# DeepPseudo : A Deep learning approach based on Pseudo values for Competing Risk Analysis

Md Mahmudur Rahman
mrahman6@umbc.edu
University of Maryland, Baltimore County
Baltimore, Maryland, USA

Shinyu Matsuzaki
zacky_s@gyne.med.osaka-u.ac.jp
Osaka University
Osaka, Japan

Koji Matsuo
Koji.Matsuo@med.usc.edu
University of Southern California
Los Angeles, California, USA

Sanjay Purushotham
psanjay@umbc.edu
University of Maryland, Baltimore County
Baltimore, Maryland, USA

## ABSTRACT

Competing Risk Analysis (CRA), an important problem in survival analysis, aims to estimate the probability of occurrence of an event in the presence of competing events. Many of the statistical approaches developed for CRA are limited by strong assumptions about the underlying stochastic processes. To overcome these issues and to handle censoring, machine learning approaches for CRA have designed specialized cost functions. However, these approaches are not generalizable and are computationally expensive. This paper formulates CRA as a cause-specific regression problem, and proposes a simple and effective feed-forward deep neural network model, *DeepPseudo*, to predict the cumulative incidence function using Aalen-Johansen estimator based pseudo values. *DeepPseudo* models the time-varying covariate effect on cumulative incidence function while handling the censored observations. Experiments on real and synthetic datasets show that DeepPseudo obtains promising and statistically significant results compared to the previous state-of-the-art CRA approaches.

## CCS CONCEPTS

• **Applied computing** → **Life and medical sciences**; • **Computing methodologies** → *Neural networks.*

## KEYWORDS

Competing risk analysis, deep learning, pseudo values, regression

**ACM Reference Format:**
Md Mahmudur Rahman, Shinyu Matsuzaki, Koji Matsuo, and Sanjay Purushotham. 2020. DeepPseudo : A Deep learning approach based on Pseudo values for Competing Risk Analysis. In *2020 KDD Workshop on Applied Data Science for Healthcare, August 24, 2020, San Diego, CA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/1122445.1122456


*KDD '20, August 24, 2020, San Diego, CA*

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

## 1 INTRODUCTION

Competing Risk Analysis (CRA) is a special type of survival analysis - *time-to-event analysis* - that aims to estimate the probability of occurrence of an event in the presence of competing events. CRA is common in medical settings [10, 16] since a patient can experience more than one type of a certain event. For example, a patient may experience death (event) at a time $t$ from one of the following causes; cardiovascular disease, breast cancer, or kidney damage. Individuals who die of cardiovascular disease are no longer at risk of dying of breast cancer or kidney damage. These causes of failure are referred to as *competing events*, and the probability of these events are referred to as *competing risks*. CRA has received substantial attention in statistics, and machine learning literature. Popular statistical approaches [8, 12] widely used in medical setting have been extended to CRA [1, 9]. However, these models are limited by the underlying parametric, linearity, and/or proportional hazards assumptions. Recently, machine learning and deep learning models have been developed for CRA [2, 5, 6, 11, 15]. These models, designed to overcome the drawbacks of statistical approaches, can capture nonlinear relationships between covariates and the risk of an event. Among all these models, *DeepHit* [15] model - a deep learning approach which makes no underlying assumption on the stochastic process - has shown state-of-the-art performance for CRA. However, it consists of a very sophisticated network and relies on a specialized objective function to handle censoring - which is inherent in survival data. Moreover, it uses a large model (i.e., a large number of parameters) to achieve good predictive performance for CRA, and as a result, does not provide easily explainable results. To address these drawbacks, in this paper, we formulate CRA as a cause-specific regression analysis problem and propose a simple and effective deep feed-forward neural network based model called **DeepPseudo**, to estimate cumulative incidence function (CIF) [18] using *Aalen-Johansen* estimator based pseudo values [14]. DeepPseudo models the non-linear time-varying covariate effect on cumulative incidence function and handles the complexity of censored data by using pseudo values. We show that a *small* DeepPseudo model obtains similar or better results compared to existing models and it's predictions can be explained by use of explanation methods such as LRP [17].

