Toward a Well-Calibrated Discrimination via Survival Outcome-Aware Contrastive Learning

Dongjoon Lee[1*], Hyeryn Park[1*], Changhee Lee[2]

[1] Chung-Ang University [2] Korea University

Motivation & Challenges

Motivation

- ✓ Improving discriminative in survival models often compromises calibration.
- ✓ A new approach is need to enhance discrimination without sacrificing calibration, using the embedding space through contrastive learning.

Challenges

- ✓ Combining NLL with ranking loss improves discrimination but misaligns model outputs with the actual risk distribution, negatively affecting calibration and clinical applicability.

Preliminaries

Discrete-Time Survival Analysis

- ✓ Survival function S represents the probability that the event occurs after time t for a patient with features $\mathbf{x}$.
- ✓ Hazard function λ is the instantaneous risk of the event at time t given feature $\mathbf{x}$

$$S(t|\mathbf{x}) = \mathbb{P}(T > t|\mathbf{x}) = \prod_{t' \le t} (1 - \lambda(t'|\mathbf{x}))$$

$$\mathcal{L}_{NLL} = -\sum_{i=1}^{N} [\delta_i \log \hat{p}(\tau_i|\mathbf{x}_i) + (1-\delta_i) \log \hat{S}(\tau_i|\mathbf{x}_i)]$$

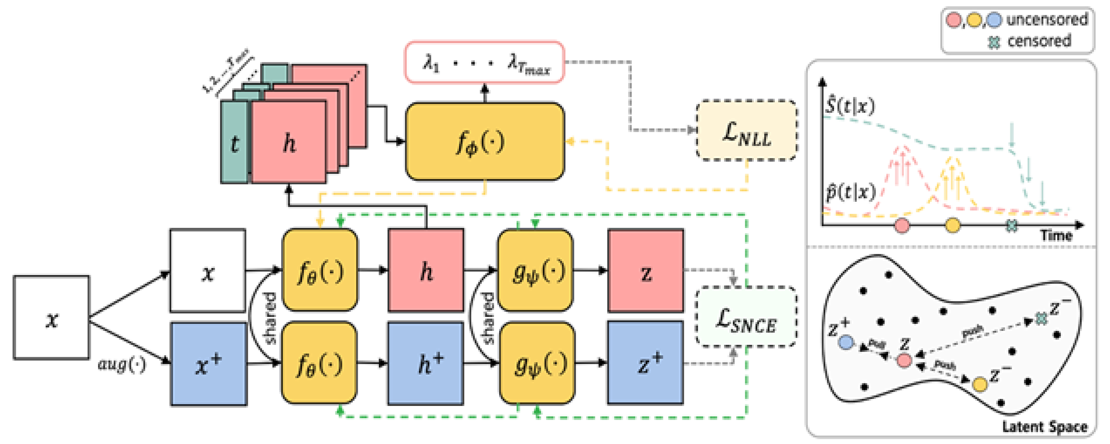
where $\hat{p}$ represents the estimate for the probability of an event occurring at time t

Ranking Loss for Survival Analysis

Aim to maximize a relaxed proxy of the concordance index.

$$\mathcal{L}_{Rank} = \sum_{i \ne j} A_{i,j} \cdot \eta\left(\hat{R}(\tau_i|\mathbf{x}_i), \hat{R}(\tau_i|\mathbf{x}_j)\right)$$

Network Description



- ✓ The encoder, $f_\theta : \mathcal{X} \to \mathcal{H}$, takes features $x \in \mathcal{X}$ as input and outputs latent representation, i.e., $h = f_\theta(x)$.
- ✓ The projection head, $f_\psi : \mathcal{H} \to \mathbb{R}^d$, maps latent representation h to the embedding space where contrastive learning is applied, i.e., $z = f_\theta(h)$.
- ✓ The hazard network, $f_\phi : \mathcal{H} \times \mathcal{T} \to [0,1]$, predicts the hazard rate at each time point $t \in \mathcal{T}$ given input latent representation $\mathbf{h}$, i.e., $\hat{\lambda}(t|\mathbf{x}) = f_\phi(f_\theta(\mathbf{x}), t)$.

