A New Deep Learning Approach

for Predicting Survival Processes in the Presence of Semi-Competing Risks

Stephen Salerno, MS [1] Yi Li, PhD [1]

June 21, 2022

[1]Department of Biostatistics, University of Michigan

Motivation: Lung Cancer Prognosis

- Lung cancer remains one of the leading causes of cancer-related deaths to date, with a 5-year survival rate of approximately 1 in 5

- Prognosis varies greatly and depends on several individualized risk factors including smoking status, genetic variants, and other comorbid conditions

- Patients diagnosed with lung cancer may experience a disease progression, go into remission, or have a recurrence prior to death

Semi-Competing Risks

- Mortality is often studied without consideration of competing events, or composite endpoints such as progression-free survival are constructed, which measure the time to the first of multiple events

- When progression and death do not correlate well, particularly for cancers with long post-progression survival, the effects of certain risk factors may differ across 'states' of a patient's disease trajectory

- Many survival processes involve a non-terminal (e.g., disease progression) and a terminal (e.g., death) event, which form a *semi-competing* relationship [2]

Disease Prognostication

- Disease prognostication is a complex task, as it often relies on the unique risk factors and health events spanning a patient's entire clinical course to predict outcomes with any accuracy

- Deep learning has emerged as a powerful tool for survival prediction; however, limited work has been done to predict multi-state or competing risk outcomes, let alone semi-competing

- We propose a new deep learning framework for semi-competing outcomes based on a compartment-type model [3, 6, 5]

Semi-Competing Data and Notation

Let T_{i1} denote the time to the non-terminal event, T_{i2} the time to the terminal event, and C_i the censoring time for the ith individual. We observe:

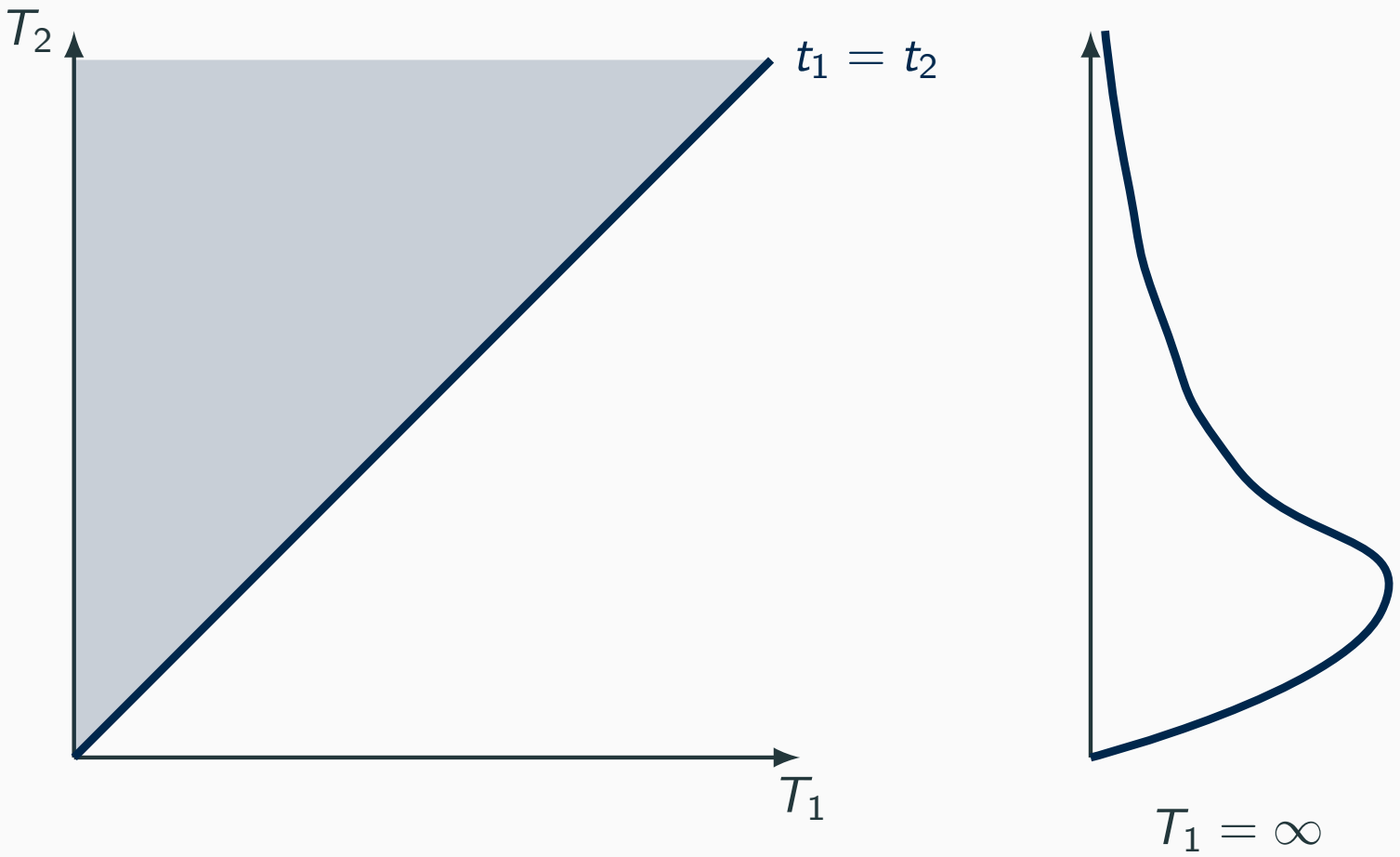
$$\mathcal{D} = \{(Y_{i1},\ \delta_{i1},\ Y_{i2},\ \delta_{i2},\ x_i);\ i = 1, \ldots, n\}$$