## 2 CRA USING PSEUDO VALUES

A survival dataset with $K$ competing risks is a collection of time-to-event information of the patients along with their corresponding event status during a follow up period. For an individual $i$, competing risk data is a tuple $(T_i, \delta_i, X_i)$, where $i = 1, \ldots, N$. $T_i$ is the survival time and $X_i = (X_{i1}, ..., X_{ip})$ is a $p$ dimensional vector of observed covariates for $i^{th}$ individual. $\delta_i$ is the event indicator, where

$$\delta_i = \begin{cases} j & \text{, if the } i^{th} \text{ individual is uncensored and event} \\ & \text{occurred due to cause } j;\ j = 1, 2, ..., K \\ 0 & \text{, if the } i^{th} \text{ individual is censored} \end{cases}$$

Most of the CRA methods [19] model the effects of covariates on the CRA outcomes through the cause-specific hazard function or sub-distribution hazard function. The independent censoring assumption for the competing risks used in cause-specific hazard models does not hold in general, and hence, researchers have used cumulative incidence function (CIF) [14] for CRA. There is a direct relationship between cause-specific hazard rate and CIF when there is only one event of interest. In the presence of competing risks, there is no such direct relationship as the CIF depends on the crude hazard rate of all the causes of the event. Therefore, directly modeling the effect of covariates on the CIF is needed. *Fine et. al.* [9] introduced a proportional hazard model to achieve this by using a sub-distribution hazard function. An alternative regression approach based on **Pseudo values** was proposed by Klein et. al [14] for directly modeling the effect of covariates on CIF. Pseudo values are calculated for both censored and uncensored subjects for all causes of an event at a specified time point. For the $i^{th}$ subject, a Jackknife pseudo value, based on the Aalen–Johansen estimate of the CIF [13], is computed for cause $k$ at time horizon $t^*$ as

$$\hat{F}_{ik}(t^*) = n\hat{F}_k(t^*) - (n-1)\hat{F}_k^{-i}(t^*)$$

where, $\hat{F}_k(t^*)$ is the Aalen–Johansen estimate of the CIF for cause $k$ based on a sample with $n$ subjects and $\hat{F}_k^{-i}(t^*)$ is the Aalen–Johansen estimate of the CIF for cause $k$ based on leave-one-out sample with $(n-1)$ subjects, obtained by omitting the $i^{th}$ subject. Pseudo values are good at handling right censored data, and thus, it has been extensively studied by other researchers [10, 16, 20].

## 3 OUR PROPOSED MODEL - DEEPPSEUDO

Pseudo values based approaches [7] have shown promising results for CRA but deep learning based models such as **DeepHit** [15] have achieved superior results as the deep models can capture non-linear relationships among covariates and the competing risks while making limited assumptions. However, DeepHit relies on a specialized objective function to handle censoring. To address this issue, we propose a deep learning model called **DeepPseudo**, which uses a simple deep feed-forward neural network to perform regression analysis based on pseudo values for CRA. We propose four variants of the DeepPseudo model based on how the pseudo values are calculated and how cause-specific events are modeled.

*Marginal DeepPseudo Model:* We feed the covariates as input and treat the pseudo values for the marginal CIF as the output of our deep feed-forward neural network. The outputs of this model are the cause-specific prediction of pseudo values at $M$ evaluation time points $(\tau_1, \tau_2, ...., \tau_M)$.

*Conditional DeepPseudo Model:* We first divide the discrete follow-up time into $M$ intervals and calculate the cause-specific pseudo values for the conditional CIFs for $M$ intervals. We consider the intervals and causes as categorical covariates and convert them into dummy variables. In the initial interval, all the patients are considered. However, in the next intervals, some patients are not considered due to failure or censoring. Therefore, the patients might have different numbers of observations in the training data. We predict the pseudo values for all of the causes at each of the intervals. The model's output layer has a single node (neuron), which predicts the pseudo value for the CIF for a cause at a particular time point.

*Cause-specific (CS) Marginal DeepPseudo Model:* We extend the Marginal DeepPseudo Model with cause-specific sub-networks to predict the pseudo values for each cause separately. It also has a shared sub-network to learn the shared representation of the competing events. The cause-specific sub-networks take the output of the shared network as input and predict the pseudo values for the specific causes at $M$ evaluation time points.

