LongSurv: Bridging Short-Term Data and Causal Priors for Longitudinal Survival Modelling

Shivaram
Indian Institute of Technology Madras
Velai Health Analytics Pvt Ltd
shivaramvelayutham3223@gmail.com

Anuva Kapoor
All India Institute of Medical Sciences Nagpur
Velai Health Analytics Pvt Ltd
anuvakapoor19@gmail.com

Kangan Maria
All India Institute of Medical Sciences Rishikesh
mariakangan@gmail.com

Preetha Balasubramanian
Revvity Inc
preet3384@gmail.com

Bhanu Duggal
All India Institute of Medical Sciences Rishikesh
bhanuduggal2@gmail.com

Mona Duggal
Indian Council of Medical Research
monaduggal2@gmail.com

Balaraman Ravindran
Indian Institute of Technology Madras
ravi@dsai.iitm.ac.in

Lakshminarayanan Subramanian
New York University
Velai Health LLC
lakshmi@nyu.edu

Abstract

Accurate long-term risk estimation is critical for managing chronic diseases. Survival analysis provides a framework to quantify the causal effect of risk factors on time-to-event outcomes, but short follow-up cohorts common in clinical studies lead to heavy censoring and limit the estimation of long-term effects. Existing extrapolation methods often focus on population-level outcomes and rely on loosely defined external data, such as expert heuristics, limiting their utility for personalised risk estimation. We propose **LongSurv**, a framework that extrapolates individual-level survival trajectories from short-term EHR data by integrating epidemiological priors like hazard ratios and relative risks. These priors are incorporated in training via two loss functions: *life expectancy consistency loss* that aligns predictions with demographic expectations, and a *group-wise ranking loss* to preserve clinically valid risk orderings. Evaluated on 4595 post-PCI patients from Maharashtra, India (95% censoring), outperforms a Weibull baseline in discrimination (C-index 0.6946 vs. 0.5227) and calibration (IBS 0.0390 vs. 0.0494), while enabling "*what-if*" reasoning for personalised care. By connecting short-term observational data with long-term causal survival insights, LongSurv provides an interpretable and scalable approach to risk estimation.

1 Introduction

Continuity of care for patients with chronic diseases, such as cardiovascular diseases (CVDs) and chronic kidney diseases (CKDs), is crucial for improving long-term health outcomes, optimising resource utilisation, and reducing medical expenses through preventive and optimised care. Effective continuity of care depends on accurate long-term risk estimation, which in turn requires understanding

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: .

the influence of risk factors and treatments on patient survival trajectories. Survival analysis provides a framework for estimating the impact of risk factors on time-to-event outcomes and quantifying this impact using survival quantities, such as life expectancy. However, most healthcare cohort studies are limited to short follow-up periods (typically 1-5 years) [1], causing a phenomenon known as *censoring* [2]. *Censoring* presents a significant challenge in estimating long-term effects. These limitations are particularly acute in the Indian healthcare context, where longitudinal cohorts are scarce, making reliable extrapolation essential.

Traditional survival models such as *Kaplan-Meier* [3] and *Cox-proportional hazards model* [4] are inadequate for long-term extrapolation due to their lack of a baseline hazard function. While recent deep learning approaches ([5, 6, 7]) can capture complex dependencies, they remain primarily predictive, requiring large uncensored datasets and their non-parametric nature [8]. Parametric survival models, by contrast, assume an underlying distribution of hazard function such as Weibull, Log-Normal, or Gompertz, which are typically recommended for robust long-term projections ([9, 10]). Although most of the existing literature [11, 12, 13] emphasises population-level extrapolation, typically comparing survival between a treatment arm and a control arm (such as in cost-effectiveness analysis), there is limited focus on individual-level extrapolation, as personalised long-term risk assessment depends on the ability to predict survival trajectories at the individual level.

In this work, we present **LongSurv**, a framework for learning parametrised survival models from short-term, heavily censored cohorts by integrating demographic statistics and clinical evidence to ensure population-level compliance and causal consistency. Applied to a real-world cohort of 4,595 post-PCI patients from Maharashtra, India (95% censoring, 1-year follow-up), LongSurv improves predictive accuracy and enables plausible long-term extrapolation.it generates causally informed *what-if* risk projections, where the model simulates longitudinal survival under hypothetical modifications of one or more risk factors while respecting the causal ordering implied by external priors.

2 Methodology

The key challenge in modelling survival from short-term, highly censored cohorts is extrapolating far beyond the observed horizon without violating known clinical or demographic patterns. Conventional models trained on limited follow-up often overfit short-term trends, yielding implausible long-term estimates. **LongSurv** addresses this by embedding external population statistics and causal priors into the learning process, using them to regularise and guide extrapolation rather than to perform causal identification, thereby steering predictions toward clinically coherent behaviour. The framework is model-agnostic and can be paired with any base survival model (here, a Weibull AFT model), enabling consistent integration of causal and population-level constraints across formulations.