with $Y_{i2} = T_{i2} \wedge C_i$, $\delta_{i2} = I(T_{i2} \leq C_i)$, $Y_{i1} = T_{i1} \wedge Y_{i2}$, $\delta_{i1} = I(T_{i1} \leq Y_{i2})$, x_i is a p-vector of covariates, and $I(\cdot)$ denotes the indicator function

- Define $T_{i1} = \infty$ if the terminal event occurs before the non-terminal
- Let $f(t_1, t_2); 0 \leq t_1 \leq t_2$ denote the joint PDF of (T_1, T_2), which assigns probability mass to the 'upper wedge' on which $T_1 < T_2$
- We attribute the balance of probability to the line $t_1 = \infty$ with density $f_\infty(t_2);\ t_2 > 0$ [7, 4]

Joint Distribution of (T_1, T_2)

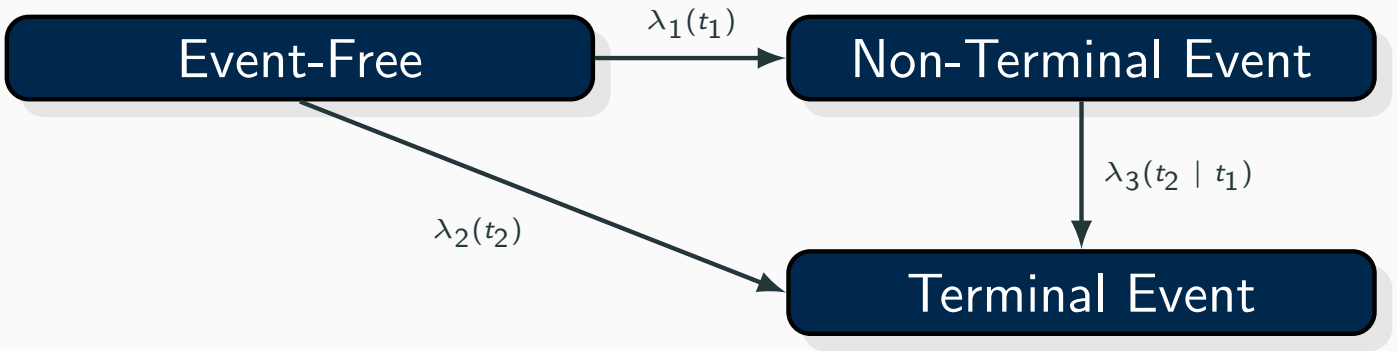
Figure 1: Graphical representation of the joint distribution of $(T_1,\ T_2)$ based on the illness–death model



The Illness-Death Model

We formulate our approach based on the illness-death model, a compartment-type model for the rates at which individuals transition between states:

Figure 2: Illness-Death Model Framework



$$\lambda_1\left(t_1 \mid \gamma_i, x_i\right) = \gamma_i \lambda_{01}\left(t_1\right) \exp\left\{h_1(x_i)\right\}; \quad t_1 > 0 \tag{1}$$

$$\lambda_2\left(t_2 \mid \gamma_i, x_i\right) = \gamma_i \lambda_{02}\left(t_2\right) \exp\left\{h_2(x_i)\right\}; \quad t_2 > 0 \tag{2}$$

$$\lambda_3\left(t_2 \mid t_1, \gamma_i, x_i\right) = \gamma_i \lambda_{03}\left(t_2 \mid t_1\right) \exp\left\{h_3(x_i)\right\}; \quad 0 < t_1 < t_2 \tag{3}$$

Method

Integrating out the frailty term, γ_i, in the conditional likelihood based on (1) - (3), we derive the following objective function:

$$\begin{aligned}\mathcal{L}(\theta,\ h_g(\cdot) \mid \mathcal{D}) = &\sum_{i=1}^{N} \delta_{i1}\{\log \lambda_{01}(y_{i1}) + h_1(x_i)\} \\ &+ \delta_{i2}(1-\delta_{i1})\{\log \lambda_{02}(y_{i2}) + h_2(x_i)\} \\ &+ \delta_{i1}\delta_{i2}\{\log \lambda_{03}(y_{i2}-y_{i1}) + h_3(x_i) + \log(1+\theta)\} - (\theta^{-1} + \delta_{i1} + \delta_{i2}) \\ &\times \log[1 + \theta\{\Lambda_{01}(y_{i1})e^{h_1(x_i)} + \Lambda_{02}(y_{i1})e^{h_2(x_i)} + \Lambda_{03}(y_{i2}-y_{i1})e^{h_3(x_i)}\}]\end{aligned} \tag{4}$$