Method

- ✓ Propose a novel contrastive learning approach for deep survival model
- ✓ Deviate from directly ranking the model outcome in the form of risk/survival functions
- ✓ Goal : Aligns with our inductive bias that patients with similar survival outcomes should share similar clinical status, which manifests through similar representations

Noise Contrastive Estimation (NCE)

To learn mapping $f = g_\psi \circ f_\theta$ utilizing a positive sample $\mathbf{x}^+ \sim p_{X^+}$,and negative samples $\mathbf{x}^- \sim q$

$$\mathbb{E}_{\substack{\mathbf{x} \sim p_X \\ \mathbf{x}^+ \sim p_{X^+}}} \left[-\log \frac{e^{s(\mathbf{x}, \mathbf{x}^+)}}{M \cdot \mathbb{E}_{\mathbf{x}^- \sim q}[e^{s(\mathbf{x}, \mathbf{x}^-)}]} \right]$$

Weighted Distribution q for Time-to-Event Differences

To reflect the difference in the time-to-events in the embedding space, we design a novel distribution q by utilizing the available information from survival outcomes.

$$w(\tau^-; \tau) = 1 - e^{|\tau - \tau^-|/\sigma}$$

- ✓ Hence, given an anchor $(\mathbf{x}, \tau)$ and a negative $(\mathbf{x}^-; \tau^-)$, we define the weight function, $\sigma > 0$ is a temperature coefficient.
- ✓ This function assigns larger weights to samples with large differences in time-to-event outcomes, and smaller weights to samples with small differences.

$$q(\mathbf{x}^-; \mathbf{x}) = \frac{1}{Z} w(\mathbf{x}^-; \mathbf{x}) p(\mathbf{x}^-)$$

- ✓ Designing q based on the following inductive bias similar patients are more likely to experience the event at similar time points than the ones who are not.

Importance Sampling Using Survival Outcomes

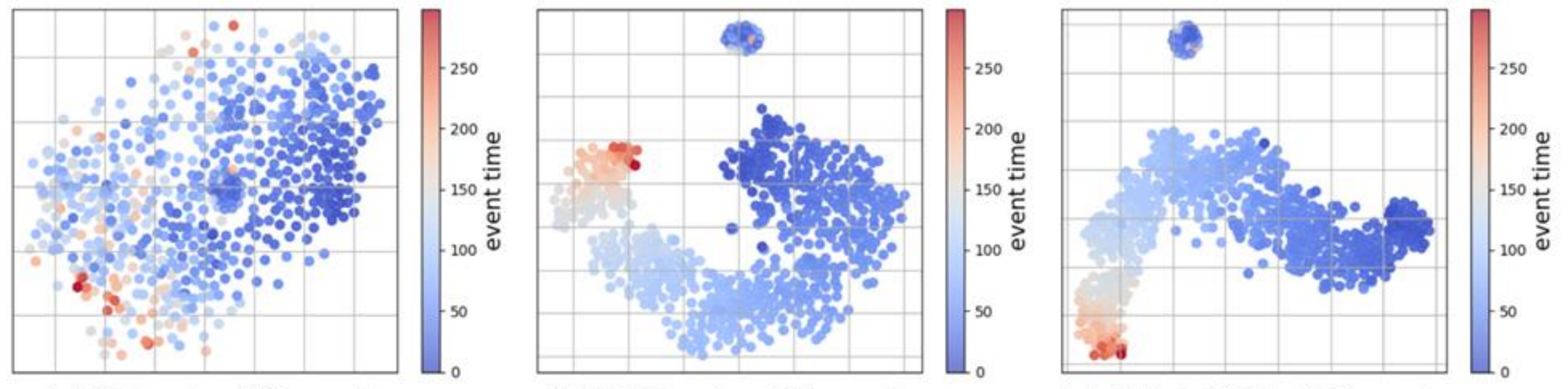
$$E_{x^- \sim q}\left[e^{s(x,x^-)}\right] = E_{x^- \sim p}\left[\left(\frac{q(x^-; x)}{p(x^-)}\right) \cdot e^{s(x,x^-)}\right]$$

$$= E_{x^- \sim p}\left[\left(\frac{w(x^-; x)}{Z}\right) \cdot e^{s(x,x^-)}\right]$$