2.1 Life Expectancy as an Impact Metric

Life expectancy, defined as the area under the survival curve, provides a natural and interpretable summary statistic for long-term survival. At the demographic level, reliable life table estimates are available across regions and populations. We impose a *life expectancy loss* ($\mathcal{L}_{LE}$) that aligns extrapolated survival predictions with these reference values, thereby anchoring the model to realistic long-term outcomes. This provides a bridge between short-term cohort observations and long-term population-level survival trends.

2.2 Feature-Level Effects on Survival

While demographic-level life expectancy offers a broad population anchor, the absolute contribution of individual risk factors (e.g., smoking, diabetes, hypertension) is rarely quantified directly due to methodological challenges in isolating their effects on survival. Epidemiological studies instead report relative risks (RR) or hazard ratios (HR), reflecting adjusted causal associations with mortality (Table 3). Let HR_j denote the prior hazard ratio for covariate x_j, derived from studies that control for key confounders, including age. These priors are incorporated as monotonic or ranking constraints in the loss function, encouraging predictions consistent with established causal directions (e.g., higher smoking exposure implies higher hazard) [14].

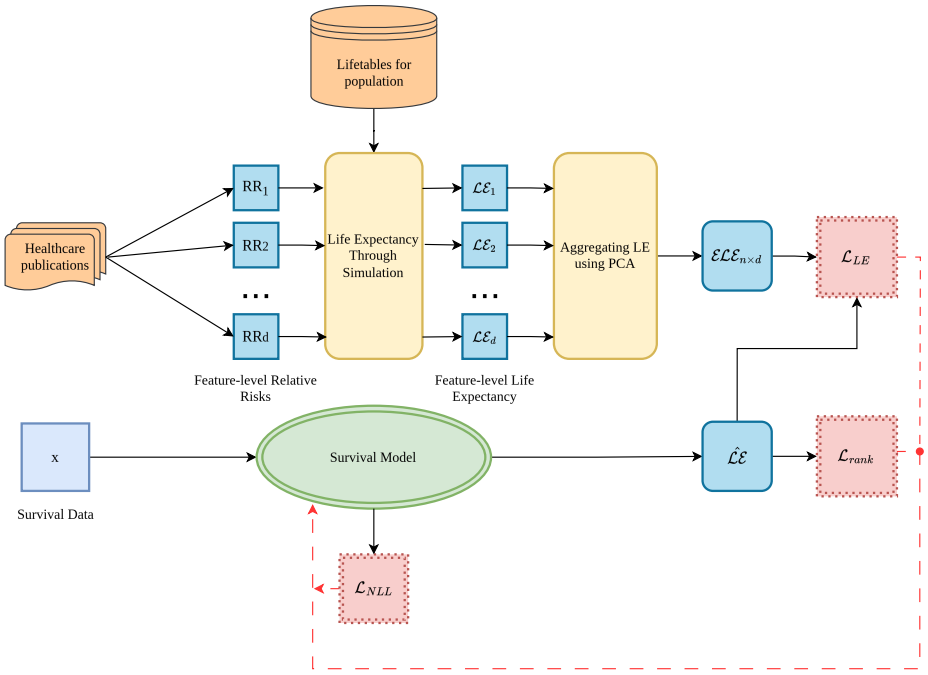

Figure 1: **Integrating Simulated Life Expectancy into Survival Modeling.**
The diagram shows how long-term survival estimates are derived from short-term cohort data. Feature-level relative risks (RR_i) from healthcare studies and lifetables are used to simulate life expectancy ($\mathcal{LE}_i$), which is aggregated via PCA into $\mathcal{ELE}_{n\times d}$. A survival model, trained on covariates (x), incorporates this aggregated life expectancy through an auxiliary loss term ($\mathcal{L}_{\mathcal{LE}}$) alongside the negative log-likelihood ($\mathcal{L}_{NLL}$) to enhance extrapolation beyond observed data.

We translate these priors into feature-specific adjusted life expectancies using life table methods on a synthetic population. Baseline age-specific mortality rates are modified by reported RRs or HRs. Since these estimates are age-adjusted, we assume their effects remain approximately age-insensitive across the follow-up horizon. To align with the target cohort, a *prevalence scaling* step calibrates effect sizes to observed feature prevalence, and principal component analysis (PCA) is applied to disentangle correlated risk factors and prevent double-counting. These calibrated priors yield population-consistent, clinically interpretable life expectancy estimates that guide long-term extrapolation.