*Cause-specific (CS) Conditional DeepPseudo Model:* This model uses cause-specific sub-networks to predict the pseudo values for each cause separately. The input of this model is the covariates, along with a mask variable for evaluation times. The output of each cause-specific network is a single node, which is the predicted pseudo value for the causes at a particular time point.

The pseudo values can be less than 0 and greater than 1 in the presence of censoring and, thus, not interpretable. Therefore, we transform the predicted pseudo values to [0, 1] range by using the clipping formula: *Transformed Pseudo values* = $min(1, max(0,$ *Predicted pseudo values*$))$. We trained all the above models by minimizing the mean squared error loss function, which minimizes the squared differences between the true pseudo values and predicted pseudo values.

## 4 EXPERIMENTS

We evaluate the performance of our model based on the cause-specific time-dependent concordance index (C-index) by performing a set of experiments on two real-world and one synthetic dataset. We compare the performance of our proposed models with many baseline and state-of-the-art CRA models. We conduct our experiments on different censoring settings to evaluate our model's performance in handling right censoring in the survival data.

**Datasets**

*SEER:* The Surveillance, Epidemiology, and End Results (SEER) Program provides information on cancer statistics to reduce the cancer burden among the United States. We extracted a cohort of 28366 patients out of which 23.2% died due to cervical cancer (cause 1), 8.4% died of other causes (Cause 2), and 68.4% patients are right-censored. We considered 13 features/covariates, including age at diagnosis, race, marital status, histology record, Grade, tumor size, cancer stages (TNM staging system), surgery record, cancer therapies, histology etc., for our analysis.

*WIHS:* We selected a cohort of 1164 women enrolled in WIHS [4] study who were alive, infected with HIV, and free of clinical AIDS during the study period December 1995- September 2006. The dataset contains two competing risks (Highly active antiretroviral therapy (HAART) initiation (Cause 1) & AIDS/Death before HAART (Cause 2) as well as right censoring. The dataset included

**Table 1: *Model Performance Comparisons using time dependent cause-specific C-index (mean and* 95% *confidence interval)***

