From the Department of Medical Epidemiology and Biostatistics
Karolinska Institutet, Stockholm, Sweden

MACHINE LEARNING METHODS FOR PRECISION MEDICINE

Pablo Gonzalez Ginestet

Karolinska Institutet

Stockholm 2022

Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2022

ISBN 978-91-8016-633-1

Machine learning methods for precision medicine

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Pablo Gonzalez Ginestet

The thesis will be defended in public at Petrén, Nobels väg 12B, 171 65 Solna,

Friday, 19 August 2022

Principal Supervisor:
Docent Erin Gabriel
Karolinska Institutet
Department of Medical Epidemiology
and Biostatistics

Co-supervisor(s):
Docent Arvid Sjölander
Karolinska Institutet
Department of Medical Epidemiology
and Biostatistics

Associate Professor Mattias Rantalainen
Karolinska Institutet
Department of Medical Epidemiology
and Biostatistics

Docent Ujjwal Neogi
Karolinska Institutet
Department of Laboratory Medicine

Opponent:
Assistant Professor Helga Westerlind
Karolinska Institutet
Department of Medicine

Examination Board:
Associate Professor Karin Leander
Karolinska Institutet
Institute of Environmental Medicine

Professor Henrik Hult
KTH Royal Institute of Technology
Department of Mathematics

Docent Ola Larsson
Karolinska Institutet
Department of Oncology-Pathology

“Amar la trama más que el desenlace”

ABSTRACT

In precision medicine, predicting the risk of an event during a specific period may help, for example, to identify patients that need early preventive treatment. Modern machine learning (ML) techniques are therefore ideal for building these predictions. However, medical datasets often suffer from right-censoring of the outcome of interest posing an obstacle to the direct applicability of ML algorithms. The aim of this thesis work is to develop and advance methods for prediction in settings of right-censoring, and in some settings also including competing risks. Specifically, in Project I, we developed an approach that combines inverse probability of censoring weighting (IPCW) with bagging as a pre-processing step to enable the application of all existing ML methods for classification in settings of right-censoring and competing risks, and we propose a procedure to combine optimally a set of single IPCW bagged methods. In Project II, we developed an extension of Project 1 to combine optimally not only over ML procedures for the same outcome but combining survival outcomes such as Cox regression model and continuous outcome such as pseudo-observations-based regression. In Project III, we integrated pseudo-observations into Convolutional Neural Network to predict the cumulative incidence using images and structured clinical data. In Project IV, we applied the methods developed in Project 1-2 to build a flexible risk prediction model to predict the risk of any cancer diagnosis using a Swedish population-based register among sarcoidosis patients.

In the last project, Project V, we explored the utility of a dynamic prediction model in a setting of complete data as decision support tool for public health to manage future pandemics. Specifically, we applied two state-of-the-art batch reinforcement learning algorithms to learn the best face covering policy response at the national level with the goal of reducing the spread of COVID-19.

LIST OF SCIENTIFIC PAPERS

I. Pablo Gonzalez Ginestet, Ales Kotalik, David Vock, Julian Wolfson and Erin Gabriel. *Stacked inverse probability of censoring weighted bagging: a case study in the InfCareHIV Register*. Journal of the Royal Statistical Society Series C, 2021,70:51-65.
https://doi.org/10.1111/rssc.12448

II. Pablo Gonzalez Ginestet, Erin Gabriel and Michael Sachs. *Survival stacking with multiple data types using pseudo-observation-based-AUC loss*. Journal of Biopharmaceutical Statistics, 2022.
https://doi.org/10.1080/10543406.2022.2041655

III. Pablo Gonzalez Ginestet, Philippe Weitz, Mattias Rantalainen and Erin Gabriel. *A deep convolutional neural network approach for predicting cumulative incidence based on pseudo-observations*.
Manuscript

IV. Elizabeth Arkema, Pablo Gonzalez Ginestet, Erin Gabriel and Michael Sachs. *Predicting risk of cancer among sarcoidosis patients: a nationwide, register-based, cohort-study.*
Manuscript

V. Pablo Gonzalez Ginestet, Erin Gabriel, Ziad El-Khatib and Ujjwal Neogi. *Batch deep reinforcement learning for policy responses to the COVID pandemic.*
Manuscript

CONTENTS

1 INTRODUCTION 1
2 RESEARCH AIMS 3
2.1 Project specific aims 3
3 MATERIALS AND METHODS 5
3.1 Supervised learning in survival data 5
3.1.1 Notation 5
3.1.2 Supervised ML 6
3.1.3 Approaches to adapt ML methods to survival data 7
3.1.4 Evaluation 9
3.2 RL in oberservational data 12
4 RESULTS 16
4.1 Project 1 16
4.2 Project 2 17
4.3 Project 3 18
4.4 Project 4 18
4.5 Project 5 20
5 DISCUSSION 21
6 ACKNOWLEDGEMENT 23
7 REFERENCES 25

LIST OF ABBREVIATIONS

AUC	Area under the ROC curve (AUC).
CNN	Convolutional Neural Network
CoxCNN	CNN with Cox loss
IPCW	Inverse Probability of Censoring Weighting
KM	Kaplan Meier
MAE	Mean Absolute Error
ML	Machine Learning
MSE	Mean Squared Error
PH	Proportional Hazard
PO	Pseudo-Observation
TCGA	The Cancer Genome Atlas
TD	Temporal Difference
RL	Reinforcement Learning
ROC	Receiver Operating Characteristic

1 INTRODUCTION

Precision medicine is an emerging approach for disease prevention and management based on delivering treatment strategies that are deeply tailored to each individual patient (Collins & Varmus, 2015). Unlike the standard strategy “one-size-fits all”, where the same treatment is assigned to all patients that suffer a disease, precision medicine aims at assigning the right treatment for the right patient given at the right time. In this new medical paradigm, the development of machine learning methods, jointly with large real-world databases, hold promise in providing a personalized medicine to each patient.

Machine learning (ML) can be broadly defined as "computational methods using experience to improve performance or to make accurate predictions" (Mohri, Rostamizadeh, & Talwalkar, 2021). Here, experience is in the form of observational data. The components of any ML problem is data, where we can learn from; a model, that tells us how to transform the data; a loss function, that measures the badness of the model, and an optimization algorithm to update the model parameter in the direction that reduces the loss function.

There are many kinds of ML problems. The research projects that support this proposal plan towards the doctoral degree focus only on supervised learning and reinforcement learning. Project 1, Project 2, Project 3 and Project 4 fall into supervised learning where the experience is survival data that suffer from right-censoring and, in some cases, competing risks. On the other hand, Project 5 falls into reinforcement learning using observational data.