2.3 Learning Objective

LongSurv augments a parametric survival model (here, a Weibull AFT model [15]) with two regularisation terms: a *life expectancy loss* that aligns predicted life expectancies $\hat{\mu}_i$ with demographic targets $\mathcal{LE}_i$, and a *feature-wise ranking loss* that enforces monotonic effects of known risk/protective factors reflecting causal relationships obtained from medical studies. The complete objective is

$$\mathcal{L}_{LE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{\mu}_i - \mathcal{LE}_i)^2 \qquad \mathcal{L}_{\text{rank}} = \sum_{j\in\mathcal{J}} \sigma\left(\frac{d_j(\mu_0^j - \mu_1^j)}{\sqrt{\sigma_1^2+\sigma_0^2}}\right) \qquad \mathcal{L} = \mathcal{L}_{NLL} + w\cdot\mathcal{L}_{LE} + v\cdot\mathcal{L}_{\text{rank}}$$

Here, $d_j \in \{+1, -1\}$ encodes whether feature j is risk-increasing or protective, and μ_0^j, μ_1^j denote group-wise predicted life expectancies. Where $\mathcal{L}_{NLL}$ is the negative log-likelihood of censored data, and w, v control the trade-off between fitting observed data and enforcing external causal structure.

3 Result

Evaluating survival extrapolations beyond the cohort period is challenging, as it involves validating the survival projections outside the available data. We validate LongSurv at 2 levels: model fitness

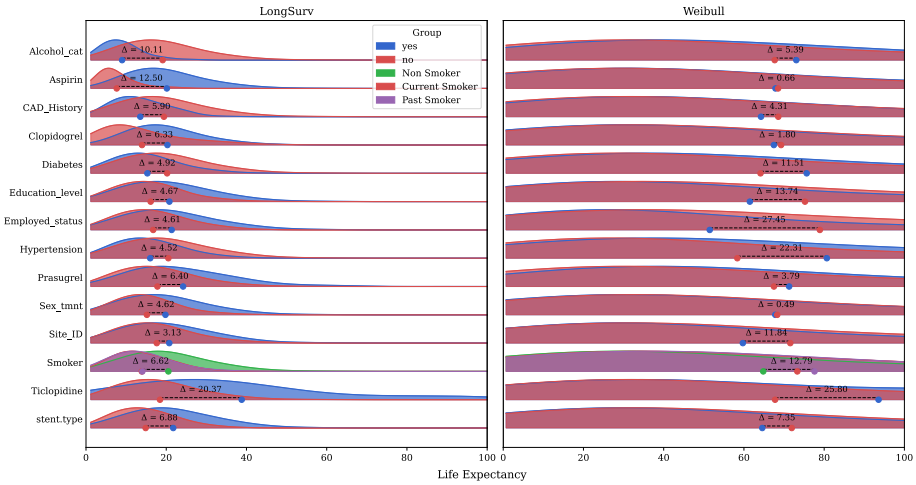

Figure 2: Feature-wise distributions of predicted life expectancy and the *mean* difference between groups under **LongSurv** (left) and a Weibull model (right). LongSurv captures clear and clinically consistent differences across categorical subgroups (e.g., smoking status, comorbidities, medications), while the Weibull model fails to separate most group-level effects. For features representing disease conditions (e.g., hypertension, diabetes, CAD history), *yes* indicates presence of the condition and *no* its absence. For medication features (e.g., aspirin, prasugrel, ticlopidine, clopidogrel), *yes* denotes patients on the medication and *no* those not receiving it. For demographic and lifestyle attributes: *yes* in sex indicates female (no = male); *yes* in education corresponds to higher education (no = lower education); *yes* in employment denotes currently employed (no = unemployed); *yes* in alcohol indicates current consumption (no = abstinent). For stent type, *yes* corresponds to drug-eluting stent (DES) and *no* to bare-metal stent (BMS). Smoking status is represented by multi-class color labels (non-smoker, current smoker, past smoker) for clarity.

within the cohort period and validity of survival extrapolation beyond the cohort period. As LongSurv is a loss augmentation framework applied over a baseline parametric survival model, we evaluate and compare it against the parametric model of choice (Weibull) in this study.

3.1 Model Fitness within Cohort Period

We evaluated survival models on discrimination (*C-index*, [16]) and calibration (*Integrated Brier Score (IBS)*, [17]) using an 80/20 train-test split (148 and 46 events, respectively). As Table 1 shows, LongSurv substantially improves discrimination (0.6946 vs. 0.5227) and modestly improves calibration (0.0390 vs. 0.0494), indicating more accurate survival predictions within the cohort period.

Method	**C-Index**	**IBS**
Weibull	0.5227	0.0494
LongSurv(Weibull)	0.6946	0.0390

Table 1: Discrimination (C-index) and calibration (IBS) within the cohort period.

3.2 Evaluating Clinical Plausibility by Sub-Group based Analysis

To assess clinical plausibility, we stratified patients by categorical features (e.g., smoking, alcohol, comorbidities, medications) and compared their extrapolated life expectancies (Figure 2). LongSurv produced clinically coherent trends, non-smokers and non-drinkers outliving counterparts, comorbidi-

ties lowering survival, and medications improving outcomes, while the baseline Weibull model failed to separate most subgroups.