| Dataset | Cause of the Event | Evaluation Time | Statistical Models | | | Machine Learning Models | | Deep Learning Models | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Cause-specific Hazard | Fine & Gray | GEE (Pseudo) | RSF | DMGP | Deephit | Marginal DeepPseudo | CS-Marginal DeepPseudo | Conditional DeepPseudo | CS-Conditional DeepPseudo |
| SEER | Cause 1 | 1 year | 0.8649 *** (0.8620, 0.8678) | 0.8625 *** (0.8594, 0.8655) | 0.8675 *** (0.8646, 0.8704) | 0.8677 *** (0.8652, 0.8702) | 0.8713 (0.8683, 0.8743) | 0.8761 (0.8726, 0.8796) | 0.8767 (0.8736, 0.8798) | **0.8773 (0.8743, 0.8802)** | 0.8659 (0.8627, 0.8691) | 0.8647 (0.8612, 0.8681) |
| | | 5 years | 0.7962 *** (0.7939, 0.7986) | 0.7984 *** (0.7960, 0.8009) | 0.8038 *** (0.8016, 0.8061) | 0.7973 *** (0.7947, 0.7999) | 0.8030 *** (0.7998, 0.8062) | 0.8080 (0.8051, 0.8109) | 0.8122 (0.8095, 0.8148) | **0.8130 (0.8106, 0.8155)** | 0.8055 (0.8032, 0.8077) | 0.8027 (0.7997, 0.8058) |
| | Cause 2 | 1 year | 0.8291 * (0.8180, 0.8402) | 0.7756 *** (0.7641, 0.7871) | 0.8005 *** (0.7898, 0.8111) | 0.8045 *** (0.7962, 0.8128) | 0.7634 *** (0.7502, 0.7767) | 0.8458 (0.8351, 0.8565) | **0.8520 (0.8431, 0.8608)** | 0.8412 (0.8320, 0.8503) | 0.8376 (0.8285, 0.8466) | 0.8451 (0.8357, 0.8545) |
| | | 5 years | 0.7871 *** (0.7812, 0.7931) | 0.7751 *** (0.7689, 0.7813) | 0.7854 *** (0.7788, 0.7921) | 0.7618 *** (0.7560, 0.7677) | 0.7684 *** (0.7633, 0.7734) | 0.8028 (0.7969, 0.8087) | 0.8077 (0.8016, 0.8138) | 0.8089 (0.8028, 0.8150) | 0.7991 (0.7907, 0.8074) | **0.8138 (0.8078, 0.8197)** |
| WIHS | Cause 1 | 1 year | 0.7225 (0.7021, 0.7429) | 0.6918 . (0.6700, 0.7135) | 0.7041 (0.6818, 0.7264) | 0.6947 (0.6738, 0.7157) | 0.7222 *** (0.7023, 0.7421) | 0.7207 (0.6986, 0.7428) | 0.7310 (0.7111, 0.7509) | **0.7342 (0.7120, 0.7565)** | 0.7318 (0.7128, 0.7509) | 0.7225 (0.7003, 0.7446) |
| | | 5 years | 0.6248 ** (0.6139, 0.6358) | 0.6193 *** (0.6076, 0.6310) | 0.6432 (0.6308, 0.6556) | 0.6073 *** (0.5967, 0.6179) | 0.6257 (0.6130, 0.6385) | 0.6076 *** (0.5931, 0.6222) | 0.6189 (0.6064, 0.6315) | 0.6156 (0.6040, 6272) | **0.6536 (0.6436, 0.6636)** | 0.6381 (0.6257, 0.6505) |
| | Cause 2 | 1 year | 0.6638 . (0.6471, 0.6805) | 0.6465 *** (0.6295, 0.6636) | 0.6717 (0.6526, 0.6908) | 0.6869 (0.6704, 0.7034) | 0.6805 * (0.6663, 0.6947) | 0.6802 (0.6623, 0.6982) | 0.7015 (0.6855, 0.7175) | **0.7019 (0.6858, 0.7179)** | 0.6982 (0.6820, 0.7143) | 0.7004 (0.6869, 0.7140) |
| | | 5 years | 0.6295 *** (0.6170, 0.6420) | 0.6336 *** (0.6208, 0.6464) | 0.6609 . (0.6480, 0.6738) | 0.6518 ** (0.6401, 0.6636) | 0.6761 (0.6678, 0.6845) | 0.6556 * (0.6437, 0.6674) | 0.6707 (0.6589, 0.6826) | 0.6681 (0.6552, 0.6811) | **0.6835 (0.6716, 0.6953)** | 0.6653 (0.6532, 0.6775) |
| Synthetic | Cause 1 | 1 year | 0.5923 *** (0.5882, 0.5963) | 0.5931 *** (0.5891, 0.5971) | 0.5811 *** (0.5772, 0.5850) | 0.6293 *** (0.6247, 0.6338) | 0.7508 ** (0.7479, 0.7537) | 0.7532 . (0.7500, 0.7564) | 0.7554 (0.7526, 0.7581) | 0.7490 (0.7461, 0.7519) | **0.7606 (0.7579, 0.7633)** | 0.7529 (0.7501, 0.7556) |
| | | 5 years | 0.5784 *** (0.5752, 0.5815) | 0.5785 *** (0.5754, 0.5817) | 0.5576 *** (0.5549, 0.5602) | 0.5849 *** (0.5813, 0.5886) | 0.6761 *** (0.6726, 0.6796) | 0.6824 *** (0.6784, 0.6865) | 0.6707 (0.6678, 0.6736) | 0.6805 (0.6774, 0.6836) | **0.7028 (0.7000, 0.7056)** | 0.6903 (0.6872, 0.6933) |
| | Cause 2 | 1 year | 0.5954 *** (0.5920, 0.5988) | 0.5973 *** (0.5941, 0.6006) | 0.5849 *** (0.5817, 0.5882) | 0.6273 *** (0.6238, 0.6308) | 0.7483 *** (0.7442, 0.7525) | 0.7516 * (0.7477, 0.7555) | 0.7543 (0.7513, 0.7573) | 0.7487 (0.7454, 0.7520) | **0.7598 (0.7566, 0.7631)** | 0.7571 (0.7548, 0.7594) |
| | | 5 years | 0.5809 *** (0.5774, 0.5844) | 0.5810 *** (0.5775, 0.5846) | 0.5587 *** (0.5549, 0.5626) | 0.5841 *** (0.5806, 0.5876) | 0.6736 *** (0.6700, 0.6772) | 0.6788 *** (0.6730, 0.6846) | 0.6638 (0.6604, 0.6672) | 0.6746 (0.6719, 0.6773) | **0.6989 (0.6959, 0.7018)** | 0.6895 (0.6867, 0.6922) |

