Deep Learning Methods for Survival Analysis with Longitudinal Data

Group Meeting, November 2, 2022

Dan Coster
Outline

• Deep Parametric Time-to-Event Regression with Time-Varying Covariates
• Dynamic-DeepHit: A Deep Learning Approach for Dynamic Survival Analysis With Competing Risks Based on Longitudinal Data
• Preliminary Results
Intro to SA (Survival Analysis)

**Basic Idea** - Patients accrue over a period of months (or years) - After recruitment, patients are followed until failure event (death\cancer diagnosis time etc.) or the end of the study.
- The actual survival times will be observed for some patients.
- Some will still be alive at the end of the study.

![Graph showing calendar time and patient time](image)
Intro to SA

• **X - Time-to Event** - The time from entry into a study until a subject had an outcome (e.g. death, cancer diagnosis, equipment breakdown), a non-negative random variable (rv).

• **Right Censoring (RC)** - all is known that the subject is still free of the event at the end of the study.

• **Left Truncation (LT)** – When the event time \( T \) is the age of the subject and persons are not observed from birth but rather from some other time \( V \) corresponding to their entry into the study.
Survival Function

Notations:
A patient $i$ is characterized by:
- $x_i \in \mathcal{X}$ - a $d$-dimensional vector of covariates
- $T_i \in \mathbb{R}^+$ which represents the time until occurrence of the event of interest.
- $\delta_i = \mathbb{I}(T_i < C_i)$ that indicates the type of event observed, where $C_i$ is the censoring time of the patient.
- Observational data set $\mathcal{D}$ comprising $n$ patients assumed to be drawn $i.i.d.$ from the random tuple $\{X_i, \delta_i, \delta_i T_i + (1 - \delta_i)C_i\}$.

Objective:
- To estimate the survival function $S: (\mathcal{X}, \tau) \to [0,1]$ which represents the prob. of event occurrence after time $t$ as a function of time $t$ and patient covariates $x_i$.

\[ S(t|x_i) = \mathbb{P}(T_i > t|x_i) \]
Intro to SA

Regression Models

➢ **Time-independent covariates** - gender, race etc.
➢ **Time-dependent covariates** - whether some intermediate event has occurred by time x, serial measurement of covariates.
Motivation

• **Growing number of EHRs** have been deployed in hospitals and comprises longitudinal measurements (chronic diseases).

• Enables better **understanding of disease progression by incorporate longitudinal measurements** of biomarkers\risk factors into a model.

• Standard extensions to **survival models for longitudinal data**:  
  • Joint Modeling  
  • Landmarking

• Modern machine learning approaches to model time-varying covariates are relatively understudied
Kaplan-Meier Estimator for LTRC data

Survival Function \(S(t)\): probability that a patient is event-free up to time \(t\).

Patients at Risk \(Y_j\): 1 2 3 4 7

Failure Events \(d_j\): 0 0 0 0 1

Survival Prob. \(\hat{S}(t)\): 1 1 1 1 1

Notations
- \(L_i\): Left Truncation time
- \(C_i\): Right Censoring time
- \(X_i\): Failure Event time
- \(T_i = \min(X_i, C_i)\): Observed time
- \(Y_j = \sum_{i=1}^{n} \mathbb{I}(L_i \leq t_j \leq T_i)\)
- \(d_j\) = # of patients that experienced a failure event at time \(t_j\)

Kaplan Meier (KM) Estimator
\[
\hat{S}(t) = \prod_{t_i \leq t} \left[1 - \frac{d_j}{Y_j}\right]
\]

Our objective: To estimate the probability for being free of the failure event at least until time \(t\) based on the subject's covariates until the latest record before that time.
Deep Parametric Time-to-Event Regression with Time-Varying Covariates

Chirag Nagpal, Vincent Jeanselme, Artur Dubrawski

AAAI Spring Symposium 2021 on Survival Prediction: Algorithms, Challenges and Applications
Intro

• 
  A fully parametric approach to model censored time-to-event outcomes with time varying covariates

• Describing the conditional event distribution as a mixed mixture of parametric distributions

• Assumption - once the representations are obtained, the event arrival times are distributed as a mixture of underlying parametric distribution

• The parameters of these underlying distributions are also assumed to be functions of the obtained representations, and are learned jointly with the recurrent neural architectures.
Preliminaries

• Consider a dataset of \( N \) subjects
• The full data: \( D = \{(X_i, t_i, \delta_i)\}_{i=1}^{N} \)
• \( t_i \) - the time at which an event of interest took place
• \( \delta_i \in \{0,1\} \) an indicator for censoring (\( \delta_i = 0 \)) or failure event (\( \delta_i = 1 \)) at time \( t_i \)
• Subject \( i \) had \( j \) records at times: \( \tau_i^1, \tau_i^2, ..., \tau_i^j \)
• \( X_i \) - the set of features sampled over time along with the corresponding timestamp:
  • \( X_i = \{(x_i^1, \tau_i^1), (x_i^2, \tau_i^2), ..., (x_i^j, \tau_i^j)\} \), \( x_i \) has \( d \) covariates
  • \( X_i^j \) - the set of all covariates observed before time-step \( j \)
• remaining time-to-event at a time-step \( j \): \( r^j = t - \tau^j \)
• The remaining time-to-event distribution \( T(j) \) at a timestep \( j \)
Model

- **Main assumption** - the conditional survival distribution of an individual with covariates $x$ is a mixture of a fixed-size parametric distributions like the Weibull or Log-Normal.

$$
\mathbb{P}(T = t, X = x) = \sum_k \mathbb{P}(Z = k | X = x) \mathbb{P}(T = t | X = x, Z = k)
$$

- $\mathbb{P}(Z = k | X = x) = \text{softmax} \left( f(\Phi(x)) \right)$, $f$ is a linear function

- $\mathbb{P}(T = t | X = x, Z = k)$ is Weibull or Log-Normal with shape and scale parametrized as function of the representation ($\Phi(x)$).