3.3 Action Plans for Improving Patient Longevity

The proposed LongSurv framework enables dynamic simulation of personalised *"what-if"* scenarios by modifying actionable covariates and quantifying the resulting gain in life expectancy (ΔLE), while maintaining the causal ordering implied by the external priors. Actionable factors include smoking, alcohol, hypertension control, and Dual Antiplatelet Therapy (DAPT) regimens.

For a 46-year-old male patient (smoker, alcohol consumer, hypertensive, on aspirin and clopidogrel), Table 2 summarises predicted life expectancy under counterfactual interventions. Alcohol cessation and switching to Ticlopidine-based DAPT yield the largest gains ($\sim$4.8–4.9 years), while smoking cessation and hypertension control provide smaller improvements. This allows the construction of a ranked, clinically relevant action plan for longevity.

Intervention	**LE (years)**	**Δ LE (years)**
Baseline Patient	6.54	0.00
Quit Smoking	6.58	0.04
Quit Alcohol	11.37	4.83
Controlled Hypertension	7.79	1.25
DAPT (Aspirin+Clopidogrel)	6.54	0.00
DAPT (Aspirin+Prasugrel)	10.22	3.68
DAPT (Aspirin+Ticlopidine)	11.48	4.93

Table 2: Predicted life expectancy and estimated improvements under causally informed *what-if* intervention simulations.

4 Discussion and Limitations

LongSurv leverages causal knowledge as soft constraints rather than formal causal identification. Hazard-ratio priors derived from age-adjusted epidemiological studies encode approximate causal relationships, serving as weak supervision to enforce directionally consistent links between risk factors and survival outcomes. This enables causally informed "what-if" projections that remain *simulations*, not true counterfactuals, since not all confounders are modelled.

Key limitations include the absence of longitudinal data to empirically validate extrapolations, reliance on plausibility checks against external evidence, and dependence on a proxy target (**Expected Life Expectancy**) that may reflect sampling or contextual biases. Despite age adjustment, residual age effects may persist for certain risk factors (e.g., hypertension, diabetes). Future work will incorporate formal sensitivity and causal adjustment analyses to mitigate these uncertainties.

5 Conclusion

In summary, LongSurv integrates causal priors with short-term data to enable robust, interpretable survival extrapolation under heavy censoring. By treating these priors as causally informed constraints, it supports meaningful "what-if" analyses and offers a practical path toward causally grounded, evidence-aware survival modelling in healthcare.

References

[1] SL Rohde, M Matebele, P Pohlner, D Radford, D Wall, and JF Fraser. Excellent cardiac surgical outcomes in paediatric indigenous patients, but follow-up difficulties. *Heart, Lung and Circulation*, 19(9):517–522, 2010.

[2] Anthony Joe Turkson, Francis Ayiah-Mensah, and Vivian Nimoh. Handling censoring and censored data in survival analysis: a standalone systematic literature review. *International journal of mathematics and mathematical sciences*, 2021(1):9307475, 2021.

[3] Edward L Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.

[4] David R Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202, 1972.

[5] Changhee Lee, William Zame, Jinsung Yoon, and Mihaela Van Der Schaar. Deephit: A deep learning approach to survival analysis with competing risks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

[6] Jared L Katzman, Uri Shaham, Alexander Cloninger, Jonathan Bates, Tingting Jiang, and Yuval Kluger. Deepsurv: personalized treatment recommender system using a cox proportional hazards deep neural network. *BMC medical research methodology*, 18:1–12, 2018.

[7] Rajesh Ranganath, Adler Perotte, Noémie Elhadad, and David Blei. Deep survival analysis. In Finale Doshi-Velez, Jim Fackler, David Kale, Byron Wallace, and Jenna Wiens, editors, *Proceedings of the 1st Machine Learning for Healthcare Conference*, volume 56 of *Proceedings of Machine Learning Research*, pages 101–114, Northeastern University, Boston, MA, USA, 18–19 Aug 2016. PMLR.

[8] Xenia Miscouridou, Adler Perotte, Noémie Elhadad, and Rajesh Ranganath. Deep survival analysis: Nonparametrics and missingness. In *Machine Learning for Healthcare Conference*, pages 244–256. PMLR, 2018.

[9] Nicholas R. Latimer. Survival analysis for economic evaluations alongside clinical trials—extrapolation with patient-level data: Inconsistencies, limitations, and a practical guide. *Medical Decision Making*, 33(6):743–754, 2013.

[10] Claire Davies, Andrew Briggs, Paula Lorgelly, Göran Garellick, and Henrik Malchau. The "hazards" of extrapolating survival curves. *Medical Decision Making*, 33(3):369–380, 2013.

[11] Zhaojing Che, Nathan Green, and Gianluca Baio. Blended survival curves: a new approach to extrapolation for time-to-event outcomes from clinical trials in health technology assessment. *Medical Decision Making*, 43(3):299–310, 2023.

