# CR-SAVAE: A Parametric Method for Survival Analysis with Competing Risks

Patricia A. Apellániz, Juan Parras, Santiago Zazo
Information Processing and Telecommunications Center
Universidad Politécnica de Madrid
Madrid, Spain
Email: *patricia.alonsod@upm.es*, *j.parras@upm.es*, *santiago.zazo@upm.es*

***Abstract*—Competing risks in survival analysis pose a significant challenge in healthcare, but few methods effectively address this problem. Moreover, the use of deep learning techniques remains limited, with the most prevalent approach, DeepHit, being non-parametric. This often leads to limitations in interpretability and statistical inference, which are crucial aspects. We propose a new alternative that harnesses the power of variational autoencoders and deep learning to tackle survival analysis with competing risks within a parametric framework. Our model, CR-SAVAE, allows direct interpretation of covariate effects on survival outcomes and enables researchers to perform robust statistical analysis compared to non-parametric approaches, essential for understanding risk mechanisms and making informed clinical decisions. It provides personalized medicine insight by accurately estimating the cumulative incidence function and avoiding the need for proportional hazards assumptions inherent in other models. Through comprehensive experiments on datasets with varying degrees of censoring and competing risks, we demonstrate the potential of our approach to achieve performance comparable to that of DeepHit based on the concordance index and integrated Brier score. This study highlights the potential of CR-SAVAE to advance survival analysis, improve interpretability, and enable more accurate and personalized clinical decision making in healthcare settings.**

## I. Introduction

Survival Analysis (SA) stands out as a key statistical technique widely used in a variety of disciplines, from medical research to engineering. In essence, SA techniques focus on predicting the time to the occurrence of an event of interest. This area is critical in clinical research, as it allows us to assess the efficacy of treatments, to understand the prognosis of the patient, and to identify risk factors. Obtaining information from SA is crucial to developing clinical guidelines, improving patient care, and guiding future research.

However, traditional SA models, such as the Kaplan-Meier (KM) estimator [1] or the Cox model [2], struggle with the complexities of Competing Risks (CR). CR scenarios (e.g., different causes of death), in which patients are susceptible to multiple potential events with interconnected probabilities, are common in medical studies. Ignoring CR can lead to biased estimates and erroneous interpretations, especially in the evaluation of risk factors and event probabilities [3]. Therefore, specialized methodologies are crucial to effectively address CR and ensure accurate inferences from survival data.

Consequently, the main approach is based on the Cumulative Incidence Function (CIF) [4], and the Fine-Gray (FG) [5] model is the most widely used. However, the FG model, similar to the Cox model, relies on strong assumptions about the underlying stochastic processes and the form of the hazard function, which may not consistently align with the complexities of real-world data. These constraints have motivated the exploration of Deep Learning (DL) techniques within the field of SA.

Despite the advances facilitated by DL in SA, as demonstrated by DeepSurv [6] and SAVAE [7], its application in the context of CR remains limited. A recent review by [8] that analyzed 69 DL-based SA models highlighted that only nine could handle CR. The most popular method is DeepHit [9], which circumvents the proportional hazards assumption and outperforms classical methods like FG. However, DeepHit's non-parametric nature results in estimated CIF curves for each patient that lack an analytical expression and are purely numerical, hindering statistical computation. Furthermore, these CIF curves can exhibit significant variability between patients. Although the existing literature includes another fully parametric DL alternative [10], it focuses on estimating individual patient survival functions, not CIFs, which are crucial in CR SA. As described in [11], using survival functions instead of CIF introduces bias.

In light of the limitations mentioned and the early stage of DL research in CR, we introduce a novel model, CR-SAVAE, that addresses these critical gaps. CR-SAVAE empowers clinicians with precise and personalized survival predictions, facilitating informed clinical decision making and customized treatment plans. Our key contributions are:

- Parametric approach for personalized medicine: Patient-specific survival parameters for deeper insight and complete statistical calculation (unlike DeepHit).
- Accurate CIF estimation: Direct estimation of CIF for improved prediction accuracy (differentiating it from [10]).
- Flexible handling of time data: Efficiently handles continuous and discrete survival times, avoiding data loss from discretization (unlike DeepHit, DeepComp [12], and CRESA [13]).
- Robustness and flexibility: Models survival times and

covariates with diverse distributions (independent of the proportional hazards assumption), leading to more accurate and robust predictions, unlike DeepHit, which uses Concordance Index (C-index) training.
- Competitive performance: Demonstrates competitive performance in terms of the C-index and the integrated Brier Score (iBS) against DeepHit on real datasets.