- The shape and scale parameters of the underlying distributions, as well as the mixing weights, are implemented as a function of the input covariates using neural networks.
Network Architecture

$\Phi(\cdot) = \text{RNN}(\cdot)$

$X^1 \rightarrow X^2 \rightarrow \ldots \rightarrow X^j$

$\tilde{X}^1 \rightarrow \tilde{X}^2 \rightarrow \ldots \rightarrow \tilde{X}^j$

Output Representations

Deep Survival Machines Head

$\text{act}(\cdot) = \text{SELU}$

$\text{Log-Normal}: \text{act}(\cdot) = \text{Tanh}$

$P(T > t|X)$

$P(T = t|X)$

Weibull: $\text{act}(\cdot) = \text{SELU}$
Model

• $\mathbb{P}(T > r^j | \theta, \mathcal{X}^j)$ and $\mathbb{P}(T = r^j | \theta, \mathcal{X}^j)$ are function of the input representation.
• $\mathbb{P}(T = t, X = x) = \sum_k \mathbb{P}(Z = k | X = x) \mathbb{P}(T = t | X = x, Z = k)$

$$\mathbb{P}(T = r^j | \theta, \mathcal{X}^j) = \sum_k \text{softmax} \left( w^T f \left( \text{RNN}(\mathcal{X}^j) \right) \right) \mathbb{P}(T = r^j | \text{RNN}(\mathcal{X}^j), Z = k)$$

**Full log likelihood of $\mathcal{D}$:**

$$\mathcal{L}(\mathcal{D}; \theta) = \sum_{i} \sum_{j} (1 - \delta) \ln \mathbb{P}(T > r^j_i | \theta, \mathcal{X}^j) + \delta \ln \mathbb{P}(T = r^j_i | \theta, \mathcal{X}^j)$$
Model

1) \( x_i \sim D \)
   We draw the co-variates of the individual, \( x_i \)
2) \( w, \xi, \zeta \sim N(0, 1/\lambda) \)
   The parameters of the model are drawn from a zero mean Gaussian distribution.
3) \( z_i \sim \text{Discrete}(\text{softmax}(\Phi_\theta(x_i)^T w)) \)
   Conditioned on the covariates, \( x_i \) and the parameters, \( w \) we draw the latent \( z_i \)
4) \( \log \tilde{\beta}_k \sim N(\beta_0, 1/\lambda) \)
    \( \log \tilde{\eta}_k \sim N(\eta_0, 1/\lambda) \)
    The set of parameters \( \{\tilde{\beta}_k\}_{k=1}^K \) and \( \{\tilde{\eta}_k\}_{k=1}^K \) are drawn from the prior \( \beta_0 \) and \( \eta_0 \).
    \( t_i \sim \text{PRIMITIVE}(\beta_k, \eta_k) \)
5) \[ \beta_k = \tilde{\beta}_k + \text{act}(\Phi_\theta(x_i)^T \xi) \]
    \[ \eta_k = \tilde{\eta}_k + \text{act}(\Phi_\theta(x_i)^T \xi) \]
    Finally, the event time \( t_i \) is drawn conditioned on \( \beta_{z_i} \) and \( \eta_{z_i} \).

<table>
<thead>
<tr>
<th>PDF(t)</th>
<th>Weibull</th>
<th>Log-Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t/\beta ) ( \eta^{-1} e^{-(t/\beta)^\eta} )</td>
<td>( \frac{1}{t/\beta \sqrt{2\pi}} e^{-\left(\frac{\ln t - \eta}{\sqrt{2}\beta}\right)^2} )</td>
<td></td>
</tr>
<tr>
<td>CDF(t)</td>
<td>( e^{-(t/\beta)^\eta} )</td>
<td>( \frac{1}{2} \text{erfc}\left(-\frac{\ln t - \eta}{\sqrt{2}\beta}\right) )</td>
</tr>
</tbody>
</table>
Loss Functions

• **Uncensored Loss** - the maximum likelihood estimator for the uncensored data
• **Censoring Loss** - the maximum likelihood estimator for the censored data
• **Prior Loss** - the strength of the prior on the $\beta_k, \eta_k$

### Hyperparameters

<table>
<thead>
<tr>
<th>Type of RNN</th>
<th>{RNN, LSTM, GRU}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Batch Size</td>
<td>{125, 250}</td>
</tr>
<tr>
<td>Learning Rate</td>
<td>{10^{-3}, 10^{-4}}</td>
</tr>
<tr>
<td>Number of Hidden Layers</td>
<td>{1,2,3}</td>
</tr>
<tr>
<td>Hidden Layer Dim</td>
<td>{50,100,200}</td>
</tr>
<tr>
<td>number of underlying parametric</td>
<td>{3,4,6}</td>
</tr>
</tbody>
</table>

*Adam optimizer*
Results

Dataset
• MIMIC III – A large publicly available dataset of ICU inpatients with 24,430 patients, Use MIMIC-extract
• First 30 hours of data from admission to ICU

Outcomes
• Length-of-Stay (LOS) – Failure Event = Discharge, Censoring = Death
• In-hospital mortality – Failure Event = Death, Censoring = Discharge

Features
• All measurements including vital signs and medications administered, sampled every hour

Evaluation
• $\Delta t = \{1,3,7\}$ days
• Data Splitting: 70% Train, 10% Validation, 20% Test (Bootstrapped CIs)

Performance Metrics:
• AuROC, Brier Score
• Metrics were measured on the agglomerated risks computed every hour during the first 24 hours after admission

Benchmarks
• DeepSurv, DeepHit
## Results

### Area under ROC Curves

<table>
<thead>
<tr>
<th>Model</th>
<th>Death&lt;1</th>
<th>Death&lt;3</th>
<th>Death&lt;7</th>
<th>LOS&gt;1</th>
<th>LOS&gt;3</th>
<th>LOS&gt;7</th>
</tr>
</thead>
<tbody>
<tr>
<td>DeepSurv</td>
<td>0.897 (0.004)</td>
<td>0.859 (0.003)</td>
<td>0.841 (0.002)</td>
<td>0.740 (0.002)</td>
<td>0.715 (0.002)</td>
<td>0.756 (0.003)</td>
</tr>
<tr>
<td>DeepHit</td>
<td>0.897 (0.004)</td>
<td>0.859 (0.003)</td>
<td>0.798 (0.003)</td>
<td>0.851 (0.001)</td>
<td>0.724 (0.001)</td>
<td>0.786 (0.002)</td>
</tr>
<tr>
<td>DSM</td>
<td>0.898 (0.004)</td>
<td>0.863 (0.002)</td>
<td>0.846 (0.002)</td>
<td>0.819 (0.002)</td>
<td>0.737 (0.001)</td>
<td>0.801 (0.001)</td>
</tr>
<tr>
<td>RDSM</td>
<td><strong>0.923 (0.003)</strong></td>
<td><strong>0.890 (0.002)</strong></td>
<td><strong>0.872 (0.002)</strong></td>
<td><strong>0.864 (0.002)</strong></td>
<td><strong>0.740 (0.002)</strong></td>
<td>0.796 (0.003)</td>
</tr>
</tbody>
</table>