[12] Patricia Guyot, Anthony E Ades, Matthew Beasley, Béranger Lueza, Jean-Pierre Pignon, and Nicky J Welton. Extrapolation of survival curves from cancer trials using external information. *Medical Decision Making*, 37(4):353–366, 2017.

[13] A. Vickers. An evaluation of survival curve extrapolation techniques using long-term observational cancer data. *Medical Decision Making*, 39(8):926–938, 2019.

[14] Mohammad Taha Bahadori, Krzysztof Chalupka, Edward Choi, Robert Chen, Walter F Stewart, and Jimeng Sun. Causal regularization. *arXiv preprint arXiv:1702.02604*, 2017.

[15] Enwu Liu, Ryan Yan Liu, and Karen Lim. Using the weibull accelerated failure time regression model to predict time to health events. *Applied Sciences*, 13(24):13041, 2023.

[16] Frank E Harrell, Robert M Califf, David B Pryor, Kerry L Lee, and Robert A Rosati. Evaluating the yield of medical tests. *Jama*, 247(18):2543–2546, 1982.

[17] Erika Graf, Claudia Schmoor, Willi Sauerbrei, and Martin Schumacher. Assessment and comparison of prognostic classification schemes for survival data. *Statistics in medicine*, 18(17-18):2529–2545, 1999.

[18] Bhanu Duggal, Jyothi Subramanian, Mona Duggal, Pushpendra Singh, Meeta Rajivlochan, Sujata Saunik, Koundinya Desiraju, Archana Avhad, Usha Ram, Sayan Sen, et al. Survival outcomes post percutaneous coronary intervention: Why the hype about stent type? lessons from a healthcare system in india. *PloS one*, 13(5):e0196830, 2018.

[19] Joseph T Lariscy, Robert A Hummer, and Richard G Rogers. Cigarette smoking and all-cause and cause-specific adult mortality in the united states. *Demography*, 55:1855–1885, 2018.

[20] Mangesh S Pednekar, Genevieve Sansone, and Prakash C Gupta. Association of alcohol, alcohol and tobacco with mortality: findings from a prospective cohort study in mumbai (bombay), india. *Alcohol*, 46(2):139–146, 2012.

[21] Jae Jeong Yang, Danxia Yu, Wanqing Wen, Eiko Saito, Shafiur Rahman, Xiao-Ou Shu, Yu Chen, Prakash C Gupta, Dongfeng Gu, Shoichiro Tsugane, et al. Association of diabetes with all-cause and cause-specific mortality in asia: a pooled analysis of more than 1 million participants. *JAMA network open*, 2(4):e192696–e192696, 2019.

[22] Dagfinn Aune, Wentao Huang, Jing Nie, and Yafeng Wang. Hypertension and the risk of all-cause and cause-specific mortality: an outcome-wide association study of 67 causes of death in the national health interview survey. *BioMed Research International*, 2021(1):9376134, 2021.

[23] Robert J Mentz, Bradley D Allen, Mary J Kwasny, Marvin A Konstam, James E Udelson, Andrew P Ambrosy, Angela J Fought, Muthiah Vaduganathan, Christopher M O'Connor, Faiez Zannad, et al. Influence of documented history of coronary artery disease on outcomes in patients admitted for worsening heart failure with reduced ejection fraction in the everest trial. *European journal of heart failure*, 15(1):61–68, 2013.

[24] Hui Hu, Wen-jun Chen, Zi-yi Xiong, Lin-fei Luo, Chuang Sun, and Jun-ping Xie. Association of prophylactic low-dose aspirin use with all-cause and cause-specific mortality in cancer patients. *Scientific Reports*, 14(1):25918, 2024.

[25] Charlotte Andersson, Stig Lyngbæk, Cu Dinh Nguyen, Mia Nielsen, Gunnar H Gislason, Lars Køber, and Christian Torp-Pedersen. Association of clopidogrel treatment with risk of mortality and cardiovascular events following myocardial infarction in patients with and without diabetes. *JAMA*, 308(9):882–889, 2012.

[26] Michael Gent, J Donald Easton, VladimirC Hachinski, Edouard Panak, Jane Sicurella, JohnA Blakely, DavidJ Ellis, JohnW Harbison, RobinS Roberts, AlexanderG G Turpie, et al. The canadian american ticlopidine study (cats) in thromboembolic stroke. *The Lancet*, 333(8649):1215–1220, 1989.

[27] Matthew T Roe, Paul W Armstrong, Keith AA Fox, Harvey D White, Dorairaj Prabhakaran, Shaun G Goodman, Jan H Cornel, Deepak L Bhatt, Peter Clemmensen, Felipe Martinez, et al. Prasugrel versus clopidogrel for acute coronary syndromes without revascularization. *New England journal of medicine*, 367(14):1297–1309, 2012.

