Federated learning for competing risk analysis in healthcare

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

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

ABSTRACT

Performing survival analysis on distributed healthcare data is an important research problem, as the existing privacy laws and emerging data-sharing regulations prohibit the sharing of sensitive patient data across multiple institutions. The distributed healthcare survival data is typically heterogeneous, non-uniformly censored, and comes from patients with multi-morbidities (or competing risks), which may lead to biased and inaccurate risk predictions. To address these challenges, we propose federated learning for survival analysis with competing risks. Specifically, (a) we propose a simple algorithm to estimate consistent federated pseudo values for survival analysis with competing risks and censoring; and (b) we introduce a novel and flexible federated pseudo-value-based deep learning framework named **FedCRA**, where we employ a transformer-based model; named **TransPseudo**, to enable subject-specific prediction of the marginal risk of an event while preserving the data privacy. Extensive experiments on two real-world distributed healthcare datasets with non-IID and non-uniform censoring properties and on synthetic data with different censoring settings demonstrate that our FedCRA framework with the TransPseudo model performs better than the federated learning framework with state-of-the-art survival models for competing risks analysis (CRA).

CCS CONCEPTS

• **Mathematics of computing** → **Survival analysis**; • **Computing methodologies** → **Distributed artificial intelligence**; • **Security and privacy**;

KEYWORDS

Survival analysis, Federated Learning, Competing Risks, Censoring

1 INTRODUCTION

Multimorbidity, i.e., the presence of two or more chronic conditions in a person, is a prevalent and urgent problem in healthcare [7, 17], especially among older patients. In the United States, the prevalence of more than 2 morbidities was 59.6%, whereas the percentage was 92% among individuals over 65 years during 2013-14 [12]. Multimorbid patients face the risk of adverse outcomes, such as mortality, due to different clinically significant diseases like cancer or heart disease. These outcomes, such as death from cancer or death from heart disease, are considered competing events, and their risks are referred to as competing risks [14, 21]. The occurrence of one competing event precludes the occurrence of other competing events, thus, influencing the risk of other competing events. The standard survival analysis, also known as time-to-event analysis, ignores the competing risks or treats the competing events as censoring in the marginal risk prediction of the event of interest, leading to biased and inaccurate risk predictions. Recently, machine learning
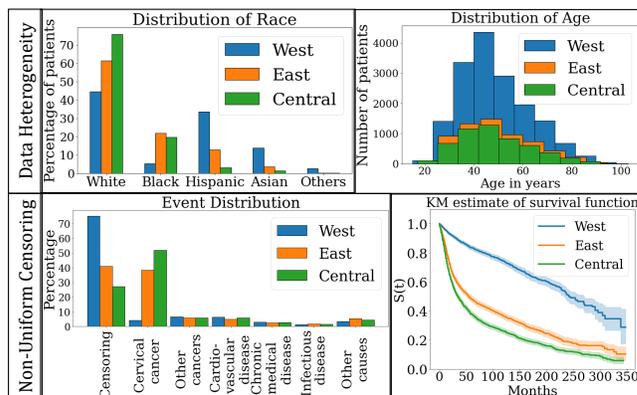
Figure 1: Challenges of Federated CRA in SEER dataset

models have been developed for competing risk analysis (CRA) [11, 14, 21, 26], which have shown promising improvement over traditional statistical CRA models [6, 7] and achieved state-of-the-art performance. However, a limited amount of survival data are typically collected by a single medical center due to resource and privacy constraints, which is insufficient to develop an efficient machine learning-based survival model for accurate risk predictions. The National Cancer Institute took a great initiative to collect large-scale survival data on registered cancer patients from the hospitals of several regions in the USA through the Surveillance, Epidemiology, and End Results (SEER) [10] program. However, such initiatives are expensive and time-consuming, and far fewer. On the other hand, while collaborations across multiple medical centers to gather harmonized large-scale datasets is feasible, such collaborations are hindered by the strict privacy laws and regulations on user data sharing, such as Health Insurance Portability and Accountability Act (HIPAA) and European Union's General Data Protection Regulation (GDPR). To overcome data sharing limitations from multiple institutions, Federated Learning (FL) [16] has been proposed as a viable solution, where models are shared and trained among multiple institutions instead of sharing data. In this paper, we study the solution of competing risk analysis problem under the federated learning settings (with the assumption that data sharing is infeasible) and propose a federated competing risk analysis (FedCRA) framework. We are inspired by the recent success of federated survival analysis [2, 19, 20, 25] in achieving performance close to the gold-standard centralized training.

Even though CRA is a well-studied problem [11, 14, 21, 26], we found no work on modeling competing risks in an FL setting, especially on some real-world challenges of Federated CRA including

non-IID data, and non-uniform censoring. In this paper, we investigate the following challenges in Federated CRA and develop novel methods to provide potential solutions to the problems.

