survival tree has the lowest median across all median comparison, followed by median of discrete-time survival model.

© University of Venda

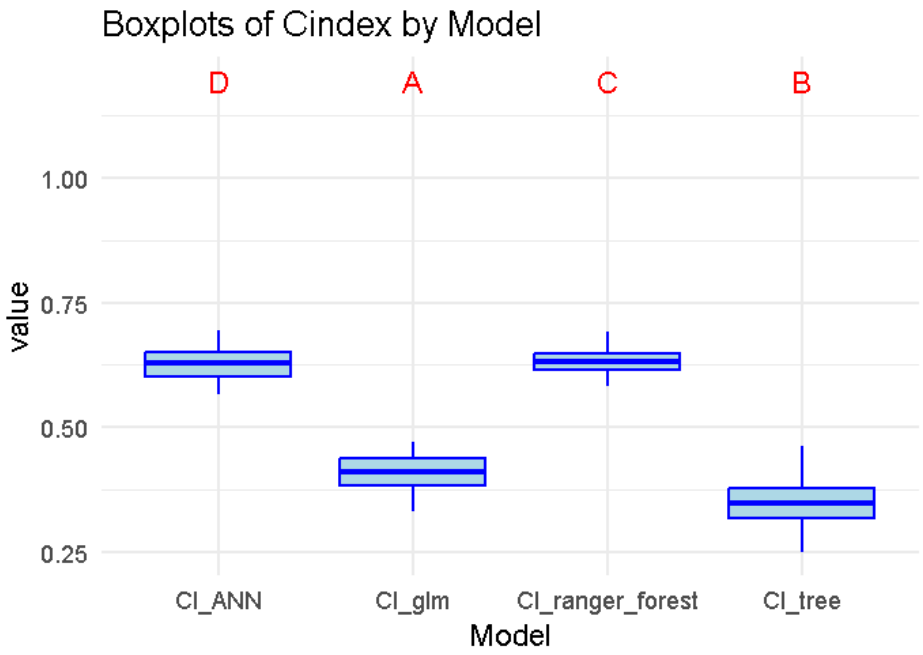

Figure 4.9: Boxplot for alcohol C-Index

Figure 4.9 displays the box-plots of the models C-Index obtained from dunn-test. The figure shows that box-plot of discrete-time random survival forest has the highest median as compared to rest of the models, followed by median boxplot of discrete-time neural network. Box-plot of discrete-time survival tree has the lowest median box-plot as compared to all models, followed by median box-plot of discrete-time survival model. The discrete-time random survival forest has the overall highest C-Index.

4.3 Crash-2

CRASH-2 trial data are available in the package Hmisc in r-statistical software. Additionally the data is available in the discSurv package. The data are available in discrete-time intervals, death times were discretized into intervals of five days.

© University of Venda

University of Venda
Creating Future Leaders

4.3.1 Hazard function

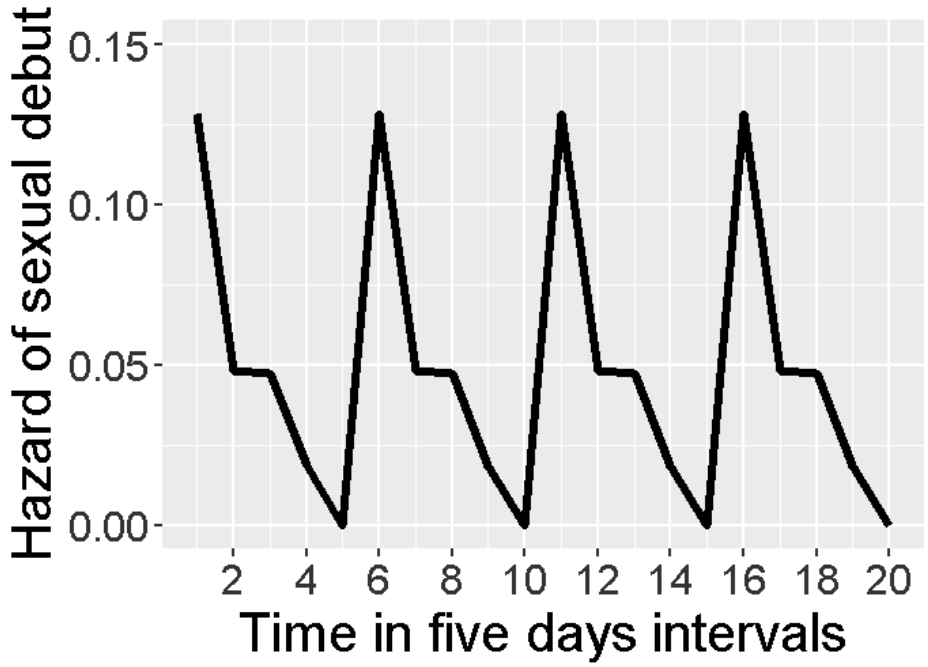

Figure 4.10: Crash2 Hazard function

Figure 4.10 shows the hazard for the crash-2 data recorded in five days intervals . The plot shows that the highest risk of dying is at interval one and then there's a decline. The risk of dying increase again from interval five to six, interval ten to twelve, and from interval thirteen to sixteen. The plot shows that the risk of dying for the CRASH-2 contains four peak points at interval one, six, eleven and sixteen.