### Brier score

<table>
<thead>
<tr>
<th>Model</th>
<th>Death&lt;1</th>
<th>Death&lt;3</th>
<th>Death&lt;7</th>
<th>LOS&gt;1</th>
<th>LOS&gt;3</th>
<th>LOS&gt;7</th>
</tr>
</thead>
<tbody>
<tr>
<td>DeepSurv</td>
<td>0.005 (&lt;0.001)</td>
<td>0.027 (&lt;0.001)</td>
<td>0.058 (&lt;0.001)</td>
<td>0.115 (&lt;0.001)</td>
<td>0.215 (0.001)</td>
<td>0.096 (0.001)</td>
</tr>
<tr>
<td>DeepHit</td>
<td>0.005 (&lt;0.001)</td>
<td>0.028 (&lt;0.001)</td>
<td>0.168 (&lt;0.001)</td>
<td>0.084 (0.001)</td>
<td>0.224 (&lt;0.001)</td>
<td>0.100 (0.001)</td>
</tr>
<tr>
<td>DSM</td>
<td>0.005 (&lt;0.001)</td>
<td>0.027 (&lt;0.001)</td>
<td>0.059 (&lt;0.001)</td>
<td>0.080 (0.001)</td>
<td>0.198 (&lt;0.001)</td>
<td><strong>0.088 (0.001)</strong></td>
</tr>
<tr>
<td>RDSM</td>
<td><strong>0.005 (&lt;0.001)</strong></td>
<td><strong>0.026 (0.001)</strong></td>
<td><strong>0.059 (0.002)</strong></td>
<td><strong>0.073 (&lt;0.001)</strong></td>
<td><strong>0.193 (0.001)</strong></td>
<td>0.091 (0.001)</td>
</tr>
</tbody>
</table>
Dynamic-DeepHit: A Deep Learning Approach for Dynamic Survival Analysis With Competing Risks Based on Longitudinal Data

Changhee Lee, Jinsung Yoon, Mihaela van der Schaar

IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING (2020)
Intro to Competing Risks

• Competing risks are said to be present when a patient is at risk of more than one mutually exclusive event, such as death from different causes, and the occurrence of one of these will prevent any other event from ever happening.

Gichangi & Vach (2005)
Slides from Survival Analysis for junior researchers course, by Prof. Sally R. Hinchliffe, Lenciter University
Preliminaries

- Consider a dataset of $N$ subjects
- Subject $i$ had $R_i$ records at times $t_1^i < ... < t_{R_i}^i$
- The $d$ covariates measured at time $t_j^i$ are denoted by the vector $x^i(t_j^i)$
- $X^i = \{x^i(t_1^i), ..., x^i(t_{R_i}^i)\}$ summarizes the longitudinal data of subject $i$
  - $X^i(t) = \{x^i(t_j^i) : 0 \leq t_j^i \leq t \}$
- $M^i = \{m_1^i, ..., m_{R_i}^i\}$ is a sequence of mask vectors
- Time interval to the next record: $\epsilon_j^i = t_{j+1}^i - t_j^i$ for $1 \leq j < R_i$ ($\epsilon_{R_i}^i = 0$), $E^i = \{\epsilon_1^i, \epsilon_2^i, ..., \epsilon_{R_i}^i\}$
- $\tau^i$ is the last time point of subject $i$ was at risk
- $\delta^i \in \{0, 1\}$ an indicator for censoring ($\delta^i = 0$) or failure event ($\delta^i = 1$) at time $\tau^i$
- $T = \{0, 1, ..., T_{max}\}$ denote the set of possible timestamps
- The full data: $D = \{(X^i, M^i, E^i, \tau^i, \delta^i)\}_{i=1}^N$

**Our (formal) objective:** Let $t^*_i = \max\{t_j^i < t \}$, wish to estimate the survival function: $S(t | X^i(t^*_i)) = \mathbb{P}(\tau^i > t | X^i(t^*_i))$
Settings - CIF (Cumulative Incidence Function)

• CIF represents the probability that a particular event \( k^* \in \mathcal{K} \) occurs on/before time \( \tau^* \) conditioned on the history of longitudinal data measurements \( \mathcal{X}^* \).

• The fact that longitudinal measurements have been recorded up to \( t^*_M \) implies survival of the subject up to this time point.

\[
F_{k^*}(\tau^*|\mathcal{X}^*) \triangleq \mathbb{P}(T \leq \tau^*, k = k^*|\mathcal{X}^*, T > t^*_M) = \sum_{m \leq \tau^*} \mathbb{P}(T = m, k = k^*|\mathcal{X}^*, T > t^*_M)
\]

• The survival probability of a subject at time \( \tau^* \) given \( \mathcal{X}^* \) is the probability that a subject doesn’t experience any event before or at time \( \tau^* \).

\[
S(\tau^*|\mathcal{X}^*) \triangleq \mathbb{P}(T > \tau^*|\mathcal{X}^*, T > t^*_M) = 1 - \sum_{k \neq \emptyset} F_k(\tau^*|\mathcal{X}^*)
\]

• Objective: estimated CIF \( \overline{F}_{k^*}(\tau^*|\mathcal{X}^*) \)
Illustration
Intro to GRU and RNN

\[ h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t) \]

- \( h_t \) size is a parameter of the model.
- RNNs have the ability to capture long-term temporal dependencies and variable-length observations and has strong prediction performance.
GRU (Gate Recurrent Unit)

For each $j$-th hidden unit, we define these functions:

For each $j$-th hidden unit, we define these functions:

\[
\begin{align*}
    r_j &= \sigma(W_r h_{j-1} + U_r [x_j m_j \epsilon_j] + b_r) \\
    z_j &= \sigma(W_z h_{j-1} + U_z [x_j m_j \epsilon_j] + b_z) \\
    \tilde{h}_j &= \tanh(W_h (r_j \odot h_{j-1}) + U_h [x_j m_j \epsilon_j]) + b_h \\
    h_j &= (1 - z_t) \odot h_{j-1} + z_j \odot \tilde{h}_j
\end{align*}
\]

- **Reset Gate**: controls what parts of previous hidden state are used to compute new content.
- **Update Gate**: controls what parts of hidden state are updated vs. preserved

**New hidden state content**: reset gate selects useful parts of prev hidden state. Use this and current input to compute new hidden content.

**Hidden state**: update gate simultaneously controls what is kept from previous hidden state, and what is updated to new hidden state content

$\sigma$ – element-wise sigmoid function, $\odot$ - element-wise multiplication.