## II. Background

### A. Survival Analysis

SA predicts the time to an event, often death in clinical settings. Given a dataset of $N$ patients $\{x_i, t_i, l_i\}_{i=1}^{N}$, each patient $i$ has a set of covariates $x_i$ (such as clinical, demographic or genomic data) used for prediction, the time of event occurrence $t_i$, and $l_i$ is the event label indicating right-censoring.

In single-risk SA, $l_i$ is the event or censoring, represented as 0 or 1, respectively. Censoring may indicate that the patient has not experienced the event during the time when the data were collected, that the patient has left the study, etc. The survival function $S(t|x) = P(T > t|x) = 1 - F(t|x)$ predicts the probability of an event given the covariates $x$, where $F$ is the Cumulative Distribution Function (CDF), and $T$ the random variable that models the survival time (we use capital $T$ for the random variable and $t$ for its realizations). It meets three key properties: (1) $S(t = 0|x) = 1$, that is, there is no event at the study start; (2) $\lim_{T\to\infty} S(T|x) = 1$, all patients experience the event eventually; (3) $S$ is monotonically decreasing, i.e., it decreases as time increases (fewer patients remain). Note that the second condition is hindered by the bias introduced by censoring, since we cannot observe patients indefinitely.

However, a more realistic scenario involves CR, where patients face multiple possible events. Here, $l_i \in \{0, 1, ..., L\}$ with $L$ possible risks and censoring marked by 0. CR SA aims to predict the survival time for each of the risks, which are said to "compete" among them because if one risk happens, the rest cannot take place. Note that the Kaplan-Meier estimator is known to be biased [11] in the presence of CR, so instead of estimating $S(t|x)$, CIF must be used:

$$CIF_k(t|x) = P(T \leq t, l = k|x). \tag{1}$$

The CIF represents the probability of experiencing an event $k \in \{1, 2, ..., L\}$ before time $t$ and any other event $l \neq k$. The sum of CIFs for all risks equals the incidence of all competing events, and unlike single-risk settings, individual CIFs do not need to reach 1 as time increases.

### B. SAVAE

SA is heavily based on Cox-based models, but their limitations, particularly the proportional hazards assumption, restrict their applicability in complex real-world scenarios. While non-parametric methods like DeepHit overcome this limitation, they lack interpretability and analytical traceability. We propose SAVAE as a compelling alternative, providing flexibility from non-parametric models while retaining the essential aspects of parametric models. Unlike DeepHit, SAVAE allows for fewer modeling restrictions and offers parametric benefits by selecting suitable survival distributions. Therefore, in our research, we use SAVAE to harness its potential to address the limitations of existing models.

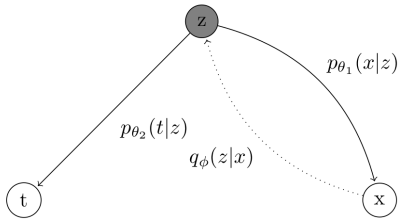


Fig. 1. SAVAE model. The shadowed circle refers to the latent variable and white circles refer to the observables. Note that the probabilities $p_{\theta_1}(x|z)$ and $p_{\theta_2}(t|z)$ denote the generative models and $q_\phi(z|x)$ denotes the variational approximation to the posterior, since the true posterior $p(z|x)$ is unknown.