Supervised learning is the most common problem in machine learning. It addresses the task of predicting a response, also called target or label, given the covariates, also called features. Let y denote the response and X the covariates. Given a dataset of N observations of the pair covariates and response denoted by $\{X_i, y_i\}_{i=1}^{N}$, supervised learning aims at constructing a prediction function f_θ that maps the inputs to the prediction function $f_\theta(X)$. Moreover, we are interested in the following task in supervised learning: predicting the individual risk of an event ($y_i = 1$ if the event happened otherwise $y_i = 0$) before a specific time given X_i. This quantity plays an important role in precision medicine to identify and target those individuals at higher risk of experience an adverse event for early preventive strategy and thus minimizing their future risk. Cox proportional hazards regression model (Cox, 1972) is standard

method to build predictive models in setting of censored time-to-event data. However, such modeling strategy often lack the necessary flexibility. Machine learning methods can be much more flexible, making them more suitable to handle high-level interactions and higher-order associations, which may result in predictions that are more accurate. However, survival data poses an obstacle to the direct applicability of machine learning algorithms due to missing events times for censored individuals (for instance, y_i is not observed for those who are censored). Project 1, 2 and 3 develop methods that enable the application of supervised ML to survival data.

Reinforcement learning (RL) provides a mathematical framework to study problems that involve the task of learning to make a sequence of decisions so as to optimize an outcome (called reward) (Sutton & Barto, 1998). Unlike supervised learning, RL algorithms collect information by interacting with the environment through a sequence of actions. At every decision point, the RL algorithm chooses an action and it receives a new observation of the environment (state) and an immediate reward. In settings such as healthcare, this online interaction is dangerous or unfeasible. Batch or offline RL is the task of learning from an observational data (fixed dataset) without further interaction with the environment (Lange, Gabel, & Riedmiller, 2012). Batch RL has only the fixed dataset of transitions in common to supervised learning, still not having an external supervisor. In the context of healthcare, the meaning of state, action and reward could be defined as patient covariates and treatment history; treatment options and clinical response after treatment, respectively. A policy, or similarly, a decision rule, specifies what action to take at any time step given the state. The goal in batch RL is to learn a policy that maximize the sum of rewards received at each time decision point. This optimal policy, which is a tailored treatment recommendation, operationalizes precision medicine.

2 RESEARCH AIMS

The overall goal of the projects is to develop and advance methods for prediction and medical decision-making. For instance, the methods developed in these projects may be used to target treatments or therapies to those patients most likely to benefit from them and potentially enhance the clinical benefits, reduce side effects and increase the treatment compliance.

Project 1, 2, 3 and 4 use the cumulative incidence as a tool for personalizing treatment. The information given by quantifying the proportion of patients, who could experience any of the causes, that fail from cause k before to time τ is relevant for choosing treatment options. On the other hand, Project 5 explores the potential usefulness of reinforcement learning methods in public policy as a decision-supporting tool.

2.1 PROJECT SPECIFIC AIMS

- **Project 1:** To develop ensemble machine learning methods for right censored data, with or without competing risk, using IPCW bagging methods.
- **Project 2:** To extend ensemble machine learning methods for right-censored data using pseudo-observation based loss functions.
- **Project 3:** To develop improved Convolutional Neural Network (CNN) methods using pseudo-observations.
- **Project 4:** To apply methods developed in Project 1 and Project 2 to be able to predict risk of any cancer diagnosis in sarcoidosis patients using a Swedish population-based register.
- **Project 5:** To apply deep reinforcement learning methods to discover the best face covering policy with the aim at reducing the spread of COVID-19.

3 MATERIALS AND METHODS

3.1 SUPERVISED LEARNING IN SURVIVAL DATA

3.1.1 Notation

Let T_i be the true event time, C_i be the censoring time, $\tilde{T}_i = \min\{T_i, C_i\}$ be the observed right-censored event time and $\Delta_i = 1\{T_i < C_i\}$ be the event indicator for individual $i = 1, \dots, N$. If there is competing risks, let $\delta \in \{1, \dots, K\}$ be the event-type and $\tilde{\delta} = \Delta_i \, \delta$ be the event indicator. In the absence of competing risks $K = 1$ and $\tilde{\delta} = 0$ implies that an individual is right-censored. In addition, we observe a set of covariates for each individual given by $X_i = (X_{i,1}, \dots, X_{i,p})$. In Project 3, in addition to the structured covariates, we observe an image data that is a three dimensional given by its spatial and channel dimension. We denote the image data as I. The goal is to predict the individual risk of experiencing the main event before a specific time τ using the information of the patient, X (Project 1 and 2) and I (Project 3).

Specifically, in Project 1, we were interested in estimating the cumulative incidence under the presence of competing risks

$$P(T \leq \tau, \delta = k | X)$$

based on the following available data $D_i = \{E_{i,k}, \Delta_i, \delta_i, X_i\}$ for $i = 1, \dots, N$ where

$$E_{i,k} = \begin{cases} 1 & if\ \tilde{T}_i \leq \tau\ and\ \delta = k \\ 0 & if\ \tilde{T}_i \leq \tau\ and\ \delta \notin \{0, k\}\ or\ if\ \tilde{T}_i > \tau \\ NA & otherwise \end{cases}$$

In Project 2, we were interested in estimating the cumulative incidence quantity:

$$P(T \leq \tau | X)$$

based on $D_i = \{\tilde{T}_i, y_i, X\}$ for $i = 1, \dots, N$ where $y_i = \Delta_i 1\{\tilde{T}_i \leq \tau\}$.

In Project 3, we were interested in estimating the cumulative incidence quantity

$$P(T \leq \tau | X, I)$$

using $D_i = \{\tilde{T}_i, y_i, X_i, I_i\}$ for $i = 1, \dots, N$.

Using directly the outcome $E_{i,k}$ or y_i in a predictive model to estimate the corresponding target quantity would lead to biases due to not taking account the

presence of right-censoring. Two approaches to overcome this problem, other than the naïve method of excluding censored observations in the analysis, are IPCW and pseudo-observations.

3.1.2 Supervised ML

Our goal is to apply ML methods to survival data in order to estimate the quantity cumulative incidence to make predictions. We assume that there is a relationship between our quantity of interest, $E[1(T \leq \tau)]$ or $E[1(T \leq \tau, \delta = k)]$, and X that is summarized by the unknown function *f*. We want to estimate *f* using the training data given by D_i. In this thesis, the ML methods are treated as a black box. We are concerned about the accuracy of the predictions.

In Project 1, we applied ML methods for classification (for example, logistic regression and more flexible methods such neural networks) to estimate *f* using observation $E_{i,k}$ and X. In Project 2, we applied ML methods for binary and continuous outcome. In Project 3, we applied convolutional neural network using images.

