

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

Md Mahmudur Rahman,¹ Sanjay Purushotham,¹

¹ 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. Existing FSA methods often rely on deep learning models, which require substantial data and communication rounds for optimal solutions. 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, particularly on small, decentralized datasets within federated learning (FL) frameworks. However, these approaches have yet to address the challenges specific to FSA, such as data heterogeneity, non-uniform censoring, and the presence of competing events across clients. To address the challenges specific to FSA, we propose a random forests-based FL framework called **FedPRF**. In FedPRF, we introduce federated pseudo values (FPV) to replace the incompletely observed outcomes caused by censoring. FPVs effectively tackle issues related to non-uniform censoring, data heterogeneity, and competing risks in FSA. The FedPRF framework is computationally efficient, requiring only two rounds of communication across clients: (i) computing federated pseudo values and (ii) aggregating a subset of trees trained locally on clients using our proposed **RFpseudo** model. 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. Moreover, our FL framework, FedPRF, effectively preserves the interpretability inherent in centrally trained survival models.

Introduction

Survival analysis is a special technique that analyzes the time duration until a specific event occurs, such as death due to cardiovascular disease, in order to predict the risk of the event over time. It has a wide range of applications from finance (Dirick, Claeskens, and Baesens 2017; Blumenstock, Lessmann, and Seow 2022) to education (Ameri et al. 2016), especially in the field of healthcare (Rahman and Purushotham 2022; Lee et al. 2018), which is the focus of this paper. The conventional survival analysis assumes data is centrally available, either collected by a single, small-sized medical center or multiple medical insti-

tutions or hospitals, such as the Surveillance, Epidemiology, and End Results (SEER) dataset (NCI 2023a). However, challenges arise when dealing with limited data from a single and small medical center, potentially resulting in sub-optimal and less confident predictions. On the other hand, while large-scale data can be obtained through a multi-institution collaboration for better survival predictions, strict data privacy laws and data sharing regulations, such as GDPR and HIPAA, hinder multi-institution collaboration. 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, thus preserving data privacy.

The critical challenges that arise in traditional survival analysis, such as censoring and competing risks, are further exacerbated when transitioning to a federated approach. Censoring, also referred to as the presence of incomplete or partial event information, can occur when a study ends before patients experience the event of interest, when they are lost to follow-up, or when they withdraw from the study. Notably, censoring and event distributions are often non-uniform across different medical centers due to demographic variations, discrepancies in healthcare facilities, or other contextual factors, which pose a critical challenge during federated training. On the other hand, competing risks arise due to the presence of multiple chronic conditions within a patient, commonly referred to as comorbidities, which influence the risk of the primary event of interest. Competing risks are often overlooked in the traditional survival analysis, resulting in biased and inaccurate risk predictions (Rahman et al. 2021; Lee et al. 2018; Danks and Yau 2022). Transitioning to a federated setting amplifies this challenge, primarily due to the imbalanced distribution of competing events across different hospitals. This imbalance introduces additional complexities during the convergence and training phases, further exacerbating bias in the predictions. Moreover, FSA brings its specific challenges, such as data heterogeneity in terms of features and events, imbalanced sample size, and resource constraints among participating medical institutions, which demand innovative techniques and practical solutions.

Recent studies on FSA have developed approaches that either adopt Cox proportional hazards (CoxPH) models or

deep survival models for federated training, which often make underlying assumptions on stochastic processes, such as proportional hazards. Moreover, these approaches do not investigate competing risk problems, the impact of censoring on predictions, and the non-uniformity of censoring rigorously. In addition, they require several communications between the server and institutions or clients during each iteration of iterative computation, resulting in a communication burden. Some recent works addressed the problem by developing approaches that require a single round of communication between the server and clients using the Cox-based model (Imakura et al. 2023) and Random survival forest models (Archetti and Matteucci 2023). However, they lack a censoring handling mechanism and do not account for the competing risks, resulting in sub-optimal performance in survival prediction.

To tackle the aforementioned challenges in FSA, we introduce 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. It adopts the popular machine learning model, Random Forests, employing FPV as response variables, an innovation termed **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 communicationally efficient, requiring only two rounds of communication between the server and clients. Notably, our FedPRF framework employing client-specific RFpseudo models particularly shows improvement in cases where clients possess limited data and operate within resource-constrained environments. FedPRF demonstrates its effectiveness by outperforming the traditional iterative FSA approaches in survival analysis and CRA tasks with less communication overhead. Furthermore, our FPV-based FedPRF framework shows significant improvement over our closest counterpart, the RSF-based federated framework, FedSurF (Archetti and Matteucci 2023), especially under various censoring settings. Additionally, FedPRF provides feature importance in predictions, which is useful for identifying risk factors and better 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 single event and multiple competing events and on synthetic data with different censoring mechanism demonstrates that our FedPRF framework consistently outperforms the state-of-the-art deep survival models as well as random survival forests trained within a federated framework, particularly under various censoring settings.

