# Data Reconstruction Method for Learning to Censored Survival Data

Sang-Kyun Ko [1,2] and Yung-Kyun Noh [1,3]

[1] Department of Computer Science, Hanyang University, Seoul 04763, Republic of Korea
[2] Ainex Co., Ltd., Seoul 06779, Korea
[3] School of Computational Sciences, Korea Institute for Advanced Study, Seoul 02455, Republic of Korea

highsg19101@gmail.com, nohyung@hanyang.ac.kr

## Abstract

Because of the incomplete target values of right-censored data, classical machine-learning approach is limited in employing censored training data; The censored data only indicates that survival time exceeds a specified threshold. However, this threshold can still be used to compare survival times with uncensored data. It enables the utilization of censored data for evaluating the concordance index (c-index), a metric for assessing the order of survival times. This study focuses on the use of right-censored data in the c-index. We propose a novel method for training incomplete data by reconstructing data into pairs used in the c-index. The strategy treats paired data as matrices, allowing the classifier to classify row data with higher survival time. Our approach can directly optimize the c-index via intuitive data reconstruction and by using right-censored data to estimate the ordering of survival time. We evaluated quantitative performance using three real-world data to demonstrate the efficiency of data reconstruction strategies. The proposed method demonstrated competitive performance with partial likelihood, rank methods, and Wasserstein metric, which have been proven effective in survival analysis and show no statistically significant difference (t-test, p-value$>$0.05). Furthermore, to explore the impact of using right-censored, we report performance improvement by increasing the ratio of censored in the training set.

***Keywords***— *Survival Analysis, Machine Learning*

## I. INTRODUCTION

Survival analysis, or time-to-event analysis, is a statistical methodology employed to examine the duration, called survival time $T$, until the incidence of a specific event within a population, considering covariates $\mathbf{x}$ that can impact event occurrence. This analysis is commonly used in medical statistics to estimate the onset of a patient's disease and survival rates over time. In a business environment, it predicts machine defects or customer departures.

Constructing data samples in survival analysis requires an observation period that follows objects of interest and records when events arise. The objects can only provide meaningful data when an event occurs during the finite observation period. However, in survival analysis, it is common for events not to occur during data collection, leading to the termination of observations. Therefore, the data's survival time $T$ can often not be fixed to an exact value. This case, referred to as "right censored," can only confirm that this data's survival time exceeds a specified threshold [16]. Because right-censored data has incomplete target values, it cannot be employed as training data for standard regression methods in traditional statistical or machine learning methods. Furthermore, the survival data is relatively expensive because this analysis must have an observation period. Therefore, the problem of data missing owing to right-censored can be fatal in survival analysis, primarily when dealing with small-scale datasets.

However, treating the right-censored data as missing is also inappropriate. The censored data can be used as test data in the commonly employed evaluation metric for survival analysis, known as the concordance index (c-index). The c-index evaluates how concordance of the forecasted order of predictive model $f$ and actual survival time values order of two data (can compare the survival time), i.e., if the survival time $T$ of the paired data $(\mathbf{x}_i, \mathbf{x}_j)$ is $T_i < T_j$ and predictive model $f$ for both data is $f(\mathbf{x}_i) < f(\mathbf{x}_j)$, thus can get the score cause predicted ordering is the concordance of ground truth. We know that the right-censored data survived to at least the threshold, therefore, survival time comparison is feasible if the survival time value of the intact data is below that of the censored data. This study focuses on these c-index characteristics, which use censored data for comparison. We propose novel methods that reconstruct the data into pairs as if they are computing the c-index and include right-censored in the training set.

Continuous research has been conducted to utilize

right-censored data for analysis while optimizing the c-index. Cox proposed a proportional hazard model (Cox PH model) [7] in 1972, considering censored data. This method is still frequently used as a survival prediction model as it can process certain right-censored data using partial likelihood. Subsequently, a study by Rakar [18] revealed that Cox's partial likelihood method approximates the maximization of the c-index. In addition, it demonstrated that survival analysis is naturally treated as a ranking problem [4, 6]. This study used right-censored, optimizing the c-index by applying previously proposed ranking methods. Furthermore, a study by Luck and Sylvain [12, 19] proposed a novel approach to applying the Wasserstein distance. Unlike treating the problem definition of existing survival analysis as a regression problem, this study approaches right censored as a classification method. In addition, studies [20, 5, 1, 2] that directly optimize or explore the c-index have distinct approaches.

In this study, we propose a matrix c-index that reconstructs data into a matrix form and evaluates it for neural network training of right-censored data. First, to surrogate the proposed matrix c-index as the objective function of the neural network, we derivate the discrete matrix c-index into a differentiable c-index. Subsequently, we conduct redefining to a generally used error function form. The proposed method can use censored data and can directly optimize the evaluation indicators of the predictive model. This approach draws inspiration from direct optimization metrics in particle physics research [10] .

The rest of this paper is organized as follows. Section 2 explains the evaluation metric c-index using right-censored data. Section 3 discusses data reconstruction and introduces a surrogate objective function for neural networks. Finally, in Section 4, we present quantitative experimental results by comparing the data reconstruction methods with other state-of-the-art methods that have already been validated.