- **Challenge 1: Data Heterogeneity:** The dissimilarities in patient demographics, event distributions, and clinical histories among various collaborative medical centers result in non-IID data.
- **Challenge 2: Non-Uniform Censoring (NUC):** Censoring, i.e., partial information of subjects' event status, is a key challenge in survival analysis that leads to biased and inaccurate risk predictions [23]. This bias exacerbates in FL due to the non-uniformity of censoring distributions across clients, as shown in Figure 1 (Subplot: KM estimate of the survival estimate). Moreover, the censoring rate also varies across different medical centers, leading to heterogeneous data distributions and sub-optimal performance of the local survival models.

Proposed Solutions: To address the challenges in federated CRA, we propose a first-of-its-kind federated pseudo-values-based deep learning framework, **FedCRA**, to solve CRA in an FL manner. We also propose a novel client-specific Transformer-based CRA model, **TransPseudo**. FedCRA jointly trains TransPseudo models in a federated framework to learn a global updated model without sharing raw data, which is further used to predict the probability of an event at or before time t due to cause k, i.e., cause-specific cumulative incidence function (**CIF**), given the covariates for a patient in a client. *We introduce federated pseudo-values to efficiently handle non-uniform censoring and account for the heterogeneity in event time and censoring distribution.* Our federated pseudo values preserve patient data privacy since they are derived from aggregated summary information (containing no identifying information) instead of raw data.
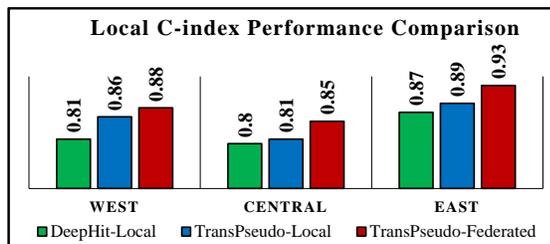
Figure 2: Benefit of participating in our FedCRA framework (TransPseudo-Federated). Our federated pseudo value-based TransPseudo model shows improvement in average C-Index performance over the local DeepHit and local TranPseudo models on distributed non-IID SEER datasets for all three clients (WEST, CENTRAL, and EAST regions).

We show in figure 2 that all decentralized clients can improve their prediction accuracy for CRA by participating in our FedCRA framework, thus improving patient outcomes. We also conduct extensive experiments on multiple realistic federated settings with real CRA data: SEER [21] and eICU [18] as well as on several synthetic datasets with different censoring settings to demonstrate the efficacy of our FedCRA framework to improve the model's

performance while preserving data privacy and to address data heterogeneity and non-uniform censoring. We show that our proposed FedCRA framework achieves close performance to the gold-standard centralized training on centrally aggregated data. Moreover, FedCRA performs better than the FL framework with SOTA CRA models.

2 PROPOSED FEDCRA FRAMEWORK

In this section, we will first introduce the notation, preliminaries, and the Federated CRA problem settings and then describe our proposed FedCRA Framework.

Notations and Problem Settings: Let K medical centers/clients participate in federated learning to learn a global model for conducting CRA. Each client has its own local survival data with multiple competing events, and individuals in a client are followed up to a particular study period. For an individual i in client k, we denote the CRA data (D_k) as a tuple, $D_k = \{X_{ik}, Y_{ik}, \delta_{ik}\}$, where $\mathbf{X_{ik}} = (X_{ik1}, X_{ik2}, ..., X_{ikp})$, Y_{ik}, δ_{ik} respectively are the p-dimensional vector of observed covariates, observed event time and event indicator for individual i in client k. The event indicator, $\delta_{ik} = r$, if the ith individual in client k experienced the event r, where $r \in \{1, 2, ..., L\}$ events and $\delta_{ik} = 0$, if the individual is censored. Let T_{ik} be the event time for individual i in client k. The first event of the L number of events is only observed, and therefore the event time of ith individual in client k is the earliest event time, defined as $T_{ik} = \min(T_{ik}^1, T_{ik}^2, ..., T_{ik}^L)$. The observed time for individual i in client k is the minimum of event time and censoring times C_{ik}, $Y_{ik} = \min(T_{ik}, C_{ik})$. For individual i in client k, let us define the cumulative incidence function (CIF), denoted as $F_r(t|x_{ik})$, as the probability of an occurring event r at or before time t given the covariate X_{ik}; $F_r(t|x_{ik}) = P(T < t, \delta = r|x_{ik})$. *The objective in Federated CRA problem is to learn a global model and accurately predict the CIF, $F_r(t|x_{ik})$, for each event, by utilizing the federated learning algorithms such as FedAvg [16], to aggregate the client-specific models trained on their own CRA data D_k.*

Overview of our FedCRA Framework: The recent success of pseudo value-based deep neural networks in standard survival analysis [22, 23], CRA [21] and federated survival analysis (FSA) [20] has motivated us to develop a pseudo-value-based deep learning model for solving CRA in a federated manner. However, federated CRA is a more intricate problem due to the complex interaction between covariates and competing events and the real-world challenges in federated CRA, such as non-IID data and non-uniform censoring. As a result, simple deep neural networks need more learning capacity for obtaining accurate predictions and satisfactory performance in federated CRA. On the other hand, the transformer-based model has a strong learning capacity and has achieved SOTA performance in a wide range of tasks [8, 26], which motivates us to design transformer-based models for performing complex CRA in an FL manner. Therefore, we propose an FL framework for CRA; we call it **FedCRA**, where we first derive the pseudo values in a federated fashion and use them as response variables (ground truth) in our proposed client-specific Transformer-based models, **TransPseudo**. Then we conduct federated training with the TransPseudo models for learning the global model parameters.

