

# Communication-Efficient Pseudo Value-Based Random Forests for Federated Survival Analysis

Md Mahmudur Rahman, Sanjay Purushotham

Department of Information Systems, University of Maryland, Baltimore County, Baltimore, Maryland, USA  
mrahman6@umbc.edu, psanjay@umbc.edu

## Abstract

Federated Survival Analysis (FSA) is an emerging technique for analyzing decentralized survival data while preserving data privacy and providing more generalized survival predictions. Existing FSA methods often rely on deep learning models, which can be computationally expensive and require substantial data and communication rounds. Recent research has demonstrated that ensemble-based approaches like random survival forests can achieve comparable performance to deep learning models with a single communication round within a federated learning (FL) framework, especially when dealing with small, decentralized datasets. However, these approaches have yet to address the challenges specific to FSA, such as data heterogeneity, non-uniform censoring, and competing risks. To address these challenges, we propose **FedPRF**, an FL framework for survival analysis based on Federated Pseudo Values (FPV) based random forests model, we call it **RFpseudo**. FedPRF introduces FPV to handle issues related to censoring and effectively address the unique challenges of FSA. FedPRF is computationally efficient, requiring only two communication rounds across clients: (i) computing FPV and (ii) aggregating a subset of trees trained locally at clients. Extensive experiments on distributed survival data with a single event and multiple competing events demonstrate that FedPRF achieves performance close to the gold-standard centralized training setting and outperforms deep learning-based FSA approaches. Importantly, FedPRF maintains the interpretability of centrally trained survival models. Furthermore, it is scalable to large-scale data and highly distributed settings with numerous clients.

## Introduction

Survival analysis is employed to analyze the time duration until a specific event occurs, such as death due to cancer, in order to predict the risk of the event over time. Survival data is often collected by a single, small-sized medical center with a small number of patients or limited resources. The potential of machine learning-based survival models can not be fully exploited while training is performed on limited data. The models result in sub-optimal and less confident survival predictions due to the small sample size. Moreover, data in a single medical center are often biased toward particular demographic groups, e.g., white Westerners, rather

than other groups, which leads to systematic biases and non-suitability of medical treatment to all demographic groups (Schork 2015; Rahman and Purushotham 2023b). The National Cancer Institute took a great initiative in conducting the Surveillance, Epidemiology, and End Results (SEER) program to collect large-scale cancer data from hospitals in different regions of the USA. However, conducting such programs by sharing data to a central server or cloud is expensive, time-consuming, and increases the risk of cyber-attacks (Hauschild et al. 2022). On the other hand, access to sufficient or large-scale data through multi-institute collaboration is hindered by strict data privacy laws and data-sharing regulations, such as GDPR and HIPAA. Federated Survival Analysis (FSA) is an emerging Federated Learning (FL) paradigm for performing survival analysis at multiple medical institutions by jointly training their models without accessing their raw survival data. FSA preserves the privacy of patients' sensitive information, addresses biases by making generalized predictions, and leads toward more accurate precision medicine.

The recent FSA approaches, most of which are based on deep neural networks, require several communications between the server and clients for exchanging model parameters during the training phase (McMahan et al. 2017), leading to a communication burden. Moreover, these approaches do not rigorously investigate the impact of censoring and competing risks in FSA. Censoring and competing risks lead to biased and inaccurate risk predictions (Rahman et al. 2021; Lee et al. 2018; Danks and Yau 2022), and they are amplified when transitioning to a federated setting, primarily due to the imbalanced distribution of competing events and non-uniform censoring across different medical centers. Moreover, FSA brings its specific challenges, such as data heterogeneity in terms of features and events, imbalanced sample size, and uneven resource distributions among participating medical institutions, which demand innovative techniques and practical solutions.

To tackle the aforementioned challenges in FSA, we propose an innovative FSA framework, **FedPRF**, that leverages federated pseudo values (FPV) as proposed in our prior works (Rahman and Purushotham 2023b,a) to handle censoring in distributed datasets while preserving privacy and to enable federated training for distributed right censored data. Within the FedPRF frameworks, we employ FPV-

based Random Forests models, termed as **RFpseudo**, for analyzing decentralized survival datasets. In contrast to deep learning-based FSA frameworks that often require a substantial number of communications for convergence, FedPRF stands out as computationally and communication efficient, requiring only two rounds of communication between the server and clients. Notably, our FedPRF framework particularly shows improvement in scenarios where clients possess limited data and operate within resource-constrained environments. FedPRF demonstrates its effectiveness by outperforming the traditional iterative FSA approaches for survival analysis and competing risk analysis (CRA) tasks with less communication overhead. Furthermore, FedPRF shows significant improvement over our closest counterpart, the RSF-based federated framework, FedSurF (Archetti and Matteucci 2023). Additionally, FedPRF can provide feature-importance in predictions due to their tree-based structure, which is useful for identifying risk factors and enhancing interpretability for better medical decision-making. The interpretation provided by FedPRF aligns with the interpretation obtained by the gold-standard centralized model. Extensive experiments on survival data with both a single event and multiple competing events and on synthetic datasets with various censoring mechanisms demonstrate that our FedPRF framework has the same predictive ability as gold-standard centralized approaches. FedPRF consistently outperforms the state-of-the-art deep survival models as well as random survival forests trained within a federated framework, especially under various censoring settings. Empirically, we show that FedPRF is scalable with respect to number of clients and works well for clients with limited data.

### Problem Statement

This paper focuses on horizontal federated survival analysis (FSA) and federated competing risk analysis (FedCRA) with  $K$  clients, where all the participating clients/silos (e.g., hospitals) share the same feature space but a different sample space in their local survival dataset  $D_k$ .  $D_k$  is a triplet, represented by  $D_k = \{\mathbf{X}_{ik}, Y_{ik}, \delta_{ik}\}$ , where  $i$  refers to a subject in client  $k$ , and  $k$  contains a total of  $n_k$  subjects, with  $1 \leq i \leq n_k$ . For a subject  $i$  in client  $k$ ,  $\mathbf{X}_{ik}$  is a  $p$ -dimensional covariates or features vector,  $Y_{ik}$  is the observed time and is equal to the failure time  $T_{ik}$ , if the subject  $i$  is uncensored, and  $C_{ik}$ , if the subject  $i$  is censored.  $\delta_{ik}$  is the event or censoring indicator, where  $\delta_{ik} = r$ , if the subject  $i$  in client  $k$  is uncensored and experienced an event  $r$ ; where  $r = 1, 2, \dots, R$  and  $\delta_{ik} = 0$ , if the subject is censored. In the traditional survival analysis, it is typically assumed that there is only one event of interest, i.e.,  $r = 1$ . Survival analysis aims to estimate the conditional survival function  $S(t|\mathbf{X}_{ik}) = Pr(T_{ik} > t|\mathbf{X}_{ik})$ , at time  $t$  given the covariates  $\mathbf{X}_{ik}$  and CRA aims to estimate the conditional cumulative incidence function (CIF),  $F_r(t|x_{ik}) = Pr(T_{ik} \leq t, \delta_{ik} = r|\mathbf{X}_{ik})$ , for event  $r$ , at time  $t$  given the covariates  $\mathbf{X}_{ik}$ . However, the survival function or CIF is incompletely observed due to the censoring, and thus, we can not directly perform regression for the estimation. To address the problem, we derive subject-specific federated pseudo values and use them as a natural replacement