## II. Concordance index

This section discusses the process of pairing right-censored data into comparable two data, namely "acceptable pairs" [19] in the c-index, and their utilization as test data. The c-index proposed by Harrell [8] is a performance evaluator of the concordance between survival time and the forecasted order. All test data are paired into acceptable pairs that can be used to compare the survival times. Subsequently, it is evaluated whether the predictive model can correctly order the survival time of the two data.

First, we define the notation for the description of acceptable pairs; Survival data can be denoted by a $D$-dimensional random variable vector $\mathbf{x}$ that concerns the occurrence of an event, $\delta \in \{0,1\}$ that indicates whether an event occurs, and an event occurrence time $T \in \mathbb{R} > 0$. Where $\delta = 0$ indicates no event has occurred, and $T$ of the right-censored data can represent the survival time value that last observed to survive. Finally, $N$ survival data can be denoted as $\mathscr{D} = \{\mathbf{x}_i, \delta_i, T_i\}_{i=1}^{N}$.

Second, we present conditions for acceptable pairs that allow survival time comparisons. For instance, in two tied data $(\mathbf{x}_i, \mathbf{x}_j), i, j = 1, ..., N$, the first data $\mathbf{x}_i$ is uncensored, and the second data $\mathbf{x}_j$ can be censored or uncensored. Accordingly, acceptable pairs $(\mathbf{x}_i, \mathbf{x}_j)$ are satisfied with the following conditions; 1) if $\delta_i = 1$ and $\delta_j \in \{0,1\}$, $T_i < T_j$. 2). For $T_i = T_j$ with the same survival time values, both data must be uncensored.

The set of acceptable pairs is defined as $\mathscr{A}$ through the above conditions. $\mathscr{A}$ can be quickly assembled by sorting the data in ascending order based on survival time and bundling the data with small values. These forms can be represented as a graph in Figure 1. As the (a) graph shows, $\mathbf{x}$ with the shortest survival time can pair to all nodes that can arrive along the arrow. The five data in Figure 1 are listed as a collection of comparable pairs as follows.

- (a):$(\mathbf{x}_1, \mathbf{x}_2), (\mathbf{x}_1, \mathbf{x}_3), (\mathbf{x}_1, \mathbf{x}_4), (\mathbf{x}_1, \mathbf{x}_5), (\mathbf{x}_3, \mathbf{x}_4), (\mathbf{x}_3, \mathbf{x}_5)$.
- (b):$(\mathbf{x}_2, \mathbf{x}_3), (\mathbf{x}_2, \mathbf{x}_4), (\mathbf{x}_2, \mathbf{x}_5), (\mathbf{x}_3, \mathbf{x}_4), (\mathbf{x}_3, \mathbf{x}_5)$.
- (c):$(\mathbf{x}_1, \mathbf{x}_2), (\mathbf{x}_1, \mathbf{x}_3), (\mathbf{x}_1, \mathbf{x}_4), (\mathbf{x}_1, \mathbf{x}_5), (\mathbf{x}_2, \mathbf{x}_3), (\mathbf{x}_2, \mathbf{x}_4), (\mathbf{x}_2, \mathbf{x}_5)$.

The c-index of the survival time predictive model $f$ for the set $\mathscr{A}$ is defined as follows [19, 8, 18]:

$$c(\mathbf{x}_i, T_i, \delta_i) = \frac{1}{|\mathscr{A}|} \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \mathscr{A}} \mathbf{1}_{f(\mathbf{x}_i), f(\mathbf{x}_j)}. \tag{1}$$

Where $|\mathscr{A}|$ is the number of elements in the set $\mathscr{A}$, and the indicator function $\mathbf{1}$ is defined as follows by receiving the results $f(x_i), f(x_j)$ of the predictive model:

$$\mathbf{1}_{f(\mathbf{x}_i), f(\mathbf{x}_j)} = \begin{cases} 1, & \text{if } T_i < T_j \text{ and } f(\mathbf{x}_i) < f(\mathbf{x}_j) \\ 0.5, & \text{if } T_i = T_j \text{ and } f(\mathbf{x}_i) = f(\mathbf{x}_j) \\ 0, & \text{otherwise} \end{cases}. \tag{2}$$

The indicator function $\mathbf{1}$ returns one when the predictive model $f$ concordance the ordering, and zero otherwise. However, if the survival time is the same, there is no basis for determining the order in the c-index that evaluates the order. Thus, pairs with the same survival time are treated randomly, and the indicator returns a value of 0.5.

## III. Data Reconstruction Method

### A. Data Reconstruction to Matrix

To achieve acceptable pair training, we need a reconstruction of the entire sample data into pairs used in the c-index and subsequently reshape them in matrix form. By transforming the data structure, the problem is redefined. We aim to predict the data index with the highest survival time from a given data pair (matrix) via reconstruction. For instance, the $i, j$ pair data satisfying condition 1) can