Method: FedPRF

Notation and Problem Setting: We consider a Federated Survival Analysis (FSA) setup with K clients, each with its own survival dataset denoted by D_k . The dataset for client k is represented as $D_k = \{X_{ik}, Y_{ik}, \delta_{ik}\}$, where i refers to a subject in client k , and k contains a total of n_k sub-

jects, with $1 \leq i \leq n_k$. The covariates for each subject are denoted by $\mathbf{X}_{ik}$, which is a p -dimensional vector. The failure time is denoted by T_{ik} , which represents the time until the occurrence of an event (e.g., death). However, not all subjects experience the event during the study; some may be right-censored. The censoring time for subject i in client k is denoted by C_{ik} , representing the time the observation ends for that subject. The observed event time Y_{ik} is given by $Y_{ik} = \min(T_{ik}, C_{ik})$, which means that Y_{ik} takes the value of T_{ik} if the event occurred before or at the censoring time; otherwise, it takes the value of C_{ik} . Additionally, an event indicator δ_{ik} is defined as $\delta_{ik} = \mathbf{I}(T_{ik} \leq C_{ik})$, where $\mathbf{I}$ is the indicator function. δ_{ik} takes the value 1, 2, 3.... each corresponding to the available competing events, and 0 otherwise. The goal of the FSA is to estimate the conditional survival function $S(t|\mathbf{X}_{ik})$, i.e., the probability that the event time T_{ik} is greater than a specific time t given the covariates $\mathbf{X}_{ik}$ and the cumulative incidence function (CIF), $F_r(t|x_{ik})$, for each event, by utilizing the federated learning algorithms such as FedAvg (McMahan et al. 2017) for aggregating client-specific models trained on their own CRA data D_k .

Overview of FedPRF

We propose a first-of-its-kind FL framework, **FedPRF**, employing federated pseudo value-based random forest model; we call it **RFpseudo** for FSA. Pseudo values, widely studied in statistics and ML community (Andersen et al. 2010; Graw et al. 2009; Binder et al. 2014; Mogensen and Gerds 2013; Rahman et al. 2021), are a natural replacement for the incompletely observed random variable, such as survival function or CIF, and thus, can handle censoring and has been shown promising results when incorporated for modeling survival data (Rahman and Purushotham 2022; Rahman et al. 2021, 2022). However, deriving pseudo values for FSA is challenging. Locally computed Jackknife pseudo values on local clients' data show local consistency (Graw et al. 2009) under the IID assumption; however, the IID assumption often gets violated in federated setup due to the non-IID and non-uniform censoring properties in survival data. Computing pseudo values from aggregated local data in a server can solve the issues; however, strict data privacy laws and regulations hinder data sharing. To address the challenges, we propose federated pseudo values (FPV) for survival analysis, which can easily be extended for CRA. To derive FPVs, the local clients share the non-identifying summary information instead of raw data, such as the number of events and censored observation at a particular time t , to the server, and the server sends the global quantities back to clients to compute FPVs at the client side. Our FedPRF framework is structured into three stages; 1) Federated pseudo values derivation, 2) Local training, and 3) Federated Learning. Next, we briefly describe each stage.

1. Federated pseudo values derivation: Clients first convert the event status into a non-identifying summary table, L_k with the following components; 1) Number of unique time points τ_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, for CRA, the number of events

due to cause r at $t \in \tau_k$, $d_{r,k,t}$ is included. Then clients send the summary table, L_k , to the server. The server creates a global summary 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 table has partial information on the number at risk, i.e., information only available at 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 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 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 the survival function, $J_{ik}(t)$ and CIF, $J_{ikr}(t)$, are computed as $J_{ik}(s) = n\hat{S}_G(s) - (n-1)\hat{S}_G^{G,-ik}(s)$ and $J_{ikr}(s) = n\hat{F}_r^G(s) - (n-1)\hat{F}_r^{G,-ik}(s)$, respectively.

2. Local Training: Once the federated pseudo values are computed, the new dataset for model training is transformed to $D'_k = \{X^k, J_{ik}(s) \text{ (for survival function) or } J_{ikr}(s) \text{ for CIF}\}$. Each client $k \in K$ independently train their corresponding local RFpseudo models, M_k , consisting of an ensemble of trees, denoted as $T_1, T_2, \dots, T_{N_k}$. To obtain the optimal hyperparameters, such as the number of local trees N_k , criterion, max features, maximum tree depth, minimum samples per split, and the number of estimators, we fine-tune the RFpseudo models using the validation set.

3. Federated Learning: The global server first pre-specify the desired number of trees, N_S , and decide the number of trees that each client should send to obtain N_S . The clients send the information about the number of trees, N_k , in their corresponding local model, M_k , to the server so that the server can perform N_S iterations, incrementing a tree counter, $N'_k \leq N_k$, for a randomly chosen client with a probability proportional to the size of their dataset, $|D_k|$. The global server initiates the process by specifying the desired total number of trees, denoted as N_S . Next, it determines the number of trees that each client needs to send to obtain N_S trees. Each client returns the information about the number of trees, N_k , within their respective local model M_k , to the server. With this information, the server performs N_S iterations, incrementing a tree counter, represented as $N'_k \leq N_k$, for a randomly selected client following a uniform distribution. The probability of a client being chosen is proportional to their dataset size, denoted as $|D_k|$. This ensures that clients with larger datasets are more likely to be selected, contributing more trees to the global model. Each client sends N'_k from their corresponding model M_k

to the server to obtain total pre-specified N_S trees. Finally, the server aggregates the trees that are sent by the clients to update a global model.

Experiments

Extensive experiments were conducted to answer the following research questions.

RQ1: To what extent does our proposed FedPRF outperform the FL framework using the state-of-the-art survival models when applied to both simulated and real-world distributed survival datasets?