for the incompletely observed survival function or CIF for the subjects in training data. Once FPV are calculated at the client side, the local training datasets  $D_{k(train)}$  are replaced by  $D'_{k(train)} = \{\mathbf{X}_{ik(train)}, J_{ik(train)}\}$ . Traditionally, it is desired to merge local training datasets  $D_{k(train)}$  to create a centralized dataset,  $D_{(train)} = D_{1(train)} \cup D_{2(train)} \cup \dots \cup D_{K(train)}$  and build survival models on  $D'_{(train)}$  to harness the diversity and decentralized datasets, which can address the challenges posed by limited data availability in survival studies. However, merging data and maintaining a centralized repository becomes impractical and ethically challenging due to the potential risk of patient-sensitive information leakage and stringent data-sharing regulations, such as those enforced by GDPR or HIPAA. Therefore, the goal of this study is to facilitate horizontal FSA to enable performing survival analysis and CRA on decentralized client datasets without sharing the actual data. In FSA, each client initially trains an independent local model using its private training data  $D_{k(train)}$ . Subsequently, a central server aggregates the local models shared by the clients to construct a global model. This innovative approach ensures that the data remains decentralized and secure while enabling the development of robust survival models by leveraging the benefit of large-scale data.

### Our Proposed Framework: FedPRF

We propose a novel federated learning (FL) framework, denoted as **FedPRF**, designed for conducting survival analysis and competing risk analysis (CRA) on decentralized datasets without compromising sensitive patient information. Our FedPRF framework employs federated pseudo value (FPV)-based random forests; we call it **RFpseudo**, for local training and federate the local models through a single round of communication between local clients and a central server. In this framework, the local RFpseudo models formulate the survival analysis and CRA problems as regression by using FPV for survival function or cumulative incidence function (CIF) as response variables in the models, which enable subject-specific predictions. Traditionally, Jackknife pseudo values (JPV) have been employed in statistics and the machine learning community to solve survival analysis and CRA problems in centralized data settings (Andersen and Pohar Perme 2010; Graw, Gerds, and Schumacher 2009; Zhao and Feng 2020; Rahman et al. 2021). However, conventional JPV are not suitable for FSA and FedCRA. They are computed locally at the client side and only exhibit local consistency under the assumption of independent and identically distributed (IID) data (Graw, Gerds, and Schumacher 2009). However, the IID assumption often gets violated in federated setups due to the non-IID and non-uniform censoring properties in survival data. Therefore, using JPV for FSA leads to biased and inconsistent predictions. Our FedPRF utilizes the FPV that account for data heterogeneity and non-uniform censoring while ensuring data privacy and enabling FL for survival analysis and CRA. FPV leverages communication between clients and the server to calculate pseudo values by exchanging summary information, such as the number of events and censored at time  $t$ , instead of

sharing the clients' patient outcomes, i.e., event or censoring time and event status. Our FedPRF framework is structured into three stages: 1) FPV derivation, 2) Local training, and 3) Tree aggregation. Next, we briefly describe each stage.

**1. FPV Derivation:** Clients first use the observed time  $Y_k$  and event status  $\delta_k$  to form a non-identifying local partial table,  $L_k$  with the following components: 1) Number of unique time points  $\tau_k$ , 2) Number at risk at  $t_0 \in \tau_k$ ,  $R_{k,t_0}$ , 3) Number of events at  $t \in \tau_k$ ,  $d_{k,t}$ , 4) Number of censored at  $t \in \tau_k$ ,  $c_{k,t}$ . Additionally, the number of events due to cause  $r$  at  $t \in \tau_k$ ,  $d_{r,k,t}$  is included for CRA. Then, clients send the local partial tables,  $L_k$ , to the server. The server creates a global partial table,  $M$ , by first sorting the union of the unique time points in the clients' summary table and assigning the aggregated value of other components to the corresponding unique time points. The global partial table has partial information on the number at risk, i.e., information only available at the starting time point  $t_0$ . To obtain the number at risk at the subsequent time points, we apply the formula:  $R_{t_j} = R_{t_{j-1}} - d_{t_{j-1}} - c_{t_{j-1}}; j = 1, 2, \dots, m$ . Once we have the global full table, we compute the global survival function for survival analysis as  $\hat{S}^G(t) = \prod_{t_j \in \tau \leq t} (1 - \frac{d_{t_j}}{R_{t_j}})$  or global CIF for CRA as  $\hat{F}_r^G(t) = \sum_{t_j \in \tau \leq t} \hat{S}^G(t) \frac{d_{r,t_j}}{R_{t_j}}$ . After that, the server sends the global partial table,  $M$ , along with the global survival function or global CIF to the clients. Clients compute the leave-out-out global full table followed by the leave-out-out global survival function and global CIF by omitting the  $i^{th}$  subjects from the shared global partial table as  $\hat{S}^{G^{-ik}}(t) = \prod_{t_j \in \tau \leq t} \frac{R_{t_j}^{-ik} - d_{t_j}^{-ik}}{R_{t_j}^{-ik}}$  and  $\hat{F}_r^{G^{-ik}}(t) = \sum_{t_j \in \tau \leq t} \hat{S}^{G^{-ik}}(t) \frac{d_{r,t_j}^{-ik}}{R_{t_j}^{-ik}}$ , respectively. Finally, the federated pseudo values for survival function,  $J_{ik}(t)$ , and CIF,  $J_{ikr}(t)$ , are computed as  $J_{ik}(s) = n\hat{S}^G(s) - (n-1)\hat{S}^{G^{-ik}}(s)$  and  $J_{ikr}(s) = n\hat{F}_r^G(s) - (n-1)\hat{F}_r^{G^{-ik}}(s)$ , respectively.

**2. Local Training:** Once the FPV are computed, the new dataset for local model training is transformed to  $D'_k = \{X^k, J_{ik}(s)$  [for survival function] or  $J_{ikr}(s)$  [for CIF]}. Each client  $k \in K$  independently train their corresponding local RFpseudo models,  $M_k$ , consisting of an ensemble of trees, denoted as  $T_1, T_2, \dots, T_{N_k}$ . The RFpseudo model is a Random Forest model that takes covariates as input and FPV as output. The optimal hyperparameters for the client-specific RFpseudo models, such as the number of local trees  $N_k$ , criterion, max features, maximum tree depth, and minimum samples per split, are obtained based on the validation performance. The pseudo values can be greater than 1 and less than 0, unlike the survival function or CIF, bounded by [0,1]. Therefore, we use the clipping transformation (Rahman et al. 2021):  $p' = \min(1, \max(0, \hat{p}))$ , for transforming predicted pseudo values to be bounded by [0,1]. Here  $\hat{p}$  and  $p'$  are predicted and transformed FPVs, respectively.