© University of Venda

University of Venda
Creating Future Leaders

4.4 Descriptive statistics

Table 4.11: Patient Characteristics by Time Since Injury

Characteristic	≤1 hour (n = 7451)	1–3 hours (n = 6033)	>3 hours (n = 6634)
Age (years)			
Mean (SD)	33.4 (13.9)	35.0 (14.0)	35.5 (14.8)
<25	2283 (30.6%)	1557 (25.8%)	1773 (26.7%)
25–34	2360 (31.7%)	1832 (30.4%)	1882 (28.4%)
35–44	1356 (18.2%)	1177 (19.5%)	1262 (19.0%)
>44	1452 (19.5%)	1467 (24.3%)	1716 (25.9%)
BP (mmHg)			
≤75	1380 (18.5%)	1012 (16.8%)	768 (11.6%)
76–89	1203 (16.1%)	1064 (17.6%)	1029 (15.5%)
>89	4857 (65.2%)	3955 (65.6%)	4821 (72.7%)
Heart rate (beats per minute)			
<77	681 (9.1%)	450 (7.5%)	603 (9.1%)
77–91	1189 (16.0%)	971 (16.1%)	1326 (20.0%)
92–107	1888 (25.3%)	1562 (25.9%)	1625 (24.5%)
>107	3637 (48.8%)	2990 (49.6%)	3059 (46.1%)
Respiratory rate (per minute)			
<10	149 (2.0%)	82 (1.4%)	77 (1.2%)
10–29	6144 (82.5%)	4992 (82.7%)	5590 (84.3%)
>29	1077 (14.5%)	901 (14.9%)	923 (13.9%)

4.4.1 Discrete-time Survival Models

This section present discrete survival model

© University of Venda

University of Venda
Creating Future Leaders

Table 4.12: Summary statistics for the discrete-time survival model

Variable	Estimate	Std Error	p-value	Odds ratio	lower CI	upper CI
age	0.01	0.01	0.001	1.02	1.01	1.02
sexfemale	0.23	0.20	0.24	1.26	0.84	1.85
injurytime	-0.09	0.03	0.01	0.91	0.84	0.97
injurytypepenetrating	0.33	0.19	0.09	1.39	0.94	2.038
injurytypeblunt and penetrating	0.31	0.21	0.14	1.36	0.89	2.06
sbp	-0.00	0.00	0.29	0.99	0.99	1.00
hr	0.00	0.00	0.19	1.00	0.99	1.01
rr	0.04	0.00	7.95e-06	1.04	1.02	1.06
cc	0.04	0.04	0.33	1.04	0.95	1.14
gcs	-0.24	0.01	2e-16	0.78	0.75	0.81

Looking at Table 4.12 we are 95% confident, that after controlling for the rest of the variables, For every 1-year increase in age, the odds of the outcome increase by 2% . The intercept is not statistically significant. It represents the baseline odds when all predictor variables are zero. Being female is associated with 26% higher odds of the outcome, but the confidence interval is between (0.84 and 1.85), that is the results are not statistically significant. A longer injury time decreases the odds of the outcome $(OR1)$. For every 1-hour increase in injury time, the odds decrease by 9%. Since the confidence interval is in between 84% and 97%, this is statistically significant. Penetrating injuries increase the odds of the outcome by 39%, but it is not statistically significant $(p\, 0.05)$. For every unit increase in respiratory rate, the odds of the outcome increase by 4%. A unit increase in GCS is associated with a 22% decrease in the odds of the outcome. Since $OR1$, higher GCS scores reduce the risk.