RQ2: How well does our proposed FedPRF perform on real-world distributed survival datasets that exhibit multiple competing events, non-independent and identically distributed (Non-IID) characteristics, and non-uniform censoring properties, compared to the FL framework utilizing state-of-the-art CRA models?

RQ3: How effectively do our federated pseudo value-based FedPRF framework handle varying types and amounts of censoring in contrast to the FL framework employing baseline models?

RQ4: How successfully do our FedPRF federated framework maintain the interpretability of the gold-standard centralized models?

Datasets

Survival Datasets with Single Event We consider four real-world centrally collected survival datasets with a single event: 1) **METABRIC** (Curtis et al. 2012), 2) **SUPPORT** (Knaus et al. 1995), 3) **GBSG** (Foekens et al. 2000; Schumacher et al. 1994), and 4) **METABRIC-HD (META-HD)** (Curtis et al. 2012) and simulate a federated environment by distributing the datasets to multiple clients. To evaluate the models on a real-world decentralized distributed dataset, we consider **TCGA** (NCI 2023b) dataset, which is collected from 7 regions across the world ((4 in the US: South, West, Midwest, Northeast), Europe, Canada, and Other).

Survival Datasets with Competing Events Competing events influence the risk of the event of interest and should be accounted for unbiased and accurate predictions. To show the effectiveness of our FedPRF framework and RFpseudo model in both federated and centralized competing risk analysis, we consider the **SEER** dataset containing 6 competing events and 28366 patients, out of which 23.2% patients died of cervical cancer (**CC**), 2.6% died due to other cancers (**OCN**), 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). The Surveillance, Epidemiology, and End Results (SEER) (NCI 2023a) program collects data from cancer patients of different geographic locations. The demographic and clinical covariates are included, such as age at diagnosis, race, histology record, Grade, tumor size, cancer stages (TNM staging system), surgery record, cancer therapies, histology, etc.

Synthetic datasets with various censoring settings We generate 12 distributed synthetic datasets for both centralized and federated training, each exhibiting distinct censoring scenarios. These censoring mechanisms include (a) time censoring (TC), (b) interim censoring (IC), (c) case censoring (CC) with varying levels of censoring (25%, 50%, and 75%) (Barrajon and Barrajon 2020), and (d) heterogeneous survival distribution with equal and unequal censoring across clients (HSDEC and HSDUC). These datasets were generated assuming a decentralized setting with 10 clients, with each client having either the same covariate distributions (IID) or different covariate distributions (non-IID). To construct these datasets, we generated survival times from an exponential distribution and 12 numerical covariates from a multivariate normal distribution with mean μ and variance σ^2 , similar to the approach in Lee et al. (Lee et al. 2018). In addition, we generated 2 binary variables from a binomial distribution with probability p .

Experimental Settings

We train the models on both centralized and federated settings and evaluate them on combined test data aggregated from clients.

Centralized Settings: The clients share their local data with the global server, and the server aggregates the datasets. The models are trained on the aggregated training data and evaluated on the aggregated test data.

Federated Settings: For survival analysis with a single event, we simulate two federated setups; **IID** (independent and identically distributed) and **Non-IID** using the real-world survival datasets and another federated setting with distributed TCGA dataset. For competing risk analysis, we simulate two distributed setups for federated training using the **SEER** dataset; 1. SEER-Region, and 2. SEER-NUC. We briefly discuss the federated setups below.

IID: In order to replicate an IID federated setup, we employ a random allocation process where an equal number of subjects are assigned to each participating client. This ensures that the samples within each client have an equal chance of belonging to any client.

Non-IID: To replicate a non-IID setting, 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. For different clients, we assume different quantiles for the skewness of event time distribution.

SEER-Region: To imitate a geographically distributed data environment for federated learning, we divide the SEER data into three clients based on the regions of the hospitals: West, Central, and East.

SEER-NUC: We induce non-uniform censoring (NUC) to the geographically distributed data environment, i.e., 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.

Baseline Models: For federated survival analysis with a single event, we compare our proposed RFpseudo model with 7 baseline survival models; 1) LinearPH (Zhang et al. 2022), 2) DeSurv (Danks and Yau 2022), 3) Neural Fine &

Gray model (NFG) (Jeanselme et al. 2023), 4) Random Survival Forests (RSF) (Ishwaran et al. 2008; Archetti and Matteucci 2023), 5) federated Neural Network-based Cox model with proportional hazard (ph) assumption - NNph (approximation of the model of (Andreux et al. 2020)), 6) Federated Neural Network-based Cox model relaxing proportional hazard (ph) assumption - NNnph (Zhang et al. 2022), 7) DeepHit (Rahimian et al. 2022; Lee et al. 2018). For competing risk analysis, we consider 3 more models along with RSF, DeSurv, NFG, and DeepHit for comparison. They are: 1) Cause-specific Cox proportional hazard model (CS-CoxPH) (Cox 1972), 2) DeepPseudo (Rahman et al. 2021), and 3) SurvTRACE (Wang et al. 2022).