**3. Tree Aggregation:** For aggregation, we consider two strategies: 1) FedPRF: Simply combine  $N_k$  trees from the local updated model,  $M_k$ , to create a global random forest,  $M_Q$ , 2) FedPRF-CI: The server initializes the total number

of trees for the global model, i.e.,  $N_S$  and informs the clients to send  $N'_k$  local trees, depending on their data size  $|D'_k|$  to ensure that clients with larger datasets have more contribution to the global model. Clients choose the  $N'_k$  trees showing the best C-index performance on the validation data to construct the global model  $M_Q$ .

---

#### Algorithm 1: FedPRF Framework

---

**Input:** Local dataset  $D_k = \{\mathbf{X}_k, Y_k, \delta_k\}$ , a vector of pre-specified time points  $\tau = \{\tau_1 < \tau_2 < \dots < \tau_M\}$ .  
**Output:** Global RFpseudo: ensemble of trees,  $M_Q$ .  
**Stage 1:** Derive FPV,  $J_k(\tau)$ , using  $Y_k$  and  $\delta_k$  and transform the local datasets  $D_k$  to  $D'_k = \{\mathbf{X}_k, J_k(\tau)\}$   
**Clients execute:**  $\leftarrow$  **Stage 2**  
**if** FedPRF **then**  
  **for**  $k \in K$  **in parallel do**  
     $M_k \leftarrow$  RFpseudo( $D'_k$ )  
  **end for**  
  Send  $M_k$  to the server  
**end if**  
**if** FedPRF-CI **then**  
  Send  $|D'_k|$  to the server.  
  **for**  $k \in K$  **in parallel do**  
     $M_k \leftarrow$  RFpseudo( $D'_k$ )  
     $M'_k \leftarrow$  Select  $N'_k$  trees from  $M_k$  based on the highest validation C-Index scores.  
  **end for**  
  Send  $M'_k$  to the server  
**end if**  
**Server executes:**  $\leftarrow$  **Stage 3**  
**if** FedPRF **then**  
   $M_Q \leftarrow \cup_{k=1}^K M_k$   
**end if**  
**if** FedPRF-CI **then**  
  Initialize: total number of trees in global RFpseudo,  $N_S$   
  **for**  $k = 1$  to  $K$  **do**  
    Compute  $N'_k = N_S * \frac{|D'_k|}{\sum_{k=1}^K |D'_k|}$   
  **end for**  
  Send  $N'_k$  to the clients  
   $M_Q \leftarrow \cup_{k=1}^K M'_k$   
**end if**  
**return**  $M_Q$

---

## Empirical Evaluation

Extensive experiments were conducted to answer the following research questions.

- **RQ1:** To what extent does our proposed FedPRF outperform the state-of-the-art survival models trained in FL settings when applied to both simulated and real-world distributed survival datasets?
- **RQ2:** How well does FedPRF perform compared to the state-of-the-art CRA models trained in federated settings when applied to real-world distributed survival datasets that exhibit multiple competing events, non-IID characteristics, and non-uniform censoring properties?

- **RQ3:** How effectively does our FPV-based FedPRF framework handle various types and amounts of censoring in contrast to the baseline models?
- **RQ4:** How successfully does our FedPRF federated framework maintain the interpretability of the gold-standard centralized models?
- **RQ5:** How does FedPRF scale to highly distributed settings with numerous clients, and how well does it reduce the computational burden?

## Federated Survival Analysis

We perform a comprehensive model evaluation in both centralized and federated settings using four widely used real survival datasets with a single event. The datasets are: METABRIC (Curtis et al. 2012), SUPPORT (Knaus et al. 1995), GBSG (Foekens et al. 2000; Schumacher et al. 1994), and METABRIC-HD (META-HD) (Curtis et al. 2012). For centralized settings, we used 20% of the entire data as test data and the remaining 80% as training data, out of which 20% are used as a validation set. For federated settings, we simulated a federated environment by distributing the datasets to multiple clients, characterized by **IID** and **non-IID** data properties. To replicate an IID federated setup, we employ a random allocation process where an equal number of subjects are assigned to each participating client. For non-IID federated setup, we assign the subjects to the clients in a way that the event time distribution in a client is skewed towards a specific quantile of the time horizon. We assume different quantiles for the skewness of event time distribution for different clients. To provide a robust evaluation of these models in a real-world distributed dataset scenario, we have incorporated the TCGA dataset (Terrail et al. 2022; Andreux et al. 2020). The TCGA data is collected from various sites of seven diverse regions worldwide, including four regions within the United States (South, West, Midwest, Northeast), Europe, Canada, and other international locations. We evaluate the models on aggregated test data collected from the participating clients. We ran the experiments five times using different seeds or random states and reported the average performance with corresponding standard deviations. We choose specific time points for calculating pseudo values and evaluating the models, e.g., the 10th to 99th percentile of the time horizon with an interval of 10.

**Baselines:** In the centralized setting, we compare our proposed RFPseudo model with 8 baseline survival models; 1) LinearPH (Zhang, Toni, and Williams 2022), 2) DeSurv (Danks and Yau 2022), 3) Neural Fine & Gray model (NFG) (Jeanselme et al. 2023), 4) NNph (Zhang, Toni, and Williams 2022; Andreux et al. 2020), 5) NNnph (Zhang, Toni, and Williams 2022), 6) DeepHit (Rahimian et al. 2022; Lee et al. 2018), 7) FedPAttN (Rahman and Purushotham 2023b), and 8) Random Survival Forests (RSF) (Ishwaran et al. 2008; Archetti and Matteucci 2023). In the federated settings, we compare our FSA frameworks, FedPRF and FedPRF-CI, with the federated variant of baseline models, 1) Fed-LinearPH, 2) Fed-DeSurv, 3) Fed-NFG, 4) Fed-NNph, 5) Fed-NNnph, 6) Fed-DeepHit, 7) FedPseudo, and 8) FedSurF.

**Evaluation Metrics:** We evaluate the models using the time-dependent concordance index (C-Index) (Antolini, Boracchi, and Biganzoli 2005), and integrated IPCW Brier score (IBS) (Graf et al. 1999). We use **pycox** (Kvamme 2023) package to compute the C-Index and IBS. The higher C-Index and lower IBS indicate better performance.