© University of Venda

University of Venda
Creating Future Leaders

4.4.2 Discrete-time survival tree

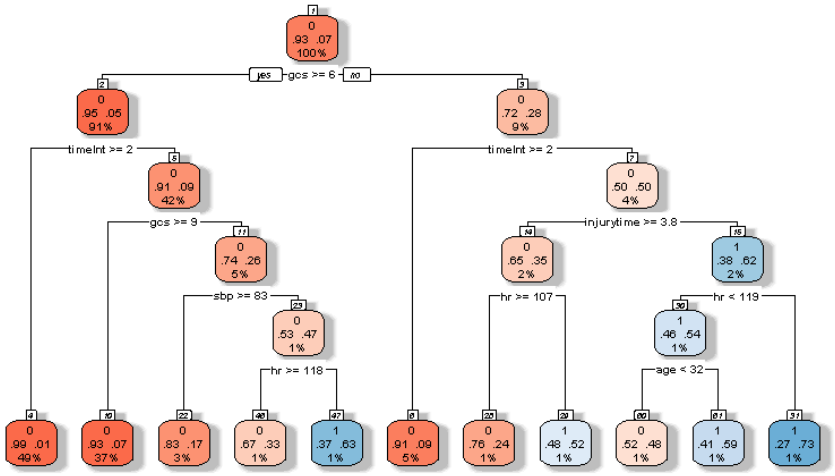

Figure 4.11: Discrete-time survival tree

The discrete-time survival tree analysis identifies Glasgow Coma Scale (GCS) as the most critical predictor of classification. Patients with a GCS score of 6.5 or higher are overwhelmingly classified as $0(93.3\%)$, suggesting a better outcome. However, if GCS is below 6.5, the probability of being classified as $1(26.9\%)$ increases, indicating a higher risk. Among this group, time interval (timeInt) between injury and medical intervention is an important factor. If timeInt is 1.5 or greater, the probability of class 0 remains high (91%), while those with timeInt below 1.5 experience a higher risk (45.2%). Further stratification based on SBP (systolic blood pressure) shows that lower SBP (< 92.5) increases the likelihood of class $1(48.4\%)$, especially when the injury time is short (< 3.5), pushing the probability of class 1 to 59.5%. The most severe cases occur when GCS is below 3.5, where the probability of class 1 jumps to 64.4%. Overall, the model highlights that lower GCS scores, shorter response times, and lower SBP significantly increase the likelihood of class 1, emphasizing the importance of rapid medical intervention and stable

© University of Venda

physiological indicators in patient outcomes.

4.4.3 Accuracy Metrics

We present the discrimination (C-nIdex) and callibration (Prediction error curves) measures used to asses the performance of four discrete-time Random forest, discrete-time survival tree, discrete-time survival model (glm), and discrete-time neural network models applied to the CRASH-2 dataset.

4.4.3.1 Prediction error curves

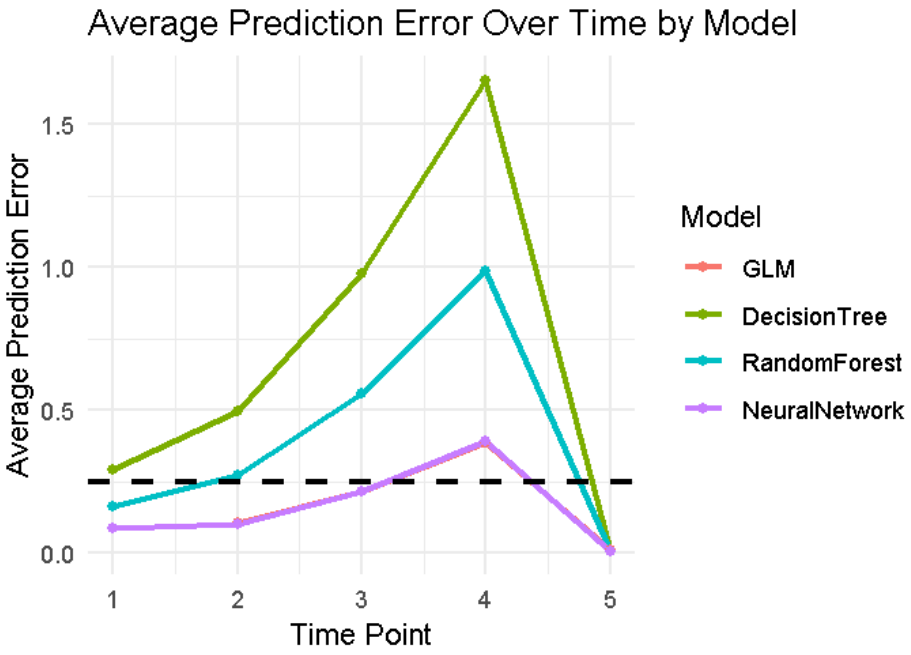

Figure 4.12: crash-2: Prediction error curves

Figure 4.12 shows the prediction error curves for the models applied to crash-2 data. The data is originally discretized into five days intervals. Figure 4.12 shows that although the prediction errors for discrete-time survival model (GLM) and

© University of Venda

University of Venda
Creating Future Leaders

discrete-time neural network appear to be above 0.25 in between time point 3 and 4, it is clear that these models have lower prediction errors which are less than 0.25 over other time points and they appear to have the same curve shape. The random forest and ranger tree have higher prediction errors which are above 0.25, with the random forest having significant predictions errors between time point 1 and 2, both also at between time point 4 and 5. The GLM curve appears to have lower prediction errors compared curves of other models, that is the GLM performs better as compared to other models followed by the neural network, and the random forest appear to perform worse compared to other models followed by the ranger tree which performs the least.