Tukey's HSD test - statistically significant codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1, (Read 0 '***' as significant at 0% level of significance)

the following features; the history of injection drug use at WIHS enrollment, race, age, and CD4 nadir prior to baseline.

*Synthetic data:* We generated a synthetic dataset as constructed in [15] with two competing risks. We generated the hitting times from exponentially distribution with a mean parameter depending on both linear and non-linear (quadratic) function. The dataset consists of 12 features that follow standard normal distribution. We generate 30000 observations, out of which 15000 observations are right censored. We use this dataset to examine the performance of the models in the presence of non-linearity in the dataset.

**Models Compared**: We evaluate the following models:

- *Statistical Models:* Cause-specific Hazard Model [19], Fine and Gray Model [9], GEE (Pseudo values) [14]
- *Machine Learning Models:* Random Survival Forests (RSF) [11], Deep Multi-task Gaussian Processes (DMGP) [2]
- *Deep Learning Models:* DeepHit [15], Our proposed models: Marginal DeepPseudo, CS Marginal DeepPseudo, Conditional DeepPseudo, CS Conditional DeepPseudo models

**Performance Metrics:** In this paper, we use cause-specific time-dependent concordance index [3] for evaluating the discriminatory ability as well as the predictive accuracy of the models, which take the time-dependency of the risks into consideration. In our experiments, we compute the time-dependent concordance index [3] for GEE (Pseudo values), DMGP model, Deephit model, and DeepPseudo models. For the cause-specific hazard model, Fine & Gray model, and Random survival forest model, we use the 'cindex' function of R package 'pec' for computing the C-index.

**Implementation:** We created 5 sets of 5-Fold cross-validation dataset for our experiments. We maintain a constant ratio of uncensored and censored individuals in each fold. We convert the categorical variables into one-hot-encoded dummy variables. We choose the best hyperparameter setting based on the average C-index as the performance metric on the validation dataset by varying the following hyperparameters: number of hidden layers [2, 3, 4, 5, 6, 7, 8], number of nodes [32, 64, 128], dropout [0.1, 0.2, 0.3, 0.4] or l2 regularization [0.01 0.001, 0.0001], 'selu' activation function, batch size 64 and ['SGD', 'Adam'] optimizer with a learning rate of [0.01, 0.001, 0.0001, 0.00001]. During hyperparameter tuning, we perform early stopping and choose the best model based on the minimum validation loss. We consider two competing risks in all the experiments, and calculate the cause-specific time-dependent concordance index at two evaluation times; 1 year and 5 years, which is of interest to most clinicians. We also perform pairwise statistical significant test (Tukey's HSD test) between the best DeepPseudo model and other baseline models.

## 5 RESULTS AND DISCUSSION

The model comparison results are shown in Table 1. For the SEER dataset, our DeepPseudo models showed statistically significant performance over all the other models except the DeepHit model in almost all the cases. Our DeepPseudo models perform similar or better than the DeepHit model in most cases. On the WIHS dataset, our DeepPseudo models give significantly better performance than all the other baseline models, especially for 5 years of evaluation time. On the Synthetic dataset, our best DeepPseudo model (i.e., Conditional DeepPseudo model) showed a very promising improvement over all the other benchmarks, and the improvement is statistically significant. It is clear that the statistical models showed the worst performance on Synthetic data as these models are limited by the linearity assumptions between covariates and risks, whereas the synthetic data was generated considering the non-linear relationship. Our model and other deep models capture both the linear and non-linear relationships present in the dataset and thus perform much better. An interesting finding is that our DeepPseudo models obtain better results over the statistical GEE approach, which also uses pseudo values for CRA. Table 2 shows that our model handles the censoring in the survival data better than all the models in almost all of the different censoring settings. It is worth noting that our DeepPseudo models use around 50 thousand parameters to obtain similar or better results than the DeepHit model, which uses >1 Million parameters.