**Performance Evaluation:** Table 1 demonstrates the superiority of the RFPseudo and FedPRF over baseline survival models in terms of C-Index and IBS metrics when evaluated on real-world survival datasets in both centralized and federated settings: Federated-IID and Federated-non-IID. In the centralized setting, our RFPseudo outperforms the baselines, showing an average improvement of up to 6% (in comparison to RSF), 2% (in comparison to LinearPH, NFG, DeepHit, and RSF), 7% (in comparison to DeepHit), and 11% (in comparison to NFG) in terms of C-Index and 12% (RSF), 8% (NFG), 8% (RSF), and 15% (NFG) in terms of IBS across the METABRIC, SUPPORT, GBSG, and META-HD datasets, respectively. In the federated settings, our FedPRF performs consistently better than the baselines up to 9% (Fed-DeSurv), 2% (Fed-DeepHit), 8% (Fed-NNnph, Fed-DeepHit), and 14% (Fed-NFG, Fed-SurF) in terms of C-Index and 11% (FedSurF), 9% (Fed-NFG), 9% (Fed-NFG), and 13% (FedSurF) in terms of IBS across the METABRIC, SUPPORT, GBSG, and META-HD datasets, respectively. It is worth noting that baseline models tend to exhibit suboptimal performance in limited-sized and high-dimensional datasets like METABRIC and META-HD, whereas FedPRF demonstrates substantial advantages under conditions of limited data availability. Figure 1 underscores the superior performance of FedPRF compared to the baselines, as evident in both the C-Index and IBS metrics, which holds across the varying number of sites from which the TCGA dataset is collected.

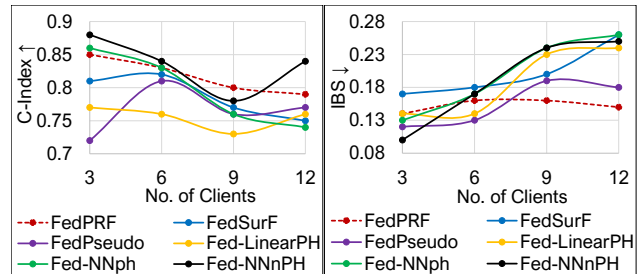


Figure 1: Comparing performance of federated frameworks with varying number of sites in TCGA dataset. The red dotted line represents our FedPRF framework.

## Federated Competing Risk Analysis

To show the effectiveness of our FedPRF frameworks and RFPseudo models in both federated and centralized competing risk analysis (CRA), we consider the Surveillance, Epidemiology, and End Results (SEER) dataset containing 6 competing events and 28366 cancer patients from different geographic locations, out of which 23.2% patients died of cervical cancer (CC), 2.6% died due to other cancers (OC),

Setup	Model	C-Index				IBS			
		METABRIC	SUPPORT	GBSG	META-HD	METABRIC	SUPPORT	GBSG	META-HD
Centralized	LinearPH	0.63 (0.00)	0.60 (0.00)	0.66 (0.01)	0.65 (0.01)	0.19 (0.01)	0.20 (0.00)	0.20 (0.01)	0.27 (0.11)
	DeSurv	0.66 (0.02)	<b>0.62 (0.00)</b>	0.67 (0.00)	0.65 (0.01)	<b>0.18 (0.00)</b>	<b>0.19 (0.00)</b>	<b>0.18 (0.00)</b>	<b>0.19 (0.00)</b>
	NFG	0.62 (0.01)	0.60 (0.00)	0.66 (0.00)	0.56 (0.01)	0.28 (0.00)	0.27 (0.00)	0.25 (0.00)	0.34 (0.01)
	NNph	0.63 (0.01)	0.61 (0.00)	0.67 (0.01)	0.63 (0.03)	0.20 (0.01)	0.20 (0.00)	0.19 (0.01)	0.25 (0.08)
	NNph	0.66 (0.01)	0.61 (0.00)	0.66 (0.01)	0.59 (0.03)	<b>0.18 (0.00)</b>	0.20 (0.00)	0.19 (0.00)	0.20 (0.00)
	DeepHit	0.65 (0.03)	0.60 (0.01)	0.61 (0.06)	0.64 (0.01)	0.19 (0.00)	0.21 (0.00)	0.23 (0.00)	0.20 (0.01)
	FedPAttn	<b>0.67 (0.00)</b>	0.61 (0.00)	0.67 (0.00)	<b>0.67 (0.01)</b>	<b>0.18 (0.00)</b>	<b>0.19 (0.00)</b>	<b>0.18 (0.00)</b>	0.21 (0.02)
	RSF	0.60 (0.00)	0.60 (0.00)	0.66 (0.01)	0.58 (0.01)	0.30 (0.00)	0.22 (0.00)	0.26 (0.00)	0.32 (0.01)
<b>RFpseudo (Ours)</b>	0.66 (0.00)	<b>0.62 (0.00)</b>	<b>0.68 (0.00)</b>	<b>0.67 (0.01)</b>	<b>0.18 (0.00)</b>	<b>0.19 (0.00)</b>	<b>0.18 (0.00)</b>	<b>0.19 (0.00)</b>	
Federated (IID)	Fed-LinearPH	0.63 (0.00)	0.60 (0.00)	0.64 (0.01)	0.65 (0.01)	<b>0.18 (0.00)</b>	<b>0.20 (0.00)</b>	0.19 (0.00)	0.22 (0.01)
	Fed-DeSurv	0.60 (0.01)	<b>0.61 (0.00)</b>	0.63 (0.01)	0.64 (0.01)	0.20 (0.01)	<b>0.20 (0.00)</b>	0.19 (0.00)	<b>0.19 (0.00)</b>
	Fed-NFG	0.62 (0.01)	0.60 (0.01)	0.66 (0.01)	0.55 (0.02)	0.28 (0.01)	0.27 (0.02)	0.26 (0.01)	0.31 (0.03)
	Fed-NNph	0.63 (0.01)	<b>0.61 (0.00)</b>	0.66 (0.01)	0.66 (0.01)	0.19 (0.00)	<b>0.20 (0.00)</b>	0.19 (0.00)	0.22 (0.00)
	Fed-NNph	0.65 (0.02)	0.60 (0.01)	0.64 (0.01)	0.65 (0.01)	<b>0.18 (0.00)</b>	<b>0.20 (0.00)</b>	0.19 (0.00)	0.23 (0.01)
	Fed-DeepHit	0.66 (0.02)	0.59 (0.01)	0.63 (0.02)	0.64 (0.03)	0.20 (0.00)	0.23 (0.00)	0.23 (0.00)	0.21 (0.03)
	FedPseudo	0.66 (0.01)	<b>0.61 (0.00)</b>	0.66 (0.01)	0.64 (0.00)	<b>0.18 (0.00)</b>	<b>0.20 (0.00)</b>	<b>0.18 (0.00)</b>	0.20 (0.00)
	FedSurF	0.62 (0.01)	<b>0.61 (0.00)</b>	<b>0.68 (0.00)</b>	0.55 (0.01)	0.29 (0.00)	0.22 (0.00)	0.25 (0.00)	0.32 (0.01)
<b>FedPRF (Ours)</b>	<b>0.67 (0.00)</b>	<b>0.61 (0.00)</b>	<b>0.68 (0.00)</b>	<b>0.69 (0.01)</b>	<b>0.18 (0.00)</b>	<b>0.20 (0.00)</b>	<b>0.18 (0.00)</b>	<b>0.19 (0.00)</b>	
<b>FedPRF-CI (Ours)</b>	<b>0.67 (0.01)</b>	<b>0.61 (0.00)</b>	<b>0.68 (0.00)</b>	<b>0.69 (0.01)</b>	<b>0.18 (0.00)</b>	<b>0.20 (0.00)</b>	<b>0.18 (0.00)</b>	<b>0.19 (0.00)</b>	
Federated (non-IID)	Fed-LinearPH	0.65 (0.00)	0.60 (0.00)	0.63 (0.01)	0.67 (0.01)	<b>0.18 (0.00)</b>	<b>0.19 (0.00)</b>	<b>0.21 (0.00)</b>	0.22 (0.01)
	Fed-DeSurv	0.61 (0.01)	<b>0.61 (0.00)</b>	0.62 (0.01)	0.66 (0.01)	0.20 (0.00)	<b>0.19 (0.00)</b>	<b>0.21 (0.00)</b>	<b>0.18 (0.01)</b>
	Fed-NFG	0.64 (0.01)	0.60 (0.00)	0.64 (0.01)	0.55 (0.02)	0.27 (0.01)	0.28 (0.00)	0.31 (0.01)	0.29 (0.01)
	Fed-NNph	0.65 (0.02)	<b>0.61 (0.00)</b>	0.60 (0.02)	0.67 (0.01)	0.19 (0.01)	0.20 (0.00)	0.22 (0.01)	0.22 (0.00)
	Fed-NNph	0.68 (0.01)	<b>0.61 (0.00)</b>	0.57 (0.01)	0.65 (0.01)	<b>0.18 (0.00)</b>	0.20 (0.00)	0.23 (0.00)	0.24 (0.01)
	Fed-DeepHit	0.68 (0.02)	0.60 (0.01)	0.57 (0.03)	0.66 (0.01)	0.20 (0.01)	0.22 (0.00)	0.23 (0.00)	<b>0.18 (0.00)</b>
	FedPseudo	0.67 (0.02)	0.60 (0.00)	0.58 (0.01)	0.61 (0.02)	<b>0.18 (0.00)</b>	<b>0.19 (0.00)</b>	0.22 (0.01)	0.19 (0.01)
	FedSurF	0.63 (0.01)	<b>0.61 (0.00)</b>	0.63 (0.02)	0.55 (0.02)	0.30 (0.00)	0.22 (0.00)	0.23 (0.00)	0.30 (0.01)
<b>FedPRF (Ours)</b>	<b>0.70 (0.00)</b>	<b>0.61 (0.00)</b>	<b>0.65 (0.02)</b>	<b>0.68 (0.02)</b>	0.23 (0.02)	<b>0.19 (0.00)</b>	0.22 (0.00)	0.27 (0.01)	
<b>FedPRF-CI (Ours)</b>	<b>0.70 (0.00)</b>	0.60 (0.00)	<b>0.65 (0.01)</b>	<b>0.68 (0.01)</b>	0.23 (0.02)	<b>0.19 (0.00)</b>	0.22 (0.00)	0.27 (0.01)	