$$\approx \frac{1}{Z \cdot M} \sum_{j=1}^{M} w(x_j^-; x) \cdot e^{s(x, x_j^-)}$$

Experiments

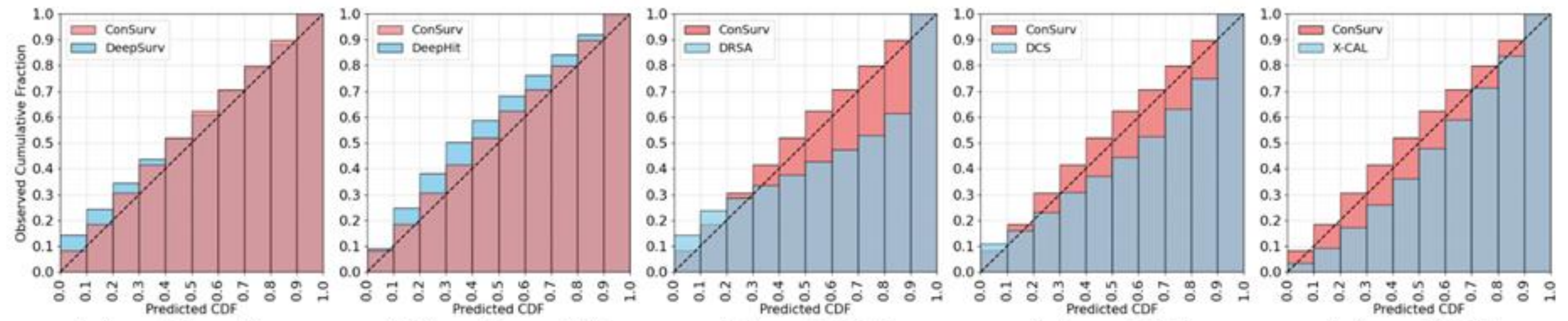
Effect of Contrastive Learning



(a) NLL only 400 epochs (b) SNCE only 400 epochs (c) NLL + SNCE 400 epochs

Figure 2: t-SNE visualization for latent representations learned with $\mathcal{L}_{NLL}$ only, $\mathcal{L}_{SNCE}$ only, and ConSurv for the METABRIC dataset, colored by event times (for uncensored samples).

Calibration Plot



(a) vs DeepSurv (b) vs DeepHit (c) vs DRSA (d) vs DCS (e) vs X-CAL

Figure 3: Calibration plots for ConSurv in comparison with benchmarks for the METABRIC dataset.

Subgroup Analysis

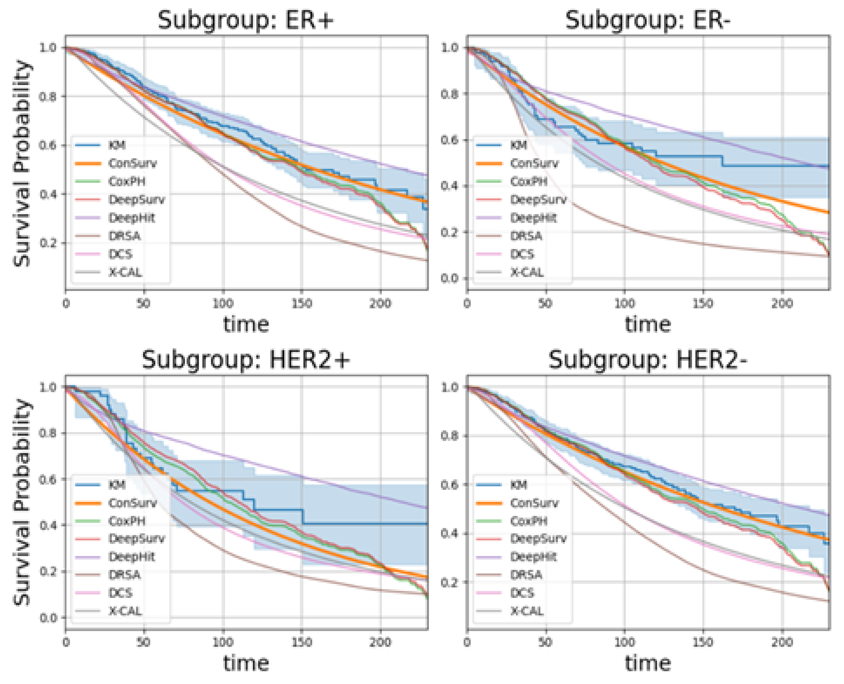


Figure 4: Comparison of the survival curves across various patient subgroups for the METABRIC dataset.