**Explaining Our Model Predictions**: Even though deep learning models provide accurate results, they are black-box models.

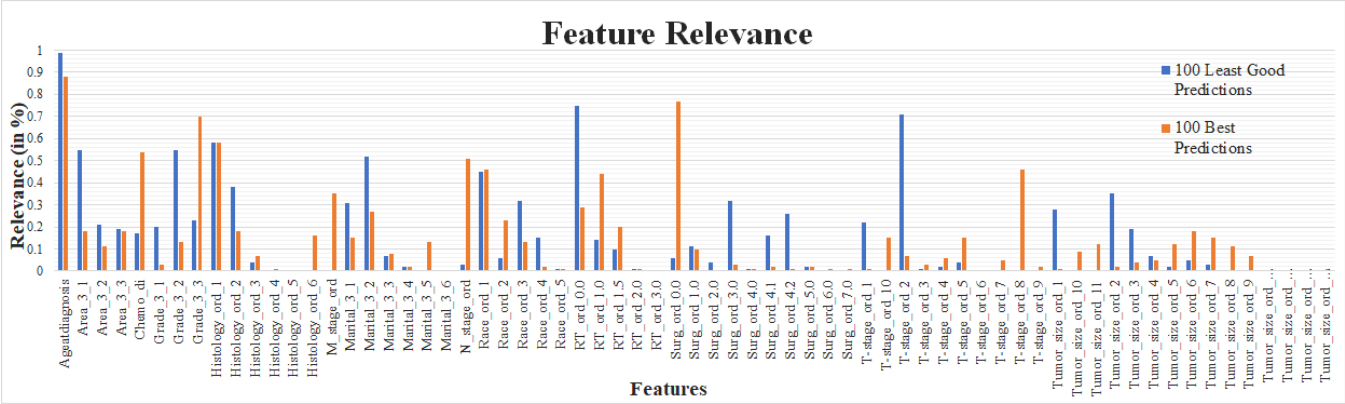


**Figure 1: Explaining DeepPseudo Predictions using feature relevance plot using LRP evaluated on SEER training data**

Thus, it becomes challenging to use them in medical and other critical applications if their predictions cannot be understood. To address this issue, we employ Layer-wise Relevance Propagation [17] approach to explain the predictions of our proposed DeepPseudo model, i.e., we explain the covariates' contribution to the prediction. We calculate the relevance score of all the features based on the 100 best predictions and the 100 worse predictions as measured by the training mean squared errors. Figure 1 shows the feature relevance distribution for these 200 predictions. In this figure, it is evident that our model can identify the important features for good and bad predictions. For instance, for the best predictions, features such as large tumor size and surgery, are chosen as important covariates for prediction, which are correct as they are highly indicative of survival risk for the patient.

## 6 CONCLUSION

This paper formulates competing risks analysis as a pseudo value based regression problem and proposes simple deep feed-forward neural network based models, referred to as DeepPseudo models, to predict the pseudo values as a substitute for the cumulative incidence function. Our proposed models do not use any special cost functions or make any strong assumptions about the relationship between the covariates and risks. Our model achieves similar or better performance than the existing CRA approaches and is apt at handling censoring. In addition, our model allows the use of off-the-self explanation approaches to provide explanations to its predictions. For future work, we will work on theoretical guarantees, and conduct extensive experiments on CRA datasets with multiple causes.