Table 1: Performance [Mean (SD)] comparisons of the models on the real survival datasets

2.4% died of cardiovascular disease (CVD), 1.1% died due to chronic medical disease (CMD), 0.6% died of infectious disease (ID), and 1.8% died due to other causes (OCS) (Rahman et al. 2021; SEER 2023). We simulate two distributed CRA data settings; 1. **SEER-Region**, and 2. **SEER-NUC**, using the centralized SEER dataset. In SEER-Region, we replicate a geographically distributed data environment for FL by partitioning the SEER data into three clients based on the hospital regions in the USA: West, Central, and East. In SEER-NUC, we induce non-uniform censoring (NUC) to the SEER-Region setting by varying the censoring percentages from the options [0.2, 0.3, 0.4, 0.50, 0.55] across clients while keeping an equal number of subjects in all clients.

**Baselines:** We compare our proposed RFpseudo model and FedCRA frameworks: FedPRF and FedPRF-CI with seven state-of-the-art CRA models and with their federated variants, respectively. The models are: 1) Cause-specific Cox proportional hazard model (CS-CoxPH) (Cox 1972), 2) DeepPseudo (Rahman et al. 2021), 3) SurvTRACE (Wang and Sun 2022), 4) RSF (Ishwaran et al. 2014), 5) DeSurv (Danks and Yau 2022), 6) NFG (Jeanselme et al. 2023), and 7) DeepHit (Lee et al. 2018).

**Performance Evaluation:** Table 2 demonstrates that our RFpseudo performs significantly better than baselines by as much as 25.5% (against NFG) in terms of event-specific C-Index when evaluated on the centralized settings of SEER-

Region dataset, except for RSF, which exhibits similar performance. On SEER-NUC data, RFpseudo achieves a significant improvement over the baselines by a minimum of 4.0% (against RSF) and a maximum of 23.3% (against DeepPseudo). In the federated settings, FedPRF frameworks consistently achieve performance gains up to 26.7% and 28.0% (against Fed-DeSurv) on SEER-Region and SEER-NUC datasets, respectively, further highlighting the effectiveness of FedPRF compared to baselines.

### FedCRA under Various Censoring Settings

We generate 12 distributed synthetic datasets for both centralized and federated training, each exhibiting distinct censoring scenarios (Rahman and Purushotham 2023a; Barrajon and Barrajon 2020). The censoring mechanisms in those datasets include time censoring (TC), interim censoring (IC), case censoring (CC) with varying levels of censoring (25%, 50%, and 75%), and heterogeneous survival distribution with equal and unequal censoring across clients (HSDEC and HSDUC). These datasets were generated assuming a distributed setting with 10 clients, with each client having either the same covariate distributions (IID) or varying covariate distributions across clients, capturing the complexities of real-world non-IID data.