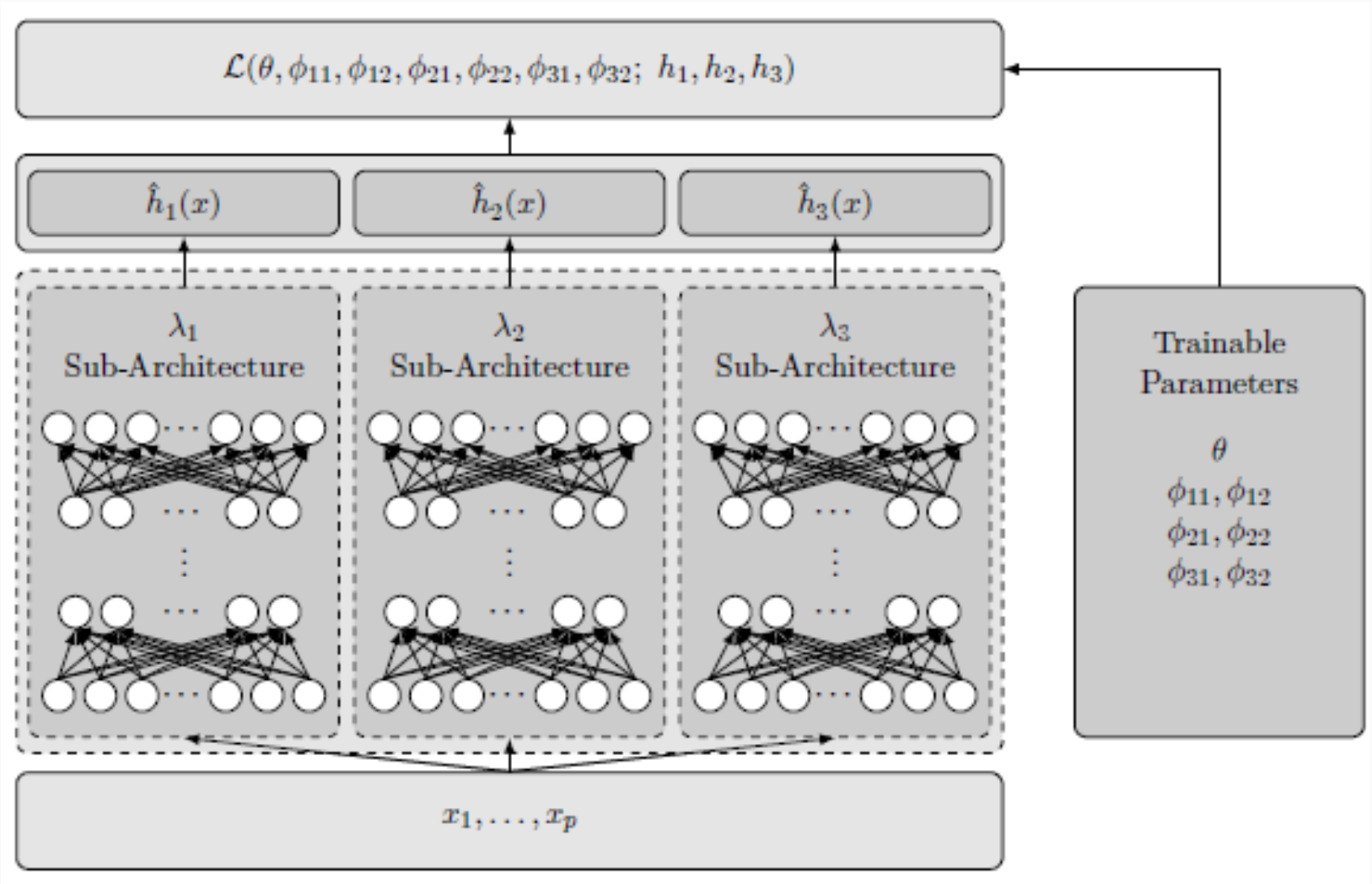
We opt for a flexible, non-parametric definition of $\hat{h}_g(x_i);\ g = 1, 2, 3$ as outputs from three fully-connected, feed-forward sub-architectures

- We implement our approach using the **TensorFlow** deep learning library in `R`, with model building done using the **Keras API**

- Finite parameter training is done via the **GradientTape API** for automatic differentiation in a custom forward pass operation

