Knowledge-Based Systems 300 (2024) 112208

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Federated semi-supervised representation augmentation with cross-institutional knowledge transfer for healthcare collaboration

Zilong Yin [a,1], Haoyu Wang [a,1], Bin Chen [a,g,*], Xin Zhang [d], Xiaogang Lin [e], Hangling Sun [b], Anji Li [c], Chenyu Zhou [f,h]

[a] School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai, China
[b] Hengtu Imalligent Technology (Shanghai) Co., Ltd., Shanghai, China
[c] Abbott Laboratories(Shanghai) Co., Ltd., Shanghai, China
[d] Tianjin University of Technology, Tianjin, China
[e] North University of China, Taiyuan, China
[f] School of Computer Science and Technology, Xinjiang University, Urumqi, China
[g] School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China
[h] Tsinghua University, Beijing, China

ARTICLE INFO

Keywords:
Healthcare collaboration
Vertical federated learning
Knowledge transfer
Semi-supervised representation augmentation

ABSTRACT

In the healthcare field, cross-institutional collaboration can fasten medical research progress. Vertical federated learning (VFL) addresses data heterogeneity across multiple medical institutions while ensuring medical data privacy, thereby enhancing the accuracy of disease diagnoses and treatments. However, traditional VFL only benefits from aligned samples, thereby limiting its applicability due to constrained sample sizes, and a large amount of non-aligned data remains untapped, resulting in wasted data. To exert full leverage on the value of all data obtained from medical institutions, this paper proposes a federated healthcare collaborative framework based on semi-supervised representation augmentation mechanism with cross-institutional knowledge transfer (CrossKT-FRA). Specifically, the developed method comprises three steps. First, the federated representations of shared data (aligned data) among medical institutions are extracted through efficient vertical federated representation learning (FRL) methods. Second, the federated knowledge contained in federated representations and potential labels derived through recurrent learning assist local shared data representations in performing supervised augmented learning. Finally, the federated knowledge is transferred indirectly from the representation augmentation module for shared data to the unsupervised representation augmentation module for local private data (non-aligned data). The experimental results show the effectiveness of the proposed knowledge transfer mechanism, whether applied independently or used to enhance VFL on medical datasets. Our findings contribute to a deeper theoretical understanding of VFL, further facilitating the utilization of high-value medical data. By promoting cross-institutional and cross-disciplinary collaboration in healthcare data sharing, our study enhances the quality efficiency of medical services, thereby accelerating the development of interdisciplinary medical research. Code is available at https://github.com/LieLieLieLieLie/CrossKT-FRA.

1. Introduction

In today's digital era, medical data have become a crucial resource for medical research and clinical practice [1,2]. However, medical institutions in underdeveloped regions can often only gather limited amounts of data [3], thereby restricting the effectiveness and comprehensiveness of their diagnostic, treatment and disease prevention efforts [4,5]. Resource scarcity in medical institutions impact the sustainability of healthcare systems in underdeveloped regions [6] and the feasibility of healthcare policy reforms [7]. Therefore, collaboration among medical institutions in various developmental stages is important to reduce data insufficiency issues and implement clinical decision-making in lagging healthcare facilities [8].

However, medical institution information pertains to patient privacy [9,10], making explicit transmission and integrated training inconvenient [11–14]. Federated learning (FL) [15,16] can satisfy the demand for jointly training medical models across heterogeneous data

[*] Corresponding author at: School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai, China.
E-mail addresses: zilong_yin@163.com (Z. Yin), ambityuki@gmail.com (H. Wang), chenbin1933@163.com (B. Chen), kxamanda@163.com (X. Zhang), lxg2630782125@163.com (X. Lin), sunnyhl2000@gmail.com (H. Sun), anjicake@gmail.com (A. Li), Zhou_cy@stu.xju.edu.cn (C. Zhou).

[1] Equal contribution.

https://doi.org/10.1016/j.knosys.2024.112208
Received 22 April 2024; Received in revised form 2 July 2024; Accepted 7 July 2024
Available online 14 July 2024

acquired from multiple medical institutions without exposing patients' raw data [17,18]. As FL research has progressed, horizontal federated learning (HFL) has dominated the subfield of FL research [19], allowing participants to share data in the same feature space. Although HFL partially addresses the problem related to independent and identically distributed (IID) medical data, the significant disparity between data derived from developed and underdeveloped medical institutions often makes HFL difficult to apply [20,21]. This challenge has driven the integration and development of vertical federated learning (VFL) to address medical data issues [22]. VFL allows participants to share data in the same sample space [23], aligning overlapping data obtained from heterogeneous data sources and extracting information value from the overlapped data to jointly train models and achieve improved model performance [24,25].

The performance of traditional VFL algorithms relies on the size and information value of the overlapping data among parties. FedMVT [26] and FedCVT [27] leverage the overlapping data among FL participants to share knowledge across different sources or views, thereby enhancing the performance of FL models. However, there is significant heterogeneity in the medical institution data produced across different regions [28]. Hospitals in developed regions collect more symptoms and have more comprehensive pathological information. Therefore, medical collaborations led by institutions in developed regions may overlook pathological information from underdeveloped regions [29]. Additionally, patients with certain regional diseases may not receive treatment in local hospitals due to local medical resource limitations, making it difficult for local hospitals to analyze and study data related to these diseases. Consequently, for traditional VFL algorithms, the heterogeneity of federated medical data not only limits the federated medical knowledge derived from overlapping data but also causes the algorithms to overlook personalized medical knowledge from private data.

Compared to other fields, healthcare data come from various sources, including hospitals, clinics, and laboratories, each with different formats and standards, significantly increasing the difficulty of data integration and analysis. The heterogeneity of data in the healthcare field presents unique challenges [30,31]. Medical data include various types such as clinical records, imaging data, and genomic data, each with its unique structure and analysis requirements, making their diversity and complexity significantly higher than in other fields [32]. There are notable differences in data formats, standards, and quality among different developed regions or specialized institutions [33], further complicating data integration and analysis. Moreover, the critical nature of medical decision-making requires models with high accuracy and reliability, posing higher demands on algorithms [34]. The healthcare field also must strictly comply with privacy protection and data security regulations [35,36], and the sensitivity of patient data makes direct data sharing impossible [37], further increasing the complexity of data processing and model training.

To address the above challenges, we propose a collaborative framework for federated healthcare based on a semi-supervised representation augmentation mechanism with cross-institutional knowledge transfer, termed CrossKT-FRA. This method leverages efficient vertical federated representation learning (FRL) to extract joint representations of aligned data across various medical institutions. Subsequently, it utilizes the federated knowledge derived from these joint representations and labels learned through cyclic learning to assist in the supervised augmentation of locally shared data representations. Finally, the federated knowledge is indirectly transferred from the representation augmentation module for shared data to the representation augmentation module for unsupervised local non-aligned data. The specific contributions of our paper are as follows:

1. We propose a novel VFL framework. Initially, each party obtains federated representations of their local shared data through federated matrix factorization. Subsequently, the parties transfer federated knowledge from their federated representations to local representations, generating and augmenting matched representations with the federated knowledge while preserving certain personalized characteristics.
2. We propose a semi-supervised representation augmentation approach. Initially, the locally shared data representations are augmented in a supervised manner by federated knowledge and latent labels. Subsequently, the federated knowledge indirectly aids in the unsupervised augmentation of local private data representations by assisting in the augmenting the representations of the shared data.
3. We propose a cyclic latent label learning method. By minimizing the difference between latent label inputs and outputs, the approach converges to generate representations that match the original data in terms of their label attributes while preserving label privacy.
4. We validate the effectiveness of our cross-institution knowledge transfer mechanism and the generalizability of federated semi-supervised representation augmentation on medical field datasets.