**Performance Evaluation:** To show the effectiveness of FPV and tree-based FL framework, FedPRF, in handling

Setup	Data	Model	Event (Death due to)					
			CC	OC	CVD	CMD	ID	OCS
Centralized	SEER-Region	CS-CoxPH	0.81 (0.00)	0.50 (0.16)	0.84 (0.00)	0.52 (0.13)	0.53 (0.04)	0.76 (0.01)
		DeepPseudo	0.72 (0.06)	0.68 (0.01)	0.81 (0.06)	0.75 (0.08)	0.69 (0.04)	0.70 (0.03)
		RSF	0.85 (0.00)	<b>0.91 (0.00)</b>	0.90 (0.00)	<b>0.85 (0.01)</b>	<b>0.87 (0.01)</b>	<b>0.89 (0.00)</b>
		NFG	0.71 (0.08)	0.60 (0.14)	0.66 (0.24)	0.61 (0.21)	0.56 (0.24)	0.61 (0.18)
		DeSurv	0.82 (0.01)	0.58 (0.06)	0.68 (0.06)	0.51 (0.05)	0.61 (0.07)	0.58 (0.06)
		DeepHit	<b>0.88 (0.00)</b>	0.86 (0.01)	0.88 (0.01)	0.75 (0.04)	0.82 (0.03)	0.80 (0.02)
		SurvTRACE	0.82 (0.00)	0.84 (0.00)	0.87 (0.01)	0.83 (0.01)	0.80 (0.01)	0.78 (0.01)
	RFPseudo (Ours)	<b>0.88 (0.00)</b>	<b>0.91 (0.00)</b>	<b>0.91 (0.00)</b>	<b>0.85 (0.00)</b>	0.85 (0.01)	<b>0.88 (0.00)</b>	
	SEER-NUC	CS-CoxPH	0.79 (0.02)	0.78 (0.01)	0.80 (0.01)	0.84 (0.02)	0.68 (0.21)	0.77 (0.03)
		DeepPseudo	0.64 (0.23)	0.69 (0.09)	0.69 (0.13)	0.70 (0.10)	0.66 (0.15)	0.63 (0.10)
		RSF	0.83 (0.01)	0.84 (0.03)	0.87 (0.02)	0.89 (0.03)	0.90 (0.04)	0.84 (0.03)
		NFG	0.69 (0.07)	0.68 (0.05)	0.80 (0.07)	0.81 (0.04)	0.78 (0.06)	0.68 (0.08)
		DeSurv	0.80 (0.01)	0.69 (0.02)	0.71 (0.05)	0.61 (0.05)	0.63 (0.07)	0.67 (0.06)
		DeepHit	0.88 (0.01)	0.85 (0.02)	0.82 (0.02)	0.76 (0.05)	0.81 (0.11)	0.82 (0.02)
SurvTRACE		0.84 (0.03)	0.83 (0.03)	0.83 (0.03)	0.79 (0.05)	0.83 (0.08)	0.80 (0.07)	
RFPseudo (Ours)	<b>0.89 (0.01)</b>	<b>0.89 (0.02)</b>	<b>0.91 (0.01)</b>	<b>0.93 (0.02)</b>	<b>0.91 (0.05)</b>	<b>0.88 (0.02)</b>		
Federated	SEER-Region	Fed-CS-CoxPH	0.81 (0.00)	0.82 (0.00)	0.84 (0.00)	0.82 (0.00)	0.73 (0.02)	0.75 (0.00)
		Fed-DeepPseudo	0.78 (0.01)	0.74 (0.01)	0.85 (0.00)	0.79 (0.01)	0.71 (0.01)	0.72 (0.01)
		FedSurF	0.83 (0.00)	<b>0.89 (0.01)</b>	0.89 (0.00)	<b>0.84 (0.02)</b>	0.83 (0.01)	0.85 (0.01)
		Fed-NFG	0.71 (0.02)	0.70 (0.03)	0.86 (0.01)	0.81 (0.01)	0.78 (0.00)	0.74 (0.01)
		Fed-DeSurv	0.79 (0.01)	0.50 (0.04)	0.57 (0.08)	0.53 (0.04)	0.57 (0.07)	0.60 (0.04)
		Fed-DeepHit	0.85 (0.00)	0.84 (0.01)	0.89 (0.01)	0.79 (0.04)	0.84 (0.02)	0.75 (0.02)
		Fed-SurvTRACE	0.82 (0.01)	0.81 (0.01)	0.87 (0.01)	0.76 (0.03)	0.77 (0.03)	0.74 (0.02)
	FedPRF (Ours)	<b>0.85 (0.00)</b>	<b>0.89 (0.00)</b>	0.89 (0.00)	0.82 (0.00)	0.85 (0.00)	<b>0.86 (0.00)</b>	
	FedPRF-CI (Ours)	<b>0.85 (0.00)</b>	<b>0.89 (0.00)</b>	<b>0.90 (0.00)</b>	0.83 (0.00)	<b>0.86 (0.00)</b>	<b>0.86 (0.00)</b>	
	SEER-NUC	Fed-CS-CoxPH	0.79 (0.01)	0.79 (0.01)	0.80 (0.01)	0.84 (0.02)	0.75 (0.03)	0.76 (0.03)
		Fed-DeepPseudo	0.72 (0.06)	0.65 (0.03)	0.77 (0.02)	0.79 (0.06)	0.73 (0.07)	0.69 (0.02)
		FedSurF	0.80 (0.01)	0.78 (0.04)	0.82 (0.01)	0.85 (0.03)	0.84 (0.06)	0.80 (0.02)
		Fed-NFG	0.67 (0.02)	0.70 (0.02)	0.81 (0.02)	0.84 (0.01)	0.79 (0.09)	0.73 (0.02)
		Fed-DeSurv	0.78 (0.01)	0.54 (0.06)	0.65 (0.07)	0.49 (0.06)	0.50 (0.07)	0.58 (0.07)
Fed-DeepHit		0.83 (0.01)	0.83 (0.03)	0.86 (0.01)	0.84 (0.05)	<b>0.87 (0.04)</b>	0.82 (0.03)	
Fed-SurvTRACE		0.82 (0.01)	0.81 (0.03)	0.82 (0.02)	0.78 (0.08)	0.84 (0.06)	0.80 (0.05)	
FedPRF (Ours)	<b>0.85 (0.01)</b>	<b>0.87 (0.03)</b>	<b>0.87 (0.02)</b>	<b>0.91 (0.01)</b>	<b>0.87 (0.07)</b>	<b>0.85 (0.02)</b>		
FedPRF-CI (Ours)	<b>0.85 (0.01)</b>	<b>0.87 (0.04)</b>	0.86 (0.02)	0.90 (0.02)	<b>0.87 (0.07)</b>	0.84 (0.02)		

Table 2: C-Index [Mean (SD)] performance comparisons on the real CRA datasets

Dataset	Centralized				Federated			
	RSF	DeepHit	SurvTRACE	RFpseudo (Ours)	FedSurF	Fed-DeepHit	Fed-SurvTRACE	FedPRF (Ours)
TC-IID	0.67 (0.01)	0.69 (0.01)	0.70 (0.00)	<b>0.80 (0.00)</b>	0.68 (0.00)	0.69 (0.01)	0.70 (0.01)	<b>0.78 (0.01)</b>
TC-Non-IID	0.65 (0.00)	0.68 (0.01)	0.72 (0.01)	<b>0.81 (0.01)</b>	0.65 (0.01)	0.66 (0.01)	0.70 (0.00)	<b>0.78 (0.01)</b>
IC-IID	0.66 (0.00)	0.68 (0.01)	0.69 (0.00)	<b>0.78 (0.01)</b>	0.67 (0.00)	0.69 (0.01)	0.69 (0.00)	<b>0.76 (0.01)</b>
IC-Non-IID	0.64 (0.00)	0.68 (0.01)	0.69 (0.02)	<b>0.81 (0.01)</b>	0.64 (0.01)	0.66 (0.01)	0.68 (0.01)	<b>0.77 (0.01)</b>
CC25-IID	0.64 (0.01)	0.77 (0.03)	<b>0.92 (0.01)</b>	0.79 (0.02)	0.67 (0.00)	0.73 (0.01)	<b>0.91 (0.03)</b>	0.84 (0.03)
CC25-Non-IID	0.64 (0.01)	0.65 (0.00)	0.67 (0.01)	<b>0.79 (0.01)</b>	0.63 (0.01)	0.65 (0.01)	0.67 (0.01)	<b>0.76 (0.01)</b>
CC50-IID	0.63 (0.02)	0.80 (0.01)	<b>0.92 (0.01)</b>	0.82 (0.00)	0.68 (0.01)	0.74 (0.01)	<b>0.91 (0.00)</b>	0.86 (0.00)
CC50-Non-IID	0.64 (0.01)	0.67 (0.01)	0.67 (0.01)	<b>0.82 (0.01)</b>	0.64 (0.01)	0.65 (0.01)	0.68 (0.01)	<b>0.75 (0.00)</b>
CC75-IID	0.61 (0.02)	0.82 (0.00)	<b>0.88 (0.01)</b>	0.83 (0.00)	0.68 (0.01)	0.73 (0.01)	<b>0.86 (0.00)</b>	<b>0.86 (0.00)</b>
CC75-Non-IID	0.64 (0.01)	0.72 (0.01)	0.67 (0.01)	<b>0.86 (0.01)</b>	0.62 (0.01)	0.65 (0.00)	0.71 (0.01)	<b>0.69 (0.02)</b>
HSDEC	0.64 (0.03)	0.66 (0.03)	0.70 (0.01)	<b>0.79 (0.02)</b>	0.60 (0.01)	0.68 (0.01)	0.68 (0.00)	<b>0.75 (0.02)</b>
HSDUC	0.66 (0.01)	0.67 (0.02)	0.70 (0.01)	<b>0.78 (0.01)</b>	0.62 (0.01)	0.68 (0.01)	0.69 (0.01)	<b>0.76 (0.01)</b>