**Table 2: Model comparisons on SEER data for different censoring settings (Cen) at 1 year evaluation time**

| Cause of the Event | Algorithms | No Cen | 1k Cen | 2k Cen | 3k Cen | 4k Cen | 5k Cen |
|---|---|---|---|---|---|---|---|
| Cause 1 | Fine & Gray | 0.7160 | 0.7186 | 0.7179 | 0.7139 | 0.7073 | 0.6929 |
| | RSF | 0.7664 | 0.7662 | 0.7608 | 0.7481 | 0.7356 | 0.7199 |
| | Deephit | 0.7735 | 0.7717 | 0.7646 | 0.7529 | 0.7399 | 0.7247 |
| | DeepPseudo | **0.7738** | **0.7737** | **0.7664** | **0.7575** | **0.7442** | **0.7250** |
| Cause 2 | Fine & Gray | 0.6103 | 0.6161 | 0.6180 | 0.6170 | 0.6381 | 0.6259 |
| | RSF | 0.6826 | 0.6838 | 0.6706 | 0.6706 | 0.6566 | 0.6097 |
| | Deephit | 0.7244 | 0.7140 | 0.7246 | 0.7117 | **0.7232** | 0.6563 |
| | DeepPseudo | **0.7417** | **0.7541** | **0.7415** | **0.7432** | 0.7192 | **0.6996** |

## ACKNOWLEDGMENTS

This work is supported by grant CRII (IIS–1948399) from the National Science Foundation.

## REFERENCES

[1] Odd O Aalen and Søren Johansen. 1978. An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics* (1978).
[2] Ahmed M Alaa and Mihaela van der Schaar. 2017. Deep multi-task gaussian processes for survival analysis with competing risks. In *NeurIPS*.
[3] Laura Antolini, Patrizia Boracchi, and Elia Biganzoli. 2005. A time-dependent discrimination index for survival data. *Statistics in medicine* (2005).
[4] Melanie C Bacon, Von Wyl, et al. 2005. The Women's Interagency HIV Study: an observational cohort brings clinical sciences to the bench. *Clin. Diagn. Lab. Immunol.* (2005).
[5] Alexis Bellot and Mihaela van der Schaar. 2018. Multitask boosting for survival analysis with competing risks. In *NeurIPS*.
[6] Alexis Bellot and Mihaela van der Schaar. 2018. Tree-based bayesian mixture model for competing risks. In *AISTATS*.
[7] Nadine Binder, Thomas A Gerds, and Per Kragh Andersen. 2014. Pseudo-observations for competing risks with covariate dependent censoring. *Lifetime data analysis* (2014).
[8] David R Cox. 1972. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)* (1972).
[9] Jason P Fine and Robert J Gray. 1999. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American statistical association* (1999).
[10] Frederik Graw, Thomas A Gerds, and Martin Schumacher. 2009. On pseudo-values for regression analysis in competing risks models. *Lifetime Data Analysis* (2009).
[11] Hemant Ishwaran, Thomas A Gerds, Udaya B Kogalur, Richard D Moore, Stephen J Gange, and Bryan M Lau. 2014. Random survival forests for competing risks. *Biostatistics* (2014).
[12] Edward L Kaplan and Paul Meier. 1958. Nonparametric estimation from incomplete observations. *Journal of the American statistical association* (1958).
[13] John P Klein et al. 2008. SAS and R functions to compute pseudo-values for censored data regression. *Computer methods and programs in biomedicine* (2008).
[14] John P Klein and Per Kragh Andersen. 2005. Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics* (2005).
[15] Changhee Lee et al. 2018. Deephit: A deep learning approach to survival analysis with competing risks. In *AAAI*.
[16] Ulla B Mogensen and Thomas A Gerds. 2013. A random forest approach for competing risks based on pseudo-values. *Statistics in medicine* (2013).
[17] Grégoire Montavon et al. 2019. Layer-wise relevance propagation: an overview. In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*.
[18] Margaret Sullivan Pepe and Motomi Mori. 1993. Kaplan—meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Statistics in medicine* (1993).
[19] Hein Putter, Marta Fiocco, and Ronald B Geskus. 2007. Tutorial in biostatistics: competing risks and multi-state models. *Statistics in medicine* (2007).
[20] Lili Zhao and Dai Feng. 2019. DNNSurv: Deep Neural Networks for Survival Analysis Using Pseudo Values. *arXiv preprint arXiv:1908.02337* (2019).