4.4.3.2 c-Index

Table 4.13: ANOVA: CRASH-2 cindex

	Df	Sum Sq	Mean Sq	F-vale	Pr(F)
Model	3	4.377	1.4591	62.77	2e-16
Residuals	196	4.556	0.0232		

Table 4.13 shows the ANOVA for the CRASH-2 c-Index. The $p - value < 2e - 16$, thus we fail to accept the null hypothesis and conclude that there's some difference in the means of the models. Further we perform the post-hoc test for the pairwise comparison, but first we check if the ANOVA assumption are satisfied using the following

© University of Venda

University of Venda
Creating Future Leaders

Table 4.14: K-S, Bartlett and krustal-Wallis test: crash2 c-Index

Test	P.adj
Asymptotic Kalmogorov test	$< 2.2e-16$
Bartlett test of homogeneity of variances	$< 2.2e-16$
Krustal-Wallis test	$< 2.2e-16$

Table 4.14 $p-value < 2.2e-16$ (Kolmogorov-Smirnov test), we fail to accept the null hypothesis and conclude that the data is normally distribution. Also the Bartlett test of homogeneity of variances results in a $p-value < 2.2e-16$, hence we reject the null hypothesis and conclude the variances are not equal across the different models. The $p-value < 2.2e-16$ of the Krustal-wallis test, we reject the null hypothesis and conclude that atleast one group has a different variance.

Table 4.15: Comparison of C-index Values

Comparison	Z	P.unadj	P.adj
Discrete-time neural network - discrete-time survival model	-3.86	0.00	0.00
Discrete-time neural network - discrete-time random survival forest	4.99	0.00	0.00
Discrete-time survival model - discrete-time random survival forest	8.85	0.00	0.00
Discrete-time neural network - discrete-time survival tree	8.32	0.00	0.00
Discrete-time survival model - discrete-time survival tree	12.18	0.00	0.00
Discrete-time random survival forest - discrete-time survival tree	3.34	0.00	0.00

Table 4.15 shows the pairwise median comparison for breast cancer models C-Index. The Table 4.15 presents strong evidence $(p < 0.00)$ across all C-Index median pairwise comparison. The results shows that discrete-time survival model has the highest median throughout the pairwise comparison, followed by

© University of Venda

discrete-time neural network. discrete-time survival tree has the lowest median across all median comparison, followed by median of discrete-time random survival forest.

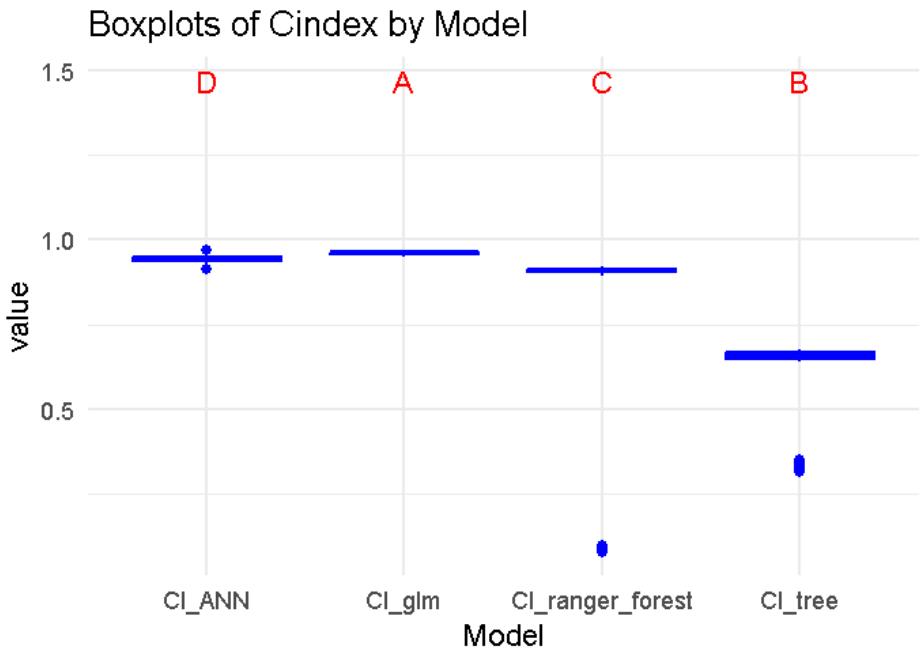

Figure 4.13: Boxplot for crash2 cindex

Figure 4.13 displays the box-plots of the models C-Index obtained from dunn-test. The graph shows that box-plot of discrete-time survival model has the highest median as compared to rest of the models, followed by median boxplot of discrete-time neural network. Box-plot of discrete-time survival tree has the lowest median box-plot as compared to all models, followed by median box-plot of discrete-time random survival forest . The discrete-time survival model has the overall highest C-Index.