SAVAE is a single-risk approach that leverages Variational Autoencoders (VAEs) to predict the time-to-event distribution based on covariates. It combines a latent variable $z$ with observed covariates $x$ and time-to-event $t$. Using conditional independence, two models, $p_{\theta_1}(x|z)$ and $p_{\theta_2}(t|z)$, are used to jointly model the variable distribution. Effectively, knowing $z$ allows the independent generation of $x$ and $t$. Through a variational distribution $q_\phi(z|x)$, SAVAE estimates the predictive distribution based on covariates, approximating the true posterior $p(z|t, x)$. The chosen distribution family and optimized parameters determine the quality of the approximation. SAVAE is depicted in Fig. 1. It has three loss components:

- $D_{KL}(q(z|x)||p(z))$. *Kullback-Leibler divergence*: Measures the similarity between the latent space distribution $q(z|x)$ and a prior isotropic Gaussian distribution $p(z)$, where $p$ and $q$ denote probability density functions.
- *Covariate reconstruction loss*: Represents the log-likelihood for each of the $m$ covariates, $p(x^m|z)$, requiring an appropriate distribution selection. SAVAE trains using the negative log-likelihood for differentiability.
- *Time reconstruction loss*: Accounts for censoring by considering only the uncensored time points $(l_i = 1)$ for log-likelihood calculation and censored time points $(l_i = 0)$ for survival function estimation:

$$\log p(t_i|z_i) = \mathbb{I}(l_i = 1) \log p(t_i|z_i) + \mathbb{I}(l_i = 0) \log S(t_i|z_i) \tag{2}$$

where $\mathbb{I}$ is the indicator function. SAVAE is parametric, which means that, given the covariates of a patient $x_i$, it outputs the parameters of the chosen time distribution for that patient. Hence, it is possible to analytically obtain the log-likelihood of that distribution, as well as the survival function and any statistic that depends on the estimated parameters.

Thus, the total loss of SAVAE for $N$ patients is expressed as:

$$\mathcal{L}_S = -\frac{1}{N}\sum_{i=1}^{N}\Bigg[-D_{KL}(q(z_i|x_i)||p(z)) + \sum_m \log p(x_i^m|z_i) + \mathbb{I}(l_i=1)\log p(t_i|z_i) + \mathbb{I}(l_i=0)\log S(t_i|z_i)\Bigg] \tag{3}$$

## III. Competing Risks SAVAE

SAVAE can be extended to the CR setting by estimating the joint survival function based on Eq. 1 and the Total Probability Theorem:

$$\begin{aligned} S(t|x) =& 1 - F(t|x) = 1 - P(T \le t|x) = 1 - \sum_{k=1}^{L} CIF_k(t) \\ =& 1 - \sum_{k=1}^{L} P(T \le t, l=k|x) \\ =& 1 - \sum_{k=1}^{L} P(T \le t|l=k,x)\cdot P(l=k|x), \end{aligned} \tag{4}$$

which requires: (1) the conditional survival probability, $P(T \le t|l=k,x)$, estimated using separate decoders for each risk, trained with Eq. 2 for both censored and uncensored data; and (2) the probability of risk $P(l=k|x)$, estimated jointly with the latent variable $z$ using the SAVAE architecture.

Information about censored data is incorporated by exploiting the fact that no event occurred before the censoring time $t$. This is reflected in Eq. 5 for censored patients.

$$\log\left(S(t_i|x_i)\right) = \log\left(1 - \sum_{k=1}^{L} CIF_k(t_i)\right) \tag{5}$$

The CIF for each risk (Eq. 6) is then estimated by combining the conditional survival probability and the risk probability.

$$CIF_k(t) = P(T \le t|l=k,x)\cdot P(l=k|x) \tag{6}$$

To estimate the risk probability $P(l=k|x)$, CR-SAVAE uses a neural network classifier that takes the latent representation $z$ of the SAVAE architecture as input. The classifier outputs $p_\theta(l|z)$, representing the probability of each risk given the latent representation of the patient. The classifier assumes a categorical distribution and is trained using the log-likelihood of this distribution, considering only uncensored patients (Eq. 7).

$$\frac{1}{N}\sum_{i=1}^{N}\mathbb{I}(l_i \neq 0)\log p(l_i|z_i) \tag{7}$$

Fig. III visually compares the architectures of SAVAE and CR-SAVAE, highlighting the additional components of CR-SAVAE to handle competing risks. Notably, SAVAE becomes a special case of CR-SAVAE in the single-risk setting, requiring only a single decoder for time and omitting the risk classifier.