Federated Pseudo Values for CIF: Due to censoring, i.e., incomplete event status or ground truth, the direct application of standard regression or classification techniques becomes infeasible. Pseudo values can be considered the natural replacement for the incompletely observed CIF [4], thus, can be used to efficiently handle censoring [1, 21, 27]. The traditional Jackknife pseudo values, computed on local client data, exhibit local consistency but suffer from global inconsistency due to heterogeneity in the event and censoring distributions of the clients. Computing pseudo values on merged data in a central server from clients can address the problem; however, data privacy laws and regulations make it infeasible. Moreover, the pseudo values, requiring leave-one-out computation for each subject in a sample, become computationally expensive and infeasible for federated CRA as the number of clients and sample size increases. To overcome these challenges, we introduce a novel federated pseudo values derivation approach for federated CRA, which uses the summary information from the clients instead of raw data that do not disclose the patient identifying information and, thus, preserve the data privacy. The leave-one-out computation required for pseudo values derivation is performed in parallel on the clients, reducing computational complexity and enabling scalability for FL. Furthermore, our federated pseudo values are directly derived from the estimate of global CIF, incorporating global information on the time-to-event distribution to account for the heterogeneity in client data. The federated pseudo values derivation approach is described as follows.

Federated Pseudo Values Derivation: First, each client transforms the inputs, i.e., event time and status, into summary information, such as the number of subjects at risk at time t_0; (R_{0k}), number of events (d_k), number of subjects experienced event r; (d_{rk}) and number of censored (c_k) at a vector of unique time points (τ_k) in the local data. Clients send the summary information to the global server, where the server aggregates the summary information at the union of the vector of unique time points of clients, $\tau = \cup_k \tau_k$, to form a global partial table. The global partial table contains the total number of subjects at risk at time $t_0 \in \tau$, $R_0 = \sum_{k=1}^{K} R_{0k}$, the total number of events, $d = \sum_{k=1}^{K} d_k$, the total number of subjects who experienced event r; $d_r = \sum_{k=1}^{K} d_{rk}$ and the total number of censored, $c = \sum_{k=1}^{K} c_k$ at the vector of unique time points τ. The server fills up the partial table by computing the number of subjects at risk at subsequent time points ($t_1, t_2, ..$) using the formula: $R_{t_j} = R_{t_{j-1}} - R_{t_{j-1}} - d_{t_{j-1}} - c_{t_{j-1}}$. Then, the server compute the global survival function as $\hat{S}^G(t) = \prod_{t_j \in \tau \leq t}(1 - \frac{d_{t_j}}{R_{t_j}})$ and the global CIF as $\hat{F}_r^G(t) = \sum_{t_j \in \tau \leq t} \hat{S}^G(t)\frac{d_{rt_j}}{R_{t_j}}$. The global server sends the global partial table and the global CIF to the clients. Clients first create leave-one-out global partial tables by omitting the ith subjects from the risk set $R(t_0)$ and from d, d_r, c at which time point the event or censoring occurred (denoted as $d^{-ik}, d_r^{-ik}, c^{-ik}$). Then clients fill the risk set at the subsequent time points in the leave-one-out global partial table using the following formula: $R_{t_j}^{-ik} = R_{t_{j-1}}^{-ik} - d_{t_{j-1}}^{-ik} - c_{t_{j-1}}^{-ik}$, where $t_j \in \tau$. Using the complete table, clients compute the leave-one-out global CIF as $\hat{F}_r^{G-ik}(t) = \sum_{t_j \in \tau \leq t} \hat{S}^{G-ik}(t)\frac{d_{rt_j}^{-ik}}{R_{t_j}^{-ik}}$. Finally, each client computes

the pseudo values for their subjects using the following equation: $J_{ikr}(t) = n\hat{F}_r^G(t) - (n-1)\hat{F}_r^{G-ik}(t); i = 1, 2, .., n_k, k = 1, 2, .., K$. Here, n is the total number of subjects of all clients, i.e., $n = \sum_{i=1}^{K} n_k$ and t can be a pre-specified single time point or a vector of time points, Υ (provided by the user). $\hat{F}_r^G(t)$ is the AJ estimate of the global CIF for event r at time t and $\hat{F}_r^{G-ik}(t)$ is the leave-one-out AJ estimate of the global CIF, obtained by omitting i^{th} subject from client k. For a subject i in client k, pseudo values are calculated for all R causes at a vector of pre-specified time points, Υ. Our federated pseudo values are directly derived from the consistent estimate of global cumulative incidence function $F_r^G(t)$ [5] and can be shown to be consistent by following the lemma 2 in [9].