© University of Venda

University of Venda
Creating Future Leaders

4.5 Discussion

From the analysis of breast cancer data set the results shows that the discrete-time neural network has lower prediction error and the highest C-Index median as compared to the of rest of the models. Hence, we conclude that discrete-time neural network had better discrimination and calibration prediction accuracy as compared to the other models when applied to the breast cancer data. For age at first alcohol intake the results shows that the discrete-time random forests, discrete-time neural network, and discrete-time survival model appear to have consistent lower prediction over each time point. The discrete-time neural network appears to have lower prediction errors as compared to other models and the highest C-Index median as compared to the of rest of the models. The discrete-time random forest outperformed the rest of the models for the age at first alcohol intake in terms of discrimination and callibration measures.

For the CRASH-2 data the results shows that the discrete-time surval model (glm) appear to have consistent lower prediction errors over each time point as compared to the rest of the models and the highest C-Index as compared to the rest of the models. Hence, we conclude that the discrete-time survival model outperforms the other models in terms of their discrimination and calibration measures. The findings from the three dataset shows that the performance of the models varies based on the dataset employ. This findings are consistent with Suresh et al. (2020) and (Cuthbert et al., 2022). Using the hellinger distance as the splitting criteria in ranger has improved the performance of the discrete-time random survival forest. This findings agree with the results from Schmid et al. (2022).

© University of Venda

University of Venda
Creating Future Leaders

Chapter 5

Conclusion and recommendations

5.1 Introduction

This section presents the conclusion of the study based on the findings.

5.2 Conclusion

This study presented comparison of the calibration and discrimination capabilities of the discrete-time survival model alongside various machine learning models, including the discrete-time survival tree, discrete-time survival forest, and discrete-time neural networks. The analysis was conducted using three distinct secondary data sets: the breast cancer data, the age at first alcohol intake data, and the CRASH-2 data. The analysis of the breast cancer data indicates that the discrete-time neural network demonstrated superior performance compared to other models, evidenced by their discrimination and calibration measures, a lower prediction error curve and the highest concordance index as determined through median pairwise comparison. Hence we conclude that discrete-time random forest best predicts survival of breast cancer compared to other models, in terms of

© University of Venda

University of Venda
Creating Future Leaders

discrimination and calibration measures.

The analysis of the age at first alcohol intake indicate that the discrete-time neural network demonstrated the lowest prediction errors through the inspection of the visualised prediction error curves when compared to the other models. The discrete-time neural network also demonstrated the highest concordance index as determined through median pairwise comparison. The discrete-time neural network demonstrated superior performance compared to the other models regarding the age at first alcohol intake. The findings derived from the CRASH-2 data set indicate that the discrete-time survival model demonstrate consistently lower prediction errors across each time point. The discrete-time survival model demonstrates the lowest prediction errors when compared to the other models. The discrete-time survival model (GLM) exhibits the highest C-index in comparison to the other models. The discrete-time survival model demonstrates superior performance compared to the machine learning models when evaluated using discriminative and calibration measures. The results are both informative and meaningful; however, we conclude that performance does indeed vary based on the data set utilized. Hence we encourage more studies to be conducted using discrete-time survival analysis, particularly with the healthcare set up, to contribute valuable insights to the field and potentially influencing future choices.

5.3 Recommendations

Future research could involve a comparative analysis of traditional statistical models and machine learning models. Additionally, incorporating extensions of generalized linear models into this comparison with machine learning models may enhance the effectiveness of predicting survival in oncology. A study utilizing a primary health data set will be advantageous, as it will provide clinicians with insights into the

© University of Venda

University of Venda
Creating Future Leaders

analyzed problems and facilitate the development of improved solutions moving forward.

© University of Venda

University of Venda
Creating Future Leaders

References

P. Allison. Discrete-time methods for the analysis of event histories. *Sociological Methodology*, 13:61–98, 1982.

E. B. Andersen. Discrete models for the analysis of time series. *Journal of the American Statistical Association*, 65(332):9–33, 1970.

L. Antolini, P. Boracchi, and E. Biganzoli. A time-dependent discrimination index for survival data. *Statistics in Medicine*, 24(24):3927–3944, 2005.

E. Baek, H. Yang, and S. Kim. Survival time prediction by integrating cox proportional hazards network and distribution function. *BioMed Central Bioinformatics*, 22:15, 2021.

A. Bender, D. Rügamer, F. Scheipl, and B. Bischl. A general machine learning framework for survival analysis, June 2020. URL https://arxiv.org/abs/2006.15442.

A. Bere, G. Sithuba, C. Mabvuu, R. Mashabela, C. Sigauke, and K. Kyei. Regularisation in discrete survival models: A comparison of lasso and gradient boosting. *South African Statistical Journal*, 55(1):29–44, 2021.

M. Berger and M. Schmid. Semiparametric regression for discrete time-to-event data. *Statistical Modelling*, 18(3–4):322–345, 2018.