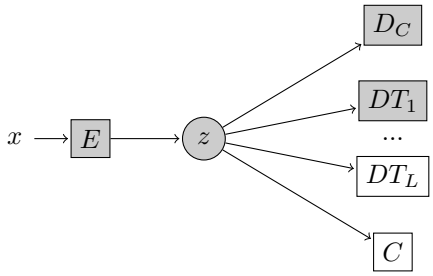


Fig. 2. Schema comparing SAVAE and CR-SAVAE. Rectangles represent Neural Networks, and shadowed elements are common to SAVAE and CR-SAVAE. $E$ means Encoder, $z$ is the latent space, $D_C$ is the covariate decoder, $DT_k$ is the time decoder for risk $k$ estimation $P(T \le t|l=k,x)$, and $C$ is the classifier estimating $P(l=k|x)$. Note that SAVAE is CR-SAVAE with a single risk and without classifier.

Hence, the total loss for CR-SAVAE is:

$$\begin{aligned} \mathcal{L}_{CRS} =& -\frac{1}{N}\sum_{i=1}^{N}\Bigg[-D_{KL}(q(z|x_i)||p(z)) + \sum_m \log p(x_i^m|z_i) \\ &+ \sum_{k=1}^{L}\Big(\mathbb{I}(l_i=k)\log p(t_i|l_i=k,z_i)\Big) \\ &+ \mathbb{I}(l_i=0)\log\left(1 - \sum_{k=1}^{M} CIF_k(t_i)\right) \\ &+ \mathbb{I}(l_i\neq 0)\log p_\theta(l_i|z_i)\Bigg] \end{aligned} \tag{8}$$

where we note that the differences to SAVAE lie in time estimation, as the KL term and the covariates reconstruction loss are identical to SAVAE loss (3). Finally, note that according to (1), once we have estimated $P(T \le t|l=k,x)$ and $P(l=k|x)$, we can estimate the CIF as in (6), which will be needed to calculate the performance metrics.

## IV. Empirical validation

To test the performance of CR-SAVAE, we run a benchmark on three publicly available datasets. The code and datasets needed to replicate our results are available at https://github.com/Patricia-A-Apellaniz/cr-savae. These datasets are:

- **Melanoma**. It contains 205 patients who had malignant melanoma and underwent surgery at the University Hospital of Odense (Denmark) between 1962 and 1977. For each patient, the dataset collects 7 covariates, such as age, sex, tumor thickness, and whether it was ulcerated or not. There are two competing risks: dying from melanoma and dying from other causes.
- **MGUS2**. This dataset contains $1,348$ patients who had monoclonal gammopathy of unknown significance (MGUS). Each patient has 5 covariates and there are two competing risks: evolution of the situation to plasma cell malignancy, or death related to other causes.
- **EBMT**. It contains $8,966$ transplanted patients from the European Society for Blood and Marrow Transplantation

TABLE I
DATASETS USED FOR BENCHMARKING

| Name | CR | Risk proportion |
|---|---|---|
| Melanoma | 2 | 0.65/0.28/0.07 |
| MGUS2 | 2 | 0.29/0.08/0.63 |
| EBMT | 6 | 0.63/0.13/0.09/0.02/0.02/0.01/0.1 |

The study includes diverse CR scenarios (quantity and proportion), alongside censored patients (first position proportion), demonstrating CR-SAVAE's performance across different data settings.

(EBMT), each with 5 covariates. We randomly sampled $1,000$ patients for our benchmark. There are six competing risks in this dataset, which are death from relapse, Graft-Versus-Host-Disease, bacterial infection, viral infection, fungal infection, or other cause of death.