Furthermore, in Project 1 and Project 2 we applied stacking (Breiman L. , 1996), an ensemble method that combines predictions from several ML methods into one with the goal of increasing predictive accuracy. For instance, if one consider the linear combination as the stacking procedure (as in Project 1 and Project 2), the final prediction is given by

$$\beta_1 \hat{f}_1 + \cdots + \beta_A \hat{f}_A$$

where $\hat{f}_a$ denotes the a-th ML method in the library that was fitted in the training set. The library contains a total of A algorithms and the coefficients β_j, for $j = 1, \ldots, A$, weights each algorithm in the final prediction. These weights are unknown a priori. We uses V-fold cross validation to find the optimal contribution of each candidate algorithm to the final prediction. Implementation of the V-fold cross validation using the linear stacking procedure is as follows:

Step 1. Split the data into V-folds

Step 2. For each fold *v* = {1,…,V},

a) The observations in the v-th fold are used as validation set and the remaining of the observations as training set.

b) Train each candidate algorithm on the training set and used the trained algorithm to compute predictions for the validation set

Step 3. Collect the predictions of each algorithm across the validation sets

Step 4. Form the weighted linear combination

$$\beta_1 \widehat{f_1} + \cdots + \beta_A \widehat{f_A}$$

and select the weights that minimize a desired loss function (for example, one minus the area under the ROC curve (AUC)).

3.1.3 Approaches to adapt ML methods to survival data

3.1.3.1 Application of IPCW in existing ML methods

Inverse probability of censoring weighting (IPCW) is a technique for dealing with censored data (Robins & Finkelstein, 2000). IPCW works reweighting the individuals who are not censored by the inverse probability of remaining uncensored. Individuals who are censored are not included as observations in the analysis and are represented by those weighted individuals. In mathematical terms, IPC weight is defined as

$$w_i = \begin{cases} \dfrac{1}{G(\min\{T_i, \tau\}\,|X)} & if\ \min\{T_i, \tau\} \leq\ C_i \\ 0 & otherwise \end{cases}$$

where $G(t|X) = P(C > t|X)$ is the censoring survivor function, an unknown function that must be estimated from the data. Estimation methods go from the simplest method as Kaplan-Meier (KM) in the case of independent censoring to procedures that allow for covariates to account for dependent censoring (Robins & Finkelstein, 2000; Satten, Datta, & Robins, 2001) such as Cox regression model. In Project 1, we estimated the IPC weights using Cox regression model and a more flexible model such as boosted Cox regression (Binder, 2013). In Project 2, the IPC were estimated in the simplest way using KM. In Project 3, we estimated the IPC using Cox regression model.

IPCW has been incorporated in ML methods in order to account for right-censoring in survival data (Vock, 2016; Wolfson, 2015; Bandyopadhyay, et al., 2015; Molinaro, Dudoit, & Laan, 2004; Hothorn, Buhlmann, Dudoit, Molinaro, & Van Der Laan, 2006).

For instance, (Vock, 2016) shows how to incorporate weights in different ML methods such as naive Bayes; bayesian networks; k-nearest neighbors; generalized additive logistic model, and support vector machine. These adaptations are specific to the method and some of them do not have software implementations like the neural network package in R.

3.1.3.2 IPCW bagging

IPCW bagging combines IPCW and bagging (Breiman L. , 1996). Bagging or bootstrap aggregation is a procedure used to reduce the variance of a ML method and increase its accuracy. It works generating B training samples from the original training set, then training the ML on b-th bootstrap sample ($\hat{f}^b$) and finally averaging all predictions to obtain the final prediction ($\hat{f}$). In mathematical terms,

$$\hat{f} = \frac{1}{B}\sum \hat{f}^b(x)$$

In IPCW bagging, IPC weights are used to generate a weighted sample from the training set. That allows training any ML on a right-censored data and removes the need to directly adapt any of ML. IPCW bagging (Hothorn, Bühlmann, Dudoit, Molinaro, & Van Der Laan, 2006) is extended in Project 1.

3.1.3.3 Other proposals that not rely on IPCW

There are other approaches that are specific to the method too and do not use weights to handle censoring. For example, some authors suggested modifying the splitting criteria for decision trees and random forest for survival data (Gordon & Olshen, 1985; Hothorn, Lausen, Benner, & Radespiel-Tröger, 2004; Ishwaran, et al., 2014). Other works have adapted deep neural networks (Faraggi & Simon, 1995; Katzman, et al., 2018; Ching, Zhu, & Garmire, 2018) and convolutional neural networks (Mobadersany, et al., 2018; Zhu, Yao, & Huang, 2016; Li, et al., 2019) replacing the linear representation of the log of hazard by the output of the network, which is a non-linear, and using the negative log partial likelihood to train the network. However, they maintained the proportional hazard assumption. Other works such as (Luck, Sylvain, Cardinal, Lodi, & Bengio, 2017) avoided the proportional hazard assumption using a sophisticated loss function for censored data.

3.1.3.4 Pseudo-Observations

Pseudo-observation (PO) is a technique used with the goal of transforming a censored problem into an uncensored one and thus apply standard statistical methods developed for complete data (Andersen & Perme, 2009). PO replaces the censored response variable of each individual by the individual jackknife. For instance, the PO cumulative incidence for individual i, as in Project 3, is:

$$\widehat{CI}_i(\tau) = N\,\widehat{CI}(\tau) - (N-1)\widehat{CI}^{-i}(\tau)$$

where $\widehat{CI}(\tau) = 1 - \widehat{KM}(\tau)$ and $\widehat{KM}(\tau)$ is computed using the entire sample and $\widehat{CI}^{-i}(\tau)$ is computed removing the i-th individual from the data. Note that in the absence of competing risks, the cumulative incidence is equal to one minus the survival function. If competing risks were present, we would use one minus the Aalen-Johansen estimator (Aalen & Johansen, 1978).

The theory of PO approach requires that $\widehat{CI}_i(\tau)$ be a consistent estimator and that censoring be independent of the event time (Graw, Gerds, & Schumacher, 2009). When censoring is covariate-dependent and conditionally independent of the event time given the covariates, the KM estimator is biased. This bias can be corrected using IPCW in order to compute weighted PO for each individual (Xiang & Murray, 2012). In Project 3, we modeled the censoring using a Cox's PH model but options that are more flexible are valid options (for example, random forest).

Once the POs are computed, the analysis is carried out based on the new data that contains the PO as the new response variable. For example, in Project 3 the new data $D_{i,po} = \{\widehat{CI}_i(\tau), X_i, I_i\}$ for $i = 1, \dots, N$ is used to train the CNN model.