[28] Shuqi Zhang, Mithlesh Chourase, Nupur Sharma, Sujata Saunik, Mona Duggal, Goodarz Danaei, and Bhanu Duggal. The effects of dual antiplatelet therapy (dapt) adherence on survival in patients undergoing revascularization and the determinants of dapt adherence. *BMC Cardiovascular Disorders*, 22(1):238, 2022.

[29] Hironori Hara, Masafumi Ono, Hideyuki Kawashima, Norihiro Kogame, Michael J Mack, David R Holmes, Marie-Claude Morice, Piroze M Davierwala, Friedrich W Mohr, Daniel JFM Thuijs, et al. Impact of stent length and diameter on 10-year mortality in the syntaxes trial. *Catheterization and Cardiovascular Interventions*, 98(3):E379–E387, 2021.

[30] Duk-Woo Park, Sung-Cheol Yun, Seung-Whan Lee, Young-Hak Kim, Cheol Whan Lee, Myeong-Ki Hong, Sang-Sig Cheong, Jae-Joong Kim, Seong-Wook Park, and Seung-Jung Park. Stent thrombosis, clinical events, and influence of prolonged clopidogrel use after placement of drug-eluting stent: data from an observational cohort study of drug-eluting versus bare-metal stents. *JACC: Cardiovascular Interventions*, 1(5):494–503, 2008.

A Appendix

A.1 Preliminaires: Survival Analysis

Consider a time-to-event dataset with N patients, denoted as $\mathcal{D} := \{(x^i, \delta^i, \tau^i)\}_{i=1}^N$. Each patient i is defined by a set of D input covariates $x^i \in \mathcal{X}$ where $\mathcal{X}$ is the input space. $\tau^i \in \mathcal{T}$ indicates the time until the occurrence of the event of interest (death) or the time of the last recorded follow-up, in cases where the event has not been observed (right censoring). Censored time need not always be equal to the cohort duration, $\mathcal{T}_C$. The event indicator, $\delta^i \in \{0, 1\}$, takes a value of 1 if the event is observed and 0 if the datapoint is right censored. As we are handling the problem of extrapolation of survival curves, we assume the time horizon to be infinite, $\mathcal{T} = \{1, ..., \mathcal{T}_C, ...\}$. Given this notion of the key variables, we define the elementary quantities: survival function, hazard function and negative log-likelihood using the probability density function $f_{\mathcal{T}|\mathcal{X}}$ and cumulative density function $F_{\mathcal{T}|\mathcal{X}}$.

The survival function indicates the probability that the event of interest has not occurred yet by time t. Formally, if $\mathcal{T}$ denotes the time until death, the survival function $S_{\mathcal{T}|\mathcal{X}}(t|x) : \mathcal{X} \times \mathcal{T} \to [0, 1]$ is defined as:

$$S_{\mathcal{T}|\mathcal{X}}(t|x) \stackrel{\text{def}}{=} 1 - F_{\mathcal{T}|\mathcal{X}}(t|x) = P(\mathcal{T} > t|x)$$

The conditional hazard function $h_{\mathcal{T}|\mathcal{X}}(t|x) : \mathcal{X} \times \mathcal{T} \to [0, 1]$ is defined as the probability that the event will occur with an infinitesimal window, given that the event has not occurred before time t.

$$h_{\mathcal{T}|\mathcal{X}}(t|x) \stackrel{\text{def}}{=} \lim_{h \to 0} P(\mathcal{T} \in [t, t + h] \big| \mathcal{T} \geq t, \mathcal{X}) = \frac{f_{\mathcal{T}|\mathcal{X}}(t|x)}{S_{\mathcal{T}|\mathcal{X}}(t|x)}$$

Thus, using these quantities, we achieve maximum likelihood estimates for the hazard function $\hat{h_{\mathcal{T}|\mathcal{X}}}(t|x)$ by minimising the following negative log-likelihood loss:

$$\mathcal{L}_{\text{NLL}} = -\sum_{i=1}^{N} \Big[\delta^i \log \left(\hat{p}_{\mathcal{T}|\mathcal{X}}(\tau^i \mid x^i) \right) + (1 - \delta^i) \log \left(\hat{S}_{\mathcal{T}|\mathcal{X}}(\tau^i \mid x^i) \right) \Big]$$

where $\hat{p}_{\mathcal{T}|\mathcal{X}}(\tau|x) = \hat{h}_{\mathcal{T}|\mathcal{X}}(\tau|x)\hat{S}_{\mathcal{T}|\mathcal{X}}(\tau|x)$ estimate the probability of an event occurring at time t. In the negative log-likelihood formulation, two distinct terms account for different groups of data points. The first term corresponds to instances where the event has occurred ($\delta^i = 1$), capturing the likelihood of event times. The second term applies to censored observations ($\delta^i = 0$), modelling the probability of survival beyond the observed time τ^i.