A summary of each dataset can be seen in Table I, where we emphasize that they have been chosen to represent different conditions in real medical datasets: having two or more risks, different levels of censoring and a marked imbalance between risks. We use 5-fold cross-validation for every dataset, using an $80\%$ of the data to train the algorithms, and another $20\%$ to obtain the results in the article. For CR-SAVAE, we model the time using a Weibull distribution. This means that each time decoder outputs the scale parameter $\lambda > 0$ and the shape parameter $\alpha > 0$. The log-likelihood and survival function for the Weibull distribution are:

$$\log p(t) = \log(\alpha) - \log(\lambda) + (\alpha - 1)\log\left(\frac{t}{\lambda}\right) - \left(\frac{t}{\lambda}\right)^{\alpha}$$
$$\log S(t) = -\left(\frac{t}{\lambda}\right)^{\alpha} \tag{9}$$

We remark that CR-SAVAE is a parametric method, where it is possible to change the time distribution to any other distribution, as long as the log-likelihood and the survival function are differentiable. Although we have chosen the Weibull distribution because it does not assume proportional hazards, other distributions like Log-Normal, Inverse Gaussian, and more can be used. This highlights the advantage of CR-SAVAE: being fully parametric, we can change the time distribution to the one desired and obtain parameter estimates for each patient.

As a concrete example with the Weibull time distribution, after training CR-SAVAE, we can input the covariates of a patient $x_i$ and obtain $(\alpha_i, \lambda_i)$ for each of the $k$ risks, as well as $\theta_k$, the probability of each risk for that concrete patient. Then, we can compute the analytical, personalized CIF as $CIF_k(t) = (1 - e^{(-t/\lambda_i)^{\alpha_i}})\theta_k$. Since CR-SAVAE estimates the distribution parameters, we can obtain not only the CIF curve, but any statistic that depends on these parameters, yielding personalized predictions for each individual patient.

Regarding the VAE parameters, we use a latent dimension of 5 and hidden layers consisting of $64$ neurons. We train using minibatches of 32 patients and opted for the Adam optimizer with a learning rate of $1e-3$. We train using a maximum of 2000 iterations, but if the validation loss does not improve in 30 epochs, we early stop the training.

TABLE II
MELANOMA. RESULTS OBTAINED COMPARING DEEPHIT AND CR-SAVAE

| Risk | Method | C-index | iBS |
|---|---|---|---|
| 1 | CR-SAVAE | $0.6318 \pm 0.0731$ | $0.2415 \pm 0.0291$ |
| 1 | DeepHit | $0.666 \pm 0.0303$ | $0.2929 \pm 0.0198$ |
| 2 | CR-SAVAE | $0.632 \pm 0.0963$ | $0.322 \pm 0.1432$ |
| 2 | DeepHit | $0.6848 \pm 0.1243$ | $0.1876 \pm 0.0469$ |

Data is average $\pm$ standard deviation. A two-sided unequal variance T-Test [17] comparing the means of every metric and the risk yields a $p$-value higher than 0.01, which means that neither of the two methods tested is significantly better than the other.

To evaluate the performance of CR-SAVAE we compared it with DeepHit in terms of the time-dependent C-index [14] and the iBS [15].

*1) The C-index* assesses the concordance between the CIF predicted by a patient at event time and that of unexperienced patients, considering the event timing and censoring. The C-index is calculated as

$$C_{index} \approx \frac{\sum_{i \neq j} A_{k,i,j} \cdot \mathbb{I}\Big(CIF_k(t|x_i) \leq CIF_k(t|x_j)\Big)}{\sum_{i \neq j} A_{k,i,j}}, \tag{10}$$

where $A_{k,i,j}$ is the indicator function of comparable pairs $(i, j)$ for the event $k$.

*2) The iBS* measures the overall discrepancy between the predicted and observed CIFs for each patient over time, considering the possibility of multiple events. Being $G(\cdot)$ the survival function corresponding to censoring and $1/G(t)$ the Inverse Probability of Censoring Weighting (IPCW) [16], the iBS is calculated as $iBS(t_{max}) = \frac{1}{t_{max}} \int_0^{t_{max}} BS(t)dt$, where $BS(t)$ is defined as

$$BS(t) = \frac{1}{N} \sum_{i=1}^{N} \Bigg[ (1 - CIF_k(t|x_i))^2 / G(t_i) \cdot \mathbb{I}(t_i < t, d_i = 1) + (CIF_k(t|x_i))^2 / G(t) \cdot \mathbb{I}(t_i \geq t) \Bigg] \tag{11}$$

The results of CR-SAVAE compared to DeepHit can be seen in Tables II, III and IV, where we report the average metric across folds, plus the standard deviation. For each triplet of dataset, risk and metric, we run an unequal variance T-test [17] to check whether the means of DeepHit and our model were