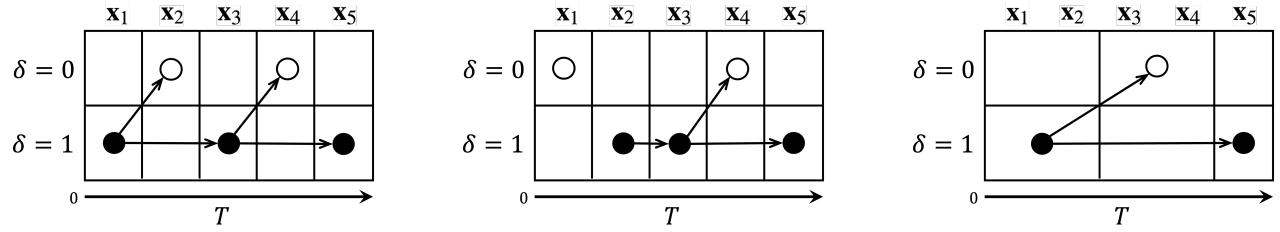


Fig. 1. The figure shows the order graph form of the acceptable pair over intact and right-censored data. The black-filled circles represent data points where the event of interest occurred, indicating the $\delta = 1$ column. The right-censored data signify an empty circle with the $\delta = 0$ column. The arrow represents whether they are comparable to manufacturing a test set for computation of the c-index. From the bottom left, nodes that can arrive along the arrow can be an acceptable pair. The arrow can only be directed to the right or in the upper right diagonal direction. The case of (c) symbolizes the case with the same survival time value.

be reconstructed into matrix data $M_{i,j}$ as follows.

$$M_{i,j} = \begin{pmatrix} \mathbf{x}_i \\ \mathbf{x}_j \end{pmatrix}_{i,j \in \mathscr{A}}, \; y_{i,j} = 1, \; i,j = 1,...,N. \tag{3}$$

We reconstruct the pair into a matrix $M_{i,j} \in \mathbb{R}^{2\times D}$ to solve the above-redefined problem. Consequently, predictive models can return two outputs $f(M_{i,j}) \in \mathbb{R}^{2\times 1}$ of predicting the row's index with the highest survival time. Based on the conditions defined above, the data with the largest survival time value in all acceptable pairs corresponds to the second row of data $\mathbf{x}_j$. Hence, the result of survival model $f : \mathbb{R}^{2\times D} \to [0,1]^2$, which processes defined data as shown in Eq. (3), should constantly be one in an ideal case. Thus, the biased learning of the predictive model is meaningless, and data pairs with the same survival time cannot be regarded. To address this, considering the redefined problem, we propose a data reconstruction approach that allows for the interchange of the order between $i$ and $j$ in pairs of data with the same survival time as follows ($i, j \in \mathscr{A}$ and $j, i \notin \mathscr{A}$):

$$M_n = \begin{pmatrix} \mathbf{x}_k \\ \frac{1}{2}(\mathbf{x}_i + \mathbf{x}_j) \\ \mathbf{x}_k \end{pmatrix}_{k \in \{i,j\}}, y_n \in \{0,1,2\}, \; n = 1,...,\mathbf{N}. \tag{4}$$

The reason $(j,i)$ pairs can be reconstructed as matrices is that we redefined the survival analysis problem into predicting the index of the data row with the highest survival time value. Furthermore, to predict the same survival time of the two data, multiplying the sum of the two random variables by 0.5 was adopted; however, as a row is arbitrarily inserted, it is not standard and permits applied differently. The target value $y_n \in \{0,1,2\}$ is an index of the vector data with the highest survival time in a matrix, and $\mathbf{N}$ is the number of matrices. Finally, the $D$-dimensionality survival data $\mathscr{D}$ can be reconstructed as follows:

$$\mathscr{D}^M = \{M_n, y_n\}_{n=1}^{\mathbf{N}}. \tag{5}$$

The proposed approach includes the converted case of $(i,j)$, then the number of matrix data is $2|\mathscr{A}|$. The maximum number $\mathscr{A}_N$ of the set $\mathscr{A}$ of $N$ survival data $\mathscr{D}$ can be expressed by a recurrence relation as follows: where $|\mathscr{A}|_1 = 0, |\mathscr{A}|_2 = 1$.

$$|\mathscr{A}|_N = |\mathscr{A}|_{N-1} + (N-1). \tag{6}$$

Consequently, if all the data are acceptable pairs, the number of matrix data $\mathbf{N}$ reconstructed from $N$ data $\mathscr{D}$ be as $2|\mathscr{A}|_N$.

### B. *Classification Model*

Our time ordering model is not a commonly used regression model in survival analysis, but it is a classifier. To predict the target value $y \in \{0,1,2\}$ of the matrix data $M \in \mathbb{R}^{3\times D}$, the survival model is output as $f : \mathbb{R}^{3\times D} \to [0,1]_0, [0,1]_1, [0,1]_2$. We can train the $f$, which returns three result values, to maximize the output $f_y$ of $y$-th index. The $n$-th prediction result $\hat{y}_n$ of the model for the reconstructed matrix $M_n \in \mathbb{R}^{3\times D}$ is determined as follows:

$$\hat{y}_n = \underset{k\in\{0,1,2\}}{\arg\max} f_k(M_n), \quad n = 1,\ldots,\mathbf{N} \tag{7}$$

The classifier is the multilayer neural network, and the sum of the model's last layer output is summed to be $\sum_{k=0}^{2} f_k = 1$ using the softmax function [3].