Proposed TransPseudo Model: Our TransPseudo model adapts the FT-Transformer (Feature Tokenizer + Transformer) architecture [8] and uses covariates as input and predicts CIF via federated pseudo values as response variables (ground truth). First, a Feature Tokenizer transforms the inputs X (both numerical and categorical covariates) into embeddings $Q \in R^{P \times q}$. The embeddings of all covariates (both numeric and categorical) are stacked to create an embedding matrix $\mathbf{Q}$. Then, a transformer module first appends the output token [OT] embedding to the embedding matrix $\mathbf{Q}$. After that a stack of L Transformer layers $H_1, H_2, ..., H_L$ are applied as: $Q_i = H_i(Q_{i-1})$ where $Q_0 = stack[[OT], Q]$. TransPseudo model predicts the CIF using the final representation of the output token [OT] as $\hat{F}(\tau) = Sigmoid(Linear(ReLU(LayerNorm(H_L^{[OT]}))))$. See the paper [8] for the details of the transformer module. The architecture of our proposed TransPseudo model is shown in figure 3.

We introduce a pseudo-value-based binary cross-entropy (PBC) loss for Federated CRA with R competing events. The loss $L_k^{PBC}(t)$ for client k at time t is defined as,

$$L_k^{PBC}(t) = \frac{1}{n_k * R} \sum_{r=1}^{R} \sum_{i=1}^{n_k} -\Big[\mathbb{I}(J_{ikr}(t) > 0.5)log\hat{F}_r(t|x_{ik})$$
$$+ (1 - \mathbb{I}(J_{ikr}(t) > 0.5))log(1 - \hat{F}_r(t|x_{ik}))\Big]$$

where $\hat{F}_r(t|x_{ik})$ and $J_{ikr}(t)$ respectively are the predicted CIF and the pseudo values at time point t for i^{th} individual in client k. Note that t can be a prespecified single time point or a vector of time points, Υ, where pseudo values are calculated based on the research interest. If we calculate the loss for a vector of time points, then the final loss is $\sum_t L_k^{PBC}(t)$.

Federated Training: Our FedCRA framework employs our proposed client-specific TransPseudo models that communicate with a global server. During each communication round, the global server sends the clients a global TransCRA model represented by w^v. The local clients then update their local models by incorporating the global model parameters and training their models using their respective local data. The newly trained local models, denoted as Δw_k^v, are then sent back to the global server. Using a standard FL algorithm, FedAvg [16], the global server aggregates the updates from the local models to update the global model. Subsequently, the updated global model is sent back to the local clients by the global server. This process is repeated for a specified number of communication rounds, denoted as V. Once the V rounds are completed, the globally updated model is utilized to make personalized CIF predictions.
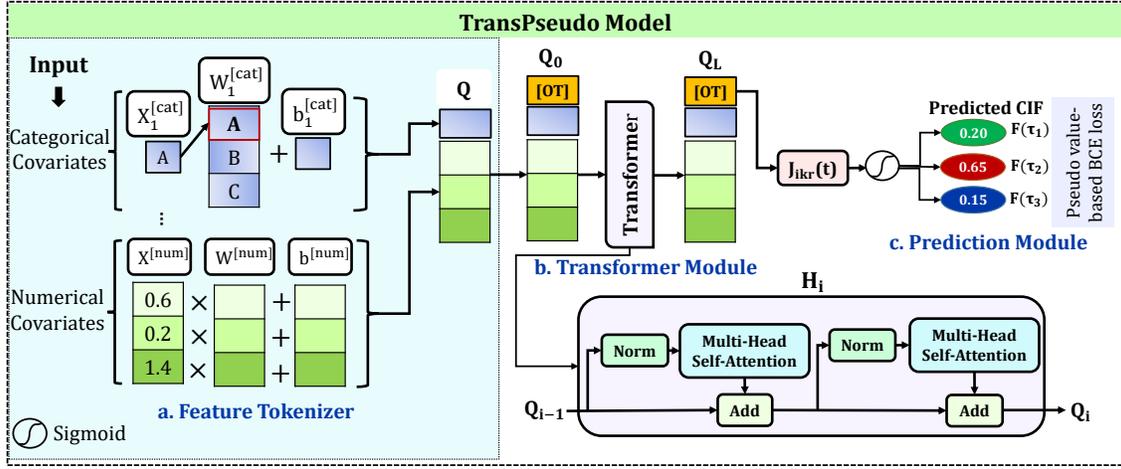
Figure 3: Architecture of TransPseudo model. Here, $X_j^{[cat]}$ and $X_j^{[num]}$ are the categorical and numeric covariates, respectively, and $W_j^{[cat]}$ and $W_j^{[num]}$ are their corresponding weights. b_j are the biases.