TABLE III
MGUS2. RESULTS OBTAINED COMPARING DEEPHIT AND CR-SAVAE

| Risk | Method | C-index | iBS |
|---|---|---|---|
| 1 | CR-SAVAE | $0.6369 \pm 0.0446$ | $0.3403 \pm 0.0972$ |
| 1 | DeepHit | $0.5919 \pm 0.0478$ | $0.2686 \pm 0.0617$ |
| 2 | CR-SAVAE | $0.5672 \pm 0.0427$ | $0.3505 \pm 0.027$ |
| 2 | DeepHit | $0.6359 \pm 0.0174$ | $0.1865 \pm 0.0823$ |

Data is average $\pm$ standard deviation. A two-sided unequal variance T-Test [17] comparing the means of every metric and the risk yields a $p$-value higher than 0.01, which means that neither of the two methods tested is significantly better than the other.

TABLE IV
EMBG. RESULTS OBTAINED COMPARING DEEPHIT AND CR-SAVAE

| Risk | Method | C-index | iBS |
|---|---|---|---|
| 1 | CR-SAVAE | $0.5645 \pm 0.0333$ | $0.2459 \pm 0.0231$ |
| 1 | DeepHit | $0.57 \pm 0.029$ | $0.257 \pm 0.0257$ |
| 2 | CR-SAVAE | $0.5737 \pm 0.0612$ | $0.2592 \pm 0.0184$ |
| 2 | DeepHit | $0.5248 \pm 0.0263$ | $0.2489 \pm 0.024$ |
| 3 | CR-SAVAE | $0.5867 \pm 0.096$ | $0.1987 \pm 0.0135$ |
| 3 | DeepHit | $0.7106 \pm 0.1197$ | $0.1647 \pm 0.0155$ |
| 4 | CR-SAVAE | $0.5729 \pm 0.0567$ | $0.2006 \pm 0.0176$ |
| 4 | DeepHit | $0.7108 \pm 0.1246$ | $0.1811 \pm 0.0229$ |
| 5 | CR-SAVAE | $0.7411 \pm 0.1111$ | $0.2 \pm 0.0092$ |
| 5 | DeepHit | $0.6007 \pm 0.0796$ | $0.1649 \pm 0.0219$ |
| 6 | CR-SAVAE | $0.5519 \pm 0.0214$ | $0.2363 \pm 0.02$ |
| 6 | DeepHit | $0.5946 \pm 0.0408$ | $0.2683 \pm 0.0208$ |

Data is average $\pm$ standard deviation. A two-sided unequal variance T-Test [17] comparing the means of every metric and the risk yields a $p$-value higher than $0.01$, which means that neither of the two methods tested is significantly better than the other.

significantly different, and we found no significant evidence of this (i.e. $p$-values higher than $0.01$). This means that CR-SAVAE is a viable alternative to DeepHit in the CR SA setting, since it provides similar metrics and is also a parametric method, such as [10]. Since DeepHit is non-parametric, it may provide higher metrics results compared to a parametric alternative, as highlighted in [9]. In fact, parametric methods are considered to provide good results when they match the performance of DeepHit, as shown in [10]. Thus, our proposed CR-SAVAE provides very promising results, matching the performance of a non-parametric method as DeepHit, with all the advantages of being parametric: it facilitates hypothesis testing, confidence interval estimation, and other statistical analysis. DeepHit's approach lacks this interpretability.

## V. CONCLUSIONS

CR-SAVAE, a parametric competing risk survival analysis model, demonstrates performance comparable to DeepHit, a widely used non-parametric model, on metrics like the C-index and iBS. However, the parametric nature of CR-SAVAE offers significant advantages. CR-SAVAE supports rigorous statistical analysis, including hypothesis testing, confidence interval estimation, and other procedures essential for research and decision-making. Moreover, our model accurately estimates the cumulative incidence function and robustly handles both continuous and discrete time data. This flexibility expands its potential in real-world healthcare settings. Furthermore, its parametric structure also offers computational advantages compared to non-parametric approaches. Future research aims to leverage the VAE architecture of CR-SAVAE for broader healthcare applications. Potential directions include using the latent space for patient clustering and synthesized patient data generation, further expanding the utility of the model beyond traditional SA. Furthermore, investigating explainability using standard techniques such as SAGE [18] would significantly improve the interpretability of the model, a critical priority in artificial intelligence in healthcare.