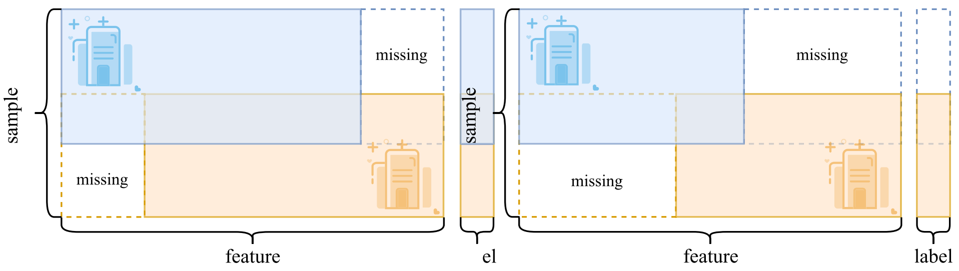
2. Related work

2.1. Federated learning in healthcare

Federated Learning (FL) enables multiple participants to jointly train models on a central server while preserving data privacy. FL, as a distributed artificial intelligence (AI) paradigm, has been identified as a promising solution in the field of intelligent healthcare [38,39]. CAFL [40] is employed in intelligent healthcare development and deployment, allowing fair evaluation of FL participants' contributions to model performance without exposing private data. It permits the distribution of the best-performing intermediate model to participants for FL training to enhance training protocols. Flop [41] is utilized in medical image classification applications, where clients only need to share part of the model with the server for federated averaging, while the remaining layers of the neural network can remain private. FedDis [42] is particularly beneficial for medical institutions sharing health and abnormal data. To mitigate medical data heterogeneity, it decomposes the parameter space into global (shape) and local (appearance) components. It jointly trains shape parameters across four institutions to simulate healthy brain anatomy. Meanwhile, each institution locally trains appearance parameters to enable client-specific personalization of globally invariant features. Cov-Fed [43] introduces the Multi-ECA mechanism, which enhances the feature extraction capability of deep learning models by focusing on fundamental features in chest X-ray scans, thereby improving classification performance without compromising privacy. Existing federated healthcare collaboration frameworks seldom address the imbalance, insufficiency, and heterogeneity of health data among medical institutions from the perspective of VFL, which is precisely our research focus.

2.2. Vertical federated learning

As a distributed machine learning approach, FL allows participants with heterogeneous data to collaboratively train models. Depending on the degree of data heterogeneity, FL can be categorized into HFL and VFL. As shown in Fig. 1(a), HFL involves datasets that share a similar feature space but have different sample spaces. For instance, medical institutions in developed regions have patient records with similar sampling indicators, although the patients are different, due to the availability of advanced medical resources like CT and MRI. Conversely. As illustrated in Fig. 1(b), VFL deals with datasets that have both different feature spaces and different sample spaces. For example, medical institutions in developed regions and underdeveloped regions have patient records from their respective areas, but due to limited



(a) Horizontal Federated Learning(HFL). (b) Vertical Federated Learning(VFL).

Fig. 1. Two main categories of FL.

medical resources, patients in underdeveloped regions have fewer and less varied sampling indicators compared to those in developed regions.

HFL focuses on sharing model parameters among devices with similar data distributions [44]. However, in regions with different developmental statuses, medical institutions exhibit significant data heterogeneity, with only partial data overlap. VFL, as a framework for cross-heterogeneous data collaboration, effectively utilizes limited overlapping data for collaborative modeling and learning. However, most existing VFL approaches only utilize overlapping data [45], leading to resource wastage due to a substantial amount of non-overlapping data being disregarded. Additionally, the majority of VFL works are based on supervised datasets, including trees [46,47], linear and logistic regression [48], and neural networks [49,50]. These supervised FL approaches rely on labels, which are costly to acquire and require field expertise.

2.3. Semi-supervised learning

As an intermediary machine learning approach between supervised and unsupervised learning, semi-supervised learning leverages a small amount of labeled data along with a large amount of unlabeled data to improve model performance. Semi-supervised learning is particularly advantageous in VFL scenarios where data acquisition is challenging and labeling costs are high. For instance, FedTG [51] utilizes multiple high-probability pseudo-labels instead of a single pseudo-label to fully exploit the knowledge from unlabeled data. Semi-HFL [52] proposes a “multi-teacher to multi-student” semi-supervised learning model by dividing deep models into smaller sub-models. It pre-trains these multi-branch models with a small amount of labeled data on the server and generates pseudo-labels for local training on the client side. SFLEDS [53] introduces a prototype-based method to address issues such as label scarcity, concept drift, and privacy protection. Despite these advancements, semi-supervised federated frameworks still exhibit dependency on labels, and strategies like multi-model segmentation and prototype utilization increase implementation complexity and computational overhead.

2.4. Privacy protection

Most HFL models utilize aggregation of model parameter updates [44,54]; however, it is worth noting that model parameters can potentially be reverse-engineered to infer original data [55]. While VFL mitigates privacy leakage by aggregating local gradients, existing research also suggests that adversaries can infer each other’s gradients in VFL parameter exchange [56–59]. Therefore, the concept of differential privacy has gained popularity [60], offering a quantifiable measure of privacy [61,62], and many existing algorithms can be modified to achieve differential privacy, including when integrated with FL [63,64]. However, these methods are not directly applicable to VFL, requiring more rigorous data protection measures, especially for sensitive fields such as the healthcare industry.

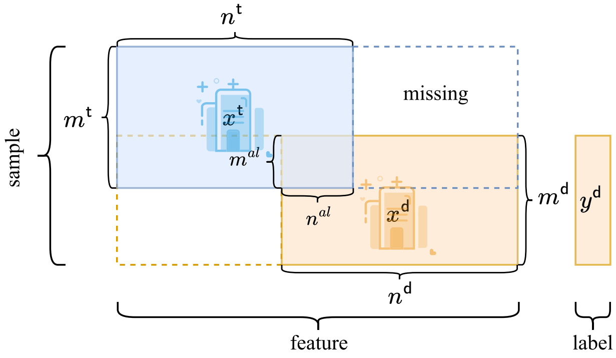


Fig. 2. Virtual views of datasets in VFL.

2.5. Data augmentation

Some researchers have begun exploring methods for data augmentation in VFL. FedTrans models joint features from multiple parties to extract joint representations of shared samples. Then, it learns a locally representative refinement module for each hospital, transferring knowledge from the representation of shared samples to enrich the representation of local samples. Finally, it utilizes enriched representations to accomplish downstream machine learning tasks for each hospital [65]. FedMVT and FedCVT estimate representations of missing features, predict pseudo-labels for unlabeled samples, and then jointly train three classifiers based on different sources or views of the extended training set to filter out invalid samples [26,27]. While existing data augmentation methods are effective, they may suffer from drawbacks such as computational costs and limited generality when integrated with FL.