3 EXPERIMENTS

We conduct extensive experiments to answer the following research questions.

- **RQ1:** What are the advantages of employing federated pseudo values as opposed to traditional Jackknife pseudo values in the FedCRA framework?
- **RQ2:** How does our FedCRA perform on real-world distributed CRA data with non-independent and identically distributed (Non-IID) and Non-uniform censoring properties compared to the FL framework with the state-of-the-art CRA approaches?
- **RQ3:** How robust is the FedCRA framework under different types and amounts of censoring?

Synthetic datasets with different censoring mechanisms: To replicate different censoring scenarios in FL, we generate 5 distributed synthetic CRA datasets with different censoring mechanisms, such as (a) time censoring (TC), (b) interim censoring (IC), (c) case censoring (CC) with 25%, 50%, and 75% censoring [20], considering 10 decentralized clients assumed each client to have different covariate distributions (non-IID). To construct these datasets, we generate 12 numerical covariates from a multivariate normal distribution with mean *mu* and variance σ^2. Additionally, we generate two binary variables from a binomial distribution with probability *p*. Survival times are generated from exponential distribution considering the linear and nonlinear interaction among covariates.

SEER Data: The Surveillance, Epidemiology, and End Results (SEER) [10] program of the National Cancer Institute collected data from breast cancer patients registered at multiple hospitals in the United States to provide cancer statistics in the United States. The dataset contains 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**) [21]. To replicate a realistic distributed CRA data scenario with

non-uniform censoring (NUC) for FL, we first partitioned the SEER data into 3 clients based on the regions of the hospitals: West, Central, and East. Next, we chose a fixed number of subjects for all clients based on the minimum number of censored and uncensored subjects in the clients. Then, we varied the censoring percentages chosen from [0.2, 0.3, 0.4, 0.50, 0.55] for each client and adjusted the number of uncensored subjects by subtracting the number of selected censored subjects from the total fixed number of subjects. This setup enables us to evaluate FL models in a geographically distributed data environment and under non-uniform censoring settings.

eICU Data: The eICU dataset is a widely used public clinical dataset obtained from the eICU Collaborative Research Database [18], where data are collected from patients admitted to the ICU setting and gathered from multiple hospitals in the United States. We extracted 17342 patients who were diagnosed with one or more of the following four diseases: cancer (**CN**), liver disease (**LV**), immunosuppression (**IM**), and diabetes (**DI**). Death from each disease is considered a competing event, and we only consider the death of patients diagnosed with a single disease as an event. We imputed the missing values using Multivariate Imputation by Chained Equations (MICE) [24] separately for each client. To simulate real-world non-IID distributed CRA data, we partitioned the eICU dataset into 4 clients based on the region of the hospital: (1) Midwest, (2) Northeast, (3) South, and (4) West.

Table 1: Comparing federated and Jackknife pseudo values using TransPseudo model

Dataset	Combined test data		local test data			
	Federated Training		local Training		Federated Training	
	Jackknife	Fed. Pseudo	Jackknife	Fed. Pseudo	Jackknife	Fed. Pseudo
SEER	0.72 (0.032)	**0.83 (0.019)**	0.80 (0.011)	**0.86 (0.007)**	0.80 (0.018)	**0.91 (0.017)**
eICU	0.74 (0.006)	**0.83 (0.008)**	0.75 (0.035)	**0.83 (0.016)**	0.79 (0.019)	**0.87 (0.017)**