## ACKNOWLEDGMENT

This research was supported by GenoMed4All and SYNTHEMA projects. Both have received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 101017549 and 101095530, respectively. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## REFERENCES

[1] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *Journal of the American statistical association*, vol. 53, no. 282, pp. 457–481, 1958.

[2] D. R. Cox, "Regression models and life-tables," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 34, no. 2, pp. 187–202, 1972.

[3] N. A. Schuster, E. O. Hoogendijk, A. A. Kok, J. W. Twisk, and M. W. Heymans, "Ignoring competing events in the analysis of survival data may lead to biased results: a nonmathematical illustration of competing risk analysis," *Journal of clinical epidemiology*, vol. 122, pp. 42–48, 2020.

[4] R. J. Gray, "A class of k-sample tests for comparing the cumulative incidence of a competing risk," *The Annals of Statistics*, vol. 16, no. 3, pp. 1141–1154, 1988.

[5] J. P. Fine and R. J. Gray, "A proportional hazards model for the subdistribution of a competing risk," *Journal of the American statistical association*, vol. 94, no. 446, pp. 496–509, 1999.

[6] J. L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, and Y. Kluger, "Deepsurv: personalized treatment recommender system using a cox proportional hazards deep neural network," *BMC medical research methodology*, vol. 18, no. 1, pp. 1–12, 2018.

[7] P. A. Apellániz, J. Parras, and S. Zazo, "Savae: Leveraging the variational bayes autoencoder for survival analysis," *arXiv preprint arXiv:2312.14651*, 2023. Unpublished.

[8] S. Wiegrebe, P. Kopper, R. Sonabend, and A. Bender, "Deep learning for survival analysis: A review," *ArXiv*, vol. abs/2305.14961, 2023.

[9] C. Lee, W. Zame, J. Yoon, and M. Van Der Schaar, "Deephit: A deep learning approach to survival analysis with competing risks," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, 2018.

[10] C. Nagpal, X. Li, and A. Dubrawski, "Deep survival machines: Fully parametric survival regression and representation learning for censored data with competing risks," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 8, pp. 3163–3175, 2021.

[11] P. C. Austin, D. S. Lee, and J. P. Fine, "Introduction to the analysis of survival data in the presence of competing risks," *Circulation*, vol. 133, no. 6, pp. 601–609, 2016.

[12] Y. Li, W. Jia, Y. Kang, T. Chen, X. Li, X. Du, J. Dong, C. Ma, F. Wang, and G. Xie, "Deepcomp: Which competing event will hit the patient first?," in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 629–636, IEEE, 2020.

[13] G. Gupta, V. Sunder, R. Prasad, and G. Shroff, "Cresa: a deep learning approach to competing risks, recurrent event survival analysis," in *Advances in Knowledge Discovery and Data Mining: 23rd Pacific-Asia Conference, PAKDD 2019, Macau, China, April 14-17, 2019, Proceedings, Part II 23*, pp. 108–122, Springer, 2019.

[14] L. Antolini, P. Boracchi, and E. M. Biganzoli, "A time-dependent discrimination index for survival data," *Statistics in Medicine*, vol. 24, 2005.

[15] G. W. BRIER, "Verification of forecasts expressed in terms of probability," *Monthly Weather Review*, vol. 78, no. 1, pp. 1 – 3, 1950.

[16] J. M. Robins, "Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers," in *Proceedings of the Biopharmaceutical Section, American Statistical Association*, vol. 24, p. 3, San Francisco CA, 1993.

[17] B. L. Welch, "The generalization of 'student's'problem when several different population varlances are involved," *Biometrika*, vol. 34, no. 1-2, pp. 28–35, 1947.

[18] I. Covert, S. Lundberg, and S.-I. Lee, "Understanding global feature contributions with additive importance measures," 2020.