A.2 Data

The dataset was collected through a government-funded cohort study initiated in 2012 by the Government of Maharashtra, India, as part of a state-sponsored insurance scheme aimed at improving access to high-cost medical care for socioeconomically disadvantaged populations [18]. The cohort consisted of **4,595** patients who underwent percutaneous coronary intervention (PCI) at various participating centres across Maharashtra. Out of the 4595 patients, only **197 (4.92%)** patients experienced the event, which shows the highly censored nature of the data. Patients were followed for one year, with outcomes including all-cause mortality, repeat PCI, rehospitalisation, and recurrence of angina being documented. The dataset encompasses a range of patient characteristics, including **demographic factors** (age, sex, employment status, education level, and locality), **clinical conditions** (presence of diabetes, hypertension, and a history of coronary artery disease), **behavioural factors** (smoking and alcohol consumption), and **medication** (Aspirin, Clopidogrel, Ticlopidine, Prasugrel, and Dual-Antiplatelet Therapy [DAPT]). Additionally, the dataset includes details related to the **PCI procedure** itself, such as stent length, total number of stents, type of stent used, and stent positioning.

A.3 Curated Hazard Ratios (HR) and Relative Risks (RR)

Covariate	Study	Population	RR / HR
Smoking	[19]	United States	*Male*: Current Smoker: 2.24 Past Smoker: 1.30 *Female*: Current Smoker: 2.29 Past Smoker: 1.35
Alcohol	[20]	Mumbai	HR: 1.22
Diabetes	[21]	Mumbai	HR: 1.67
Hypertension	[22]	United States	HR: 1.25
History of CAD	[23]	United States and UK (EVEREST)	HR: 1.12
Aspirin	[24]	United States (NHANES)	HR: 0.647
Clopidogrel	[25]	Danish	Diabetic, HR: 0.89 Non-Diabetic, HR: 0.75
Ticlopidine	[26]	Canada	RR: 0.698
Prasugrel	[27]	Multiple Regions	HR: 0.96 (Prasugrel vs Clopidogrel)
Dual Anti Platelet Therapy	[28]	Mumbai	HR: 0.52
Total Number of Stents	[28]	Mumbai	HR: 1.01
Total Length	[29]	Europe and North America	HR: 1.05 per 10mm
Stent Position	[28]	Mumbai	LCX, HR: 0.43, RCA, HR: 0.42, Multiple, HR: 0.75 (ref: LAD)
Stent Type	[30]	Korea	HR: 0.60 (BMS as ref.)

Table 3: Compilation of hazard ratios and relative risks from the literature, adjusted for relevant covariates and meeting statistical significance criteria ($p < 0.005$)

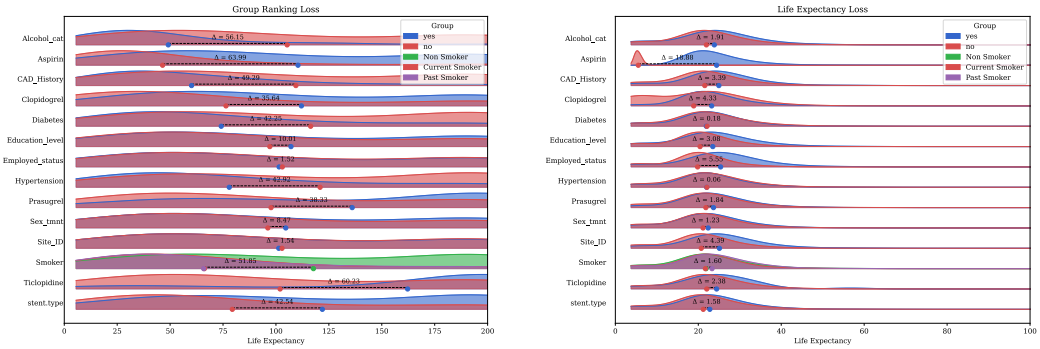

Figure 3: **Effect of loss components on feature-wise life expectancy separation. (a)** Application of only the group ranking loss ($\mathcal{L}_{\text{rank}}$) leads to exaggerated separation in life expectancy distributions across feature-defined groups. This often results in implausibly wide ranges (e.g., $\Delta > 60$ years), undermining clinical credibility. **(b)** Application of only the life expectancy loss ($\mathcal{L}_{\text{LE}}$) produces more constrained and clinically plausible distributions. However, the suppression of variance can mask the impact of important but less dominant features, yielding minimal and wrong group separation in several covariates.

A.4 Training and Hyperparameters

For training and evaluation, we used an 80:20 train–test split, resulting in 3,152 samples for training and 788 samples for testing. Model training was performed using the `torchsurv` library, employing the negative log-likelihood loss for the Weibull model. In addition, we implemented the auxiliary losses introduced in this work.