Weight matrices $W_z, W_r, W_h, U_z, U_r, U_h$ and vectors $b_r, b_z, b_h$ which parameterize the shared subnetwork.

---

Temporal Attention Mechanism

- The temporal attention mechanism on the hidden states helps our network decide which parts of the previous longitudinal measurements to pay attention to.
- Formally, it outputs a context vector $c$ - an weighted sum of the previous hidden states as follows: $c = \sum_{j=1}^{J-1} a_j h_j$
- $a_j$ represents the importance of the j-th measurements: $a_j = \frac{\exp(e_j)}{\sum_{l=1}^{J-1} \exp(e_l)}$
- To score the importance of the $j$-th measurement, using $f_a(\cdot)$ - a two-layer feed forward that its input is the last measurement, $(x_j, m_j)$ and $h_j$ and it outputs a scalar $e_j$ as follows: $e_j = f_a(h_j, x_j, m_j)$.
- The temporal mechanism is jointly trained with all the other components of our network.

Model

Based on the Dynamic Deep-Hit competing risks model (Lee et al., 2020).

Subject $i$:

<table>
<thead>
<tr>
<th>$x_i^1, m_i^1, e_i^1$</th>
<th>$x_i^2, m_i^2, e_i^2$</th>
<th>$(x_i^{R(i-1)}, m_i^{R(i-1), R(i-1)})$</th>
<th>$(x_i^R, m_i^R)$</th>
</tr>
</thead>
</table>

Standardization: $\tilde{x}_j^i = \frac{(x_j^i - \mu)}{\sigma}$

L3 - Prediction Loss

$y_i^1, y_i^2, \{h_j^i\}_{j=1}^{R_i}$

RNN

Temporal Attention

$e_j = f_a(h_j, x_R^i, m_R^i)$

$a_j = \frac{\exp(e_j)}{\sum_{l=1}^{R(i-1)} \exp(e_l)}$

$c = \sum_{j=1}^{R(i-1)} a_j h_j$

Total Loss = L1+L2+L3

Feed – Forward (Fully Connected)

Softmax

L1 - Log Likelihood Loss

$L_1$ = $\mathbb{P}(T = \tau, k = K_p | x^i)$

$L_2$ - Ranking Loss

$L_2$ = $\mathbb{P}(T = \tau, k = K_p | x^i)$

Subject $i$:

<table>
<thead>
<tr>
<th>$o_{1,K_1}$</th>
<th>$o_{2,K_1}$</th>
<th>$o_{3,K_1}$</th>
</tr>
</thead>
</table>

$L_3$ - Prediction Loss

$L_3$ = $\mathbb{E}(L(y, o))$

$\mathbb{E}$ is the expected value.

$L = L_1 + L_2 + L_3$

Total Loss = $L_1 + L_2 + L_3$
L3 - Prediction Loss

This loss function try to incorporate the prediction error on the trajectories of time-varying covariates as an **auxiliary task** in the RNN.

- \( y_j^i \) is the step ahead prediction of the covariates \( x_{j+1}^i \) and used in order to regularize the RNN such that the hidden representations preserve information for the step-ahead predictions.

- Taking account missing measurements into consideration, the prediction loss is defined as follows:

- \( \beta \) is an hyper-parameter, and. The authors defined the function: \( \zeta(a, b) = \sum_{d \in \mathcal{J}} \zeta_d(a_d, b_d) \), where \( \mathcal{J} \) is a set of all the time-varying covariates.

- In case of **continuous** variables: \( \zeta(a, b) = |a - b|^2 \)

- In case of **binary** covariates: \( \zeta(a, b) = a \log(b) + (1 - a) \log(1 - b) \)

- \[
L_3 = \beta \cdot \sum_{i=1}^{N} \sum_{j=0}^{R - 1} m_{j+1}^i \zeta(x_{m+1}^i, y_m^i)
\]
**$L_1$ - Log Likelihood Loss (Dynamic Deep Hit)**

- The CIF of cause $k^*$:
  \[
  \hat{F}_{k^*}(\tau^*|X^*) = \frac{\sum_{t_{R^*}^*<\tau\leq\tau^*} O_{k^*,\tau}}{1 - \sum_{k\neq\emptyset} \sum_{n\leq t_{R^*}^*} O_{k,n}^*}
  \]

- The negative log-likelihood of the joint distribution of the first hitting time of the relevant event type in a right-censored setting, conditioned by measurements recorded until the last observation.

\[
L_1 = -\sum_{i=1}^{N} \left[ \mathbb{I}_{k_i \neq \emptyset} \cdot \log \left( \frac{O_{k_i,\tau_i}^i}{1 - \sum_{k\neq\emptyset} \sum_{n\leq t_{R_i}^i} O_{k,n}^i} \right) + \mathbb{I}_{k_i = \emptyset} \cdot \log \left( 1 - \sum_{k\neq\emptyset} \hat{F}_k(\tau_i^i|X_i^i) \right) \right]
\]

- minimize of the probability to have other type of event before the $k_i$ event
- minimize the probability that the subject will have any kind of event
A generalization of the concordance index concept, which assume that a patient who dies at time $t$ should have a higher risk at time $\tau$ than a subject which survive longer than $\tau$.

- The longitudinal measurements of subjects can begin at any point in their lifetime or disease progression, and this makes direct comparison of the risks at different time points difficult to assess.
- Suggestion – To compare the risks of subjects at times elapsed since their last measurements
- For subject $i$ defined $s^i = \tau^i - t_R^i$ (instead of $\tau^i$).
- In order to measure the concordance Index, a permissible pair $(i, j)$ is defined if an event $k$ had occurred to subject $i$ at time $s^i$ but subject $j$ didn’t experienced any type of event including censoring yet ($s^i < s^j$).
- We count a pair into the concordance score if in the same period of time ($s^i$) from the last measurement ($t_{M_i}^i$ or $t_{M_j}^j$) the CIF satisfies: $\hat{F}_k(s^i + t_{M_i}^i|X^i) > \hat{F}_k(s^i + t_{M_j}^j|X^j)$.
- We define $A_{k,i,j} \triangleq \mathbb{I}_{k = k,s^i < s^j}$ as indicator function for a permissible pair, $\alpha_k \geq 0$ is a hyper-parameter per $k^{th}$ event type, and $\eta(a, b) = \exp(-\frac{a-b}{\sigma})$ is a differentiable loss function which penalizes in case of a non-concordant pair.