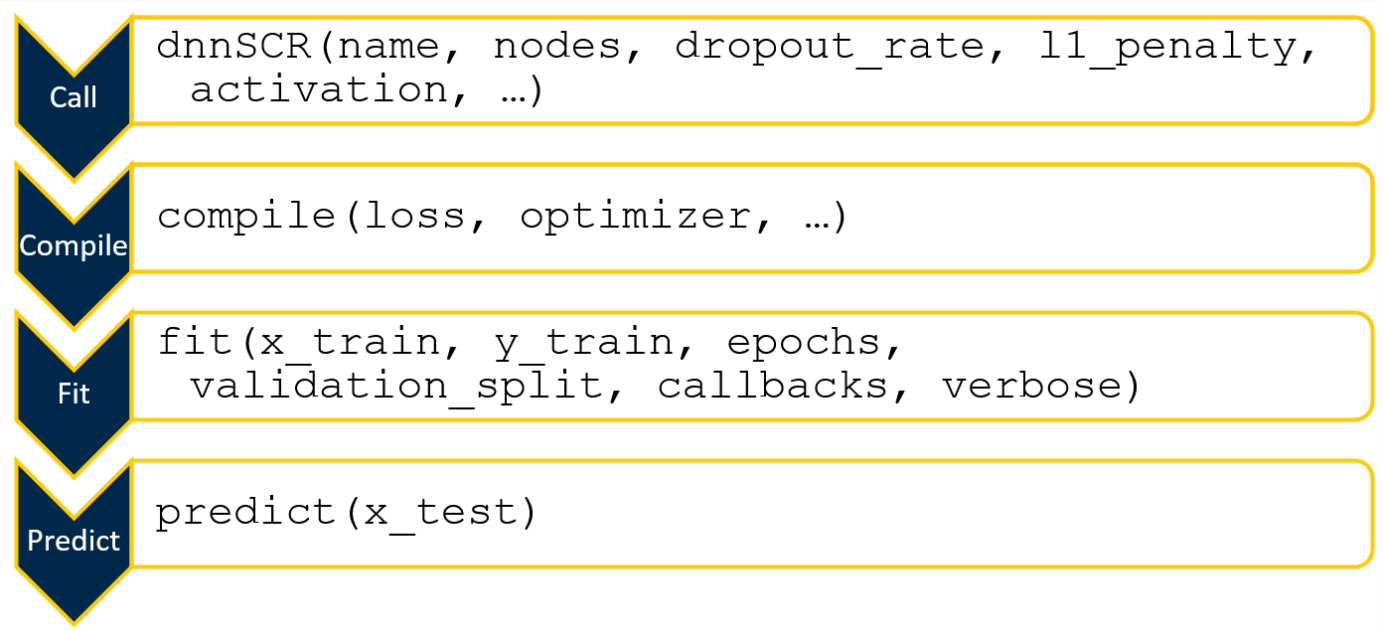
Architecture

Figure 3: Deep Neural Network for Semi-Competing Risks Architecture



Implementation and Usage

- Taking advantage of Keras' **progressive disclosures of complexity**, we implement a custom model in a **standard, user-friendly manner**
- The user **instantiates** the DNN-SCR model with the custom model wrapper function, then proceed with the **typical workflow**



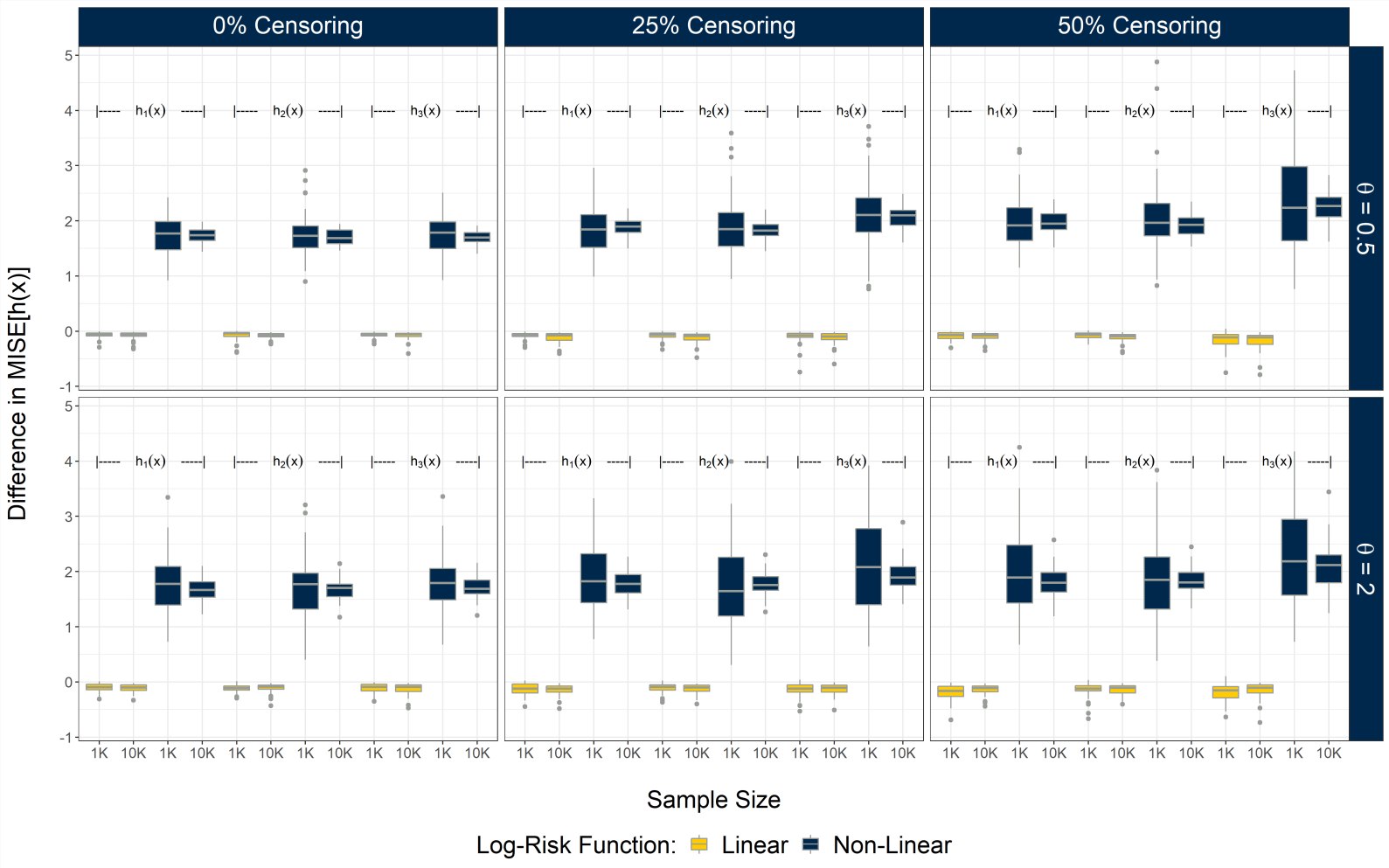
Simulation Study

We generated 50 independent datasets from (4) for each setting, fixing $\beta_g = [1, 1]^T$, $g = 1, 2, 3$, and $x_i \sim N_2(0, I_2)$, and varying:

- Sample Sizes (n): 1,000 and 10,000
- Frailty Variances (θ): 0.5 and 2
- Censoring Rates: 0%, 25%, and 50%
- Log-Risk Functions:
  - Linear: $h_g(x_i) = x_i^{\mathrm{T}} \beta_g$
  - Non-Linear: $h_g(x_i) = \log(|x_i|^{\mathrm{T}} \beta_g)$

Simulation Results

Figure 4: Difference in mean integrated squared errors (MISE) of $E||\hat{h_g} - h_g||_2^2$ $g = 1, 2, 3$ for Classical MLE - DNN-SCR



Bivariate Brier Score

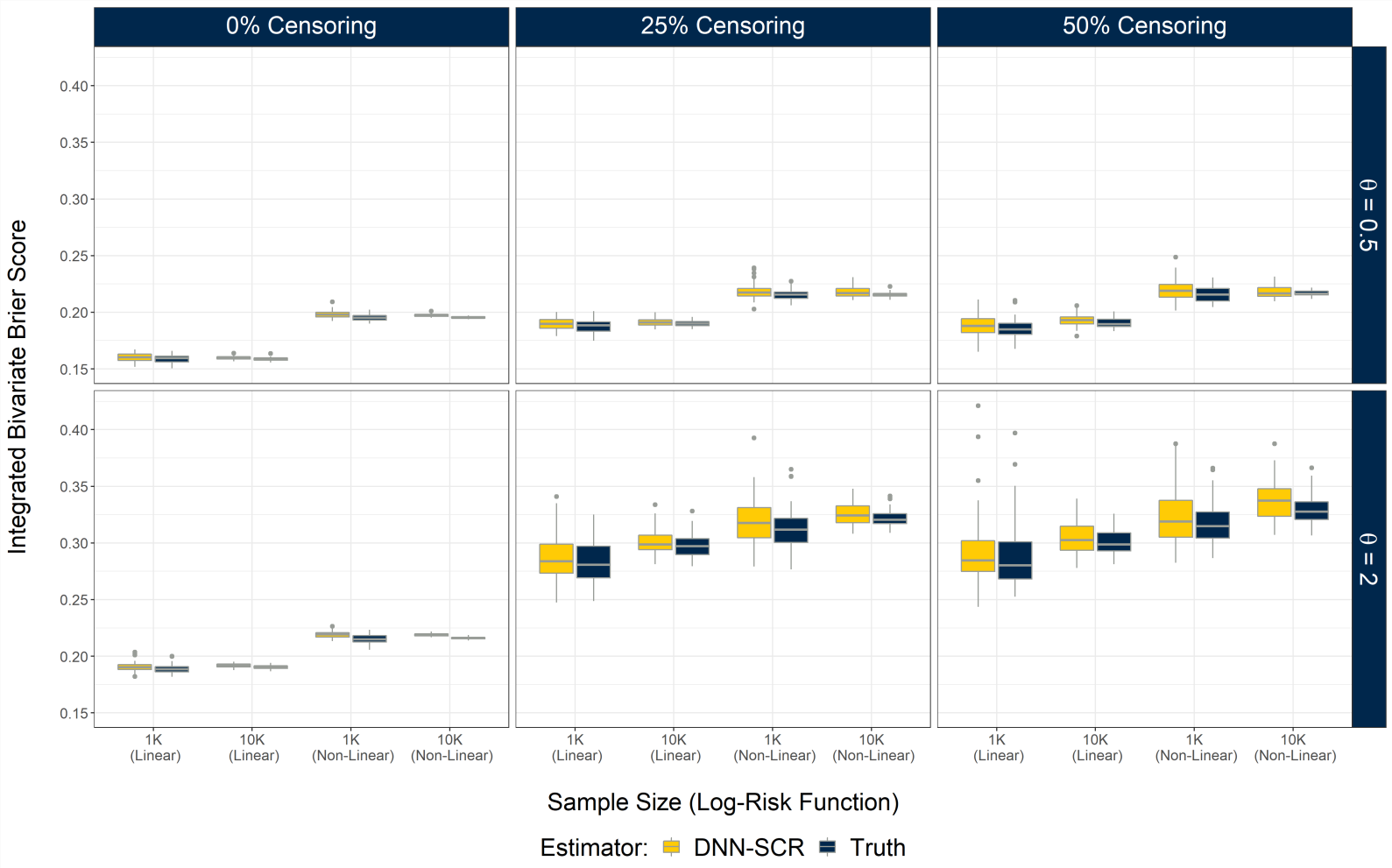
As evaluating predictive performance under semi-competing risks has not yet been explored, we extend the Brier Score for right-censored data to the bivariate survival function:

$$
\begin{aligned}
BBS_c = & \frac{\pi_i(t)^2 \cdot \mathbb{I}\{Y_{i1} \leq t,\ \delta_{i1} = 1,\ Y_{i1} \leq Y_{i2}\}}{G_i(Y_{i1})} \\
& + \frac{\pi_i(t)^2 \cdot \mathbb{I}\{Y_{i1} \leq t,\ Y_{i2} \leq t,\ \delta_{i1} = 0,\ \delta_{i2} = 1,\ Y_{i1} \leq Y_{i2}\}}{G_i(Y_{i2})} \\
& + \frac{[1 - \pi_i(t)]^2 \cdot \mathbb{I}\{Y_{i1} > t,\ Y_{i2} > t\}}{G_i(t)}
\end{aligned} \tag{5}
$$

We calculate the integrated Bivariate Brier Score for 1-year survival over a sequence of 100 evenly spaced time points in simulation

Bivariate Brier Score Results

Figure 5: Integrated Bivariate Brier Score for DNN-SCR versus the true bivariate survival function



Boston Lung Cancer Study Cohort

Our study includes 5,296 patients with non-small cell lung cancer, diagnosed between June 1983 and October 2021 [1]

We investigate time to disease progression and death, where progression might be censored by death or the study endpoint

Figure 6: BLCSC Study Outcomes

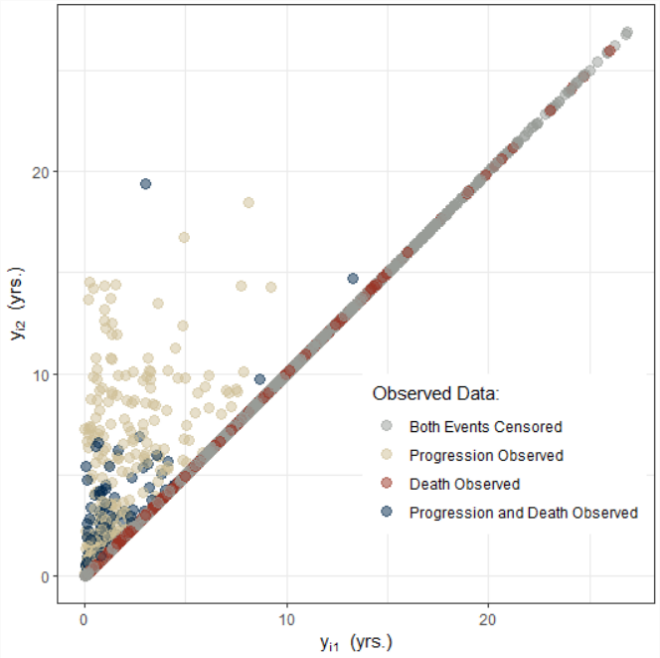


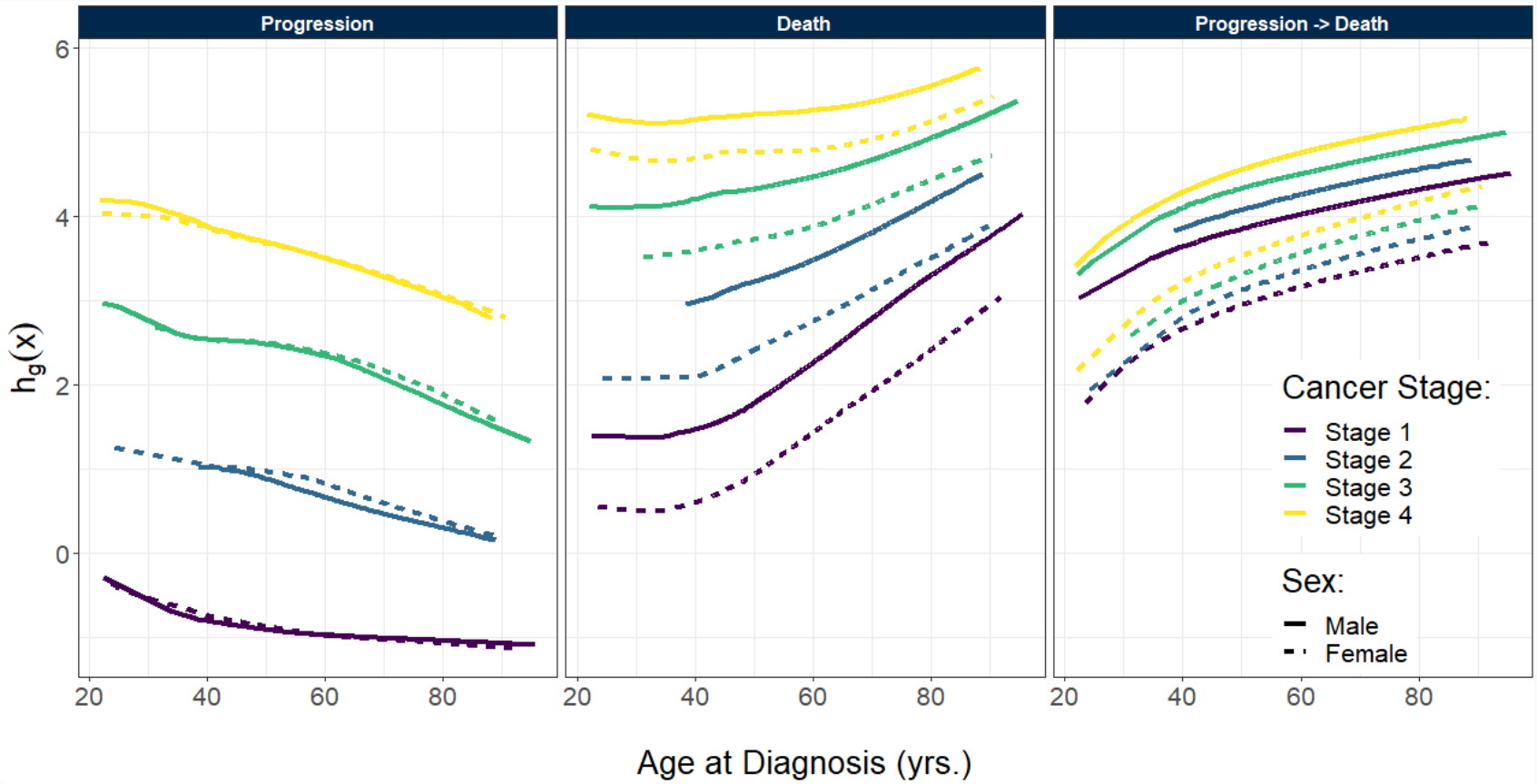
Table 1: Observed Outcomes in the BLCSC Study