I. Bou-Hamad, D. Larocque, and H. Ben-Ameur. Discrete-time survival trees and

© University of Venda

University of Venda
Creating Future Leaders

forests with time-varying covariates: Application to bankruptcy data. *Statistical Modelling*, 11, 10 2011a. doi: 10.1177/1471082X1001100503.

I. Bou-Hamad, D. Larocque, and H. Ben-Ameur. A review of survival trees. *Statistics Surveys*, 5:44–71, 2011b.

L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees*. Chapman and Hall/CRC, New York, 1984.

S. F. Brown, A. J. Branford, and W. Moran. On the use of artificial neural networks for the analysis of survival data. *IEEE Transactions on Neural Networks*, 8(5): 1071–1077, 1997. ISSN 1045-9227. doi: 10.1109/72.623200.

W. Chai. A timeline of machine learning history. https://towardsdatascience.com/a-timeline-of-machine-learning-history, 2020. Blog post.

Y. Chen, Z. Jin, D. Mercola, and X. Xie. A gradient boosting algorithm for survival analysis via direct optimization of concordance index. *Computational and Mathematical Methods in Medicine*, page 8 pages, 2013. doi: 10.1155/2013/873595.

Y.-C. Chen. Survival analysis of lifetime data. *Applied Statistics with Computing (AST)*, 2018.

C. Chi, W. N. Street, and W. H. Wolberg. Application of artificial neural network-based survival analysis on two breast cancer datasets. *PLOS One or PMC Journal*, 2007:130–134, 2007. Publisher: PubMed.

M. Churpek, T. Yuen, C. Winslow, D. Meltzer, M. Kattan, and D. Edelson. Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. *Critical Care Medicine*, 44:7, 2016.

© University of Venda

University of Venda
Creating Future Leaders

T. G. Clark, M. J. Bradburn, S. B. Love, and D. G. Altman. Survival analysis part i: Basic concepts and first analyses. *British Journal of Cancer*, 89:232–238, 2003.

D. Cox. Regression models and life tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202, 1972.

D. R. Cox. Partial likelihood. *Biometrika*, 62(2):269–276, 1975.

A. R. Cuthbert, L. C. Giles, G. Glonek, L. M. Kalisch Ellett, and N. L. Pratt. A comparison of survival models for prediction of eight-year revision risk following total knee and hip arthroplasty. *BMC Medical Research Methodology*, 22:164, 2022. doi: 10.1186/s12874-022-01644-3.

Y. Deng, L. Liu, H. Jiang, et al. Comparison of state-of-the-art neural network survival models with the pooled cohort equations for cardiovascular disease risk prediction. *BMC Medical Research Methodology*, 23(1):22, 2023.

A. Dobra. *Decision tree classification*, pages 765–769. Springer US, Boston, MA, 2009. ISBN 978-0-387-39940-9. doi: 10.1007/978-0-387-39940-9_554. URL https://doi.org/10.1007/978-0-387-39940-9_554.

B. Efron. Logistic regression, survival analysis, and the kaplan-meier curve. *Journal of American Statistical Association*, 83:414–425, 1988. URL https://www.jstor.org/stable/2288857.

D. Faraggi and R. Simon. A neural network model for survival data. *Statistics in Medicine*, 14(1):73–82, 1995. doi: 10.1002/sim.4780140108. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.4780140108.

D. Finkelstein. A proportional hazards model for interval-censored failure time data. *Biometrics*, 42:845–854, 1986.

D. Finkelstein and R. Wolfe. A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics*, 41:933–945, 1985.

© University of Venda

University of Venda
Creating Future Leaders

M. D. Ganggayah, N. A. Taib, Y. C. Har, P. Lio, and S. K. Dhillon. Predicting factors for survival of breast cancer patients using machine learning techniques. *Computers in Biology and Medicine*, 115:103516, 2019.

M. Gensheimer and B. Narasimhan. A scalable discrete-time survival model for neural network. *PeerJ*, e6257:7, 2019.

T. Gerds, T. Cai, and M. Schumacher. The performance of risk prediction models. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 50(4): 457–479, 2008.

M. Glans, A. K. Ekstam, U. Jakobsson, A. Bondesson, and P. Midlöv. Risk factors for readmission in older adults within 30 days of discharge: A comparative retrospective study. *BioMed Central Geriatrics*, 20(467), 2020.

W. B. Goggins, D. M. Finkelstein, D. A. Schoenfeld, and A. M. Zaslavsky. A markov chain monte carlo em algorithm for analysing interval-censored data under the cox proportional hazards model. *Biometrics*, 54:1498–1507, 1998.

L. Gordon and R. A. Olshen. Tree-structured survival analysis. *Cancer Treatment Reports*, 69(10):1065–1069, 1985.

S. Gripp, S. Moeller, E. Bolke, G. Schmitt, C. Matuschek, A. S. Asgharzadeh, S. Roth, W. Budach, M. Franz, and R. Willers. Prognosis and survival estimates in palliative therapy. *Journal of Clinical Oncology*, 34(27):3313–3320, 2016.