### C. *Matrix Concordance Index*

The performance of $f$, which deals with reconstructed $M$, cannot be calculated directly with the c-index. Rather than evaluating performance by ordering the survival time value of two data points, we can evaluate it by identifying which data point has the greater survival time value. Then we can evaluate the c-index with the same meaning. Because in equations both (1) and (8), the predictive model gives scores when predicting values with larger survival times (evaluation purpose is the same). Hence, we propose the matrix c-index $c^M$ for the matrix data $M$.

$$c^M = \frac{1}{\mathbf{N}} \sum_{n=1}^{\mathbf{N}} \mathbf{1}_{\hat{y}_n, y_n}. \tag{8}$$

The indicator function is as follows:

$$\mathbf{1}(y_n,\hat{y}_n) = \begin{cases} 1, & \text{if } T_i \neq T_j \text{ and } y_n = \hat{y}_n \\ 0.5, & \text{if } T_i = T_j \text{ and } y_n = \hat{y}_n \\ 0, & \text{otherwise} \end{cases}. \quad (9)$$

This is similar to Eq. (2) and returns one when the hit of the case is an unequal survival time pair and returns 0 in the failure case. Likewise, it returns 0.5 when the hit is an equal survival time pair.

### D. Matrix Concordance Index Optimization

As previously introduced, directly maximizing the c-index as an objective function in survival analysis is an efficient way to use right-censored data naturally. Likewise, the objective of the approach proposed in this study is to directly optimize the matrix c-index as an objective function during neural network training. However, maximizing the gradient methods is challenging because matrix c-index $c^M$ is a discrete function undifferentiated. Therefore, we propose a surrogate objective function that appropriately relaxes the maximizable matrix c-index.

Suppose an ideal classifier with a $c^M$ of 1.0 for any dataset $\mathscr{D}^M = \{M_n, y_n\}_{n=1}^{\mathbf{N}}$ consisting of matrix data $M_n$ and its corresponding target value $y_n$. The optimal classifier always assigns a score of "one" to the $y_n$-th element $f_{y_n}(M_n)$ in the prediction result for test data $M_n$. That is, if the predictive model's $y_n$-th row value is a score, maximizing this value to 1 may be considered. Accordingly, we can maximize the continuous function $f_{y_n}(M_n), n = 1, ..., \mathbf{N}$ and approximate it to $c^M$ through the reconstruction data and gradient ascent method. We propose the $\tilde{c}^M$ that relaxation $c^M$ as follows:

$$\tilde{c}^M = \frac{1}{\mathbf{N}} \sum_{n=1}^{\mathbf{N}} f_{y_n}(M_n). \quad (10)$$

Eq. (9) can be optimized via the gradient descent method. However, to transform it into a function of error form generally used in machine learning, we propose a loss function

$$\mathscr{L} = \frac{1}{\mathbf{N}} \sum_{n=1}^{\mathbf{N}} (1 - f_{y_n}(M_n)), \quad (11)$$

which can be gradient descent. The surrogate loss $\mathscr{L}$ is feasible to optimize through the reconstructed data $\mathscr{D}^M$. Furthermore, we can utilize the entire right-censored data included in the acceptable pair for the training.

## IV. Experiments

In this section, we quantitatively evaluate the proposed method by comparing and assessing its performance against other methods with demonstrated efficient performance in the field. This study experimentally evaluated the performance by incorporating only the proposed method and loss function, based on an approach published in [12][1]; all experimental settings were configured to be the same, including data preprocessing, hyperparameter tuning, and neural network structure used. However, the performance reported in Table 2, which is the performance to obtain statistical significance, reports a total of 10 cross-validations, not the previous set of five cross-validations. For information about the loss functions and theoretical background of the other seven compared models, please refer to the comprehensive study conducted in [19].

### A. Survival Data

We used three public survival datasets SUPPORT2, AIDS3, and COLON DEATH. SUPPORT2 refers to the data collected during the 1990 SUPPORT study and is characterized by comprising the largest number of data points among the three datasets. The data dimension is the highest at 98, and the ratio of right-censored data is also the lowest among the three. AIDS3 is a dataset from Australia that represents the AIDS survival data. It is characterized by having the lowest covariate dimension and the highest proportion of right-censored observations. The COLON DEATH dataset consists of survival analysis data related to chemotherapy treatment for colorectal cancer and is the small dataset with the lowest number of observations among the three.

### B. Implementation Details

We tuned our hyperparameters using validation data and employed a grid search approach, setting a maximum of 1000 epochs with a patience of 20; learning rates [1e-4, 1e-3, 1e-2, 1e-1] and L2 regularization [14] coefficient ($\lambda$) [1e-6, 1e-5, 1e-4, 1e-3, 1e-2]. The neural network comprises three layers of 100 nodes with an ReLU activation [13] . Furthermore, we used an optimizer Adam [11] and normalization for mini-batch [9] and drop-out [17] of 0.5 implemented by [15] . However, the drop-out used for input of the first layer is 0.2. The seven comparable models assumed a batch size of 512 in the SUPPORT2 data, 128 in the AIDS3, and 64 in the COLON DEATH dataset. The proposed reconstruction method was trained using a mini-batch size of 262144 for the SUPPORT2 data, while for the AIDS3 data, it was set to 8192, and for the COLON DEATH data, it was set to 4096.