Implementation Details: The data of each client is randomly divided into 80% for training and 20% for testing purposes. We use 20% of training data as validation sets. We ran the experiment five times using different seeds or random states and reported the average performance along with the corresponding standard deviations. To train the Random Forest Regressor with federated pseudo values (RFpseudo), we set the *friedman mse* as a criterion and choose the parameters for *maximum number of features*, *minimum samples per split*, *number of estimators* based on the validation performance. For federated learning, two rounds of communications happen between clients and servers; 1) one is for deriving federated pseudo values, and 2) another round is for federated training with random forests. We choose specific time points for calculating pseudo values and evaluating the models, such as the 10th to 99th percentile of the time horizon with an interval of 10. Only for the eICU dataset, we set the vector of pre-specified time points as [10, 20, 40, 80, 160, 320, 740].

Evaluation Metrics: We evaluate the models using the time-dependent concordance index (C-Index) (Antolini et al. 2005), integrated IPCW Brier score (iBS) (Graf et al. 1999). For competing risk analysis, we employ the cause-specific time-dependent concordance index and iBS as evaluation metrics. We use **pycox** (Kvamme 2022) package to compute the C-Index and iBS. The higher C-Index value and lower iBS value indicate better performance.

Results and Discussion

Performance Comparison in FSA: Table 1 shows the performance comparison of our FedPRF model with baseline survival models on 4 real-world survival datasets and 1 distributed survival dataset, TCGA, in terms of C-Index and iBS. FedPRF shows consistently better performance in real-world survival data compared to baselines. FedPRF achieves up to 5% (DeSurv, RSF), 1% (LinearPH, NFG, DeepHit), 6% (DeepHit), and 11% (NFG, RSF) average improvement in METABRIC, SUPPORT, GBSG, and META-HD datasets, respectively, in terms of C-Index. With respect to iBS, FedPRF obtains up to 10% (RSF), 8% (NFG), 8% (NFG), and 9% (NFG, RSF) average outperformance in METABRIC, SUPPORT, GBSG, and META-HD datasets, respectively. In the TCGA dataset, FedPRF shows better performance in terms of C-index, but both Random Forests-based models, RSF and FedPRF, show poor iBS in a centralized setting.

Table 1: Performance comparisons of the models on the real survival datasets

Metric	Dataset	Federated Setup	Baseline Models							FedPRF
			LinearPH	DeSurv	NFG	RSF	NNnph	NNph	DeepHit	
C-Index	METABRIC	Centralized	0.63 (0.001)	0.66 (0.022)	0.62 (0.005)	0.60 (0.002)	0.66 (0.009)	0.63 (0.008)	0.65 (0.034)	0.65 (0.002)
		Fed-IID	0.63 (0.004)	0.60 (0.010)	0.62 (0.009)	0.62 (0.005)	0.65 (0.019)	0.63 (0.006)	0.66 (0.017)	0.66 (0.005)
		Fed-non-IID	0.65 (0.003)	0.61 (0.013)	0.64 (0.007)	0.63 (0.006)	0.68 (0.014)	0.65 (0.020)	0.68 (0.015)	0.70 (0.007)