Table 3: C-Index [Mean (SD)] comparisons on the distributed synthetic CRA datasets with different censoring settings

various types and amounts of censoring, we compare FedPRF with state-of-the-art CRA models, DeepHit and SurvTRACE, along with another tree-based model RSF under various censoring settings. Table 3 shows that our RFpseudo model outperforms the DeepHit and SurvTRACE by 9.9% and 6.3% in the centralized setting, and our FedPRF outper-

forms their federated variants by 9.6% and 4.0% in the federated setting. Notably, RFpseudo and FedPRF perform significantly better than tree-based RSF and its federated variant FedSurF (which have no censoring handling mechanism like FPV) by 16.3% and 13.2% in the centralized and federated settings, respectively. These findings underscore the

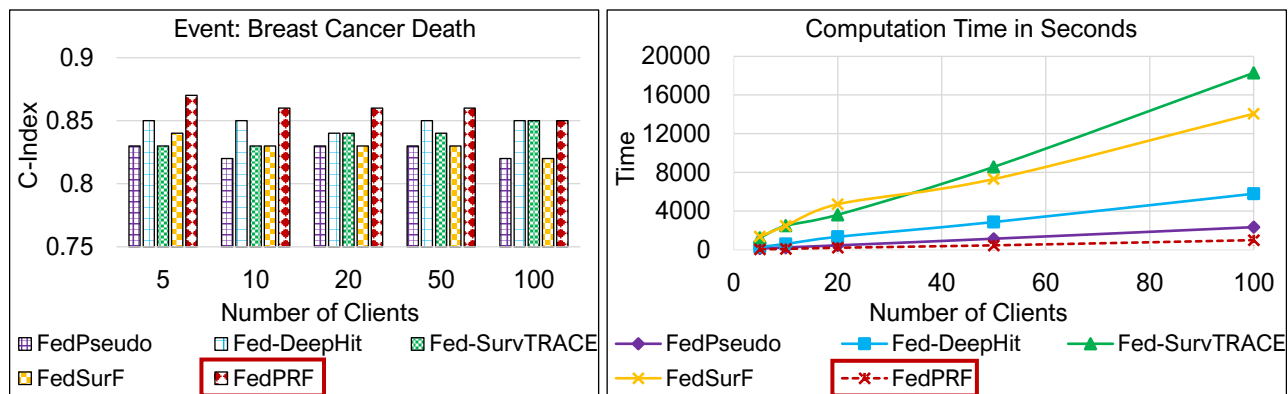


Figure 2: Comparing performance of the models with a growing number of clients in large-scale SEER data

efficiency of FPV in handling challenging scenarios characterized by various and extreme censoring.

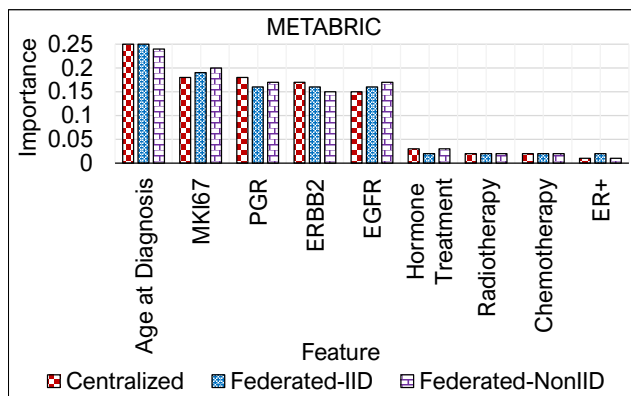


Figure 3: Comparing feature importance between RFPseudo (Centralized) and FedPRF (Federated-IID and Federated-NonIID)

### Interpretability of FedPRF

Our FedPRF offers interpretability by visualizing the importance scores of each feature within the dataset. FedPRF leverages the tree-based model to calculate feature importance by assessing the total reduction in impurity (measured through the Gini index) achieved by each feature during the node-splitting process in decision trees. The feature importance scores are normalized to sum up to 1, ensuring meaningful comparison. This interpretability is crucial for identifying significant and influential features that directly impact survival predictions, often absent in many Deep survival models. In Figure 3, We first show the feature importance of gold-standard centrally trained RFPseudo on the METABRIC dataset. Then, we compare the feature importance derived from the centralized RFPseudo with our FedPRF framework. FedPRF jointly trains the client-specific RFPseudo models in the federated (IID & non-IID) settings to learn a global model, which is further used to obtain the feature importance scores. Notably, FedPRF shows interpretability similar to the centrally trained RFPseudo model.

### Computational Efficiency and Scalability

We conducted a scalability experiment using the large-scale, highly censored (77.1%) SEER dataset (SEER 2023; Wang and Sun 2022), containing 476,746 breast cancer patients with two competing events; death due to breast cancer and heart disease. The number of clients is varied from 5 to 100. Figure 2 demonstrates that FedPRF consistently achieves the highest C-Index for the event of death due to breast cancer across smaller to highly distributed federated settings. Moreover, FedPRF requires 21.5 times less training time than the SurvTRACE and RSF models, 6.2 times less training time than DeepHit, and 2.4 times less training time than the DeepPseudo model. This significant outperformance and reduction in training time highlights the scalability and computational efficiency of FedPRF.