A. Groll and G. Tutz. Variable selection in discrete survival models including heterogeneity. *Lifetime Data Analysis*, 22(3):305–338, 2016.

S. Guo. The discrete-time models in survival analysis. Unpublished lecture notes, University of Illinois at Urbana-Champaign, 2010.

F. Harrell, R. Califf, D. Pryor, K. Lee, and R. Rosati. Evaluating the yield of medical tests. *Journal of American Medical Association*.

© University of Venda

University of Venda
Creating Future Leaders

T. Hothorn, P. Bühlmann, A. Molinaro, and M. J. van der Laan. Survival ensembles. *Biostatistics*, 7(3):355–373, 2006.

H. Ishwaran, U. Kogalur, E. Blackstone, and M. Lauer. Random survival forests. *The Annals of Applied Statistics*, 2(3):841–860, 2008.

E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481, 1958.

K. Kartsonaki. Survival analysis. *Diagnostic Histopathology*, 22(7):263–270, 2016.

J. L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, and Y. Kluger. Deepsurv: Personalized treatment recommender system using a cox proportional hazards deep neural network. *BMC Medical Research Methodology*, 18(1):24, 2018.

B. Kearns, M. D. Stevenson, K. Triantafyllopoulos, and A. Manca. Generalized linear models for flexible parametric modeling of the hazard function. *Statistical Methods in Medical Research*, 28(3):867–878, 2019.

H. Kvamme and Borgan. Continuous and discrete-time survival prediction with neural networks, October 2019. URL https://arxiv.org/abs/1910.06724. arXiv:1910.06724.

H. Kvamme and Borgan. Continuous and discrete-time survival prediction with neural networks. *Lifetime Data Analysis*, 27(4):710–736, 2021.

M. LeBlanc and J. Crowley. Survival trees by goodness of split. *Journal of the American Statistical Association*, 88:457–467, 1993.

C. Lee, W. R. Zame, J. Yoon, and M. van der Schaar. Deephit: A deep learning approach to survival analysis with competing risks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018. URL https://api.semanticscholar.org/CorpusID:19102244.

© University of Venda

University of Venda
Creating Future Leaders

C. Mabvuu, A. Bere, and C. Siguake. Variable selection in discrete survival models. *University of Venda Research Report*, pages 1–73, 2020.

B. Mahesh. Machine learning algorithms – a review, 2018. Available online: https://arxiv.org/abs/1708.07811 (accessed June 10, 2025).

N. Mantel and B. F. Hankey. A logistic regression analysis of response time data where the hazard function is time dependent. *Communications in Statistics*, 7(4): 333–347, 1978.

A. Maposa, A. Bere, C. Chimedza, C. Sigauke, and P. Nevhungoni. A comparison of the accuracy of the pareto and logit-power families of link functions in estimating covariate effects and predicting hazards for discrete survival models, 2022. ResearchGate.

S. A. Mashele, T. B. Zwane, L. Kounza, M. M. Muchengeti, and L. Motsuku. Risk factors for breast cancer among women in ekurhuleni metropolitan municipality, gauteng province of south africa, 2017–2020: A case-control study. *ecancer*, 17: 1593, 2023.

W. S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biology*, 52(1-2):99–115, 1991.

U. B. Mogensen, H. Ishwaran, and T. A. Gerds. Evaluating random forests for survival analysis using prediction error curves. *Journal of Statistical Software*, 50 (11):1–23, 2012.

J. N. Morgan and J. A. Sonquist. Problems in the analysis of survey data. *Journal of the American Statistical Association*, 58:415–434, 1963.

R. Musci, B. Fairman, K. Maysn, G. Uhl, B. Maher, D. Sistor, S. Kellam, and G. Ialongo. Polygenic moderation: An application of discrete-time survival to model the timing of first marijuana use among urban youth. *National Library of Medicine*, 19:6–14, 2018.

© University of Venda

University of Venda
Creating Future Leaders

A. Nair. A guide to model selection for survival analysis. https://towardsdatascience.com/a-guide-to-model-selection-for-survival-analysis, 2022. Blog post.

P. M. Odell and R. B. D'Agostino. Maximum likelihood estimation for interval-censored data using a weibull-based accelerated failure time model. *Biometrics*, 48:951–959, 1992.

O. M. Odeniya, A. A. Abiodun, and B. A. Oyejola. Discrete-time survival analysis of age at pregnancy among nigerian women. *Research Journal of Applied Sciences*, 8(5):215–220, 2013.

L. Ohno-Machado. Modeling medical prognosis: Survival analysis technique. *Journal of Biomedical Informatics*, 34(6):428–439, 2001.

A. Pandey. Survival analysis: Intuition & implementation in python. https://towardsdatascience.com/survival-analysis-intuition-implemention-in-python-504fde4fcf8e, 2019. Blog post, Towards Data Science.