### C. Quantitative Evaluation

The evaluation of the classifier trained by the reconstruction method converts all the data into matrix form

[1] https://github.com/Museau/Survival_Pytorch_EMD

Table 1. Table 1 shows detailed information on data evaluating the performance of the model. The table's left side shows the number of sample observations, the right-censored ratio, and the dimensions of probability variables.

| Survival Data | $N$ of data | $N$(%) of censored | $N$ of features |
|---|---|---|---|
| SUPPORT2 | 9105 | 2904 (32.2) | 98 |
| AIDS3 | 3985 | 2223 (55.8) | 19 |
| COLON DEATH | 929 | 477 (51.3) | 48 |

Table 2. It is the average of the c-index performance of other methods and our approach, and the standard deviation of the 10-fold validation results is bracketed. The performance indicated in bold corresponds to achieving the best results in the respective datasets. The performance marked in bold at the bottom is the c-index average of all models. The Cox model and Cox Efron's model are in the partial likelihood method, and the sigmoid, log-sigmoid, support vector machine, and ranking boost models belong to the ranking model. Finally, the classification approach includes two methods Wasserstein metric model and the proposed approach.

| Loss Type | Variant | SUPPORT2 | AIDS3 | COLON DEATH |
|---|---|---|---|---|
| Partial likelihood | Cox | 0.8487(0.0146) | 0.5641(0.0233) | 0.6428(0.0232) |
| Partial likelihood | Cox Efron's | 0.8495(0.0135) | 0.5638(0.0269) | 0.6471(0.0324) |
| Ranking | Sigmoid ($\sigma(z)$) | 0.8545(0.0137) | **0.5722**(0.0222) | **0.6536**(0.0296) |
| Ranking | Log-sigmoid | 0.8538(0.0147) | 0.5703(0.0251) | 0.6496(0.0279) |
| Ranking | SVM $((z-1)_+)$ | 0.8517(0.0135) | 0.5604(0.027) | 0.6453(0.0232) |
| Ranking | Boost $(1-(\exp -z))$ | 0.8537(0.0146) | 0.5714(0.0195) | 0.6391(0.038) |
| Classification | WM ($l = 1.5$) | 0.8542(0.014) | 0.5596(0.0336) | 0.6475(0.0345) |
| Classification (Ours) | Data Reconstruction | **0.8548**(0.0135) | 0.5555(0.0300) | 0.6524(0.0247) |
| Mean of All Performance | - | **0.8526**(0.0022) | **0.5647**(0.0270) | **0.647**(0.004) |

and then evaluates the prediction result in Eq. (6) as a matrix c-index in Eq. (7). To assess other results, we calculate the original c-index in Eq. (1). All the test results for the ten-fold cross-validations can be verified, as shown in Table 2; Furthermore, the complete trial results are presented in supplementary Table 4,5, and 6. Supplementary Table 4 shows the t-test p-value of ten trials. Therefore, we can confirm that there is no statistically significant difference between the data reconstruction method and the seven state-of-the-art methods in survival analysis.

The experimental results demonstrate that the ranking method achieved the best performance in terms of ranking sigmoid on both datasets. Additionally, the classification method proposed in this paper exhibited the highest performance on the SUPPORT2 data. The predictive models generated through the proposed classification method only achieved superior performance on specific datasets. Further, the performance of other datasets other than AIDS3 data was above average and comparable to the highest-level methods.

### *D. The C-index Based on the Ratio of Right-Censored*

To quantitatively test the effect of including the right-censored data during training, the performance was reported by progressively increasing the ratio of the right-censored in the training set. We selected the most negligible size of the data, the COLON DEATH dataset. This experiment was tested by splitting the survival data into train, validation, and test sets and then changing the ratio of right-censored only for the train data. In addition, owing to the limitation of the large number of trials in this experiment, five cross-validations were conducted, unlike Table 1. Figure 2 shows the performance from 0% to 90% inclusion, which does not include any censored learning data.

Among the eight methods reported in Table 1, two partial likelihood methods and two ranking models, sigmoid with the best performance in COLON DEATH data and rank boost method with the worst performance, were compared. Additionally, both classification methods were selected and tested, including the proposed model. The experimental results demonstrate the adequacy of including right-censored data. For analyzing the COLON DEATH performance, we observed that the average performance of all models was 0.612 when the ratio was 0%. However, the performance improved to 0.629 when the ratio increased to 90%.

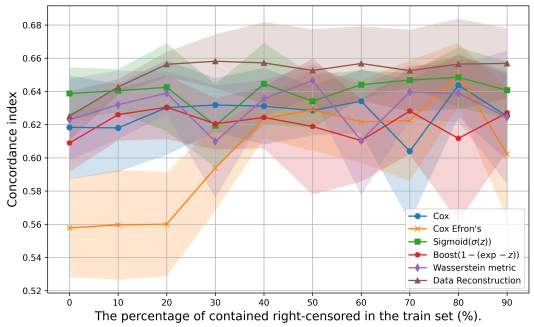


Fig. 2. The figure shows the c-index changes according to the ratio of right-center cut data. The right-censored ratio on the horizontal axis indicates that the training data include a ratio from 0% to 90%. In the figure, the range painted on the lines showing each performance is indicated in a color that matches the line with the error range (c-index-std, c-index+std) of the c-index based on each ratio.

## V. CONCLUSION

The right-censored data cannot be used as training data for a general machine learning approach as it comprises an incomplete target value. Dealing with numerous censoring proportions and limited survival data can be challenging and expensive. Consequently, this study proposes an unexplored data reconstruction method to use censored data for including the train set in survival analysis. Our strategy does not define the problem as a regression problem commonly approached in this task but casts it as a classification problem of the matrix data row. In the process, we present an approach to treat data with the same survival time value and present a matrix c-index that can evaluate reconstruction data. Furthermore, this methodology exhibits competitive performance with other proven studies, and the ten cross-validation results show no statistically significant difference. Furthermore, to demonstrate the effectiveness of employing censored data, we verified the effect by differing the proportion of censored data in the train set. Through this work, we could corroborate the usefulness of utilizing incomplete data.