3. Problem definition

Local-Data Vertical Federated Knowledge Transfer Problem: One institution, referred to as the task side (*task* side), possesses only sample features, denoted as t side. The other institution, referred to as the data side (*data* side), has both sample features and labels, denoted as d side(In the latter, we uniformly describe “institution” in terms of “side”). The samples and features of t and d only partially overlap. The goal is to design a federated knowledge transfer algorithm to augment the representation of local data in healthcare institutions as much as possible, thereby improving the performance of local models or VFL models. Fig. 2 shows the data distribution of the two parties in this scenario.

Specifically, the dataset $\mathcal{D}^{\mathrm{t}} = \left\{\left(x_i^{\mathrm{t}}\right)\right\}_{i=1}^{m^{\mathrm{t}}} \in \mathbb{R}^{|m^{\mathrm{t}}|\times|n^{\mathrm{t}}|}$ of t side consists of m^{t} samples, each with n^{t} features, where x_i^{t} denotes the feature data of the ith sample. The dataset $\mathcal{D}^{\mathrm{d}} = \left\{\left(x_i^{\mathrm{d}}, y_i^{\mathrm{d}}\right)\right\}_{i=1}^{m^{\mathrm{d}}} \in \mathbb{R}^{\left|m^{\mathrm{d}}\right|\times\left|n^{\mathrm{d}}\right|}$ of d side

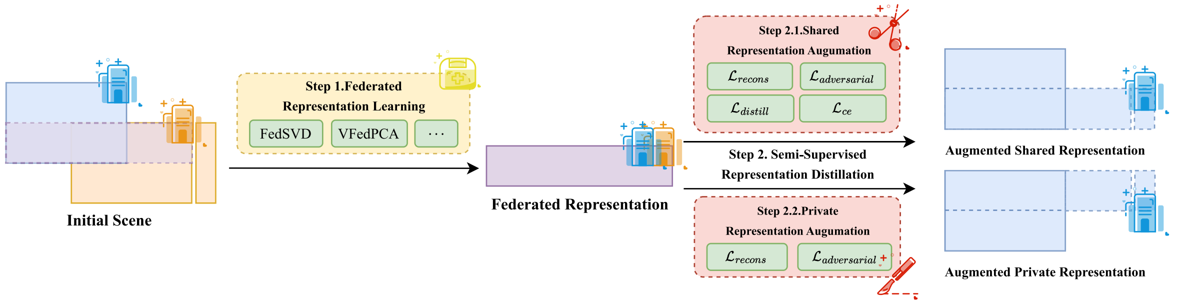


Fig. 3. Overview of CrossKT-FRA. Each healthcare institution can act as either a task institution or a data institution. In Step 1, we use FRL methods such as FedSVD [67] and VFedPCA [68] (to be detailed in Section 4.3) to obtain the federated representation. In Step 2, the task institution locally trains a semi-supervised representation augmentation module for knowledge transfer. In Step 2.1, the representation of the shared data is augmented by combining adversarial loss $\mathcal{L}_{adversarial}$, reconstruction loss $\mathcal{L}_{recons}$, distillation loss $\mathcal{L}_{distill}$ and classification loss $\mathcal{L}_{ce}$ (to be detailed in Section 4.3.2). In Step 2.2, the representation of the private data is augmented by combining adversarial loss $\mathcal{L}_{adversarial}$ and reconstruction loss $\mathcal{L}_{recons}$ (to be detailed in Section 4.3.3).

consists of m^{d} samples, each with n^{d} features, where x_i^{d} and y_i^{d} denote the feature and label data of the ith sample respectively.

Additionally, we assume the existence of a partially overlapping dataset $\mathcal{D}^{al}$, represented as $\mathcal{D}^{al} = \left\{\left(x_i^{\mathrm{t,d}}\right)\right\}_{i=1}^{m^{al}} \in \mathbb{R}^{\left|m^{al}\right|\times\left|n^{al}\right|}$, where m^{al} denotes the number of overlapping samples and n^{al} represents the number of overlapping features. The aligned dataset can be found using entity alignment techniques in privacy-preserving settings [66]. Here, we assume that t side and d side have already found aligned data. Specifically, $\mathcal{D}^{\mathrm{t},al} = \left\{\left(x_i^{\mathrm{t},al}\right)\right\}_{i=1}^{m^{al}} \in \mathbb{R}^{\left|m^{al}\right|\times\left|n^{\mathrm{t}}\right|}$ and $\mathcal{D}^{\mathrm{d},al} = \left\{\left(x_i^{\mathrm{d},al}\right)\right\}_{i=1}^{m^{al}} \in \mathbb{R}^{\left|m^{al}\right|\times\left|n^{\mathrm{d}}\right|}$ serve as the shared data ($\mathcal{D}^{al} = \mathcal{D}^{\mathrm{t},al} \cap \mathcal{D}^{\mathrm{d},al}$) after alignment for both sides, which are used for federated knowledge transfer rather than plain text exchange.

Apart from the aligned data, the remaining samples from both sides constitute their respective private data. For t side, this includes $\mathcal{D}^{\mathrm{t},nl} = \left\{\left(x_i^{\mathrm{t},nl}\right)\right\}_{i=1}^{m^{\mathrm{t},nl}} \in \mathbb{R}^{\left|m^{\mathrm{t},nl}\right|\times\left|n^{\mathrm{t}}\right|}$($\mathcal{D}^{\mathrm{t}} = \mathcal{D}^{\mathrm{t},al} \cup \mathcal{D}^{\mathrm{t},nl}$), and for d side, it includes $\mathcal{D}^{\mathrm{d},nl} = \left\{\left(x_i^{\mathrm{d},nl}\right)\right\}_{i=1}^{m^{\mathrm{d},nl}} \in \mathbb{R}^{\left|m^{\mathrm{d},nl}\right|\times\left|n^{\mathrm{d}}\right|}$($\mathcal{D}^{\mathrm{d}} = \mathcal{D}^{\mathrm{d},al} \cup \mathcal{D}^{\mathrm{d},nl}$).

Optimization Objectives of the Proposed Algorithm: Traditional VFL utilizes only overlapping data $\mathcal{D}^{al}$ to establish a federated machine learning model, where only $\mathcal{D}^{\mathrm{t},al}$ and $\mathcal{D}^{\mathrm{d},al}$ ($m^{\mathrm{t},al} = m^{\mathrm{d},al}$) participate in training. This results in the waste of a significant amount of non-aligned data $\mathcal{D}^{\mathrm{t},nl}$ and $\mathcal{D}^{\mathrm{d},nl}$. Therefore, we propose a federated representation augmentation method combining cross-institution knowledge transfer and semi-supervised generative adversarial networks (CrossKT-FRA). Without loss of generality, we solely validate the impact of knowledge transfer on the task healthcare institution t to verify that CrossKT-FRA provides strong support for healthcare institutions with limited knowledge. Our VFL knowledge transfer method is not limited to overlapping data $\mathcal{D}^{al}$, non-aligned data $\mathcal{D}^{\mathrm{t},nl}$ and $\mathcal{D}^{\mathrm{d},nl}$ also play a critical role. The goal is to obtain federated representation knowledge through cross-institution learning from $\mathcal{D}^{al}$, combined with semi-supervised generative adversarial networks to augment the representation of aligned data $\mathcal{D}^{\mathrm{t},al}$ and non-aligned data $\mathcal{D}^{\mathrm{t},nl}$, aiming to fully exploit existing data information and improve the performance of local models or VFL models. This expands the practical application range of FL, especially in healthcare-related institutions with limited knowledge.