3.1.4 Evaluation

Evaluation of the performance of the ML method to predict future data is central to the development of any prediction model. At the end, the final goal is to accurately predict the risk of an event for future patients based on their clinical measurements (for example, predict the five-year risk of any cancer diagnosis as in Project 4). Thus, we are interested in the accuracy of the predictions on the test data, the data that was not used to train the model.

The area under the ROC curve (AUC) is extensively used as a performance measure of a maker in the medical domain. Here the marker is represented by the risk

predictions $\hat{f}$. The ROC curve compares the sensitivity estimates (probability of a true positive) against the one minus the specificity (probability of a false positive) for all possible values of $\hat{f}$. A higher AUC value indicates a better $\hat{f}$ performance and it is usually assumed that a higher marker value is more indicative of more risk of experiencing the event. When the event status is time-dependent, sensitivity and specificity become time-dependent too, leading to time-dependent AUC (Heagerty & Zheng, 2005; Heagerty, Lumley, & Pepe, 2000).

3.1.4.1 Time-dependent AUC

In time-dependent AUC, cases and controls are defined in a cumulative/dynamic fashion. More precisely, at a fixed time τ an individual is classified as a case for $T_i \leq \tau$ and as control for $T_i > \tau$. Thus, the classification of the population in terms of cases and controls depends on the specific time of interest τ. For a marker value $\hat{f}$,

$$sensitivity(c, \tau) = P(\hat{f} > c | T_i \leq \tau)$$

$$specificity(c, \tau) = P(\hat{f} \leq c | T_i > \tau)$$

Thus, the AUC is given by:

$$AUC = P(\hat{f}_i > \hat{f}_j | T_i < \tau, T_j > \tau)$$

In competing risks, two definitions of time-dependent specificity can be considered depending on the control group definition considered (Zheng, Cai, Jin, & Feng, 2012). If the control group is defined as those who are free of any event at time τ, $T_i > \tau$, leads to the previous specificity.

Whereas, the second definition of control include those who experienced the competing risks before time τ, $T_i > \tau \; or \; T_i \leq \tau \; and \; \delta \neq k$, and leads to the following specificity measure:

$$specificity^*(c, \tau) = P(\hat{f} \leq c | T_i > \tau \; or \; T_i \leq \tau \; and \; \delta \neq k)$$

These two definitions of the control group result in two AUC:

$$AUC_1 = P(\hat{f}_i > \hat{f}_j | T_i < \tau, \delta_i = k, T_j > \tau)$$

$$AUC_2 = P(\hat{f}_i > \hat{f}_j | T_i < \tau, \delta_i = k, T_j > \tau \; or \; T_j \leq \tau \; and \; \delta_j \neq k)$$

In Project 1, we used this latter definition, AUC_2, for the evaluation of the predictive performance of the stacked procedure. Furthermore, we minimize the amount

$1 - AUC_2$ for selecting the optimal contribution, $\beta_1, \ldots, \beta_A$, to the stacking.

Estimation of the AUC has been approached using Bayes theorem and applying KM estimator (Heagerty, Lumley, & Pepe, 2000). For example, using Bayes theorem, one can re-write the sensitivity as

$$P\left(\hat{f} > c \middle| T_i \leq \tau\right) = \frac{P\left(T_i \leq \tau, \hat{f} > c\right)}{P(T_i \leq \tau)} = \frac{P\left(T_i \leq \tau | \hat{f} > c\right) P(\hat{f} > c)}{P(T_i \leq \tau)}$$

And recognizing that $P\left(T_i \leq \tau | \hat{f} > c\right)$ is the conditional survival function in the subset of individuals with $\hat{f} > c$, one can estimate that quantity using KM estimator. The quantity $P(\hat{f} > c)$ can be estimated using the empirical distribution, $1 - \sum \frac{1(\hat{f} \leq c)}{N}$. Other approaches have used IPCW estimates of the AUC (Blanche, Dartigues, & Jacqmin-Gadda, 2013; Uno, Cai, Tian, & Wei, 2007). In Project 1, we followed that latter approach.

In Project 2, we used a time-dependent AUC that uses PO as outcome. This is justified by the property of the PO for the cumulative incidence (Graw, Gerds, & Schumacher, 2009)

$$E\left(\widehat{CI}_l(\tau) \middle| X_i\right) = P(T_i \leq \tau | X_i) + o_p(1)$$

where $o_p(1)$ is a term that vanishes asymptotically.

This property and Bayes' rule suggest estimating the true positive as

$$sensitivity(c, \tau) = P\left(\hat{f} > c \middle| T_i \leq \tau\right) = \frac{P\left(T_i \leq \tau, \hat{f} > c\right) P(\hat{f} > c)}{P(T_i \leq \tau)} \approx \frac{\sum \widehat{CI}_l(\tau)\, 1(\hat{f} > c)}{\sum \widehat{CI}_l(\tau)}$$

In a similar way, specificity approximation using PO is:

$$specificity(c, \tau) \approx \frac{\sum (1 - \widehat{CI}_l(\tau))\, 1(\hat{f} > c)}{\sum (1 - \widehat{CI}_l(\tau))}$$

Lastly, the optimization of the AUC is challenging because the AUC is invariance to monotone transformations and non-differentiable due to the step function. There has been several approaches to deal with this identifiability problem from imposing a

constraint on the coefficients to including a penalization term on the coefficients (Fong, Yin, & Huang, 2016). In Project 1 and Project 2, we followed the latter. The step function in the AUC is usually approximated, as we did in Project 2, using a smooth version of these indicators functions such as sigmoid or ramp function (Fong, Yin, & Huang, 2016).

3.2 RL IN OBERSERVATIONAL DATA

Reinforcement learning (RL) problem has been formalized using a Markov Decision Process (MDP), which it has been used in theoretic stochastic decision-making problems. Formally, an MDP is defined by a five-tuple $< S, A, R, P, \gamma >$, where S is the state space and $s_t \in S$ is the state at time t, A is the action space and $a_t \in A$ is the action chosen at time t, $R(s_t, a_t)$ is the reward function that determines the reward r_{t+1} the agent receives after taking action a_t in state s_t and transitioning to next state s_{t+1}, $P(.|s_t, a_t)$ is the transition probability distribution that governs the transition from state s_t to state s_{t+1} after taking action a_t, and $\gamma \in [0,1]$ is the discount factor. Figure 1 shows schematically the dynamic of the MDP. In every time, the agent observes the current state s_t, takes an action a_t and observes a feedback from the environment in form of a reward r_{t+1} and then observes the next state s_{t+1}. Underlying the MDP is the Markov. The Markov property says that the future is independent of the past given the current state. That is, $P(s_{t+1}|all\ history\ up\ to\ t) = P(s_{t+1}|s_t, a_t)$ and $E(r_{t+1}|all\ history\ up\ to\ t) = E(\ r_{t+1}|s_t, a_t)$.