	SUPPORT	Centralized	0.60 (0.001)	0.62 (0.004)	0.60 (0.001)	0.60 (0.001)	0.61 (0.004)	0.61 (0.004)	0.60 (0.008)	0.61 (0.002)
		Fed-IID	0.60 (0.001)	0.61 (0.003)	0.60 (0.014)	0.61 (0.001)	0.60 (0.006)	0.61 (0.002)	0.59 (0.007)	0.61 (0.002)
		Fed-non-IID	0.60 (0.002)	0.61 (0.002)	0.60 (0.002)	0.61 (0.002)	0.61 (0.004)	0.61 (0.004)	0.60 (0.008)	0.61 (0.002)
	GBSG	Centralized	0.66 (0.005)	0.67 (0.003)	0.66 (0.002)	0.66 (0.006)	0.66 (0.008)	0.67 (0.008)	0.61 (0.055)	0.67 (0.003)
		Fed-IID	0.64 (0.008)	0.63 (0.005)	0.66 (0.006)	0.68 (0.004)	0.64 (0.011)	0.66 (0.010)	0.63 (0.016)	0.69 (0.003)
		Fed-non-IID	0.63 (0.007)	0.62 (0.014)	0.64 (0.011)	0.63 (0.022)	0.57 (0.013)	0.60 (0.019)	0.57 (0.031)	0.63 (0.015)
	META-HD	Centralized	0.65 (0.008)	0.65 (0.008)	0.56 (0.007)	0.58 (0.012)	0.59 (0.026)	0.63 (0.034)	0.64 (0.010)	0.67 (0.005)
		Fed-IID	0.65 (0.012)	0.64 (0.007)	0.55 (0.018)	0.55 (0.005)	0.65 (0.009)	0.66 (0.007)	0.64 (0.025)	0.66 (0.008)
		Fed-non-IID	0.67 (0.007)	0.66 (0.013)	0.55 (0.015)	0.55 (0.015)	0.65 (0.013)	0.67 (0.007)	0.66 (0.007)	0.65 (0.005)
	TCGA	Centralized	0.76 (0.010)	0.75 (0.011)	0.75 (0.008)	0.74 (0.001)	0.76 (0.010)	0.74 (0.012)	0.76 (0.006)	0.77 (0.001)
		Federated	0.76 (0.002)	0.76 (0.002)	0.74 (0.006)	0.75 (0.002)	0.75 (0.004)	0.76 (0.005)	0.75 (0.010)	0.76 (0.003)
iBS	METABRIC	Centralized	0.19 (0.007)	0.18 (0.004)	0.28 (0.004)	0.30 (0.004)	0.18 (0.003)	0.20 (0.009)	0.19 (0.002)	0.18 (0.001)
		Fed-IID	0.18 (0.001)	0.20 (0.005)	0.28 (0.008)	0.29 (0.003)	0.18 (0.003)	0.19 (0.003)	0.20 (0.002)	0.18 (0.002)
		Fed-non-IID	0.18 (0.002)	0.20 (0.004)	0.27 (0.010)	0.30 (0.001)	0.18 (0.003)	0.19 (0.008)	0.20 (0.005)	0.22 (0.014)
	SUPPORT	Centralized	0.20 (0.001)	0.19 (0.001)	0.27 (0.004)	0.22 (0.001)	0.20 (0.001)	0.20 (0.002)	0.21 (0.002)	0.19 (0.000)
		Fed-IID	0.20 (0.001)	0.20 (0.002)	0.27 (0.020)	0.22 (0.001)	0.20 (0.003)	0.20 (0.003)	0.23 (0.004)	0.20 (0.001)
		Fed-non-IID	0.19 (0.001)	0.19 (0.001)	0.28 (0.004)	0.22 (0.001)	0.20 (0.002)	0.20 (0.003)	0.22 (0.002)	0.19 (0.001)
	GBSG	Centralized	0.20 (0.011)	0.18 (0.001)	0.25 (0.002)	0.26 (0.003)	0.19 (0.003)	0.19 (0.007)	0.23 (0.001)	0.19 (0.001)
		Fed-IID	0.19 (0.002)	0.19 (0.001)	0.26 (0.007)	0.25 (0.003)	0.19 (0.002)	0.19 (0.003)	0.23 (0.001)	0.18 (0.001)
		Fed-non-IID	0.21 (0.004)	0.21 (0.004)	0.31 (0.007)	0.23 (0.004)	0.23 (0.004)	0.22 (0.005)	0.23 (0.002)	0.22 (0.003)
	META-HD	Centralized	0.27 (0.114)	0.19 (0.003)	0.34 (0.008)	0.32 (0.008)	0.20 (0.003)	0.25 (0.075)	0.20 (0.007)	0.19 (0.002)
		Fed-IID	0.22 (0.010)	0.19 (0.003)	0.31 (0.032)	0.32 (0.006)	0.23 (0.009)	0.22 (0.002)	0.21 (0.032)	0.19 (0.002)
		Fed-non-IID	0.22 (0.006)	0.18 (0.005)	0.29 (0.005)	0.30 (0.005)	0.24 (0.008)	0.22 (0.004)	0.18 (0.004)	0.30(0.010)
	TCGA	Centralized	0.11 (0.006)	0.15 (0.007)	0.18 (0.004)	0.17 (0.001)	0.11 (0.006)	0.13 (0.008)	0.14 (0.006)	0.16 (0.004)
		Federated	0.15 (0.001)	0.15 (0.004)	0.17 (0.003)	0.18 (0.001)	0.13 (0.006)	0.15 (0.004)	0.16 (0.003)	0.15 (0.003)