4. Methods

4.1. Overview

The paper proposes a federated healthcare collaborative framework based on semi-supervised representation augmentation mechanism with cross-institutional knowledge transfer, termed CrossKT-FRA. Fig. 3 illustrates an overview of CrossKT-FRA, which consists of two phases:

- **Federated Representation Learning**: Through privacy-preserving federated collaborative learning techniques, both the task side and the data side can extract federated representations of shared data. In essence, this representation possesses the capability to amalgamate the shared data knowledge from all sides while safeguarding the original data features of each side from disclosure.
- **Semi-Supervised Representation Augmentation**: The task side extracts knowledge from federated representations to sequentially enrich shared data representations and private data representations. Specifically, for shared data representations, knowledge is distilled from federated representations: firstly, learning the knowledge of shared data representations, and then assisting in augmenting private data representations.

1. **Shared Representation Augmentation**: Firstly, in the learning of shared data representations, the task side aims to capture the data distribution information of federated representations while retaining a portion of personalized shared data's data distribution information. Furthermore, shared data representations and federated representations are refined and extracted through the pre-trained feature extractor of the data side. In this process, the refined representation of shared data distills the refined representation of federated representations. Simultaneously, it ensures that shared data representations and federated representations exhibit comparable superior classification performance on pre-trained classifiers.
2. **Private Representation Augmentation**: Next, in the learning of private data representations, the task side aims to capture the data distribution information of shared data representations while preserving a portion of personalized data distribution information from private data.

4.2. Federated representation learning

The purpose of FRL is to extract federated representations of shared data from both the task side and the data side, serving as carriers of knowledge transfer to the task side and aiding in augmenting its representations [65]. Generally, various FRL methods can be employed to accomplish this step. As literature suggests, matrix factorization is effective for extracting meaningful latent representations in machine learning tasks [69]. Therefore, in this paper, we adopt federated representation methods based on matrix factorization, namely FedSVD [67] and VFedPCA [68]:

FedSVD [67]: Both the task side and the data side employ two random orthogonal matrices to transform their respective shared data into locally masked data, ensuring consistency in the results of the masking transformations. Subsequently, both sides upload the masked data to a third-side server, where they undergo secure aggregation and standard singular value decomposition. Finally, the task side reconstructs federated representations based on the decomposition results by unveiling the masks.

Given t side's shared data matrix $x^{\mathsf{t},al} \in \mathbb{R}^{|m^{al}| \times |n^{\mathsf{t}}|}$ and d side's shared data matrix $x^{\mathsf{d},al} \in \mathbb{R}^{|m^{al}| \times |n^{\mathsf{d}}|}$, let $S = [x^{\mathsf{t},nl} \mid x^{\mathsf{d},nl}]$ denote the combined data matrix from t side and d side. The federated representation U is learned through $S = U \Sigma V^T$ (singular value decomposition, SVD) . FedSVD employs a randomized masking approach to derive U:

1. A trusted key generator produces two random orthogonal matrices, $M \in \mathbb{R}^{|m^{al}| \times |m^{al}|}$ and $N \in \mathbb{R}^{|n^{\mathsf{t,d}}| \times |n^{\mathsf{t,d}}|}$ $\left(\left|n^{\mathsf{t,d}}\right| = \left|n^{\mathsf{t}}\right| + \left|n^{\mathsf{d}}\right| - \left|n^{\mathsf{al}}\right|\right)$. N is further divided into two parts, $N^{\mathsf{t}} \in \mathbb{R}^{|n^{\mathsf{t}}| \times |n^{\mathsf{t,d}}|}$ and $N^{\mathsf{d}} \in \mathbb{R}^{|n^{\mathsf{d}}| \times |n^{\mathsf{t,d}}|}$. Define $N^T = [(N^{\mathsf{t}})^T \mid (N^{\mathsf{d}})^T]$.
2. M and N^{t} are sent to t side, while M and N^{d} are sent to d side. Each side uses the received matrices to mask its own data matrix for local computation:
$$\tilde{S}^{\mathsf{p}} = M S^{\mathsf{p}} N^{\mathsf{p}}, \quad \forall \mathsf{p} \in \{\mathsf{t}, \mathsf{d}\} \tag{1}$$
3. t side and d side send $\tilde{S}^{\mathsf{t}}$ and $\tilde{S}^{\mathsf{d}}$ to a third-side server,[2] and the server performs SVD on the securely aggregated data matrix $\tilde{S}$, denoted as $\tilde{S} = \tilde{U} \Sigma \tilde{V}^T$, where U and V are orthogonal matrices, and $\tilde{S} = [\tilde{S}^{\mathsf{t}} \mid \tilde{S}^{\mathsf{d}}]$.
4. t side can reconstruct the federated representation $\tilde{x}^{fed}$ as follows:
$$\tilde{x}^{fed} = U = M^T \tilde{U} \tag{2}$$

Compared to FedSVD, which aims to recover U and V^T, we only need to recover U since it represents the shared result of FedSVD, serving as the manifestation of federated knowledge. Moreover, the original data's singular vectors are not required, eliminating the complex computational cost of recovering V^T due to privacy protection needs. The correctness of the aforementioned process depends on the requirement that S and $\tilde{S}$ must have the same singular values Σ.

VFedPCA[68]: Unsupervised FL is conducted between the task side and the data side to compute federated feature vectors, aiming to reduce the dimensionality of shared data and extract principal component data information. Both sides converge their own federated feature vectors to global feature vectors without needing to know each other's data. Local power iteration is employed by both sides to train local feature vectors. Then, the feature vectors from both sides are merged into a federated feature vector. Finally, the task side reconstructs the original data to obtain federated representations.

Given that the t side shares a data matrix $x^{\mathsf{t},al} \in \mathbb{R}^{|m^{al}| \times |n^{\mathsf{t}}|}$ and the d side shares a data matrix $x^{\mathsf{d},al} \in \mathbb{R}^{|m^{al}| \times |n^{\mathsf{d}}|}$. Assuming $S = [x^{\mathsf{t},nl} \mid x^{\mathsf{d},nl}]$(Data matrix combinations of t side and d side), where $S \in \mathbb{R}^{|m^{al}| \times |n^{\mathsf{t,d}}|}$ $\left(\left|n^{\mathsf{t,d}}\right| = \left|n^{\mathsf{t}}\right| + \left|n^{\mathsf{d}}\right| - \left|n^{\mathsf{al}}\right|\right)$.