Table 2: C-Index comparison of the models on the SEER and eICU datasets

Setup	↓ Model Event →	SEER						eICU			
		CC	OCN	CVD	CMD	ID	OCS	CN	LV	IM	DI
Centralized (Combined Test Data)	CS-CoxPH	0.79 (0.015)	0.78 (0.010)	0.80 (0.007)	0.84 (0.023)	0.68 (0.207)	0.77 (0.026)	0.80 (0.001)	0.71 (0.117)	0.72 (0.008)	0.80 (0.000)
	DeepHit	**0.88 (0.010)**	0.85 (0.018)	0.82 (0.018)	0.76 (0.046)	0.81 (0.107)	0.82 (0.022)	0.71 (0.022)	0.65 (0.071)	0.56 (0.058)	0.67 (0.026)
	SurvTRACE	0.84 (0.028)	0.83 (0.025)	0.83 (0.026)	0.79 (0.051)	0.83 (0.080)	0.80 (0.067)	0.77 (0.015)	0.82 (0.035)	0.76 (0.021)	0.81 (0.006)
	TransPseudo	0.87 (0.006)	**0.88 (0.027)**	**0.88 (0.018)**	**0.89 (0.065)**	**0.89 (0.064)**	**0.86 (0.010)**	**0.82 (0.018)**	**0.89 (0.028)**	**0.82 (0.034)**	**0.89 (0.014)**
Federated (Combined Test Data)	CS-CoxPH	0.79 (0.013)	0.79 (0.007)	0.80 (0.009)	0.84 (0.022)	0.75 (0.027)	0.76 (0.031)	0.79 (0.001)	0.80 (0.004)	0.70 (0.007)	**0.80 (0.000)**
	DeepHit	**0.83 (0.008)**	0.83 (0.030)	0.86 (0.014)	0.84 (0.049)	**0.87 (0.038)**	0.82 (0.028)	0.71 (0.024)	0.68 (0.022)	0.59 (0.021)	0.77 (0.019)
	SurvTRACE	0.82 (0.007)	0.81 (0.025)	0.82 (0.017)	0.78 (0.078)	0.84 (0.058)	0.80 (0.051)	0.75 (0.009)	0.80 (0.034)	0.72 (0.043)	0.78 (0.008)
	TransPseudo	0.80 (0.007)	0.82 (0.041)	0.83 (0.013)	**0.87 (0.034)**	0.86 (0.073)	**0.83 (0.021)**	**0.81 (0.021)**	**0.87 (0.027)**	**0.83 (0.012)**	0.80 (0.014)
Local Training (Local Test Data)	CS-CoxPH	0.79 (0.033)	0.79 (0.047)	0.81 (0.040)	0.84 (0.042)	0.68 (0.119)	0.78 (0.056)	0.69 (0.012)	0.63 (0.035)	0.61 (0.051)	0.80 (0.017)
	DeepHit	0.87 (0.007)	0.82 (0.028)	0.83 (0.017)	0.78 (0.076)	0.83 (0.065)	**0.83 (0.033)**	0.67 (0.039)	0.67 (0.049)	0.56 (0.063)	0.79 (0.007)
	SurvTRACE	0.84 (0.028)	0.83 (0.025)	0.83 (0.026)	0.79 (0.051)	0.83 (0.080)	0.80 (0.067)	0.70 (0.030)	0.67 (0.009)	0.71 (0.037)	0.80 (0.010)
	TransPseudo	**0.88 (0.004)**	**0.87 (0.017)**	**0.86 (0.019)**	**0.91 (0.020)**	**0.84 (0.021)**	0.81 (0.025)	**0.86 (0.029)**	**0.78 (0.061)**	**0.82 (0.014)**	**0.89 (0.010)**
Local Federated (Local Test Data)	CS-CoxPH	0.79 (0.028)	0.79 (0.040)	0.80 (0.042)	0.84 (0.037)	0.74 (0.105)	0.77 (0.052)	0.76 (0.000)	0.73 (0.005)	0.68 (0.007)	0.82 (0.001)
	DeepHit	**0.88 (0.009)**	0.85 (0.022)	0.88 (0.020)	0.86 (0.056)	0.90 (0.040)	0.86 (0.027)	0.83 (0.012)	0.79 (0.021)	0.71 (0.052)	0.80 (0.017)
	SurvTRACE	0.87 (0.002)	0.82 (0.035)	0.81 (0.026)	0.79 (0.048)	0.83 (0.091)	0.81 (0.049)	0.76 (0.059)	0.69 (0.031)	0.76 (0.042)	0.79 (0.006)
	TransPseudo	0.87 (0.014)	**0.91 (0.021)**	**0.91 (0.018)**	**0.95 (0.022)**	**0.91 (0.061)**	**0.89 (0.020)**	**0.89 (0.007)**	**0.82 (0.036)**	**0.88 (0.029)**	**0.89 (0.002)**

Table 3: Model comparison on different censoring settings

Setup	Model	TC	IC	CC25	CC50	CC75
Centralized	DeepHit	0.68 (0.009)	0.68 (0.012)	0.65 (0.004)	0.67 (0.007)	0.72 (0.007)
	SurvTRACE	0.72 (0.014)	0.69 (0.015)	0.67 (0.006)	0.67 (0.009)	0.67 (0.008)
	TransPseudo	**0.79 (0.005)**	**0.79 (0.005)**	**0.75 (0.009)**	**0.80 (0.004)**	**0.82 (0.008)**
Federated	DeepHit	0.66 (0.007)	0.66 (0.006)	0.65 (0.006)	0.65 (0.009)	0.65 (0.002)
	SurvTRACE	0.70 (0.004)	0.68 (0.005)	0.67 (0.009)	0.68 (0.007)	0.71 (0.009)
	TransPseudo	**0.72 (0.008)**	**0.71 (0.009)**	**0.68 (0.006)**	**0.71 (0.011)**	**0.76 (0.007)**

Experimental Setup: To evaluate the performance of the models, we consider three training setups: (1) *Gold standard centralized training:* models are trained on combined training data shared from clients to the server, (2) *Local training:* clients' own data are used to train their models locally, and (3) *Federated training:* clients communicate with a global server by sharing their models instead of raw training data to update a global model.