Table 2: Centralized model C-Index performance comparisons on the real CRA datasets

Setup	Dataset	Event	CS-CoxPH	DeepPseudo	RSF	DeSurv	NFG	DeepHit	SurvTRACE	FedPRF
Centralized	SEER (Region)	CC	0.81 (0.000)	0.72 (0.058)	0.85 (0.000)	0.82 (0.007)	0.65 (0.013)	0.88 (0.003)	0.82 (0.002)	0.90 (0.000)
		OCN	0.50 (0.162)	0.68 (0.010)	0.91 (0.002)	0.58 (0.064)	0.69 (0.009)	0.86 (0.007)	0.84 (0.004)	0.91 (0.002)
		CVD	0.84 (0.002)	0.81 (0.061)	0.90 (0.001)	0.68 (0.056)	0.85 (0.003)	0.88 (0.005)	0.87 (0.005)	0.91 (0.002)
		CMD	0.48 (0.125)	0.75 (0.081)	0.85 (0.005)	0.49 (0.048)	0.82 (0.005)	0.75 (0.036)	0.83 (0.009)	0.81 (0.009)
		ID	0.53 (0.042)	0.69 (0.038)	0.87 (0.005)	0.61 (0.066)	0.76 (0.006)	0.82 (0.033)	0.80 (0.012)	0.86 (0.008)
		OCS	0.76 (0.009)	0.70 (0.027)	0.89 (0.004)	0.58 (0.059)	0.73 (0.008)	0.80 (0.021)	0.78 (0.013)	0.86 (0.002)
	SEER (NUC)	CC	0.79 (0.015)	0.64 (0.229)	0.83 (0.006)	0.80 (0.012)	0.66 (0.006)	0.88 (0.010)	0.84 (0.028)	0.88 (0.006)
		OCN	0.78 (0.010)	0.69 (0.089)	0.84 (0.026)	0.69 (0.018)	0.62 (0.027)	0.85 (0.018)	0.83 (0.025)	0.89 (0.016)
		CVD	0.80 (0.007)	0.69 (0.131)	0.87 (0.018)	0.71 (0.045)	0.82 (0.021)	0.82 (0.018)	0.83 (0.026)	0.90 (0.009)
		CMD	0.84 (0.023)	0.70 (0.100)	0.89 (0.029)	0.61 (0.052)	0.83 (0.016)	0.76 (0.046)	0.79 (0.051)	0.90 (0.027)
		ID	0.68 (0.207)	0.66 (0.147)	0.90 (0.043)	0.63 (0.066)	0.75 (0.083)	0.81 (0.107)	0.83 (0.080)	0.87 (0.068)
		OCS	0.77 (0.026)	0.63 (0.100)	0.84 (0.025)	0.67 (0.061)	0.72 (0.039)	0.82 (0.022)	0.80 (0.067)	0.87 (0.013)
Federated	SEER (Region)	CC	0.81 (0.000)	0.78 (0.011)	0.83 (0.001)	0.78 (0.014)	0.64 (0.016)	0.85 (0.004)	0.82 (0.005)	0.87 (0.001)
		OCN	0.82 (0.001)	0.74 (0.013)	0.89 (0.007)	0.58 (0.031)	0.70 (0.008)	0.84 (0.008)	0.81 (0.01)	0.87 (0.005)
		CVD	0.84 (0.000)	0.85 (0.004)	0.89 (0.004)	0.67 (0.050)	0.84 (0.002)	0.89 (0.007)	0.87 (0.009)	0.90 (0.006)
		CMD	0.82 (0.002)	0.79 (0.006)	0.84 (0.015)	0.60 (0.018)	0.82 (0.005)	0.79 (0.041)	0.76 (0.029)	0.77 (0.016)
		ID	0.73 (0.018)	0.71 (0.009)	0.83 (0.009)	0.64 (0.043)	0.77 (0.005)	0.84 (0.024)	0.77 (0.029)	0.81 (0.013)
		OCS	0.75 (0.002)	0.72 (0.007)	0.85 (0.007)	0.57 (0.061)	0.74 (0.003)	0.75 (0.015)	0.74 (0.02)	0.82 (0.006)
	SEER (NUC)	CC	0.79 (0.013)	0.72 (0.062)	0.80 (0.006)	0.76 (0.011)	0.63 (0.027)	0.83 (0.008)	0.82 (0.007)	0.85 (0.005)
		OCN	0.79 (0.007)	0.65 (0.033)	0.78 (0.041)	0.66 (0.031)	0.62 (0.011)	0.83 (0.03)	0.81 (0.025)	0.85 (0.016)
		CVD	0.80 (0.009)	0.77 (0.023)	0.82 (0.013)	0.65 (0.047)	0.78 (0.016)	0.86 (0.014)	0.82 (0.017)	0.86 (0.014)
		CMD	0.84 (0.022)	0.79 (0.055)	0.85 (0.028)	0.65 (0.052)	0.82 (0.011)	0.84 (0.049)	0.78 (0.078)	0.87 (0.008)
		ID	0.75 (0.027)	0.73 (0.072)	0.84 (0.063)	0.72 (0.025)	0.73 (0.086)	0.87 (0.038)	0.84 (0.058)	0.84 (0.087)
		OCS	0.76 (0.031)	0.69 (0.024)	0.80 (0.016)	0.68 (0.050)	0.71 (0.024)	0.82 (0.028)	0.80 (0.051)	0.83 (0.023)

Performance Comparison in Federated CRA: Table 2 shows that our FedPRF performs significantly better (up to 25% (DeSurv)) than all the baselines except for the RSF, which shows similar performance in the centralized setting on SEER-Region data in terms of C-Index. On SEER-NUC data, FedPRF achieves at least 2.3% (RSF) and at most 21.7% (DeepPseudo) improvement than the baselines. FedPRF obtains up to 20% and 16.3% (DeSurv) improvements in the federated settings on SEER-Region and SEER-NUC

datasets, respectively, compared to the baselines.

Performance under Various Censoring Settings: To show the effectiveness of the federated pseudo value-based Random Forest model (FedPRF) in handling different types and amounts of censoring, we compare FedPRF with state-of-the-art CRA models, DeepHit and SurvTRACE along with another Random Forests-based model RSF under various censoring settings. Table 3 shows that our FedPRF model outperforms the DeepHit and SurvTRACE by 10.8%

Table 3: C-Index performance comparisons on the distributed synthetic CRA datasets with different censoring settings