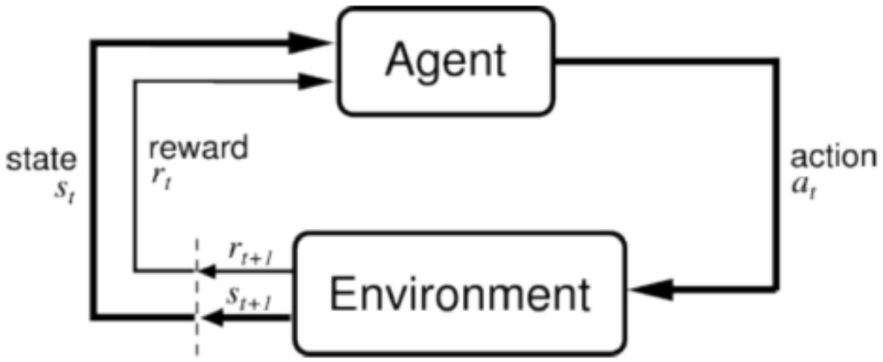


Figure 1. Figure taken from Sutton and Barto (1998).

A policy specifies what action to take at any time step t. Define the *policy*, π, as a mapping from states to actions. A policy could be deterministic, denote it as $a = \pi(s)$, or stochastic policy, where $\pi(a|s)$ represents the probability of taking action a in state s. In Project 5, we considered a deterministic policy.

Let a trajectory composed of the sequence $\{s_0, a_0, r_0, s_1, \dots, s_{T-1}, a_{T-1}, r_{T-1}, s_T\}$ generated by the policy π in a finite MDP. The return associated to the trajectory is defined as

$$R_{0:T-1} = \sum_{t=0}^{T-1} \gamma^t r_t$$

The state-value function is the expected return starting from state s, and then following policy π:

$$V^{\pi}(s) = E(R_{0:T-1}|s_0 = s)$$

and the action-value function is the expected return starting from state s, taking action a, and then following policy π:

$$Q^{\pi}(s, a) = E(R_{0:T-1}|s_0 = s, a_0 = a)$$

Both definitions are related as follows:

$$V^{\pi}(s) = E_{a\sim\pi(s)}(Q^{\pi}(s, a))$$

The goal of an RL agent is to learn an optimal policy defined as one that maximizes its expected discounted future return:

$$\pi^*(s) = argmax_{\pi} E_{a\sim\pi(s)}(Q^{\pi}(s, a))$$

The Bellman equation (Bellman, 1957) characterizes the optimal policy:

$$Q^{\pi^*}(s, a) = max_a R(s, a) + \gamma \sum_{s'} P(s'|s, a) V^{\pi*}(s')$$

The techniques to compute an optimal policy for a given MDP are model-based or model-free. Model-based RL requires to modelling the transition distribution $P(s'|s, a)$ in order to use it to find the optimal policy π^*. On the other hand, model-free RL learns an optimal policy based on the received observations and rewards without modeling $P(s'|s, a)$. In Project 5, we took this latter approach.

Algorithms such as value iteration and policy iteration have been used for solving MDPs. Value iteration starts with an initial Q-value, Q_0, and iterates the Bellman equation, updating the Q-function until Q_k and Q_{k+1} are really close and deriving the

optimal policy by $argmax_{\pi}Q_{k+1}$. On the other hand, policy iteration starts with an initial policy and the Q-value associated to that policy (policy evaluation) and then improves the policy in each iteration of the Bellman equation. These two algorithms are exact methods since they require to have access to the dynamic of the model governed by the rewards $R(s_t, a_t)$ and transition probability $P(.|s_t, a_t)$. When the dynamic of the model is not given, the optimal policy can be computed using sample based methods like Monte Carlo and Temporal Difference (TD) methods (Sutton R. , 1988) or tabular Q-learning (Watkins & Dayan, 1992). However, these methods do not scale with increase in the size of state space, something that is found in most real applications. A method proposed to overcome this limitation is Q-learning with function approximation. Q-learning approximates the Q-function using a parametric approximation $Q_{\theta}(s, a)$. For example, the Q-function could be represented using a linear function in θ or a deep neural network where θ are the parameters of the network. Q-learning algorithm learns the parameter θ. So, instead of updating the Q-function as the previous methods, Q-learning updates the estimate of θ. Thus, the Q-function can be computed for any unseen pair $(s\,, a)$ given θ. The algorithm tries to find θ such that for every pair $(s\,, a)$ the Bellman equation can be approximated well.

Batch RL assumes that the dataset is fixed. The goal is to learn the best possible policy from an observational data without online interaction (Lange, Gabel, & Riedmiller, 2012; Fugimoto, Meger, & Precup, 2019). Instead of interacting with the environment observing the state s and performing action a and then updating the policy according to the reward r and next state s', the learner receives a sample of transitions $(s\,, a\,, r\,, s')$ sampled from the retrospective dataset.

In batch deep reinforcement learning, the Q-function is approximated using a deep neural network Q_{θ} (Mnih, et al., 2015). The parameter θ is updated minimizing a loss function over batches of transitions $(s\,, a\,, r\,, s')$ where the output of the network is $Q_{\theta}(s, a)$ and the target $r + \gamma \max_{a'} Q_{\theta'}(s', a')$:

$$Loss(r + \gamma \max_{a'} Q_{\theta'}(s', a') - Q_{\theta}(s, a))$$

A potential problem of using this algorithm is that $Q_{\theta'}(s', a')$ could be estimated poorly if the selected action a' in the target policy and the next state s' are not contained in the dataset (Fugimoto, Meger, & Precup, 2019). Discrete Batch Constrained deep Q-Learning is an algorithm to learn an optimal policy that

eliminates actions that are unlikely in the dataset (Fujimoto, Conti, Ghavamzadeh, & Pineau, 2019). In Project 5, we applied this latter algorithm as well as the standard deep Q-learning.