Besides, we require to consider the growing data in this method. When using data reconstruction, training data increases exponentially. It is possible to process such a large number of data with hardware performance nowadays. However, if survival data becomes enormous in the future, the difference in train speed from other methods will significantly widen. We must consider these issues in further research. Conversely, another target for future studies is the effect of exponentially grown data on initially sparse datasets.

## REFERENCES


[1] Abdallah Alabdallah, Mattias Ohlsson, Sepideh Pashami, and Thorsteinn Rögnvaldsson. The concordance index decomposition–a measure for a deeper understanding of survival prediction models. *arXiv preprint arXiv:2203.00144*, 2022.

[2] Adam R Brentnall and Jack Cuzick. Use of the concordance index for predictors of censored survival data. *Statistical methods in medical research*, 27(8):2359–2373, 2018.

[3] John S Bridle. Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In *Neurocomputing: Algorithms, architectures and applications*, pages 227–236. Springer, 1990.

[4] Chris Burges, Tal Shaked, Erin Renshaw, Ari Lazier, Matt Deeds, Nicole Hamilton, and Greg Hullender. Learning to rank using gradient descent. In *Proceedings of the 22nd international conference on Machine learning*, pages 89–96, 2005.

[5] Yifei Chen, Zhenyu Jia, Dan Mercola, and Xiaohui Xie. A gradient boosting algorithm for survival analysis via direct optimization of concordance index. *Computational and mathematical methods in medicine*, 2013, 2013.

[6] William W Cohen, Robert E Schapire, and Yoram Singer. Learning to order things. *Advances in neural information processing systems*, 10, 1997.

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

[8] 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.

[9] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. pmlr, 2015.

[10] Cheongjae Jang, Sang-Kyun Ko, Jieun Choi, Jongwon Lim, Yung-Kyun Noh, and Tae Jeong Kim. Learning to increase matching efficiency in identifying additional b-jets in the $t\bar{t}b\bar{b}$ process. *The European Physical Journal Plus*, 137(7):870, 2022.

[11] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[12] Margaux Luck, Tristan Sylvain, Joseph Paul Cohen, Heloise Cardinal, Andrea Lodi, and Yoshua Bengio. Learning to rank for censored survival data. *arXiv preprint arXiv:1806.01984*, 2018.

[13] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010.

[14] Kamal Nigam, John Lafferty, and Andrew McCallum. Using maximum entropy for text classification. In *IJCAI-99 workshop on machine learning for information filtering*, volume 1, pages 61–67. Stockholm, Sweden, 1999.

[15] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

[16] G. Rodríguez. Chapter 7. survival analysis, lecture notes on generalized linear models.`https://grodri.github.io/glms/notes/`. *Princeton University*, 2007.

[17] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

[18] Harald Steck, Balaji Krishnapuram, Cary Dehing-Oberije, Philippe Lambin, and Vikas C Raykar. On ranking in survival analysis: Bounds on the concordance index. *Advances in neural information processing systems*, 20, 2007.

[19] Tristan Sylvain, Margaux Luck, Joseph Cohen, Heloise Cardinal, Andrea Lodi, and Yoshua Bengio. Exploring the wasserstein metric for time-to-event analysis. In *Survival Prediction-Algorithms, Challenges and Applications*, pages 194–206. PMLR, 2021.

[20] Lian Yan, David Verbel, and Olivier Saidi. Predicting prostate cancer recurrence via maximizing the concordance index. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 479–485, 2004.

## VI. SUPPLEMENTARY

Table 3. The p-value obtained from 10 trials t-tests compares the performance of the proposed methods with that of others.

| Loss Type | Variant | SUPPORT2 | AIDS3 | COLON DEATH |
|---|---|---|---|---|
| Partial likelihood | Cox | 0.3673 | 0.5038 | 0.4056 |
| Partial likelihood | Cox Efron's | 0.4163 | 0.544 | 0.6991 |
| Ranking (Sigmoid) | $\sigma(z)$ | 0.9618 | 0.7868 | 0.9252 |
| Ranking (Log Sigmoid) | Log-sigmoid | 0.8769 | 0.2697 | 0.8208 |
| Ranking (SVM) | $(z-1)_+$ | 0.6257 | 0.1989 | 0.5351 |
| Ranking (Boost) | $1-(\exp -z)$ | 0.8692 | 0.7237 | 0.3888 |
| Classification (WM) | $l=1.5$ | 0.9262 | 0.1973 | 0.7322 |

Table 4. Results of each method's c-index in the main text through 10-fold validation with SUPPORT2 data.