Evaluation Criteria: In both centralized and federated settings, we use the combined test data from clients to evaluate the models. We also use the client's local test data to evaluate the locally trained models as well as federated trained local models (training of updated global model by FL on local client data). We use the time-dependent concordance index (C-Index) [3] as our evaluation metric and use **pycox** [13] package to compute them.

Implementation Details: Each client's data is randomly split into 80% training and 20% test data. We use 10% or 20% of the training data as validation sets. We ran the experiment 5 times with different random initialization or a different set of censoring percentages (for SEER) and reported the average performance with corresponding standard deviation. We train our proposed TransPseudo models using the Adam optimizer [15] with an early stopping criterion based on the best validation loss. We use a learning rate scheduler and select the batch size from [512, 1024]. For a centralized setting, the models are trained up to 500 epochs with a patience

of 10. For the federated settings, we perform a hyperparameter tuning to select the best learning rate, the number of local epochs, and total communication rounds. Based on the hyper-parameter tuning, we choose the learning rate, the number of local epochs, and total communication rounds 0.0001, 20, and 20, respectively. To obtain the prediction of CIF, we use `Sigmoid` activation function in the final output layer. We set 10th to 99th percentile of the time horizon with an interval of 10 for SEER and Synthetic datasets and [10, 20, 40, 80, 160, 320, 740] for eICU dataset, as the pre-specified time points for calculating pseudo values and evaluating the models.

Models Comparison: We compare our proposed FL framework **FedCRA** with our proposed client-specific models, **TransPseudo**, to FL framework with three state-of-the-art CRA models: i) Statistical model, Cause-specific Cox proportional hazards model (**CS-CoxPH**) [6] ii) Deep-learning-based model, **DeepHit** [14], and iii) Transformer-based model, **SurvTRACE** [26].

4 RESULTS AND DISCUSSION

Jackknife pseudo values vs. federated pseudo values: To show the effectiveness of employing our proposed federated pseudo values in our FedCRA framework as opposed to the traditional jackknife pseudo values, we evaluate and compare the performance of federated trained and locally trained TransPseudo models using both pseudo values on the combined test set and the local test sets of distributed SEER and eICU datasets. Table 1 demonstrates that the TransPseudo model with our proposed federated pseudo values shows up to 9% improvement over the traditional Jackknife pseudo values in terms of C-Index.

Comparing model performance on real-world distributed CRA datasets: Table 2 shows that our TransPseudo model outperforms CS-CoxPH, DeepHit, and SurvTRACE by 10.2%, 5.5%, and 5.8% in the centralized setting, 8.0%, 3.5%, and 4.1% in local training, and 11.8%, 3.5% 8.5% in federated setting evaluated on the local test set of SEER data, respectively. DeepHit and TransPseudo perform

similarly in the federated setting evaluated on the combined test set of SEER data. However, TransPseudo outperforms CS-CoxPH and SurvTRACE by 4.7% and 2.3%. Our TransPseudo model, compared to CS-CoxPH, DeepHit, and SurvTRACE, respectively, obtains 9.8%, 20.8%, and 6.5% better C-Index in the centralized setting, 15.5%, 16.5%, and 11.8% in local training, 5.5%, 14% and 6.5% in federated setting evaluated on the combined test set of eICU data and 12.3%, 8.7% and 12% in federated setting evaluated on the local test set of eICU data. Our findings highlight the effectiveness of TransPseudo in improving the local performance of the models and suggest the potential of FL for CRA.

Comparing model performance on various censoring settings: Table 3 demonstrates that our TransPseudo model achieves 9.3% and 5.6% overall improvement over the DeepHit and Surv-TRACE models, respectively, in the centralized setting evaluated on the combined test set of synthetic datasets with different censoring mechanisms, such as time censoring (TC), Interim Censoring (IC) and Case Censoring (CC). In the federated settings, our TransCRA model outperforms DeepHit by 5.8% and shows similar performance as the SurvTRACE model. The results support the effectiveness of using federated pseudo values in the TransPseudo model to handle different types of censoring.

Limitations: While our TransPseudo model provides accurate predictions, it requires more computational time and resources for training than the deep learning-based and statistical CRA models, such as DeepHit and CS-CoxPH. Furthermore, it is important to note that our federated pseudo values derivation assume covariate-independent censoring. However, it may not hold true for all real-world CRA datasets.

5 CONCLUSION

In this paper, we studied the emerging problem of competing risk analysis (CRA) in federated learning. We proposed **FedCRA**, a first-of-its-kind pseudo-value-based federated framework for CRA, **FedCRA**, aimed at estimating the subject-specific CIF in the presence of competing events and censoring. Additionally, we introduced TransPseudo, a transformer-based model specifically designed for CRA. We also introduced federated pseudo values, which allow us to analyze the CRA data in a federated framework while preserving privacy. We conducted experiments on both real and synthetic distributed CRA datasets with non-IID, non-uniform censoring properties and demonstrated the superiority of our FedCRA framework over the FL framework employing state-of-the-art CRA models.