Dataset	Centralized				Federated			
	RSF	DeepHit	SurvTRACE	FedPRF	RSF	DeepHit	SurvTRACE	FedPRF
TC-IID	0.67 (0.007)	0.69 (0.006)	0.70 (0.004)	0.80 (0.007)	0.68 (0.003)	0.69 (0.006)	0.70 (0.007)	0.74 (0.005)
TC-Non-IID	0.65 (0.001)	0.68 (0.009)	0.72 (0.014)	0.81 (0.004)	0.65 (0.007)	0.66 (0.007)	0.70 (0.004)	0.75 (0.006)
IC-IID	0.66 (0.001)	0.68 (0.010)	0.69 (0.003)	0.79 (0.013)	0.67 (0.001)	0.69 (0.006)	0.69 (0.003)	0.73 (0.012)
IC-Non-IID	0.64 (0.003)	0.68 (0.012)	0.69 (0.015)	0.80 (0.007)	0.64 (0.008)	0.66 (0.006)	0.68 (0.005)	0.75 (0.009)
CC25-IID	0.64 (0.008)	0.77 (0.033)	0.92 (0.013)	0.82 (0.018)	0.67 (0.002)	0.73 (0.007)	0.91 (0.030)	0.84 (0.024)
CC25-Non-IID	0.64 (0.009)	0.65 (0.004)	0.67 (0.006)	0.78 (0.006)	0.63 (0.005)	0.65 (0.006)	0.67 (0.009)	0.73 (0.002)
CC50-IID	0.63 (0.015)	0.80 (0.005)	0.92 (0.008)	0.86 (0.002)	0.68 (0.007)	0.74 (0.006)	0.91 (0.004)	0.85 (0.003)
CC50-Non-IID	0.64 (0.009)	0.67 (0.007)	0.67 (0.009)	0.81 (0.003)	0.64 (0.006)	0.65 (0.009)	0.68 (0.007)	0.73 (0.004)
CC75-IID	0.61 (0.023)	0.82 (0.003)	0.88 (0.014)	0.89 (0.002)	0.68 (0.010)	0.73 (0.009)	0.86 (0.004)	0.85 (0.003)
CC75-Non-IID	0.64 (0.006)	0.72 (0.007)	0.67 (0.008)	0.87 (0.004)	0.62 (0.010)	0.65 (0.002)	0.71 (0.009)	0.74 (0.011)
DSDEC	0.64 (0.027)	0.66 (0.025)	0.70 (0.008)	0.78 (0.011)	0.60 (0.005)	0.68 (0.008)	0.68 (0.004)	0.71 (0.015)
DSDEC	0.66 (0.007)	0.67 (0.017)	0.70 (0.006)	0.78 (0.012)	0.62 (0.008)	0.68 (0.006)	0.69 (0.005)	0.73 (0.01)

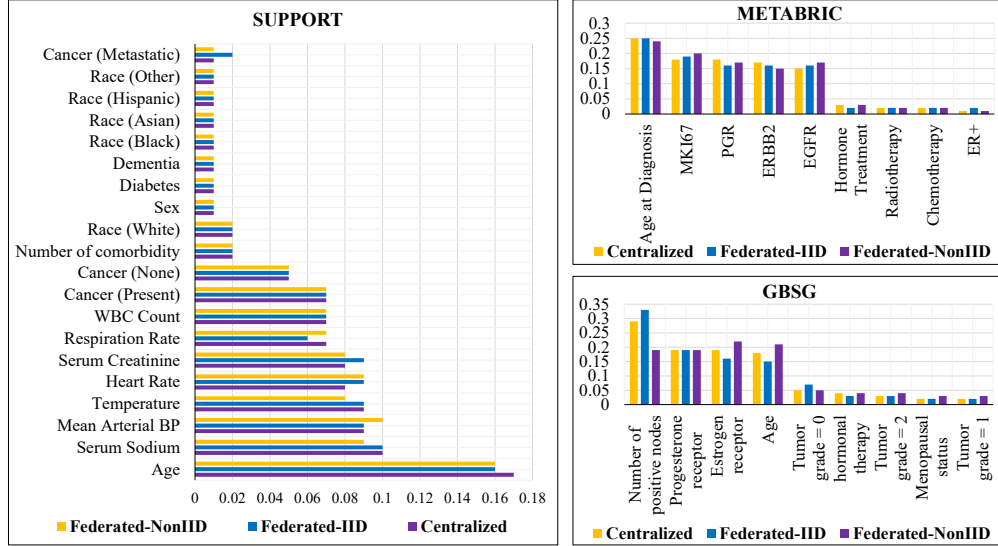


Figure 1: Comparing interpretation between RFpseudo (Centralized) and FedPRF (Federated-IID and Federated-NonIID).

and 7.2% in the centralized setting and 7.8% and 2.3% in the federated setting. Interestingly, FedPRF performs significantly better than Random Forests-based RSF (which does not have any censoring handling mechanism like federated pseudo values), by 17.3% and 11.4% in the centralized and federated settings, respectively. The results indicate that our federated pseudo values are efficient in handling different types and amounts of censoring and can improve the survival predictions under extreme censoring scenarios.

Interpretability of FedPRF: Our FedPRF provides interpretability that helps to identify important and influential features that impact the survival predictions, which most Deep survival models lack. In Figure 1, We compare the feature contributions obtained by FedPRF between Centralized and Federated (IID & non-IID) settings in real-world survival datasets; METABRIC, SUPPORT, and GBSG. Interestingly, our FedPRF not only offers interpretability but also maintains the similar interpretability of the gold-standard centrally trained model when trained in a federated learning framework. The results demonstrate that our FedPRF provides both accurate and trustworthy survival predictions in both centralized and federated settings.

Discussion

The paper introduces a communication-efficient federated framework called FedPRF for survival analysis, addressing challenges related to limited data sharing and data privacy in multi-institute medical collaborations. The experiments on real-world survival data with limited subjects demonstrate that deep survival models perform sub-optimally in limited-sized data, whereas the ensemble approach RFpseudo-based FL framework (FedPRF) showed significant advantages in survival analysis with limited data. FedPRF leverages federated pseudo-values (FPV) for handling censoring and enabling regression analysis as response variables, leading to superior performance compared to baseline approaches, including the closest architecture RSF. FedPRF was further evaluated in the presence of competing events and various censoring settings, including non-uniform censoring across clients. FedPRF successfully overcame these challenges by utilizing federated pseudo-values and the ensemble approach (random forest). In addition, experiments on large distributed data settings with competing events based on locations (SEER-Region) further confirm the model's effectiveness in challenging scenarios. Furthermore, Fed-

PRF maintained similar interpretations to gold-standard centrally trained survival models, ensuring reliable results when trained on the federated framework. The FedPRF framework offers promising applications in distributed healthcare data settings, especially in resource-constrained environments, allowing decentralized training while preserving data privacy and achieving accurate survival predictions.