4 RESULTS

4.1 PROJECT 1

In Project 1, we developed ML methods for right censored data, with or without competing risks, combining IPCW and bagging methods. The integration of IPCW and bagging is used as pre-processing step allowing the application of any developed ML methods for classification to data that suffers from right-censoring, including competing risks. Previously, there have been adaptation of ML to survival data to account for right-censoring. IPCW has been used for this purpose. However, the way in which the IPC weights are incorporated depend on the ML algorithm. There have been other adaptations that do not use IPCW but they are class specific too. For instance, survival trees accommodates the splitting and pruning criteria in a specific manner in order to account for right-censoring. Our proposal incorporates the IPC weights in the resampling step of the bagging method and that removes the need to directly adapt the specific ML algorithm. Moreover, we proposed a procedure to stack optimally predictions from any set of IPCW bagged methods. We use the IPCW time-dependent AUC, which account for censoring and competing risks, to find the optimal coefficients to combine multiple predictions and to evaluate the predictive performance of each method. Lastly, our proposal can account for dependent censoring modeling the censoring survivor function conditional on covariates.

To illustrate our developed method, we investigated the performance of our proposed IPCW bagging procedure in four simulations scenarios that differs in terms of independent/dependent censoring and competing risk, and in a real data application. In the latter, we applied our stacking method in the Swedish InfCareHIV register to predict treatment failure in maintain an undetectable viral load for at least 2 years following initial suppression. We compared our procedure to survival methods such as Cox PH model, Cox Boost and Random Forest, and to methods that allow weights as an argument in the R package function.

The results concluded that our proposed method performed similarly to the counterpart algorithms that allow natively weights and the stacked IPCW bagging performed similarly to the best single IPCW bagging. Furthermore, our proposed method allowed building a risk prediction model that accounts for censored observations based on an ML method that cannot directly incorporate weights. We have developed an R package “stackBagg” that can be found at my GitHub page.

4.2 PROJECT 2

In Project 2, we proposed an ensemble method of a wide range of algorithms that differ in survival outcomes or methods to account for right censoring with the goal of predicting an individual risk. A mix of models may form the ensemble such as IPCW classification that was developed in Project 1, PO based regression and survival methods such Cox PH. Each algorithm considered in the ensemble uses the estimation method appropriate for that type of model. Each prediction of each candidate algorithm in the ensemble were obtained using V-fold cross validation. We optimally stacked the cross-validated predictions of each individual method using the area under the PO-based time-dependent ROC curve. We illustrated the proposed method using two examples in breast cancer. In the first example, we used the breast cancer data of Royston and Altman (2013) as a training set and the German Breast Cancer Study Group (GBSG) as our external validation set. In this example, the goal was to predict death or recurrence within 5 years of primary surgery. In the second example, we used the breast cancer data from the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC). In this latter example, we considered two endpoints: overall survival and recurrence-free survival in months. Since we did not have an external validation set, we split the patients randomly into training set (70%) and validation set (30%). We trained the models in the training set and we predicted the risk of dying and of breast cancer recurrence within five years in the validation dataset. The optimized ensemble showed the best predictive accuracy or similarly to the best individual method for the 5-year risk in the validation sets. Furthermore, the prediction model demonstrated better discrimination performance in the first example than the second example using KM curves for the predicted risk categorized into quartiles and hazard ratios. The results from these two real applications showed that our proposed method could improve on single survival based methods such as Cox PH model or on other single strategies that use a pre-processing step such as only IPCW or IPCW with bagging or pseudo-observations.

4.3 PROJECT 3

In Project 3, we proposed to integrate pseudo-observations into CNN methods in order to make risk predictions based on medical images and clinical covariates in a context of right-censored outcome data. Pseudo-observations is used as a pre-processing step and allows the researcher to implement existing CNN methods with standard loss functions such as mean squared error and handling censoring without relying on complicated loss functions or Cox partial likelihood loss function under the assumption of PH that were used by previous works that have applied CNN for survival predictions. The performance of the proposed method is assessed in simulation studies that differs in terms of independent/dependent censoring based on the CIFAR-10 images and a real data example in breast cancer from The Cancer Genome Atlas (TCGA). The results are compared to the existing CNN with Cox loss (CoxCNN). Our simulation results showed that our proposed method performed similarly to CoxCNN in a small sample setting but in a large sample setting our proposal outperformed CoxCNN. Using a Cox PH model for the censoring mechanism to handle dependent censoring improved the predictive accuracy slightly when the censoring model is correct. Otherwise, there were losses in accuracy due to incorrect modelling. The results found in the application in the TCGA data were consistent with those found in the simulation.

The proposed method facilitates the application of deep CNN methods to time-to-event data with a simple and easy-to-modify loss function that contributes to modern image-based precision medicine.

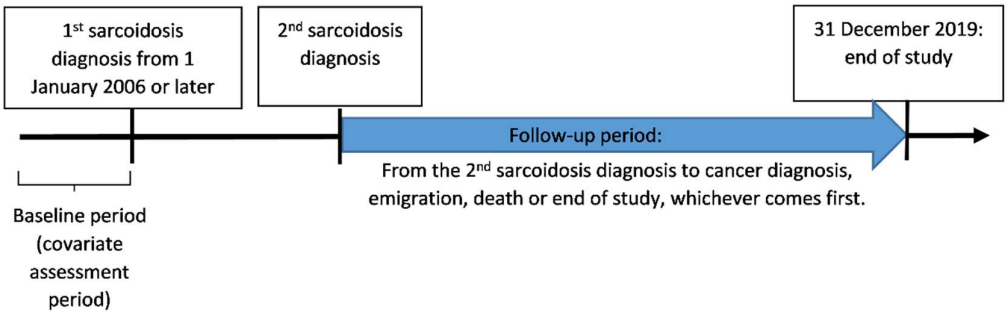
4.4 PROJECT 4

Motivated by the fact that Sweden has the largest sarcoidosis incidence and the reported association between sarcoidosis and an increased risk of cancer (Arkema, Grunewald, Kullberg, Eklund, & Askling, 2016), we aimed at building a cancer risk prediction accounting for censoring and competing risk of death and emigration.

We applied the methods developed in Project 1 and 2 with the aim of predicting the five- and ten-year risk of any cancer diagnosis among patients that were diagnosed with sarcoidosis. For that purpose, we used a large sarcoidosis cohort from the Swedish population-based register (Arkema, Grunewald, Kullberg, Eklund, & Askling,

2016) . This prediction model may be used as a tool for identifying patients at high risk of cancer who need early prevention.

In terms of the study design, we started the follow-up at the second sarcoidosis diagnosis, and we used the information on demographics, diagnoses, and prescriptions available at the time of the first diagnosis for prediction. Patients were followed up until December 31st, 2019.



We included in the prediction model the method developed in Project 1 applied to support vector machine and k-nearest neighbors, Cox model with step-wise variable selection, Cox model with lasso variable penalty, survival random forest, a boosted Cox model, a direct binomial regression and pseudo-observation based linear regression with and without penalization. We performed a 10-fold cross-validation where in each fold the models were trained on training set (70%) and optimally stacked and the performance of the resulting models is evaluated on a validation set (30%). For the training, we used a PO-based MSE instead of the PO-based AUC used in Project 2. For the validation set, we used the time-dependent AUC estimated using PO.