1. For any side $\mathsf{p} \in \{\mathsf{t}, \mathsf{d}\}$, compute the maximum eigenvalue $A^{\mathsf{p}} = \frac{1}{|m^{\mathsf{p}}|}(S^{\mathsf{p}})^T S^{\mathsf{p}}$ and a nonzero vector a^{p} corresponding to the eigenvector α^{p} ($A^{\mathsf{p}} a^{\mathsf{p}} = \alpha^{\mathsf{p}} a^{\mathsf{p}}$). Set the number of iterations to L, and both sides need to locally compute until convergence as follows:
$$a_l^{\mathsf{p}} = \frac{A^{\mathsf{p}} a_{l-1}^{\mathsf{p}}}{\left\| A^{\mathsf{p}} a_{l-1}^{\mathsf{p}} \right\|}, \quad \alpha_l^{\mathsf{p}} = \frac{A^{\mathsf{p}} \left(a_l^{\mathsf{p}}\right)^T a_l^{\mathsf{p}}}{\left(a_l^{\mathsf{p}}\right)^T a_l^{\mathsf{p}}}, \quad \forall \mathsf{p} \in \{\mathsf{t}, \mathsf{d}\}, \quad l = 1, 2, \dots, L \tag{3}$$
2. Then each medical institution uploads the eigenvectors and eigenvalues to a third-side server[3]. The server aggregates the results and generates the federated eigenvalue u:
$$u = w^{\mathsf{t}} a_L^{\mathsf{t}} + w^{\mathsf{d}} a_L^{\mathsf{d}}, \quad w^{\mathsf{p}} = \frac{\alpha_L^{\mathsf{p}}}{\sum \alpha_L^{\mathsf{p}}} \tag{4}$$
3. t side can utilize u to compute the federated representation $\tilde{x}^{fed}$:
$$\tilde{x}^{fed} = x^{\mathsf{t},al} \frac{QQ^T}{\|QQ^T\|}, \quad Q = (x^{\mathsf{t},al})^T u \tag{5}$$

4.3. Semi-supervised representation augmentation

In CrossKT-FRA, we propose a semi-supervised representation augmentation method. Simply put, after obtaining the federated representation $\tilde{x}^{fed}$ of the data shared by the data side, the task side can transfer the knowledge of the data side's shared data $x^{\mathsf{d},al}$ locally. The task side's shared data $x^{\mathsf{t},al}$ undergoes a novel semi-supervised representation augmentation strategy to distill learning from $\tilde{x}^{fed}$ while introducing new loss functions. This is done to augment the representation while retaining local personalization and without compromising classification performance. The knowledge learned from $x^{\mathsf{t},al}$ is then used to augment the representation of local private data $\tilde{x}^{\mathsf{t},nl}$.

4.3.1. Motivation

Adversarial Generation and Knowledge Transfer of Federated Representations: In our strategy, the intention is for the knowledge from the data side to transfer to the task side through federated representation, aiding the task side in enhancing its intermediate processing results. We consider using GAN as encoders for the intermediate processing results, cleverly embedding the federated representation into this encoder. In traditional GANs, the generator's objective is to produce realistic samples similar to the distribution of the original data to deceive the discriminator. This objective mainly focuses on ensuring that the generated model accurately simulates the appearance and distribution characteristics of the original data, emphasizing the model's learning of individual details from the original data. However, this objective lacks adaptability and generalization for the distributed environment of FL. In our strategy, we aim to propose a novel objective for GAN, where the goal is to generate local representations of the task side that are adversarial to the federated representation. Compared to the traditional objective of generative adversarial networks, we are more focused on extracting global knowledge from the federated representation, aiming to generate augmented local representations with global knowledge.

Leveraging Private Data to Overcome the Limitations of Shared Data: In traditional FL, training typically relies on shared data, where data samples from each participant correspond one-to-one in terms of features or labels. However, this approach has certain limitations, particularly when dealing with large datasets and uneven data distribution among participants. The high cost of obtaining shared data and the limited number of aligned samples can easily lead to model overfitting, reducing the model's generalization ability on new datasets. In contrast, private data does not require strict sample alignment, allowing for better utilization of each participant's data resources, thereby enhancing the model's generalization capability and robustness. Therefore, this paper explores the application of both shared and private data in VFL to improve the model's performance in real-world applications.

Sequential Knowledge Transfer: From Shared Data to Private Data: Therefore, in our semi-supervised representation augmentation strategy, the objective is to generate representations of the task side's data using GAN and then integrate them with "knowledge extraction from federated representation" to augment the data representation of the task side. The task side data is divided into private data and shared data. Since the federated representation is computed jointly from the shared data of both the task side and the data side, the shared data and federated representation are directly related. The representation of the task side's shared data can directly distill the

[2] The third-isde server needs to be semi-honest. Note that in FL, this secure configuration (i.e., the information aggregation server being semi-honest) is widely accepted [48].

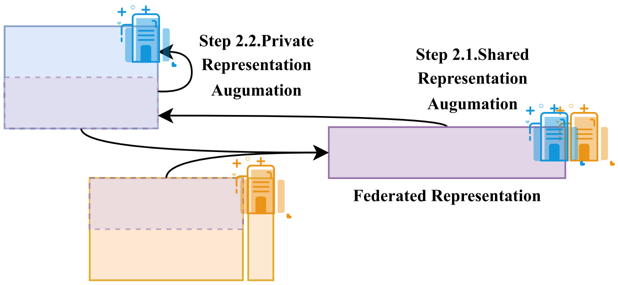


Fig. 4. Training Sequence for Representation Augmentation of Task-Side Shared and Private Data.

knowledge from the data side's shared data through learning from the federated representation. There is also a direct relationship between the task side's private data and shared data because both originate from the task side. In VFL, aligned shared data often represents only a small portion of the original data, and due to its strong association with the federated representation, the generator focuses more on extracting federated knowledge when training on shared data rather than aligned private data. Conversely, non-aligned private data often constitutes the majority of the original data, and the generator focuses more on preserving the individuality of the task-side data when training on private data. Therefore, in terms of training sequence, training on shared data precedes training on private data. In other words, the knowledge from the federated representation does not directly affect the private data. Instead, it indirectly transfers knowledge to the private data, as illustrated in Fig. 4, to achieve the augmentation of the task side's representation. Certainly, in Step 2.1, we concentrate on augmenting the representation of shared data, while in Step 2.2, the focus shifts to augmenting the representation of private data.

Balancing Needs with CGAN and GAN in Semi-Supervised Representation Augmentation: Given the differing emphases of private and shared data in the generation objectives, along with their respective label attributes, we propose the incorporation of GAN and Conditional Generative Adversarial Networks (CGAN) to augment the representations of both types of data. Firstly, while task-side data is inherently unlabeled, the federated representations computed from shared data have labels on the data side. Task side can indirectly obtain pseudo-labels equivalent to this label information and thus perform supervised representation augmentation on this portion of the data. It is important to note that we assume that task side can achieve the aforementioned under the premise of privacy protection. The limited quantity of aligned shared data on the task side determines the value of knowledge extracted from federated representations. Supervised CGAN allows us to generate augmented representations matching the shared data based on these label information, thereby stably preserving the crucial knowledge extracted from federated representations. Secondly, the significantly larger proportion of unlabeled private data on the task side contains more individualized data distribution information. This enables unsupervised GAN to embed a more diverse range of personalized knowledge. This strategy of using CGAN and GAN separately aims to balance the demands of both types of data. CGAN generates augmented representations that are more matching, while GAN retains the features and distribution consistency of individual data. Theoretically, this approach is better suited to meeting the global knowledge requirements of task-side FL while accommodating the flexibility of individual data needs. Therefore, in our strategy, as depicted in Fig. 5, the generators and discriminators for both types are uniformly designed, except for the input layer, to balance the demands of both types.