### Conclusion

In this paper, we studied some critical challenges in federated survival analysis (FSA), including competing risks, non-uniform censoring, data heterogeneity, limited data availability, and the computation and communication burden of deep learning-based FSA approaches. Our proposed framework, FedPRF, leverages FPV to handle censoring in distributed datasets and facilitating federated training for distributed right-censored data. FedPRF employs FPV-based Random Forests models, RFPseudo, showcasing computational and communication efficiency with only two rounds of communication between the server and clients. Notably, FedPRF exhibits superior performance in scenarios where clients have limited data and operate within resource-constrained environments, outperforming traditional iterative FSA approaches and our closest counterpart, FedSurF. Furthermore, FedPRF provides interpretable results, aligning with gold-standard centralized models. The promise of the scalability of FedPRF makes it a valuable tool for a wide range of medical institutions.

### Acknowledgments

This work is supported by CAREER grant #2238743 from the US National Science Foundation (NSF).

## References

- Andersen, P. K.; and Pohar Perme, M. 2010. Pseudo-observations in survival analysis. *Statistical methods in medical research*, 19(1): 71–99.
- Andreux, M.; Manoel, A.; Menuet, R.; Saillard, C.; and Simpson, C. 2020. Federated survival analysis with discrete-time cox models. *arXiv preprint arXiv:2006.08997*.
- Antolini, L.; Boracchi, P.; and Biganzoli, E. 2005. A time-dependent discrimination index for survival data. *Statistics in medicine*, 24(24): 3927–3944.
- Archetti, A.; and Matteucci, M. 2023. Federated Survival Forests. *arXiv preprint arXiv:2302.02807*.
- Barrajon, E.; and Barrajón, L. 2020. Effect of right censoring bias on survival analysis. *arXiv preprint arXiv:2012.08649*.
- Cox, D. R. 1972. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2): 187–202.
- Curtis, C.; Shah, S. P.; Chin, S.-F.; Turashvili, G.; Rueda, O. M.; Dunning, M. J.; Speed, D.; Lynch, A. G.; Samarajiwa, S.; Yuan, Y.; et al. 2012. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, 486(7403): 346–352.
- Danks, D.; and Yau, C. 2022. Derivative-based neural modelling of cumulative distribution functions for survival analysis. In *International Conference on Artificial Intelligence and Statistics*, 7240–7256. PMLR.
- Foekens, J. A.; Peters, H. A.; Look, M. P.; Portengen, H.; Schmitt, M.; Kramer, M. D.; Brunner, N.; Janicke, F.; Gelder, M. E. M.-v.; Henzen-Logmans, S. C.; et al. 2000. The urokinase system of plasminogen activation and prognosis in 2780 breast cancer patients. *Cancer research*, 60(3): 636–643.
- Graf, E.; Schmoor, C.; Sauerbrei, W.; and Schumacher, M. 1999. Assessment and comparison of prognostic classification schemes for survival data. *Statistics in medicine*, 18(17-18): 2529–2545.
- Graw, F.; Gerds, T. A.; and Schumacher, M. 2009. On pseudo-values for regression analysis in competing risks models. *Lifetime data analysis*, 15: 241–255.
- Hauschild, A.-C.; Lemanczyk, M.; Matschinske, J.; Frisch, T.; Zolotareva, O.; Holzinger, A.; Baumbach, J.; and Heider, D. 2022. Federated Random Forests can improve local performance of predictive models for various healthcare applications. *Bioinformatics*, 38(8): 2278–2286.
- Ishwaran, H.; Gerds, T. A.; Kogalur, U. B.; Moore, R. D.; Gange, S. J.; and Lau, B. M. 2014. Random survival forests for competing risks. *Biostatistics*, 15(4): 757–773.
- Ishwaran, H.; Kogalur, U. B.; Blackstone, E. H.; and Lauer, M. S. 2008. Random survival forests. *Ann. Appl. Stat.*, 2(3): 841–860.
- Jeanselme, V.; Yoon, C. H.; Tom, B.; and Barrett, J. 2023. Neural Fine-Gray: Monotonic neural networks for competing risks. In *Conference on Health, Inference, and Learning*, 379–392. PMLR.
- Knaus, W. A.; Harrell, F. E.; Lynn, J.; Goldman, L.; Phillips, R. S.; Connors, A. F.; Dawson, N. V.; Fulkerson, W. J.; Califf, R. M.; Desbiens, N.; et al. 1995. The SUPPORT prognostic model: Objective estimates of survival for seriously ill hospitalized adults. *Annals of internal medicine*, 122(3): 191–203.
- Kvamme, H. 2023. Pycox package. <https://github.com/havakv/pycox>. Accessed: 2023-11-19.
- Lee, C.; Zame, W.; Yoon, J.; and Van Der Schaar, M. 2018. Deephit: A deep learning approach to survival analysis with competing risks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.
- McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; and y Arcas, B. A. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, 1273–1282. PMLR.
- Rahimian, S.; Kerkouche, R.; Kurth, I.; and Fritz, M. 2022. Practical challenges in differentially-private federated survival analysis of medical data. In *Conference on Health, Inference, and Learning*, 411–425. PMLR.
- Rahman, M. M.; Matsuo, K.; Matsuzaki, S.; and Purushotham, S. 2021. Deeppseudo: Pseudo value based deep learning models for competing risk analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 479–487.
- Rahman, M. M.; and Purushotham, S. 2023a. Federated Competing Risk Analysis. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, 2106–2115.
- Rahman, M. M.; and Purushotham, S. 2023b. FedPseudo: Privacy-Preserving Pseudo Value-Based Deep Learning Models for Federated Survival Analysis. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 1999–2009.
- Schork, N. J. 2015. Personalized medicine: time for one-person trials. *Nature*, 520(7549): 609–611.
- Schumacher, M.; Bastert, G.; Bojar, H.; Hübner, K.; Olschewski, M.; Sauerbrei, W.; Schmoor, C.; Beyerle, C.; Neumann, R.; and Rauschecker, H. 1994. Randomized 2 x 2 trial evaluating hormonal treatment and the duration of chemotherapy in node-positive breast cancer patients. German Breast Cancer Study Group. *Journal of Clinical Oncology*, 12(10): 2086–2093.
- SEER. 2023. The Surveillance, Epidemiology, and End Results (SEER) Program. <https://seer.cancer.gov>. Accessed: 2023-11-19.
- Terrail, J. O. d.; Ayed, S.-S.; Cyffers, E.; Grimberg, F.; He, C.; Loeb, R.; Mangold, P.; Marchand, T.; Marfoq, O.; Mush-taq, E.; et al. 2022. Flamby: Datasets and benchmarks for cross-silo federated learning in realistic healthcare settings. *arXiv preprint arXiv:2210.04620*.
- Wang, Z.; and Sun, J. 2022. SurvTRACE: Transformers for survival analysis with competing events. In *Proceedings of the 13th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, 1–9.

Zhang, D. K.; Toni, F.; and Williams, M. 2022. A federated cox model with non-proportional hazards. In *Multimodal AI in healthcare: A paradigm shift in health intelligence*, 171–185. Springer.

Zhao, L.; and Feng, D. 2020. Deep neural networks for survival analysis using pseudo values. *IEEE journal of biomedical and health informatics*, 24(11): 3308–3314.