The results showed a cause-specific cumulative incidence of all cancer types of 4.67% at five years and 8.88% at ten years. The ensemble had the best predictive accuracy for the five-year risk of cancer (an AUC of 0.789) and it resulted in an AUC of 0.821 for the ten-year risk of cancer, which was slightly lower than the highest AUC, 0.823, obtained by one of the single method included in the ensemble. The prediction model showed good discriminatory power being able to differentiate those at the lowest 10% risk to those at the highest 10% risk. The feature importance resulted that the predictor age, at first and second diagnosis, were the most important predictor.

4.5 PROJECT 5

In Project 5, we investigated the utility of deep RL to discover the best face covering policy with the aim at reducing the spread of COVID-19. We used a retrospective dataset that is freely available online. The data includes historical daily records at the national level up to the date of publication. We restricted the data until 31 of March 2021 to focus on the period where the non-pharmaceutical interventions such as face coverings and lockdowns were the main actions to limit the spread of the virus and the roll-out of the vaccine was still very slow or null in most countries. The final dataset contained 140 countries and the total number of observations was 50760. We examined two batch deep RL algorithms: deep Q-learning and discrete batch-constrained Q-learning. We performed 5-fold cross-validation where in each fold the RL model was run on the training set based on 80% of the countries and its performance is evaluated on held-out test set (20%).

We considered the following state variables: new confirmed cases of COVID-19; stringency index; population density; GDP per capita; human development index; life expectancy; diabetes prevalence; cardiovascular death rate; share of the population that is 65 years and older and hospital beds.

We found that the RL algorithms tended to recommend less strict actions compared to the government. We investigated the dynamic of this looking at a few countries by continent: Australia and New Zealand in Oceania; Argentina, Brazil, Canada and Mexico in America; China and Japan in Asia; Italy, Spain and Sweden in Europe; Israel in Middle East, and Niger and South Africa in Africa. The results showed that the RL algorithms recommended similar policy to what the country implemented.

We found that the RL algorithms suggested similar dynamic but a different level of the policy like Japan. We also found cases, like in Sweden, that the RL algorithms recommended the same level of the policy for a period.

We concluded that batch RL may be a useful decision support tool for implementing public health policies to the COVID-19 and future pandemics.

5 DISCUSSION

The overall aim of this work has been to develop methods for prediction and medical decision-making. The first three methodological projects and the application in the fourth project contribute to the literature on the implementation of machine learning methods for time-to-event data with right censoring. In Project 1, we extended the idea of combining IPCW and ensemble methods proposed by Hothorn et al. (2006) to allow for all ML methods for classification to be applied for survival data that suffers from right-censoring, including in the presence of competing risk, and a procedure to stack the predictions from these ML methods. In Project 2, we extended the idea of Project 1 to consider in the ensemble the methods suggested in Project 1 for classification methods, ML methods for continuous outcomes and classical survival methods. In Project 3, we proposed a method based on PO to fit deep CNN models to survival data using standard loss functions such as MSE. These projects build the adaptation of ML methods to survival data using a pre-processing step either IPCW or PO or both.

Similar to Super Learner (van der Laan, Polley, & Hubbard, 2007), the optimally stacked prediction in Project 1 and Project 2 is guaranteed to perform, on average, at least as well as the best single method included in the stack. Our proposed method allows for an implementation of a Super Learner approach using survival data without limiting the types of ML methods in the stack.

In Project 4, we applied the methodology developed in Project 1 and Project 2 to build a risk prediction model for any cancer diagnosis among sarcoidosis patients accounting for censoring and competing risk.

We demonstrated the implementation of IPCW-AUC-based loss function for stacking in Project 1, the pseudo-observation-based AUC in Project 2 and the pseudo-observation-based MSE in Project 4. Whereas, the AUC loss function available in the Super Learner package (Polley, LeDell, Kennedy, Lendle, & van der Laan, 2019) is unweighted and thus it does not account for censoring.

Most of this thesis focuses on the AUC as performance metric. However, other performance metrics may be used such as IPCW non-negative log-likelihood in Project 1. Whereas, the use of other loss function in Project 2 is an area of future research. The use of PO in Project 3 enables us to use the MSE or MAE.

We demonstrated that under the assumption of coarsening at random (Robins & Finkelstein, 2000), our suggested methods can account for dependent censoring. Lastly, the proposed procedures are computationally intense. This may limit the number of methods that one would include in the stack. Moreover, the performance of the CNN model in Project 3 is limited to the small amount of images. The lack of investigation in detail of tuning hyper-parameters of each method may also limit the performance of the procedures proposed in this thesis.

6 ACKNOWLEDGEMENT

To my main supervisor, Erin Gabriel. Thank you for giving me the opportunity of doing this Ph.D. It has been a pleasure to work along you for these four years. I have learned a lot from you. Thank you for all your help and guidance in this journey. I also want to thank my co-supervisors Arvid Sjölander, Mattias Rantalainen and Ujjwal Neogi for your time and sharing with me your expertise in several discussions we have had. Thank you Mattias Rantalainen for lending me a high-performance computer in order to work on my Project 3.

To Michael Sachs, for the fruitful collaboration. Thank you for sharing your statistical expertise and your help and guidance through this process too.

Thank Ziad El-Khatib and Elizabeth Arkema for your generosity and your time.

Thank to my office mate and friend Nickolaos Skourlis for being there. It was a pleasure and fun to share the office along four years! Thank Ale, Abbi, Marco, Emilio and Jet! Thanks to the Pub MEB organizers!

Thanks to all my colleagues in the biostatistics corridor.

Thanks to all the administrative staff.

Thanks to all MEB!

Above all, the most important, thanks to my partner Romola, my family and friends for supporting me along this journey!

7 REFERENCES

Aalen, O., & Johansen, S. (1978). An empirical transition matrix for non-homogeneous Markov. *Scandinavian Journal of Statistics*, 141-150.

Andersen, P., & Perme, M. (2009). Pseudo-observations in survival analysis. *Statistical methods in medical research*, 71-99.

Arkema, E., Grunewald, J., Kullberg, S., Eklund, A., & Askling, J. (2016). Sarcoidosis incidence and prevalence: a nationwide register-based assessment in Sweden. *European Respiratory Journal*, 1690-1699.

Bandyopadhyay, S., Wolfson, J., Vock, D., Vazquez-Benitez, G., Adomavicius, G., Elidrisi, M., . . . O'Connor, P. (2015). Data mining for censored time-to-event data: a Bayesian network model for predicting cardiovascular risk from electronic health record data. *Data Mining and Knowledge Discovery*, 1033-1069.