$$L_2 = \sum_{k=1}^{K} \alpha_k \sum_{i \neq j} A_{k,i,j} \cdot \eta \left( \hat{F}_k(s^i + t_{M_i}^i|X^i), \hat{F}_k(s^i + t_{M_j}^j|X^j) \right)$$
Concordance Index (C-index)

1. Form all possible pairs of cases over the data.
2. Omit pairs:
   1. whose shorter survival time is censored
   2. \(i\) and \(j\) if \(T_i = T_j\) unless at least one is a failure
3. Let Permissible denote the total number of remaining pairs.
3. A permissible pair receives a value of 1 if:
   1. \(T_i \neq T_j\) and shorter survival time has worse predicted outcome
   2. \(T_i = T_j\) and their predicted outcomes are also equal
   3. \(T_i = T_j\) and the predicted outcome is worse for the observation with the death event
4. A permissible pair receives a value of 0.5 if:
   1. \(T_i \neq T_j\) and predicted outcomes are tied
   2. \(T_i = T_j\) and their predicted outcomes are not equal
   3. \(T_i = T_j\) and the predicted outcome is not worse for the observation with the death event
5. Let Concordance denote the sum over all permissible pairs.
6. The C-index, \(C\), is defined by \(C = \frac{\text{Concordance}}{\text{Permissible}}\)
7. Error rate is given by \(\text{Error} = 1 - C\) (\(0 \leq \text{Error} \leq 1\))

Harrell et al. (1982)
**Dataset**

**Cystic Fibrosis** - Retrospective longitudinal data from the UK Cystic Fibrosis

- 5,883 patients, 605 patients (10.28%) were followed until deaths with annual check-ups between 2009-2015

**Features**

- 11 static covariates
- 79 time-varying covariates

Covariates for individual CF patients including demographics, genetic mutations, bacterial infections, comorbidities, hospitalization, lung function scores and therapeutic management

**3 competing events:**

- Death from respiratory failure
- Death due to other causes
- Right Censoring
Results

Evaluation
• $\Delta t = \{1,3,5,10\}$ years
• 5 random 80/20 train/test splits: separated the data into a training set (80%) and Data Splitting: 60% Train, 20% Validation

Performance Metrics:
• AuROC, Brier Score
• Metrics were measured on the agglomerated risks computed every hour during the first 24 hours after admission

Benchmarks
• Joint model (JM)
• Cause Specific Cox Regression (cs-COX)
• Random Survival Forest for Competing Risks (RSF)
Concordance Index For Competing Risks

- A time-dependent and cause-specific concordance index $C_k(t, \Delta t)$
  - $t$ represents the time of the prediction
  - $\Delta t$ is the evaluation time
- for example – the performance in predicting $(k=)$ Resp. Failure at time $(t =)14$ years from the beginning of the study, in the next $(\Delta t =)3$ years.

$$C_k(t, \Delta t) = \mathbb{P} \left( \hat{F}_k(t + \Delta t | X^i(t)) > \hat{F}_k(t + \Delta t | X^j(t)) \mid T^i < T^j, k^i = k, T^i < t + \Delta t \right)$$
### Results

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>$\Delta t = 1$</th>
<th>$\Delta t = 3$</th>
<th>$\Delta t = 5$</th>
<th>$\Delta t = 10$</th>
<th>$\Delta t = 1$</th>
<th>$\Delta t = 3$</th>
<th>$\Delta t = 5$</th>
<th>$\Delta t = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\Delta t = 1$</td>
<td>$\Delta t = 3$</td>
<td>$\Delta t = 5$</td>
<td>$\Delta t = 10$</td>
<td>$\Delta t = 1$</td>
<td>$\Delta t = 3$</td>
<td>$\Delta t = 5$</td>
<td>$\Delta t = 10$</td>
</tr>
<tr>
<td>cs-Cox</td>
<td>0.840 ± 0.09</td>
<td>0.837 ± 0.08</td>
<td>0.837 ± 0.08</td>
<td>0.837 ± 0.08</td>
<td>0.667 ± 0.10*</td>
<td>0.664 ± 0.10*</td>
<td>0.665 ± 0.10*</td>
<td>0.665 ± 0.10*</td>
</tr>
<tr>
<td>RSF</td>
<td>0.936 ± 0.01†</td>
<td>0.932 ± 0.01</td>
<td>0.931 ± 0.02†</td>
<td>0.929 ± 0.01†</td>
<td>0.798 ± 0.04*</td>
<td>0.792 ± 0.04*</td>
<td>0.773 ± 0.05*</td>
<td>0.776 ± 0.05*</td>
</tr>
<tr>
<td>JM</td>
<td>0.882 ± 0.03*</td>
<td>0.896 ± 0.01*</td>
<td>0.896 ± 0.01*</td>
<td>0.897 ± 0.01*</td>
<td>0.760 ± 0.02*</td>
<td>0.795 ± 0.03*</td>
<td>0.802 ± 0.02*</td>
<td>0.812 ± 0.01*</td>
</tr>
<tr>
<td>JM-LC</td>
<td>0.897 ± 0.04†</td>
<td>0.894 ± 0.05†</td>
<td>0.894 ± 0.05†</td>
<td>0.894 ± 0.05†</td>
<td>0.856 ± 0.02*</td>
<td>0.855 ± 0.02*</td>
<td>0.855 ± 0.02*</td>
<td>0.855 ± 0.02*</td>
</tr>
<tr>
<td>[4]</td>
<td>0.910 ± 0.02*</td>
<td>0.907 ± 0.02*</td>
<td>0.907 ± 0.02*</td>
<td>0.907 ± 0.02*</td>
<td>0.819 ± 0.07†</td>
<td>0.831 ± 0.07†</td>
<td>0.834 ± 0.07†</td>
<td>0.839 ± 0.07†</td>
</tr>
<tr>
<td>Exponential</td>
<td>0.895 ± 0.03*</td>
<td>0.890 ± 0.03*</td>
<td>0.890 ± 0.03*</td>
<td>0.890 ± 0.02*</td>
<td>0.824 ± 0.05*</td>
<td>0.825 ± 0.05*</td>
<td>0.824 ± 0.05*</td>
<td>0.824 ± 0.05*</td>
</tr>
<tr>
<td>Proposed</td>
<td>0.948 ± 0.01</td>
<td>0.939 ± 0.01</td>
<td>0.938 ± 0.01</td>
<td>0.937 ± 0.01</td>
<td>0.924 ± 0.02</td>
<td>0.922 ± 0.02</td>
<td>0.921 ± 0.02</td>
<td>0.921 ± 0.02</td>
</tr>
<tr>
<td>FEV₁%</td>
<td>0.946 ± 0.01</td>
<td>0.937 ± 0.02</td>
<td>0.936 ± 0.02</td>
<td>0.932 ± 0.02</td>
<td>0.875 ± 0.04†</td>
<td>0.867 ± 0.05†</td>
<td>0.862 ± 0.05†</td>
<td>0.866 ± 0.05†</td>
</tr>
<tr>
<td>cause-spec.</td>
<td>0.949 ± 0.01</td>
<td>0.941 ± 0.01</td>
<td>0.942 ± 0.01</td>
<td>0.941 ± 0.01</td>
<td>0.929 ± 0.02</td>
<td>0.927 ± 0.02</td>
<td>0.925 ± 0.02</td>
<td>0.926 ± 0.02</td>
</tr>
<tr>
<td>full-fledged</td>
<td>0.924 ± 0.02</td>
<td>0.922 ± 0.02</td>
<td>0.921 ± 0.02</td>
<td>0.921 ± 0.02</td>
<td>0.921 ± 0.02</td>
<td>0.921 ± 0.02</td>
<td>0.921 ± 0.02</td>
<td>0.921 ± 0.02</td>
</tr>
<tr>
<td></td>
<td>0.842 ± 0.03*</td>
<td>0.842 ± 0.03*</td>
<td>0.842 ± 0.03*</td>
<td>0.842 ± 0.03*</td>
<td>0.748 ± 0.10*</td>
<td>0.749 ± 0.10*</td>
<td>0.749 ± 0.10*</td>
<td>0.749 ± 0.10*</td>
</tr>
<tr>
<td></td>
<td>0.888 ± 0.01*</td>
<td>0.887 ± 0.02*</td>
<td>0.886 ± 0.03*</td>
<td>0.891 ± 0.03*</td>
<td>0.803 ± 0.06†</td>
<td>0.771 ± 0.05*</td>
<td>0.749 ± 0.05*</td>
<td>0.746 ± 0.05*</td>
</tr>
<tr>
<td></td>
<td>0.906 ± 0.01†</td>
<td>0.905 ± 0.01†</td>
<td>0.908 ± 0.01†</td>
<td>0.909 ± 0.01†</td>
<td>0.818 ± 0.03*</td>
<td>0.814 ± 0.03*</td>
<td>0.813 ± 0.02*</td>
<td>0.840 ± 0.02*</td>
</tr>
<tr>
<td></td>
<td>0.911 ± 0.04†</td>
<td>0.910 ± 0.04†</td>
<td>0.910 ± 0.04†</td>
<td>0.910 ± 0.04†</td>
<td>0.851 ± 0.02*</td>
<td>0.851 ± 0.02*</td>
<td>0.850 ± 0.02*</td>
<td>0.850 ± 0.02*</td>
</tr>
</tbody>
</table>