ACKNOWLEDGMENTS

REFERENCES

[1] Per Kragh Andersen and Maja Pohar Perme. 2010. Pseudo-observations in survival analysis. *Statistical methods in medical research* 19, 1 (2010), 71–99.
[2] Mathieu Andreux, Andre Manoel, Romuald Menuet, Charlie Saillard, and Chloé Simpson. 2020. Federated survival analysis with discrete-time cox models. *arXiv preprint arXiv:2006.08997* (2020).
[3] Laura Antolini, Patrizia Boracchi, and Elia Biganzoli. 2005. A time-dependent discrimination index for survival data. *Statistics in medicine* 24, 24 (2005), 3927–3944.
[4] Nadine Binder, Thomas A Gerds, and Per Kragh Andersen. 2014. Pseudo-observations for competing risks with covariate dependent censoring. *Lifetime data analysis* 20 (2014), 303–315.
[5] Bingshu E Chen, Joan L Kramer, Mark H Greene, and Philip S Rosenberg. 2008. Competing risks analysis of correlated failure time data. *Biometrics* 64, 1 (2008), 172–179.
[6] David R Cox. 1972. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)* 34, 2 (1972), 187–202.
[7] Jason P Fine and Robert J Gray. 1999. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American statistical association* 94, 446 (1999), 496–509.
[8] Yury Gorishniy, Ivan Rubachev, Valentin Khrulkov, and Artem Babenko. 2021. Revisiting deep learning models for tabular data. *Advances in Neural Information Processing Systems* 34 (2021), 18932–18943.
[9] Frederik Graw, Thomas A Gerds, and Martin Schumacher. 2009. On pseudo-values for regression analysis in competing risks models. *Lifetime data analysis* 15 (2009), 241–255.
[10] National Cancer Institute. 2023. *The Surveillance, Epidemiology, and End Results (SEER) Program.* https://seer.cancer.gov/.
[11] Hemant Ishwaran, Thomas A Gerds, Udaya B Kogalur, Richard D Moore, Stephen J Gange, and Bryan M Lau. 2014. Random survival forests for competing risks. *Biostatistics* 15, 4 (2014), 757–773.
[12] Dana E King, Jun Xiang, and Courtney S Pilkerton. 2018. Multimorbidity trends in United States adults, 1988–2014. *The Journal of the American Board of Family Medicine* 31, 4 (2018), 503–513.
[13] Haavard Kvamme. 2022. *Pycox package.* https://github.com/havakv/pycox
[14] Changhee Lee, William Zame, Jinsung Yoon, and Mihaela Van Der Schaar. 2018. Deephit: A deep learning approach to survival analysis with competing risks. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 32.
[15] Ilya Loshchilov and Frank Hutter. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101* (2017).
[16] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*. PMLR, 1273–1282.
[17] Ulla B Mogensen and Thomas A Gerds. 2013. A random forest approach for competing risks based on pseudo-values. *Statistics in medicine* 32, 18 (2013), 3102–3114.
[18] Tom J Pollard, Alistair EW Johnson, Jesse D Raffa, Leo A Celi, Roger G Mark, and Omar Badawi. 2018. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Scientific data* 5, 1 (2018), 1–13.
[19] Shadi Rahimian, Raouf Kerkouche, Ina Kurth, and Mario Fritz. 2022. Practical challenges in differentially-private federated survival analysis of medical data. In *Conference on Health, Inference, and Learning*. PMLR, 411–425.
[20] Rahman et al. 2022. FedPseudo: Pseudo value-based Deep Learning Models for Federated Survival Analysis. *arXiv preprint arXiv:2207.05247* (2022).
[21] Md Mahmudur Rahman, Koji Matsuo, Shinya Matsuzaki, and Sanjay Purushotham. 2021. Deeppseudo: Pseudo value based deep learning models for competing risk analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 479–487.
[22] Md Mahmudur Rahman and Sanjay Purushotham. 2021. PseudoNAM: A Pseudo Value Based Interpretable Neural Additive Model for Survival Analysis. *UMBC Student Collection* (2021).
[23] Md Mahmudur Rahman and Sanjay Purushotham. 2022. Fair and interpretable models for survival analysis. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 1452–1462.
[24] Stef Van Buuren and Karin Groothuis-Oudshoorn. 2011. mice: Multivariate imputation by chained equations in R. *Journal of statistical software* 45 (2011), 1–67.
[25] Xuan Wang et al. 2022. SurvMaximin: robust federated approach to transporting survival risk prediction models. *Journal of biomedical informatics* 134 (2022), 104176.
[26] Zifeng Wang and Jimeng Sun. 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.
[27] Lili Zhao and Dai Feng. 2020. Deep neural networks for survival analysis using pseudo values. *IEEE journal of biomedical and health informatics* 24, 11 (2020), 3308–3314.