Binder, H. (2013). *CoxBoost: Cox models by likelihood based boosting for a single survival endpoint or competing risks.* R package version 1.4.

Blanche, P., Dartigues, J. F., & Jacqmin-Gadda, H. (2013). Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med.* , 5381-97.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, 123-140.

Breiman, L. (1996). Stacked regressions. *Machine Learning*, 49–64.

Ching, T., Zhu, X., & Garmire, L. (2018). Cox-nnet: An artificial neural network method. *PLOS Computational*, 1-18.

Collins, F. S., & Varmus, H. (2015). A New Initiative on Precision Medicine. *New England Journal of Medicine*, 793-795.

Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 187-202.

Faraggi, D., & Simon, R. (1995). A neural network model for survival data. *Statistics in Medicine*, 73-82.

Fong, Y., Yin, S., & Huang, Y. (2016). Combining Biomarkers Linearly and Nonlinearly for Classification Using the Area Under the ROC Curve. *Stat Med*, 3792-3809.

Fugimoto, S., Meger, D., & Precup, D. (2019). Off-policy deep reinforcement learning without exploration. *International Conference of Machine Learning.*

Fujimoto, S., Conti, E., Ghavamzadeh, M., & Pineau, J. (2019). *Benchmarking Batch Deep Reinforcement Learning Algorithms.*

Gordon, L. L., & Olshen, R. A. (1985). Tree-structured survival analysis. *Cancer treatment reports*, 1065-1069.

Graw, F., Gerds, T., & Schumacher, M. (2009). On pseudo-values for regression analysis in competing risks models. *Lifetime Data Analysis*, 241-255.

Graw, F., Gerds, T., & Schumacher, M. (2009). On pseudo-values for regression analysis in competing risks models. *Lifetime Data Anal*, 241–255.

Heagerty, P., & Zheng, Y. (2005). Survival model predictive accuracy and ROC curves. *Biometrics*, 92-105.

Heagerty, P., Lumley, T., & Pepe, M. (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*, 337-344.

Hothorn, T., Buhlmann, P., Dudoit, S., Molinaro, A., & Van Der Laan, M. (2006). Survival ensembles. *Biostatistics*, 355-373.

Hothorn, T., Bühlmann, P., Dudoit, S., Molinaro, A., & Van Der Laan, M. J. (2006). Survival ensembles. *Biostatistics*, 355–373.

Hothorn, T., Lausen, B., Benner, A., & Radespiel-Tröger, M. (2004). Bagging survival trees. *Statistics in Medicine*, 77-91.

Ishwaran, H., Gerds, T., Kogalur, U., Moore, R., Gange, S., & Lau, B. M. (2014). Random survival forests for competing risks. *Biostatistics*, 757-773.

Katzman, J. L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., & Kluger, Y. (2018). DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Medical Research Methodology*.

Lange, S., Gabel, T., & Riedmiller, M. (2012). Batch Reinforcement Learning. In M. v. In: Wiering, *Reinforcement Learning. Adaptation, Learning, and Optimization* (pp. 45–73). Berlin, Heidelberg: Springer.

Li, H., Boimel, P., Janopaul-Naylor, J., Zhong, H., Xiao, Y., Ben-Josef, E., & al, e. (2019). Deep convolutional neural networks for imaging data based survival analysis of rectal cancer. *IEEE 16th International Symposium on Biomedical Imaging*, (pp. 846–849).

Luck, M., Sylvain, T., Cardinal, H., Lodi, A., & Bengio, Y. (2017). *Deep Learning for Patient-Specific Kidney Graft Survival Analysis.* arXiv.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A., Veness, J., Bellemare, M., . . . Kumar, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 529–533.

Mobadersany, P., Yousefi, S., Amgad, M., Gutman, D., Barnholtz-Sloan, J., Velazquez Vega, J., . . . Cooper, L. (2018). Predicting cancer outcomes from histology and genomics using convolutional networks. *Proceedings of the National Academy of Sciences*, (pp. E2970-E2979).

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2021). *Foundations of Machine Learning.* The MIT Press.

Molinaro, A. M., Dudoit, S., & Laan, M. J. (2004). Tree-based multivariate regression and density estimation with right-censored data. *Journal of Multivariate Analysis*, 154 - 177.

Polley, E., LeDell, E., Kennedy, C., Lendle, S., & van der Laan, M. (2019). *Superlearner: Super learner prediction.* R package version 2.0-26.

Robins, J., & Finkelstein, D. (2000). Correcting for noncompliance and dependent censoring in an aids clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics*, 779-788.

Robins, J., & Finkelstein, D. (2000). Correcting for Noncompliance and Dependent Censoring in an AIDS Clinical Trial with Inverse Probability of Censoring Weighted (IPCW) Log-Rank Tests. *Biometrics*, 779-788.

Satten, G., Datta, S., & Robins, J. (2001). Estimating the marginal survival function in the presence of time dependent covariates. *Statistics & Probability Letters*, 397-403.

Sutton, R. (1988). Learning to predict by the methods of temporal differences. *Machine Learning*, 9-44.

Sutton, R., & Barto, A. (1998). *Introduction to Reinforcement Learning.* Cambridge, MA, USA: MIT Press.

Uno, H., Cai, T., Tian, L., & Wei, L. (2007). Evaluating Prediction Rules for t-Year Survivors With Censored Regression Models. *Journal of the American Statistical Association*, 527-537.

van der Laan, M. J., Polley, E. C., & Hubbard, A. E. (2007). Super learner. Statistical Applications. *Statistical Applications in Genetics and Molecular Biology*, 1-23.

Vock, D. M.-B. (2016). Adapting machine learning techniques to censored time-to-event health record data: a general-purpose approach using inverse probability of censoring weighting. *Journal of Biomedical Informatics*, 119-131.

Watkins, C., & Dayan, P. (1992). Q-learning. *Machine Learning*, 279-292.

Wolfson, J. a.-B. (2015). A naive Bayes machine learning approach to risk prediction using censored, time-to-event data. *Statistics in Medicine*, 2941-2957.

Xiang, F., & Murray, S. (2012). Restricted mean models for transplant benefit and urgency. *Statistics in Medicine*, 561-576.

Zheng, Y., Cai, T., Jin, Y., & Feng, Z. (2012). Evaluating prognostic accuracy of biomarkers under competing risk. *Biometrics*, 388-396.

Zhu, X., Yao, J., & Huang, J. (2016). Deep convolutional neural network for survival analysis with pathological images. *IEEE International Conference on Bioinformatics and Biomedicine*, (pp. 544–547).