Conclusion

In this paper, we studied some critical challenges in federated survival analysis (FSA), including competing risks, non-uniform censoring across clients, and data heterogeneity. To address these challenges, we introduced a pseudo-value-based federated learning framework called FedPRF. This framework employs random forest models for local training to estimate subject-specific quantities of interest in survival analysis and CRA, such as survival probabilities and cumulative incidence functions. The existing FSA approaches employing deep survival models suffer from computation and communication burdens while the FSA approach with random survival forest does not investigate the different censoring impacts on predictions and does not impose any censoring handling mechanism. Our proposed FedPRF framework effectively addresses critical challenges in FSA by leveraging federated pseudo-values. FedPRF provides accurate and trustworthy predictions in general survival analysis problems, especially in distributed settings with limited data and resources.

Acknowledgments

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

References

- Ameri, S.; Fard, M. J.; Chinnam, R. B.; and Reddy, C. K. 2016. Survival analysis based framework for early prediction of student dropouts. In *Proceedings of the 25th ACM international on conference on information and knowledge management*, 903–912.
- Andersen; et al. 2010. Pseudo-observations in survival analysis. *Statistical methods in medical research*.
- Andreux; et al. 2020. Federated survival analysis with discrete-time Cox models. *arXiv preprint arXiv:2006.08997*.
- Antolini; et al. 2005. A time-dependent discrimination index for survival data. *Statistics in medicine*.
- Archetti, A.; and Matteucci, M. 2023. Federated Survival Forests. *arXiv preprint arXiv:2302.02807*.
- Barrajon, E.; and Barrajon, L. 2020. Effect of right censoring bias on survival analysis. *arXiv preprint arXiv:2012.08649*.
- Binder; et al. 2014. Pseudo-observations for competing risks with covariate dependent censoring. *Lifetime data analysis*, 20(2): 303–315.
- Blumenstock, G.; Lessmann, S.; and Seow, H.-V. 2022. Deep learning for survival and competing risk modelling. *Journal of the Operational Research Society*, 73(1): 26–38.
- Cox, D. R. 1972. Regression models and life-tables. *Journal of the Royal Statistical Society*.
- Curtis, C.; Shah, S. P.; Chin, S.-F.; Turashvili, G.; Rueda, O. M.; Dunning, M. J.; Speed, D.; Lynch, A. G.; Samara-jjiwa, 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.
- Dirick, L.; Claeskens, G.; and Baesens, B. 2017. Time to default in credit scoring using survival analysis: a benchmark study. *Journal of the Operational Research Society*, 68: 652–665.
- Foekens, J. A.; et al. 2000. The urokinase system of plasminogen activation and prognosis in 2780 breast cancer patients. *Cancer research*, 60(3): 636–643.
- Graf; et al. 1999. Assessment and comparison of prognostic classification schemes for survival data. *Statistics in medicine*.
- Graw; et al. 2009. On pseudo-values for regression analysis in competing risks models. *Lifetime Data Analysis*.
- Imakura, A.; Tsunoda, R.; Kagawa, R.; Yamagata, K.; and Sakurai, T. 2023. DC-COX: Data collaboration Cox proportional hazards model for privacy-preserving survival analysis on multiple parties. *Journal of Biomedical Informatics*, 137: 104264.
- Ishwaran, H.; Kogalur, U. B.; Blackstone, E. H.; and Lauer, M. S. 2008. Random survival forests.
- Jeanselme, V.; Yoon, C. H.; Tom, B.; and Barrett, J. 2023. Neural Fine-Gray: Monotonic neural networks for competing risks. *arXiv preprint arXiv:2305.06703*.
- 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. 2022. Pycox package.
- Lee; et al. 2018. Deephit: A deep learning approach to survival analysis with competing risks. In *AAAI*.
- McMahan, B.; et al. 2017. Communication-efficient learning of deep networks from decentralized data. In *AISTATS*.
- Mogensen, U. B.; and Gerds, T. A. 2013. A random forest approach for competing risks based on pseudo-values. *Statistics in medicine*, 32(18): 3102–3114.
- NCI. 2023a. SEER Dataset.
- NCI. 2023b. TCGA Dataset.
- Rahimian; et al. 2022. Practical Challenges in Differentially-Private Federated Survival Analysis of Medical Data. In *CHIL*.
- Rahman; et al. 2021. Deeppseudo: Pseudo value based deep learning models for competing risk analysis. In *AAAI*.

- Rahman; et al. 2022. FedPseudo: Pseudo value-based Deep Learning Models for Federated Survival Analysis. *arXiv preprint arXiv:2207.05247*.
- Rahman, M. M.; and Purushotham, S. 2022. Fair and Interpretable Models for Survival Analysis. In *KDD*.
- Rahman, M. M.; and Purushotham, S. 2023a. Federated learning for competing risk analysis in healthcare. In *International Workshop on Federated Learning for Distributed Data Mining*.
- 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*.
- 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.
- Wang; et al. 2022. Survtrace: Transformers for survival analysis with competing events. In *ACM BCB*.
- Zhang; et al. 2022. A Federated Cox Model with Non-Proportional Hazards. *arXiv preprint arXiv:2207.05050*.