$\Delta t$ indicates p-value < 0.01, † indicates p-value < 0.05.
## Results

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Resp. Failure</th>
<th>Other Causes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\Delta t = 1$</td>
<td>$\Delta t = 3$</td>
</tr>
<tr>
<td>$t = 30$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exponential</td>
<td>$0.910 \pm 0.02^*$</td>
<td>$0.907 \pm 0.02^*$</td>
</tr>
<tr>
<td>$L_1$</td>
<td>$0.895 \pm 0.03^*$</td>
<td>$0.890 \pm 0.03^*$</td>
</tr>
<tr>
<td>$L_2$</td>
<td>$0.897 \pm 0.02^*$</td>
<td>$0.890 \pm 0.03^*$</td>
</tr>
<tr>
<td>$L_1 + L_2$</td>
<td>$0.942 \pm 0.02$</td>
<td>$0.934 \pm 0.02$</td>
</tr>
<tr>
<td>$\frac{L_1}{L_{total}}$</td>
<td>$0.907 \pm 0.01^*$</td>
<td>$0.901 \pm 0.01^*$</td>
</tr>
<tr>
<td>$t = 40$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exponential</td>
<td>$0.913 \pm 0.02^*$</td>
<td>$0.923 \pm 0.02^*$</td>
</tr>
<tr>
<td>$L_1$</td>
<td>$0.883 \pm 0.03^*$</td>
<td>$0.883 \pm 0.03^*$</td>
</tr>
<tr>
<td>$L_2$</td>
<td>$0.926 \pm 0.01^*$</td>
<td>$0.952 \pm 0.02$</td>
</tr>
<tr>
<td>$L_1 + L_2$</td>
<td>$0.930 \pm 0.06$</td>
<td>$0.958 \pm 0.01$</td>
</tr>
<tr>
<td>$\frac{L_1}{L_{total}}$</td>
<td>$0.883 \pm 0.05^*$</td>
<td>$0.907 \pm 0.02^*$</td>
</tr>
<tr>
<td>$t = 50$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exponential</td>
<td>$0.929 \pm 0.01^*$</td>
<td>$0.929 \pm 0.01^*$</td>
</tr>
<tr>
<td>$L_1$</td>
<td>$0.875 \pm 0.02^*$</td>
<td>$0.874 \pm 0.02^*$</td>
</tr>
<tr>
<td>$L_2$</td>
<td>$0.896 \pm 0.02^*$</td>
<td>$0.887 \pm 0.02$</td>
</tr>
<tr>
<td>$L_1 + L_2$</td>
<td>$0.960 \pm 0.01$</td>
<td>$0.960 \pm 0.01$</td>
</tr>
<tr>
<td>$\frac{L_1}{L_{total}}$</td>
<td>$0.864 \pm 0.01^*$</td>
<td>$0.964 \pm 0.01$</td>
</tr>
</tbody>
</table>

* indicates p-value < 0.01, † indicates p-value < 0.05
## Results

- A cause-specific time-dependent Brier Score $BS_k(t, \Delta t)$

$$BS_k(t, \Delta t) = E\left[\left(1(T^i < t + \Delta t, k^i = k) - \hat{F}_k(t + \Delta t | \mathcal{X}^i(t))\right)^2\right]$$