The overall objective function is given by:

$$\mathcal{L} = \mathcal{L}_{\text{NLL}} + \lambda_{\text{LE}} \cdot \mathcal{L}_{\text{LE}} + \lambda_{\text{rank}} \cdot \mathcal{L}_{\text{rank}},$$

where $\mathcal{L}_{\text{NLL}}$ denotes the negative log-likelihood loss, $\mathcal{L}_{\text{LE}}$ the life expectancy loss, and $\mathcal{L}_{\text{rank}}$ the ranking loss. The coefficients λ_{NLL}, λ_{LE}, and λ_{rank} control the relative contribution of each component.

In our experiments, we set $\lambda_{\text{NLL}} = 5$, $\lambda_{\text{LE}} = 0.05$, and $\lambda_{\text{rank}} = 5$ to obtain the reported results.

A.5 Interpreting the Role of Loss Components

To evaluate the contribution of each loss component, specifically the life expectancy consistency loss ($\mathcal{L}_{LE}$) and the ranking-based separation loss ($\mathcal{L}_{rank}$), toward generating interpretable and clinically meaningful longitudinal survival curves, we analyze their effects using two criteria: (i) feature-wise group separation in the predicted survival distributions, and (ii) the plausibility and coherence of the resulting life expectancy estimates across cohorts.

Figure 3 demonstrates that group ranking loss ($\mathcal{L}_{rank}$) promotes greater separation between the feature-defined groups with clinically consistent ordering, effectively pushing survival curves apart. While this enhances group-wise discrimination, it often results in excessively broad and clinically implausible life expectancy distributions. In contrast, the life expectancy loss ($\mathcal{L}_{LE}$) enforces tighter alignment of survival curves among individuals with similar age profiles, yielding more coherent population-level life expectations. However, this constraint can make the separation between subtler features disappear or become invisible. From Figure 3, we can see that the ordering of groups is also not consistent with clinical trends. The proposed LONGSURV model balances these opposing tendencies by jointly optimising both objectives. In the presented configuration, we weight $\mathcal{L}_{\text{LE}}$ and $\mathcal{L}_{\text{rank}}$ at 0.05 and 5.0, respectively, effectively harmonising clinical plausibility with interpretability through structured group-wise separations.

NeurIPS Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

- You should answer [Yes] , [No] , or [NA] .
- [NA] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
- Please provide a short (1–2 sentence) justification right after your answer (even for NA).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers, area chairs, senior area chairs, and ethics reviewers. You will be asked to also include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While "[Yes] " is generally preferable to "[No] ", it is perfectly acceptable to answer "[No] " provided a proper justification is given (e.g., "error bars are not reported because it would be too computationally expensive" or "we were unable to find the license for the dataset we used"). In general, answering "[No] " or "[NA] " is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

IMPORTANT, please:

- **Delete this instruction block, but keep the section heading "NeurIPS Paper Checklist"**,
- **Keep the checklist subsection headings, questions/answers and guidelines below.**
- **Do not modify the questions and only use the provided macros for your answers**.

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The main claims are that our method produces survival extrapolations that are both plausible and consistent with established causal relationships between survival and risk factors, relying solely on external causal priors. We substantiate this in the results section at two levels: first, by reporting commonly used within-cohort metrics to demonstrate short-term concordance; and second, by evaluating the plausibility of long-term extrapolations.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [Yes]

Justification: The approach has a few limitations in-terms lack of sensitivity analysis, transfering to new population, evaluating the extrapolation using actual data, etc., which we disucss breifly at the end.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [NA]

Justification: We don't have any theoretical analysis and proofs for this work.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: We have provided the core details of the algorithm which is sufficient to recreate; however, the code and data underlying this study are not made publicly available.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [No]

Justification: As this is a workshop paper, the code and data are not yet ready for release. We plan to make them available in future iterations of this work

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.

- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [No]

   Justification: As this is a workshop paper, space constraints prevent us from providing full training and test details; however, these will be included in an extended version of the work

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: We do not report error bars in this work, as the focus is on demonstrating the plausibility of extrapolations and reporting standard within-cohort concordance metrics

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
   - If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [No]

Justification: The experiments were conducted on a simple tabular dataset and required no special computational resources.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We foresee no harm arising from our research, and no personal information from the underlying human subject data is disclosed. This work fully adheres to the NeurIPS ethics guidelines.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: We discuss how are approach can be used for survival extrapolation for patients and get to know the impact of risk factors on their survival

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: The paper discusses simple survival models which don't pose any risk

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: All code was developed by the authors, and the data were collected as part of a previous work by one of the authors

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: The work doesn't introduce any new assest

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.

- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: No active crowdsourcing was conducted for this study; however, the work relies on data from human subjects that were collected and published previously by one of the authors.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: This study involves neither active crowdsourcing nor new research with human subjects; it relies solely on previously collected and published human subject data.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

    Justification: This work is not related to LLM research.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.