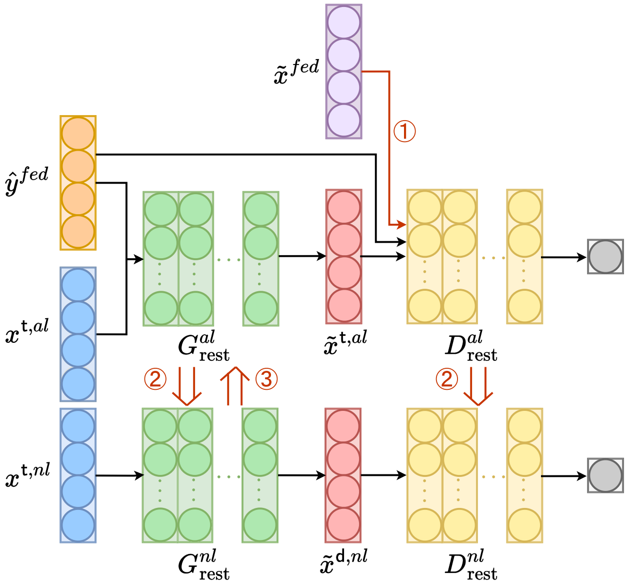


Fig. 5. The collaborative architecture design of CGAN for shared data $x^{\mathrm{t},al}$ and GAN for private data $x^{\mathrm{t},nl}$ in our semi-supervised representation augmentation strategy. *rest* represents the model excluding the input layer. In each iteration, the knowledge extracted by the federated representation $\tilde{x}^{fed}$ is first embedded into the discriminator D^{nl} of $x^{\mathrm{t},nl}$, and then into the generator G^{nl}. After training the shared data, G^{al}_{rest} and D^{al}_{rest} are respectively assigned to G^{nl}_{rest} and D^{nl}_{rest}. After training the private data, G^{nl}_{rest} is fed back to G^{al}_{rest} to start a new iteration.

Under the premise of protecting the privacy of data side, task side can acquire label data using cutting-edge data protection communication methods, such as those mentioned in [70,71]. However, it is essential to note that encryption algorithms and intermediary mechanisms like blockchain may introduce accuracy loss and computational complexity. In our strategy, within the augmentation of shared data representation, we introduce feature extractors and classifiers for all sides:

- **Classifier**: The classifier is introduced in the task side, employing a method that fits the classification results. The logits for the augmented representation's predictions are computed by the data side, and the loss is fed back. The encoding of classification results is continually updated and converged, serving as the label conditional input for the subsequent round of CGAN. As the task side's objective is to augment local representations by extracting and learning from the knowledge encoded in the federated representation, the specific classification labels are not necessary. The encoded classification results can achieve the alignment between the augmented representation of the task side's shared data and the categories of the shared data, all while maintaining the privacy of the labels associated with the shared data of the data side.
- **Extractor**: Since the federated representation is computed from the shared data of all sides, and small amounts of shared data tend to have noisy data present as well. In theory, due to the properties of matrix decomposition, the federated representation captures the data distribution of the shared data from all sides. However, it may lack effective knowledge condensation, resulting in the presence of some ineffective knowledge. As the task side lacks labels, it is challenging to distinguish and extract valid features from the federated representation. On the other hand, the data side, as labeled side, are better equipped to fulfill this requirement. Therefore, in addition to introducing separate feature extractors for each side, the data providers share their trained feature extractors with the task side, without sharing the classifiers. This approach allows the task side to extract effective knowledge from the federated representation while preserving the privacy of the data providers' label information. Consequently, this empowers the task side to achieve more accurate convergence in the encoding of classification results.

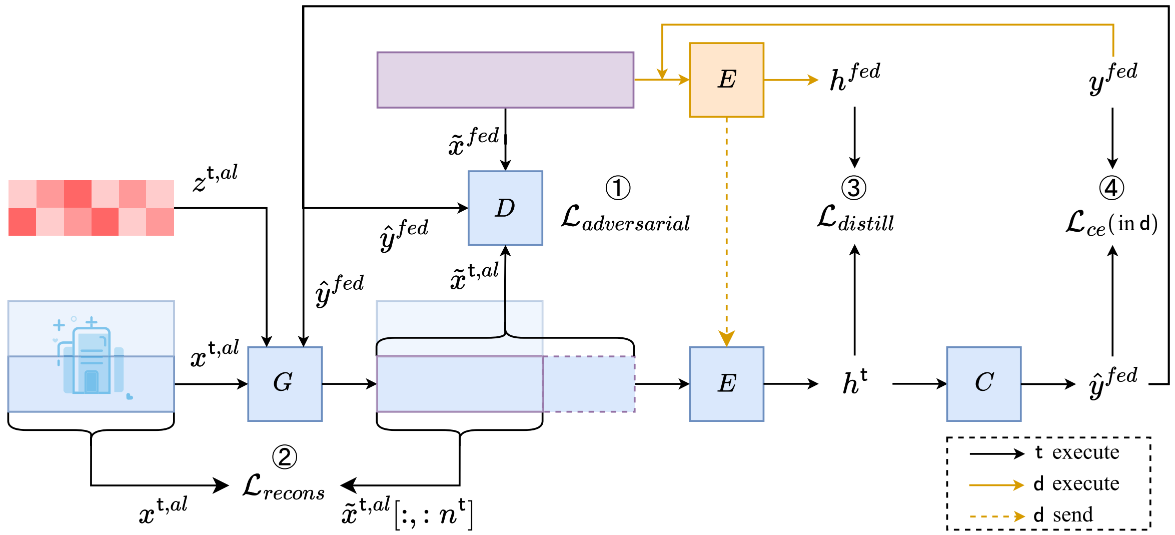


Fig. 6. The Architecture for Shared Representation Augmentation in Step 2.1.

4.3.2. *Shared representation augmentation*

The objective of Step 2.1 is to augment the representation of shared data and integrate it with “knowledge extraction from federated representation”. Fig. 6 illustrates the architecture of Step 2.1.

Prerequisites: Before commencing formal training, it is imperative to ensure that the t side has access to category encodings that correspond to real labels, as previously mentioned. To achieve this, the generated representation $\tilde{x}^{\mathrm{t},al}$ needs to possess robust feature extraction and classification capabilities. As the labeled party, d side conducts supervised training on the federated representation $\tilde{x}^{fed}$ to obtain a feature extractor E^{d} and a classifier C^{d}. Subsequently, E^{d} is shared with t as the local initial feature extractor E^{t}. Although the excellent feature extraction capabilities of E^{d} on $\tilde{x}^{fed}$ may not directly translate to $x^{\mathrm{t},al}$, but it serves as prior knowledge to embed into E^{t}. Consequently, both E^{t} and the randomly initialized C^{t} require subsequent training. The category encodings for the first round are randomly initialized.

Training Objective: t side aims to generate augmented representations of shared data $x^{\mathrm{t},al}$ that match the federated knowledge while preserving a certain degree of individuality.

Learning the Data Distribution of Federated Representations: We employ CGAN as a supervised feature extractor for this data. The discriminator D^{al} and the generator G^{al} engage in the following min–max adversarial game: D^{al} attempts to distinguish the t-side representations $\tilde{x}^{\mathrm{t},al}$ generated by G^{al} from the federated representation $\tilde{x}^{fed}$, while G^{al} endeavors to make the generated $\tilde{x}^{\mathrm{t},al}$ similar to $\tilde{x}^{fed}$ in terms of data distribution. The data distribution of $\tilde{x}^{fed}$ serves as part of the federated knowledge incorporated into the training of E^{t}.