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>$\Delta t = 1$</th>
<th>$\Delta t = 3$</th>
<th>$\Delta t = 5$</th>
<th>$\Delta t = 10$</th>
<th>$\Delta t = 1$</th>
<th>$\Delta t = 3$</th>
<th>$\Delta t = 5$</th>
<th>$\Delta t = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>cs-Cox</td>
<td>0.085±0.02</td>
<td>0.145±0.01</td>
<td>0.225±0.02</td>
<td>0.377±0.03</td>
<td>0.060±0.02</td>
<td>0.084±0.02</td>
<td>0.156±0.01</td>
<td>0.256±0.02</td>
</tr>
<tr>
<td>RSF</td>
<td><strong>0.044±0.01</strong></td>
<td>0.053±0.00</td>
<td>0.058±0.00</td>
<td><strong>0.059±0.00</strong></td>
<td>0.012±0.00</td>
<td>0.012±0.00</td>
<td>0.013±0.00</td>
<td>0.013±0.00</td>
</tr>
<tr>
<td>JM</td>
<td>0.050±0.01</td>
<td><strong>0.051±0.00</strong></td>
<td><strong>0.051±0.00</strong></td>
<td>0.066±0.01</td>
<td>0.012±0.00</td>
<td>0.012±0.00</td>
<td>0.014±0.00</td>
<td>0.018±0.00</td>
</tr>
<tr>
<td>JM-LC</td>
<td>0.053±0.01</td>
<td>0.062±0.00</td>
<td>0.065±0.00</td>
<td>0.066±0.00</td>
<td>0.012±0.00</td>
<td>0.012±0.00</td>
<td>0.013±0.00</td>
<td><strong>0.013±0.00</strong></td>
</tr>
<tr>
<td>Proposed</td>
<td>0.058±0.01</td>
<td>0.059±0.00</td>
<td>0.059±0.00</td>
<td>0.060±0.00</td>
<td><strong>0.011±0.00</strong></td>
<td><strong>0.012±0.00</strong></td>
<td>0.013±0.00</td>
<td>0.017±0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>$\Delta t = 1$</th>
<th>$\Delta t = 3$</th>
<th>$\Delta t = 5$</th>
<th>$\Delta t = 10$</th>
<th>$\Delta t = 1$</th>
<th>$\Delta t = 3$</th>
<th>$\Delta t = 5$</th>
<th>$\Delta t = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>cs-Cox</td>
<td>0.150±0.04</td>
<td>0.309±0.08</td>
<td>0.354±0.08</td>
<td>0.433±0.07</td>
<td>0.016±0.00</td>
<td>0.055±0.04</td>
<td>0.133±0.09</td>
<td>0.133±0.09</td>
</tr>
<tr>
<td>RSF</td>
<td><strong>0.057±0.00</strong></td>
<td><strong>0.051±0.00</strong></td>
<td><strong>0.054±0.00</strong></td>
<td><strong>0.056±0.00</strong></td>
<td>0.015±0.00</td>
<td>0.015±0.00</td>
<td>0.016±0.00</td>
<td>0.016±0.00</td>
</tr>
<tr>
<td>JM</td>
<td>0.058±0.00</td>
<td>0.052±0.00</td>
<td>0.055±0.00</td>
<td>0.087±0.01</td>
<td>0.015±0.00</td>
<td>0.016±0.00</td>
<td>0.018±0.00</td>
<td>0.031±0.00</td>
</tr>
<tr>
<td>JM-LC</td>
<td>0.063±0.00</td>
<td>0.064±0.00</td>
<td>0.065±0.00</td>
<td>0.067±0.00</td>
<td><strong>0.015±0.00</strong></td>
<td><strong>0.016±0.00</strong></td>
<td>0.017±0.00</td>
<td><strong>0.016±0.00</strong></td>
</tr>
<tr>
<td>Proposed</td>
<td><strong>0.067±0.00</strong></td>
<td><strong>0.063±0.00</strong></td>
<td><strong>0.062±0.00</strong></td>
<td><strong>0.059±0.00</strong></td>
<td><strong>0.015±0.00</strong></td>
<td><strong>0.016±0.00</strong></td>
<td><strong>0.017±0.00</strong></td>
<td><strong>0.017±0.00</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>$\Delta t = 1$</th>
<th>$\Delta t = 3$</th>
<th>$\Delta t = 5$</th>
<th>$\Delta t = 10$</th>
<th>$\Delta t = 1$</th>
<th>$\Delta t = 3$</th>
<th>$\Delta t = 5$</th>
<th>$\Delta t = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>cs-Cox</td>
<td>0.442±0.24</td>
<td>0.616±0.28</td>
<td>0.658±0.30</td>
<td>0.658±0.30</td>
<td>0.315±0.21</td>
<td>0.428±0.23</td>
<td>0.428±0.23</td>
<td>0.737±0.20</td>
</tr>
<tr>
<td>RSF</td>
<td><strong>0.055±0.00</strong></td>
<td><strong>0.065±0.01</strong></td>
<td><strong>0.069±0.01</strong></td>
<td><strong>0.069±0.01</strong></td>
<td>0.018±0.00</td>
<td>0.021±0.00</td>
<td>0.021±0.00</td>
<td>0.021±0.00</td>
</tr>
<tr>
<td>JM</td>
<td>0.056±0.00</td>
<td><strong>0.057±0.00</strong></td>
<td><strong>0.066±0.01</strong></td>
<td><strong>0.111±0.01</strong></td>
<td>0.017±0.00</td>
<td>0.022±0.00</td>
<td>0.028±0.00</td>
<td>0.054±0.01</td>
</tr>
<tr>
<td>JM-LC</td>
<td>0.069±0.00</td>
<td>0.072±0.00</td>
<td>0.073±0.00</td>
<td>0.075±0.00</td>
<td>0.017±0.00</td>
<td>0.017±0.00</td>
<td>0.017±0.00</td>
<td><strong>0.016±0.00</strong></td>
</tr>
<tr>
<td>Proposed</td>
<td><strong>0.074±0.00</strong></td>
<td>0.071±0.00</td>
<td><strong>0.070±0.00</strong></td>
<td><strong>0.069±0.00</strong></td>
<td><strong>0.016±0.00</strong></td>
<td><strong>0.016±0.00</strong></td>
<td><strong>0.017±0.00</strong></td>
<td><strong>0.022±0.00</strong></td>
</tr>
</tbody>
</table>
PERFORMANCE WITH VARIOUS NUMBER OF TRAINING SAMPLES

• 10 Subsets

(a) Respiratory failure

(b) Other causes
Variable Importance

- A post-processing statistic that can be used by clinicians to interpret predictions issued by Dynamic-DeepHit
- imply the averaged increase/decrease of the risk predictions (by varying the covariate from its minimum to maximum) at $\Delta t = 5$.