| Loss Type | Variant | Mean(std) | SUPPORT2(0) | SUPPORT2(1) | SUPPORT2(2) | SUPPORT2(3) | SUPPORT2(4) |
|---|---|---|---|---|---|---|---|
| Partial likelihood | Cox | 0.8487(0.0146) | 0.8287161 | 0.8440138 | 0.8328753 | 0.8520341 | 0.8710052 |
| Partial likelihood | Cox Efron's | 0.8495(0.0135) | 0.82799286 | 0.84428793 | 0.8347416 | 0.8530667 | 0.87039906 |
| Ranking (Sigmoid) | $\sigma(z)$ | 0.8545(0.0137) | 0.83587545 | 0.84677494 | 0.84109277 | 0.857242 | 0.87626034 |
| Ranking (Log Sigmoid) | Log-sigmoid | 0.8538(0.0147) | 0.8325637 | 0.8504772 | 0.8422552 | 0.85884833 | 0.8772993 |
| Ranking (SVM) | $(z-1)_+$ | 0.8517(0.0135) | 0.83005583 | 0.84796757 | 0.83980906 | 0.8537607 | 0.8731487 |
| Ranking (Boost) | $1-(\exp -z)$ | 0.8537(0.0146) | 0.8321345 | 0.84660256 | 0.8390593 | 0.8562793 | 0.877706 |
| Classification (WM) | $l=1.5$ | 0.8542(0.014) | 0.8337838 | 0.8462634 | 0.84035385 | 0.8563409 | 0.87598485 |
| Classification (Ours) | Data Reconstruction | 0.8548(0.0135) | 0.83657324 | 0.84710634 | 0.84217817 | 0.86095977 | 0.8768565 |

| Loss Type | Variant | SUPPORT2(5) | SUPPORT2(6) | SUPPORT2(7) | SUPPORT2(8) | SUPPORT2(9) |
|---|---|---|---|---|---|---|
| Partial likelihood | Cox | 0.8763474 | 0.8546999 | 0.84553784 | 0.83749115 | 0.84430283 |
| Partial likelihood | Cox Efron's | 0.87313443 | 0.85611075 | 0.8452227 | 0.84310246 | 0.84734935 |
| Ranking (Sigmoid) | $\sigma(z)$ | 0.8808763 | 0.8585521 | 0.85124063 | 0.8463339 | 0.851071 |
| Ranking (Log Sigmoid) | Log-sigmoid | 0.8813798 | 0.857956 | 0.84857583 | 0.8425179 | 0.8460878 |
| Ranking (SVM) | $(z-1)_+$ | 0.87567145 | 0.85911703 | 0.8459387 | 0.8467583 | 0.844566 |
| Ranking (Boost) | $1-(\exp -z)$ | 0.8801085 | 0.8592533 | 0.8511577 | 0.84420335 | 0.85082215 |
| Classification (WM) | $l=1.5$ | 0.880306 | 0.8610815 | 0.852233 | 0.84744966 | 0.8485365 |
| Classification (Ours) | Data Reconstruction | 0.8789487 | 0.85904855 | 0.8536769 | 0.843512 | 0.8495719 |

Table 5. Results of each method's c-index in the main text through 10-fold validation with AIDS3 data.

| Loss Type | Variant | Mean(std) | AIDS3(0) | AIDS3(1) | AIDS3(2) | AIDS3(3) | AIDS3(4) |
|---|---|---|---|---|---|---|---|
| Partial likelihood | Cox | 0.5641(0.0233) | 0.57449716 | 0.5328909 | 0.5648271 | 0.5451088 | 0.5554647 |
| Partial likelihood | Cox Efron's | 0.5638(0.0269) | 0.5830169 | 0.53866905 | 0.57303625 | 0.54474306 | 0.5572045 |
| Ranking (Sigmoid) | $\sigma(z)$ | 0.5722(0.0222) | 0.59559935 | 0.5064397 | 0.59222007 | 0.5453282 | 0.56854796 |
| Ranking (Log Sigmoid) | Log-sigmoid | 0.5703(0.0251) | 0.59581655 | 0.55053544 | 0.5851886 | 0.53593445 | 0.55657816 |
| Ranking (SVM) | $(z-1)_+$ | 0.5604(0.027) | 0.5961142 | 0.54938805 | 0.56620544 | 0.5558818 | 0.56019694 |
| Ranking (Boost) | $1-(\exp -z)$ | 0.5714(0.0195) | 0.5954626 | 0.55630326 | 0.5321998 | 0.54899585 | 0.57453287 |
| Classification (WM) | $l=1.5$ | 0.5596(0.0336) | 0.5947144 | 0.53653973 | 0.57938725 | 0.5606884 | 0.5631894 |
| Classification (Ours) | Data Reconstruction | 0.5555(0.0300) | 0.5946658 | 0.5458947 | 0.53825074 | 0.51921105 | 0.54685867 |

| Loss Type | Variant | AIDS3(5) | AIDS3(6) | AIDS3(7) | AIDS3(8) | AIDS3(9) |
|---|---|---|---|---|---|---|
| Partial likelihood | Cox | 0.56729966 | 0.59445345 | 0.58768004 | 0.5250362 | 0.59465575 |
| Partial likelihood | Cox Efron's | 0.55881894 | 0.6017751 | 0.59016097 | 0.5071766 | 0.58396727 |
| Ranking (Sigmoid) | $\sigma(z)$ | 0.5752081 | 0.5293146 | 0.5865182 | 0.50559866 | 0.59186 |
| Ranking (Log Sigmoid) | Log-sigmoid | 0.5686264 | 0.6024751 | 0.58051556 | 0.5290804 | 0.5992202 |
| Ranking (SVM) | $(z-1)_+$ | 0.5626951 | 0.6080188 | 0.5932107 | 0.55076206 | 0.5720997 |
| Ranking (Boost) | $1-(\exp -z)$ | 0.5409209 | 0.59085566 | 0.60440516 | 0.544953 | 0.5158806 |
| Classification (WM) | $l=1.5$ | 0.55759627 | 0.61517245 | 0.59336805 | 0.55186075 | 0.5697604 |
| Classification (Ours) | Data Reconstruction | 0.55926025 | 0.60442376 | 0.5776098 | 0.50353986 | 0.5656614 |