We define a prior noise variable z', which is a combination of the original shared data distribution $p_{\mathrm{t}^{al}}(x)$ and the noise distribution $p_z(z)$. This prior noise variable z' is combined with the conditional label $\hat{y}^{fed}$ as input, and is mapped through G^{al} with parameters $\theta_{g^{al}}$ to generate the representation $\tilde{x}^{\mathrm{t},al} = G^{al}(z'|\hat{y}^{fed};\theta_{g^{al}})$. Together, $\tilde{x}^{\mathrm{t},al}$, $\tilde{x}^{fed}$, and $\hat{y}^{fed}$ are input to D^{al} with parameters $\theta_{d^{al}}$, which outputs a single scalar $D^{al}(\tilde{x}^{\mathrm{t}}|y^{\mathrm{d}};\theta_{d^{al}})$.

During training, D^{al} and G^{al} are updated asynchronously. When updating D^{al}, G^{al} remains fixed. The loss function for D^{al} is as follows:

$$\begin{aligned}\mathcal{L}_D^{al} &= -\mathbb{E}_{x\sim p_{\mathrm{t}^{al}}(x)}[\log D^{al}(x \mid \hat{y}^{fed})]\\ &\quad - \mathbb{E}_{z'\sim p_{\mathrm{t}^{al}+z}(z')}[\log(1 - D^{al}(G^{al}(z' \mid \hat{y}^{fed})))]\end{aligned} \tag{6}$$

When updating G^{al}, D^{al} remains fixed. In our strategy, D^{al} aims to distinguish between $\tilde{x}^{fed}$ and $\tilde{x}^{\mathrm{t},al}$, while G^{al} determines the similarity between $\tilde{x}^{fed}$ and $\tilde{x}^{\mathrm{t},al}$. The loss function for G^{al} as an adversarial loss is as follows:

$$\mathcal{L}_{adversarial}^{al} = \mathcal{L}_G^{al} = -\mathbb{E}_{z'\sim p_{\mathrm{t}^{al}+z}(z')}[\log D^{al}(G^{al}(z' \mid \hat{y}^{fed}))] \tag{7}$$

Through the adversarial interplay between D^{al} and G^{al}, the representation $\tilde{x}^{\mathrm{t},al}$ of shared data can learn federated knowledge from the data distribution of $\tilde{x}^{fed}$:

$$\tilde{x}^{\mathrm{t},al} = G^{al}_{z'\sim p_{\mathrm{t}^{al}+z}(z')}(z' \mid \hat{y}^{fed}) \tag{8}$$

Learning Federated Feature Extraction Capabilities: Due to the noise present in the shared data from all sides, the effective knowledge that the federated representation can provide regarding the data distribution is limited. d shares the pre-trained E^{d} with t to augment t’s ability to extract features from unlabeled generated representations. Simultaneously, the transparency of E^{d} determines t’s access to the distilled representation h^{fed} extracted from $\tilde{x}^{fed}$. The representation extraction capability of E^{d} and the key federated features of h^{fed} are incorporated as part of the federated knowledge in the training of E^{t}. The loss function for E^{t} as distillation loss is given below:

$$\mathcal{L}_{distill} = 1 - \frac{h^{fed}\cdot h^{\mathrm{t}}}{\|h^{fed}\|\cdot\|h^{\mathrm{t}}\|},\quad h^{\mathrm{t}} = E^{\mathrm{t},al}\left(\tilde{x}^{\mathrm{t},al}\right) \tag{9}$$

Retaining Personalization: As participants in FL, the unique characteristics of their own data distinguish them from other parties. Therefore, the federated knowledge is not dominant in augmenting the representation $\tilde{x}^{\mathrm{t},al}$. Our representation augmentation strategy also involves balancing the weight allocation between federated knowledge and personalization. Since h^{fed} is a product of the matrix decomposition of the shared data from all sides, the $\tilde{x}^{\mathrm{t},al}$ generated by CGAN corresponds directly to the original shared data $x^{\mathrm{t},al}$ in terms of data distribution. Hence, we add a reconstruction loss to preserve personalization:

$$\mathcal{L}_{recons}^{al} = \left|\tilde{x}^{\mathrm{t},al}[:,:n^{\mathrm{t}}] - x^{\mathrm{t},al}\right| \tag{10}$$

Ensuring Labeling Correctness and Stability: In t side, the introduction of the classifier C^{al} serves two main purposes within our strategy. Firstly, unlike traditional classifiers that predict explicit classification results, C^{al} ensures that the latent labels corresponding to the generated $\tilde{x}^{\mathrm{t},al}$, derived from shared data $x^{\mathrm{t},al}$, align with the labels of shared data $x^{\mathrm{d},al}$ from d. This ensures the correctness of the generated representation. Secondly, C^{al} ensures that each generated $\tilde{x}^{\mathrm{t},al}$ consistently corresponds to the same label, ensuring the stability of the generated representation. However, the true labels y^{fed} reside with d, so the logits of t’s classification predictions are computed by d to calculate the classification loss, which is then fed back to t for updating:

$$\mathcal{L}_{ce} = -\frac{1}{m^{al}}\sum_{i=1}^{m^{al}} y_i^{fed}\cdot\log(\hat{y}_i^{fed}),\quad \hat{y}^{fed} = C^{al}\left(h^{\mathrm{t}}\right) \tag{11}$$

Here, $\hat{y}^{fed}$ serves both as an input and an output. By minimizing the objective function $\mathcal{L}_{ce}$, it balances and controls its own correctness and stability:

$$\hat{y}^{fed,s+1} = C^{al}\left(E^{\mathrm{t},al}\left(G^{al}_{z'\sim p_{\mathrm{t}^{al}+z}(z')}\left(z'|\hat{y}^{fed,s+1}\right)\right)\right), \quad s = 1, 2, \ldots, T \tag{12}$$

Therefore, for the shared data, we adopt a multi-objective optimization strategy to augment this part of the representation, integrating multiple objective loss functions as the shared data's loss function, defined as follows:

$$\mathcal{L}^{al} = \gamma^{al}_a \mathcal{L}^{al}_{adversarial} + \gamma^{al}_r \mathcal{L}^{al}_{recons} + \gamma_d \mathcal{L}_{distill} + \gamma_c \mathcal{L}_{ce} \tag{13}$$

$$\gamma^{al}_a + \gamma^{al}_r + \gamma_d + \gamma_c = 1 \tag{14}$$

Here, $\gamma^{al}_a, \gamma^{al}_r, \gamma_d, \gamma_c$ serve as weights for the adversarial loss, reconstruction loss, distillation loss, and classification loss, respectively, to coordinate the augmentation of the shared data representation. To comprehensively consider both factors, namely minimizing $\mathcal{L}^{al}$ and ensuring the sum of weights to be 1, we adopt the Lagrangian multiplier method to solve:

$$\mathcal{L}^{al}_{Lagrange} = \mathcal{L}^{al} - \lambda^{al}\left(\gamma^{al}_a + \gamma^{al}_r + \gamma_d + \gamma_c - 1\right) \tag{15}$$

Here, λ is the Lagrange multiplier for shared data. We take the partial derivative of each weight and the Lagrange multiplier with respect to the corresponding loss function and set it equal to 0.