$$\gamma_k(\Delta t, x'_d) = \mathbb{E}_{x'_{\Delta d}} \left[ \hat{F}_k(t_J + \Delta t | x'_d, x'_{\Delta d}) \right]$$

$$\approx \frac{1}{N} \sum_{i=1}^{N} \hat{F}_k(t^i_{J_i} + \Delta t | x'_d, x^i_{\Delta d})$$

$$\gamma_k(\Delta t, x'_d = x_{d,\text{min}}) - \gamma_k(\Delta t, x'_d = x_{d,\text{max}})$$

<table>
<thead>
<tr>
<th>Rank</th>
<th>Resp. Failure</th>
<th>Death Cause</th>
<th>Other Causes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>FEV$_1$ Predicted (-0.033)</td>
<td>IV ABX Days Hosp. (+0.014)</td>
<td>Gram-Negative (-0.013)</td>
</tr>
<tr>
<td>2</td>
<td>IV ABX Days Hosp. (+0.032)</td>
<td>Gram-Negative (-0.013)</td>
<td>FEV$_1$ Predicted (-0.012)</td>
</tr>
<tr>
<td>3</td>
<td>Gram-Negative (-0.029)</td>
<td>FEV$_1$ (-0.012)</td>
<td>Weight (-0.011)</td>
</tr>
<tr>
<td>4</td>
<td>FEV$_1$ (-0.026)</td>
<td>Weight (-0.011)</td>
<td>BMI (-0.010)</td>
</tr>
<tr>
<td>5</td>
<td>Weight (-0.026)</td>
<td>BMI (-0.010)</td>
<td>Oral Hypo. Agents (-0.008)</td>
</tr>
<tr>
<td>6</td>
<td>BMI (-0.025)</td>
<td>Oral Hypo. Agents (-0.008)</td>
<td>Class IV Mutation (-0.008)</td>
</tr>
<tr>
<td>7</td>
<td>Colonic Stricture (-0.024)</td>
<td>Class IV Mutation (-0.008)</td>
<td>IV ABX Days Hosp. (+0.007)</td>
</tr>
<tr>
<td>8</td>
<td>Oral Hypo. Agents (-0.019)</td>
<td>IV ABX Days Hosp. (+0.007)</td>
<td>Cancer (+0.007)</td>
</tr>
<tr>
<td>9</td>
<td>Class IV Mutation (-0.017)</td>
<td>Cancer (+0.007)</td>
<td>GI Bleed (var.) (+0.007)</td>
</tr>
<tr>
<td>10</td>
<td>B. Cepacia (+0.016)</td>
<td>GI Bleed (var.) (+0.007)</td>
<td>HypertonicSaline (-0.006)</td>
</tr>
<tr>
<td>11</td>
<td>GI Bleed (non-var.) (-0.016)</td>
<td>HypertonicSaline (-0.006)</td>
<td>Bone Fracture (-0.006)</td>
</tr>
<tr>
<td>12</td>
<td>O$_2$ Continuous (+0.015)</td>
<td>Bone Fracture (-0.006)</td>
<td>Colonic Stricture (-0.006)</td>
</tr>
<tr>
<td>13</td>
<td>Drug Dornase (-0.015)</td>
<td>Colonic Stricture (-0.006)</td>
<td>O$_2$ Nocturnal (+0.006)</td>
</tr>
<tr>
<td>14</td>
<td>IV ABX Days Home (+0.014)</td>
<td>O$_2$ Nocturnal (+0.006)</td>
<td>O$_2$ Nocturnal (+0.006)</td>
</tr>
<tr>
<td>15</td>
<td>O$_2$ Nocturnal (+0.013)</td>
<td>O$_2$ Nocturnal (+0.006)</td>
<td>O$_2$ Nocturnal (+0.006)</td>
</tr>
</tbody>
</table>

IV: intravenous, ABX: antibiotics
Illustration of dynamic prediction

Dynamic Prediction Illustration of a specific patient. At each time-stamp (gray solid line) of a measurement, the risk function is being adapted and the dashed green and red lines present different feature measurements. Yellow star denotes the time at which the respiratory failure occurred.
Model

Based on the Dynamic Deep-Hit competing risks model (Lee et al., 2020).

Subject $i$: $(x_{1i}, m_{1i}, \epsilon_{1i}) (x_{2i}, m_{2i}, \epsilon_{2i}) (x_{R_{i-1}i}, m_{R_{i-1}i}, \epsilon_{R_{i-1}i})) (x_{R_{i}i}, m_{R_{i}i})$

Quantile Normalization

RNN

Temporal Attention

$e_j = f_a(h_j, x_{R_{i}i}, m_{R_{i}i})$

$a_j = \exp(e_j) \sum_{l=1}^{R_{i}} \exp(e_l) c = \sum_{j=1}^{R_{i}} a_jh_j$

Total Loss = L3+L4

Feed – Forward
(Fully Connected)

Softmax

Single Event

Event $K_p$

$L1 – Log Likelihood Loss$

$L2 – Ranking Loss$

$L3 – Prediction Loss$

$L4 – Surv. Loss Function$
Preliminary Results – Breast Cancer

- Predictions cancer at $t_{Ri}^i + \Delta t$ given $X^i$, for $\Delta t = \{0.5, 1, 2\}$
- **Performance** evaluated using **AUROC** and **Concordance Index**
- **Benchmarks**: RDSM by Napgal et al. (2021) and **Dynamic Deep-Hit** by Lee et al. (2020)

<table>
<thead>
<tr>
<th>$\Delta t$</th>
<th>0.5 year</th>
<th>1 year</th>
<th>2 years</th>
<th>10 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk Factors</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>55.8</td>
</tr>
<tr>
<td>PRS</td>
<td>-</td>
<td>-</td>
<td>63.6</td>
<td>65.3</td>
</tr>
<tr>
<td>PRS + Risk Factors</td>
<td>-</td>
<td>-</td>
<td>65.3</td>
<td>65.3</td>
</tr>
</tbody>
</table>

- **BOADICEA**:

<table>
<thead>
<tr>
<th># Visits</th>
<th>Females</th>
<th># Visits</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9,410</td>
<td>6</td>
<td>727</td>
</tr>
<tr>
<td>2</td>
<td>4,633</td>
<td>7</td>
<td>531</td>
</tr>
<tr>
<td>3</td>
<td>2,446</td>
<td>8</td>
<td>403</td>
</tr>
<tr>
<td>4</td>
<td>1,513</td>
<td>9</td>
<td>320</td>
</tr>
<tr>
<td>5</td>
<td>1,004</td>
<td>$\geq10$</td>
<td>1,407</td>
</tr>
</tbody>
</table>

Thanks!