Table 6. Results of each method's c-index in the main text through 10-fold validation with COLON DEATH data.

| Loss Type | Variant | Mean(std) | COLON DEATH(0) | COLON DEATH(1) | COLON DEATH(2) | COLON DEATH(3) | COLON DEATH(4) |
|---|---|---|---|---|---|---|---|
| Partial likelihood | Cox | 0.6428(0.0232) | 0.6497384 | 0.64746815 | 0.65331584 | 0.6539242 | 0.64809656 |
| Partial likelihood | Cox Efron's | 0.6471(0.0324) | 0.6762081 | 0.64332247 | 0.64379513 | 0.65223277 | 0.6511916 |
| Ranking (Sigmoid) | $\sigma(z)$ | 0.6536(0.0296) | 0.65497077 | 0.64746815 | 0.6424819 | 0.652571 | 0.62952644 |
| Ranking (Log Sigmoid) | Log-sigmoid | 0.6496(0.0279) | 0.6605109 | 0.6294048 | 0.69139856 | 0.63058186 | 0.6570721 |
| Ranking (SVM) | $(z-1)_+$ | 0.6453(0.0232) | 0.67405355 | 0.6294048 | 0.6631648 | 0.6343031 | 0.6301455 |
| Ranking (Boost) | $1-(\exp -z)$ | 0.6391(0.038) | 0.6115728 | 0.6584246 | 0.619501 | 0.67320704 | 0.6103374 |
| Classification (WM) | $l=1.5$ | 0.6475(0.0345) | 0.6859034 | 0.6442108 | 0.632633 | 0.6197564 | 0.64035904 |
| Classification (Ours) | Data Reconstruction | 0.6524(0.0247) | 0.6542488 | 0.64129144 | 0.6593894 | 0.65324765 | 0.6394989 |

| Loss Type | Variant | COLON DEATH(5) | COLON DEATH(6) | COLON DEATH(7) | COLON DEATH(8) | COLON DEATH(9) |
|---|---|---|---|---|---|---|
| Partial likelihood | Cox | 0.6298488 | 0.6038186 | 0.6881759 | 0.6466974 | 0.6068966 |
| Partial likelihood | Cox Efron's | 0.62064433 | 0.62325263 | 0.6232877 | 0.72688174 | 0.61003137 |
| Ranking (Sigmoid) | $\sigma(z)$ | 0.6282051 | 0.62461644 | 0.7144917 | 0.70445466 | 0.6376175 |
| Ranking (Log Sigmoid) | Log-sigmoid | 0.62919134 | 0.6113195 | 0.6593367 | 0.70046085 | 0.6263323 |
| Ranking (SVM) | $(z-1)_+$ | 0.6045365 | 0.6283669 | 0.6777217 | 0.6697389 | 0.6413793 |
| Ranking (Boost) | $1-(\exp -z)$ | 0.6321499 | 0.62325263 | 0.59480894 | 0.7321045 | 0.6354232 |
| Classification (WM) | $l=1.5$ | 0.5963182 | 0.62052506 | 0.7101658 | 0.69155145 | 0.6335423 |
| Classification (Ours) | Data Reconstruction | 0.6263556 | 0.6157518 | 0.7 | 0.6898618 | 0.6445141 |

## SUMMARY OF THIS PAPER

### A. *Problem Setup*

Because right-censored data has incomplete target values, it cannot be employed as training data for standard regression methods in traditional statistical or machine learning methods. The censored data can be used as test data in the commonly employed evaluation metric for survival analysis, known as the concordance index (c-index). This study focuses on using right-censored data in the c-index. Then we propose a novel method for training incomplete data by reconstructing data into pairs used in the c-index.

### B. *Novelty*

The proposed approach, unlike treating the problem definition of existing survival analysis as a regression problem, this study approaches right censored as a classification method. Moreover, to the best of our knowledge, there is no way to directly optimize using the data form used in the c-index as it is. The strategy treats paired data as matrices, allowing the classifier to classify row data with higher survival time. Our approach can directly optimize the c-index via intuitive data reconstruction and by using right-censored data to estimate the ordering of survival time.

### C. *Algorithms*

First, to achieve acceptable pair training, we need a reconstruction of the entire sample dataset into pairs used in the c-index and subsequently reshape them in matrix form. Second, to surrogate the proposed matrix c-index as the objective function of the neural network, we derivate the discrete matrix c-index into a differentiable c-index. Finally, we conduct redefining to a generally used error function form.

### D. *Experiments*

We present quantitative experimental results by comparing the data reconstruction methods with other state-of-the-art methods that have already been validated from three open real-world data. The proposed method demonstrated competitive performance with methods such as partial likelihood, rank methods, and Wasserstein metric, which have been proven to be effective in survival analysis and show no statistically significant difference (t-test, p-value$>$0.05). Furthermore, to explore the impact of using right-censored, we report performance improvement by increasing the ratio of censored in the training set. Through this work, we could verify the usefulness of utilizing incomplete data.