G. Piccinini. The first computational theory of mind and brain: A close look at mcculloch and pitts's "logical calculus of ideas immanent in nervous activity". *Synthese*, 141, 08 2004. doi: 10.1023/B:SYNT.0000043018.52445.3e.

K. L. Pickett, K. Suresh, K. R. Campbell, et al. Random survival forests for dynamic predictions of a time-to-event outcome using a longitudinal biomarker. *BMC Medical Research Methodology*, 21:216, 2021.

R. Praba, G. Darshan, K. T. Roshanraj, and P. B. Surya. Study on machine learning algorithms. *Information Systems Journal*, 7(4):67–72, 2021.

R. Prentice and L. Gloeckler. Regression analysis of grouped survival data with application to breast cancer data. *Biometrics*, 34:57–67, 1978.

© University of Venda

University of Venda
Creating Future Leaders

F. Rosenblatt. A perceiving and recognizing automaton. *Cornell Aeronautical Laboratory Report*, 1957.

S. Safavian and D. Landgrebe. A survey of decision tree classifier methodology. *IEEE Transactions on Systems, Man, and Cybernetics*, 21:660–674, 1991.

A. L. Samuel. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3):210–229, 1959.

M. Schmid, H. Kuchenhoff, and A. Hoerauf. A survival tree method for the analysis of discrete event times in clinical and epidemiological studies. *Statistics in Medicine*, 35:734–751, 2016.

M. Schmid, G. Tutz, and T. Welchowski. Discrimination measures for discrete time-to-event predictions. *Econometrics and Statistics*, 7:153–164, 2018. ISSN 2452-3062. doi: 10.1016/j.ecosta.2017.03.008. URL https://www.sciencedirect.com/science/article/pii/S2452306217300321.

M. Schmid, T. Welchowski, and M. Berger. Discrete-time survival forests with hellinger distance decision trees. *Statistical Methods in Medical Research*, 34: 812–832, 2022.

T. Sharaf and C. P. Tsokos. Predicting survival time of localized melanoma patients using discrete survival time method. *Journal of Modern Applied Statistical Methods*, 13(1):1–156, 2014.

J. D. Singer and J. B. Willett. Modeling the days of our lives: Using survival analysis when designing and analyzing longitudinal studies of duration and the timing of events. *Psychological Bulletin*, 110(2):268–290, 1991.

J. D. Singer and J. B. Willett. It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational Statistics*, 18: 155–195, 1993.

© University of Venda

University of Venda
Creating Future Leaders

A. Spooner, E. Chen, and A. Sowmya. A comparison of machine learning methods for survival analysis of high-dimensional clinical data for dementia prediction. *Scientific Reports*, 10(20410):1–10, 2020.

T. Stainer. The notion of censoring in survival analysis. https://towardsdatascience.com/the-notion-of-censoring-in-survival-analysis, 2021. Blog post.

K. Suresh, C. Severn, and D. Ghosh. Survival prediction models: An introduction to discrete-time modeling. *BioMed Central Medical Research Methodology*, 2020.

W. Thompson. On the treatment of grouped observations in life studies. *Biometrics*, 33:463–470, 1977.

R. Tibshirani. The lasso method for variable selection in the cox model. *Statistics in Medicine*, 16(4):385–395, 1997.

W. Tizi and A. Berrado. Machine learning for survival analysis in cancer research: A comparative study. *Scientific African*, 21:e01880, 2023. ISSN 2468-2276. doi: 10.1016/j.sciaf.2023.e01880. URL https://www.sciencedirect.com/science/article/pii/S2468227623003356.

N. A. Tutkun, N. Koyuncu, and U. Karabey. Discrete-time survival analysis under ranked set sampling: An application to turkish motor insurance data. *Journal of Statistical Computation and Simulation*, 89(4):660–667, 2019.

G. Tutz and M. Schmid. *Modeling Discrete Time-to-Event Data*. Springer, Cham, 2016.

P. Wang, Y. Li, and C. Reddy. Machine learning for survival analysis: A survey. *arXiv preprint arXiv:1708.04649*, 10:35, 2017.

S. Wiegrebe, P. Kopper, R. Sonabend, et al. Deep learning for survival analysis: A review. *Artificial Intelligence Review*, 57(65), 2022.

© University of Venda

University of Venda
Creating Future Leaders

M. N. Wright and A. Ziegler. Ranger: a fast implementation of random forests for high dimensional data in c++ and r. *Journal of Statistical Software*, 77(1):1–17, 2017.

Z. Zhang, J. Reinikainen, K. A. Adeleke, M. E. Pieterse, and C. G. M. Groothuis-Oudshoorn. Time-varying covariates and coefficients in cox regression models. *Annals of Translational Medicine*, 6(7):121, 2018.

© University of Venda