	Progression	Censored
Death	111 (2%)	1,916 (36%)
Censored	224 (4%)	3,045 (58%)

BLCSC Study Results

We estimate the frailty variance to be 3.55, suggesting that progression is highly correlated with death. iBBS for 5-year survival was 0.178

Figure 7: Hazard functions for the effect of age at diagnosis on each state transition, stratified by sex and initial cancer stage



Extension: Neural Expectation-Maximization

- Previously, we assumed a parametric (Weibull) form for the baseline hazards, with parameters ϕ_{1g}, ϕ_{2g} which we optimized directly

- Other specifications of the objective function, particularly a fully non-parametric baseline hazard, may allow for even greater prediction accuracy

- We extend our approach to this setting with a novel neural expectation-maximization (EM) algorithm and piecewise-constant baseline hazards with jumps at unique observed failure times

Non-Parametric Baseline Hazards

- Rather than assuming a parametric form, an alternative convention is to estimate the baseline hazards non-parametrically

- Let the non-parametric MLEs (NPMLEs) of Λ_{0g} be non-decreasing step functions with jumps at unique observed failure times

- Note, the Hessian matrix for the NPMLE Newton-Raphson algorithm is not sparse, and its size increases linearly in n

- The EM algorithm is a more numerically stable approach, especially for large sample sizes

EM Algorithm

- Viewing the subject-specific frailty as a latent, random effect, the EM algorithm

- The algorithm iterates between two steps, namely the expectation (E) step and the maximization (M) step

- In the E-step, the frailties are estimated given the data and current values for the model parameters

- In the M-step, the model parameters are maximized given the current estimates for the frailties

Neural EM Algorithm

0. **Initialization**: Initialize parameter estimates with reasonable guesses

    a. Initialize the frailty variance (θ) using the maximum likelihood estimate from the semi-parametric model

    b. Initialize λ_{01}, λ_{02}, and λ_{03} using the Nelson-Aalen estimates for baseline hazard functions with no covariates

1. **E-Step**: Update the posterior mean of the frailty terms (γ_i)

2. **M-Step**: Update the estimated risk functions and baseline hazards

    a. Update the estimated log-risk functions, $\hat{h}_g(x_i)$, **with outputs from the neural network sub-architectures**

    b. Update the estimates of the baseline hazards

3. **Iteration**: Iterate steps (1) and (2) until convergence is achieved

E-Step

Given our objective function, the conditional distribution of γ_i given the observed data and current estimates for the model parameters is:

$$\text{Gamma}\left(\frac{1}{\tilde{\theta}} + \delta_{i1} + \delta_{i2},\ \frac{1}{\tilde{\theta}} + \tilde{\Lambda}_{01}\left(Y_{i1}\right)\exp\{\tilde{h}_1(x_i)\} + \tilde{\Lambda}_{02}\left(y_{i1}\right)\exp\{\tilde{h}_2(x_i)\} + \delta_{i1}\tilde{\Lambda}_{03}\left(y_{i1}, y_{i2}\right)\exp\{\tilde{h}_3(x_i)\}\right)$$

and the posterior means are:

$$\frac{1/\tilde{\theta} + \delta_{i1} + \delta_{i2}}{1/\tilde{\theta} + \tilde{\Lambda}_{01}\left(Y_{i1}\right)\exp\{\tilde{h}_1(x_i)\} + \tilde{\Lambda}_{02}\left(y_{i1}\right)\exp\{\tilde{h}_2(x_i)\} + \delta_{i1}\tilde{\Lambda}_{03}\left(y_{i1}, y_{i2}\right)\exp\{\tilde{h}_3(x_i)\}}$$

M-Step

Updates for the baseline hazards resemble Nelson-Aalen type estimators:

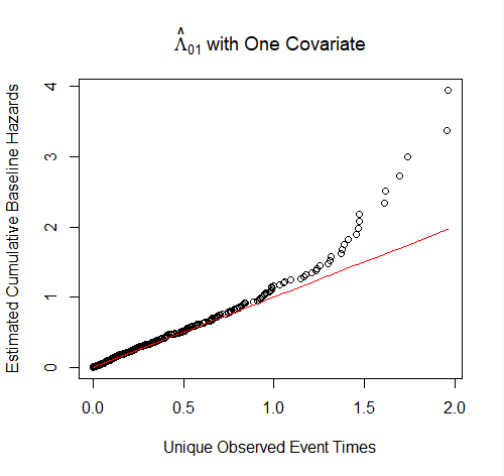
$$\tilde{\lambda}_{01}(y_{i1}) = \frac{\delta_{i1}}{\sum_{j=1}^{n} I(y_{i1} \leq y_{j1}) E\left[\gamma_j \exp\{\tilde{h}_1(x_i)\} \mid \mathcal{D}\right]}$$

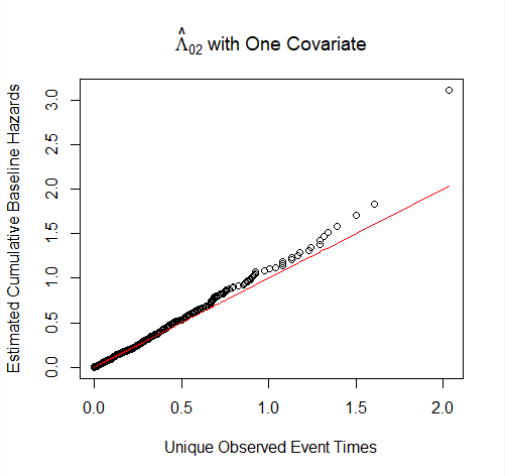
$$\tilde{\lambda}_{02}(y_{i1}) = \frac{(1-\delta_{i1})\delta_{i2}}{\sum_{j=1}^{n} I(y_{i2} \leq y_{j1}) E\left[\gamma_j \exp\{\tilde{h}_2(x_i)\} \mid \mathcal{D}\right]}$$

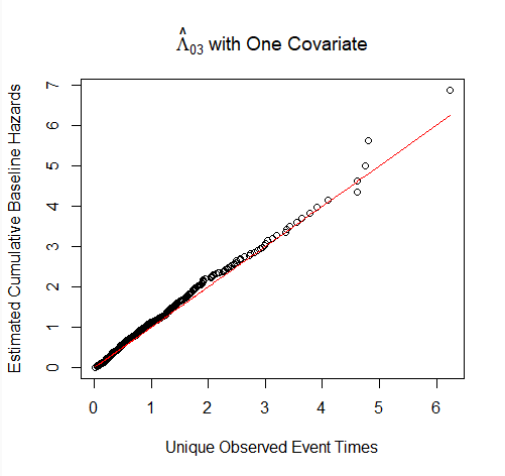
$$\tilde{\lambda}_{03}(y_{i2}) = \frac{\delta_{i1}\delta_{i2}}{\sum_{j=1}^{n} I(y_{j1} < y_{i2} \leq y_{j2}) E\left[\gamma_j \exp\{\tilde{h}_3(x_i)\} \mid \mathcal{D}\right]}.$$

Proof of Concept

We simulate data from our objective function, as described previously, now assuming a constant baseline hazard equal to one for all transitions:







Results show consistent estimation of the cumulative baseline hazards using the neural EM approach

Next Steps

- We will first validate the method with a more comprehensive simulation study
- We will continue to study the efficiency of the neural EM approach in more depth, in contrast to a fully deep learning approach explored previously

- We will apply this updated approach to the BLCSC data and include higher dimensional covariates to study the prognostic accuracy of our proposed method

Other Directions and Open Areas

- Our approach fits nicely in a Bayesian paradigm, which would facilitate formulating this as a Bayesian neural network, with individualized risk prediction intervals

- Alternatively, we can consider treating this as a classification problem, predicting survival probabilities directly with a single, sigmoidal output

Conclusions

- We have proposed a novel deep learning approach in the presence of semi-competing risks, a currently unexplored area

- Our method can recover non-linear relationships and potentially higher order interactions between disease progression, survival, and high-dimensional risk factors

- Utilizing existing paradigms for machine learning in `R`, we implement our method in a user-friendly workflow

- Extending beyond current algorithms, we further propose a novel Neural EM-Algorithm for non-parametric learning

References

David C. Christiani.
The Boston lung cancer survival cohort.
http://grantome.com/grant/NIH/U01-CA209414-01A1, 2017.
[Online; accessed November 27, 2018].

Jason P Fine, Hongyu Jiang, and Rick Chappell.
On semi-competing risks data.
Biometrika, 88(4):907–919, 2001.

Evelyn Fix and Jerzy Neyman.
A simple stochastic model of recovery, relapse, death and loss of patients.
Human Biology, 23(3):205–241, 1951.

Sebastien Haneuse and Kyu Ha Lee.
Semi-competing risks data analysis: accounting for death as a competing risk when the outcome of interest is nonterminal.
Circulation: Cardiovascular Quality and Outcomes, 9(3):322–331, 2016.

Questions?

salernos@umich.edu