Quantitative Analysis

Method	CI ↑ (METABRIC)	IBS ↓ (METABRIC)	DDC ↓ (METABRIC)	D-CAL (METABRIC)
CoxPH	$0.645_{\pm 0.019}$	$\mathbf{0.175_{\pm 0.028}}$	$0.111_{\pm 0.024}$	**25**
DeepSurv	$0.625_{\pm 0.025}$	$0.183_{\pm 0.029}$	$0.103_{\pm 0.026}$	**25**
DeepHit	$0.604_{\pm 0.019}$	$0.204_{\pm 0.018}$	$0.292_{\pm 0.017}$	0
DRSA	$0.604_{\pm 0.032}$	$0.249_{\pm 0.038}$	$0.178_{\pm 0.060}$	0
DCS	$0.612_{\pm 0.029}$	$0.206_{\pm 0.043}$	$\mathbf{0.054_{\pm 0.039}}$	2
X-CAL	$0.632_{\pm 0.027}$	$0.182_{\pm 0.023}$	$0.065_{\pm 0.037}$	2
$\mathcal{L}_{NLL}$	$0.642_{\pm 0.022}$	$0.197_{\pm 0.030}$	$0.077_{\pm 0.020}$	13
$\mathcal{L}_{NLL}$ & $\mathcal{L}_{NCE}$	$0.659_{\pm 0.020}$	$0.193_{\pm 0.029}$	$0.080_{\pm 0.022}$	21
$\mathcal{L}_{NLL}$ & $\mathcal{L}_{Rank}$	$0.652_{\pm 0.022}$	$0.247_{\pm 0.030}$	$0.177_{\pm 0.020}$	0
ConSurv	$\mathbf{0.665_{\pm 0.023}}$	$0.186_{\pm 0.021}$	$0.110_{\pm 0.024}$	23

Method	CI ↑ (GBSG)	IBS ↓ (GBSG)	DDC ↓ (GBSG)	D-CAL (GBSG)
CoxPH	$0.662_{\pm 0.179}$	$0.181_{\pm 0.007}$	$0.183_{\pm 0.037}$	**25**
DeepSurv	$0.653_{\pm 0.042}$	$0.182_{\pm 0.009}$	$0.153_{\pm 0.066}$	24
DeepHit	$0.633_{\pm 0.032}$	$0.205_{\pm 0.006}$	$0.342_{\pm 0.023}$	3
DRSA	$0.668_{\pm 0.016}$	$0.278_{\pm 0.018}$	$0.402_{\pm 0.055}$	0
DCS	$\mathbf{0.677_{\pm 0.017}}$	$0.181_{\pm 0.008}$	$\mathbf{0.124_{\pm 0.025}}$	10
X-CAL	$0.675_{\pm 0.017}$	$0.181_{\pm 0.010}$	$0.166_{\pm 0.020}$	8
$\mathcal{L}_{NLL}$	$0.668_{\pm 0.019}$	$0.179_{\pm 0.006}$	$0.154_{\pm 0.029}$	0
$\mathcal{L}_{NLL}$ & $\mathcal{L}_{NCE}$	$0.669_{\pm 0.020}$	$0.179_{\pm 0.007}$	$0.155_{\pm 0.026}$	0
$\mathcal{L}_{NLL}$ & $\mathcal{L}_{Rank}$	$0.687_{\pm 0.019}$	$0.280_{\pm 0.007}$	$0.263_{\pm 0.025}$	0
ConSurv	$\mathbf{0.677_{\pm 0.020}}$	$\mathbf{0.179_{\pm 0.007}}$	$0.160_{\pm 0.026}$	18

- ✓ Significantly improves the alignment of representations with event time information.
- ✓ Compares survival plots of the models with the Kaplan-Meier curve to confirm calibration.
- ✓ Consurv outperforms all benchmarks in discrimination and maintains comparable calibration.