$$\frac{\partial \mathcal{L}^{al}_{Lagrange}}{\partial \gamma^{al}} = 0, \quad \forall \gamma^{al} \in \left\{\gamma^{al}_a, \gamma^{al}_r, \gamma_d, \gamma_c\right\} \tag{16}$$

$$\frac{\partial \mathcal{L}^{al}_{Lagrange}}{\partial \lambda^{al}} = 0 \tag{17}$$

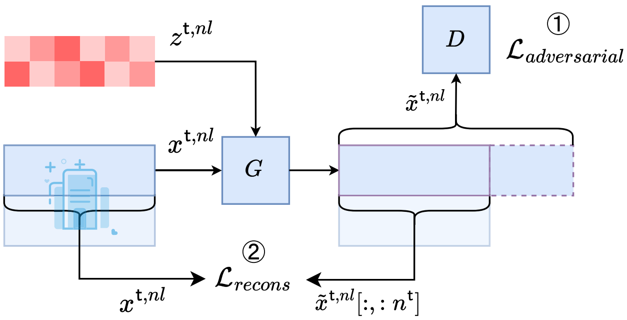
Solving this set of equations yields the gradient for each weight used in the shared data.

Algorithm 1 **Shared Representation Augmentation**

```
**Input:** dataset $\mathcal{D}^{\mathrm{t},al} = \left\{\left(x^{\mathrm{t},al}_i\right)\right\}^{m^{al}}_{i=1} \in \mathbb{R}^{|m^{al}|\times|n^{\mathrm{t}}|}$; $\hat{y}^{fed}$ in d; generator $G^{al}$; discriminator $D^{al}$; extractor $E^{\mathrm{t},al}$; classifier $C^{al}$.
**Output:** $G^{al}, D^{al}$.
⇒ **Run on the** t**-side.**
$\mathcal{B}$ divide $\mathcal{D}^{\mathrm{t},al}$ into batches of batch size.
**for** batch in $\mathcal{B}$ **do**
    $\hat{y}^{fed} = C^{al}\left(E^{\mathrm{t},al}\left(G^{al}_{z'\sim p_{\mathrm{t}^{al}+z}(z')}\left(z'|\hat{y}^{fed}\right)\right)\right)$;
    send $\hat{y}^{fed}$ to d**-side**;
    get $\nabla_{\theta_{g^{al}}}\left(\mathcal{L}_{ce}\right)$ from d**-side**;
    $\theta_{d^{al}} \leftarrow \theta_{d^{al}} - \eta_{d^{al}}\nabla_{\theta_{d^{al}}}\left(\mathcal{L}^{al}_D\right)$;
    $\nabla_{\theta_{g^{al}}}\left(\mathcal{L}^{al}\right) = \nabla_{\theta_{g^{al}}}\left(\mathcal{L}^{al}_G\right) + \nabla_{\theta_{g^{al}}}\left(\mathcal{L}^{al}_{recons}\right) + \nabla_{\theta_{g^{al}}}\left(\mathcal{L}_{distill}\right) + \nabla_{\theta_{g^{al}}}\left(\mathcal{L}_{ce}\right)$;
    $\theta_{g^{al}} \leftarrow \theta_{g^{al}} - \eta_{g^{al}}\nabla_{\theta_{g^{al}}}\left(\mathcal{L}^{al}\right)$;
    $\nabla_{\theta_{e^{al}}}\left(\mathcal{L}^{al}\right) = \nabla_{\theta_{e^{al}}}\left(\mathcal{L}^{al}_G\right) + \nabla_{\theta_{e^{al}}}\left(\mathcal{L}^{al}_{recons}\right) + \nabla_{\theta_{e^{al}}}\left(\mathcal{L}_{distill}\right)$;
    $\theta_{e^{al}} \leftarrow \theta_{e^{al}} - \eta_{e^{al}}\nabla_{\theta_{e^{al}}}\left(\mathcal{L}^{al}\right)$;
    $\nabla_{\theta_{c^{al}}}\left(\mathcal{L}^{al}\right) = \nabla_{\theta_{c^{al}}}\left(\mathcal{L}^{al}_G\right) + \nabla_{\theta_{c^{al}}}\left(\mathcal{L}^{al}_{recons}\right)$;
    $\theta_{c^{al}} \leftarrow \theta_{c^{al}} - \eta_{c^{al}}\nabla_{\theta_{c^{al}}}\left(\mathcal{L}^{al}\right)$;
    $\nabla_{\gamma^{al}}\left(\mathcal{L}^{al}_{\mathrm{Lagrange}}\right) = \frac{\partial\mathcal{L}^{al}}{\partial\gamma^{al}} - \lambda^{al}, \quad \forall\gamma^{al} \in \left\{\gamma^{al}_a, \gamma^{al}_r, \gamma_d, \gamma_c\right\}$;
    $\gamma^{al} = \gamma^{al} - \eta_{\gamma^{al}}\nabla_{\gamma^{al}}$;
    $\nabla_{\lambda^{al}}\left(\mathcal{L}^{al}_{\mathrm{Lagrange}}\right) = \sum_{\gamma^{al}}\gamma^{al} - 1$;
    $\lambda^{al} = \lambda^{al} - \eta_{\lambda^{al}}\nabla_{\lambda^{al}}$.
**end for**
```

The gradient updates for D^{al} in the augmentation of shared data representation are as follows:

$$\nabla_{\theta_{d^{al}}}\left(\mathcal{L}^{al}_D\right) = \frac{\partial}{\partial\theta_{d^{al}}}[\log D^{al}_{x\sim p_{\mathrm{t}^{al}}(x)}(x \mid \hat{y}^{fed})$$



$z^{\mathrm{t},nl}$
D
①
$\mathcal{L}_{adversarial}$
$\tilde{x}^{\mathrm{t},nl}$
$x^{\mathrm{t},nl}$
G
②
$\mathcal{L}_{recons}$
$x^{\mathrm{t},nl}$
$\tilde{x}^{\mathrm{t},nl}[:, : n^{\mathrm{t}}]$

Fig. 7. The Architecture for Privated Representation Augmentation in Step 2.2.

$$+ \log(1 - D^{al}(G^{al}_{z'\sim p_{\mathrm{t}^{al}+z}(z')}(z' \mid \hat{y}^{fed})))] \tag{18}$$

$$\theta_{d^{al}} \leftarrow \theta_{d^{al}} - \eta_{d^{al}}\nabla_{\theta_{d^{al}}}\left(\mathcal{L}^{al}_D\right) \tag{19}$$

The gradient update for G^{al} in the augmentation of shared data representation is as follows:

$$\nabla_{\theta_{g^{al}}}\left(\mathcal{L}^{al}_G\right) = \frac{\partial}{\partial\theta_{g^{al}}}[\log(D^{al}(G^{al}_{z'\sim p_{\mathrm{t}^{al}+z}(z')}(z' \mid \hat{y}^{fed})))] \tag{20}$$

$$\nabla_{\theta_{g^{al}}}\left(\mathcal{L}^{al}_{recons}\right) = \frac{\partial}{\partial\theta_{g^{al}}}[\left|G^{al}_{z'\sim p_{\mathrm{t}^{al}+z}(z')}(z' \mid \hat{y}^{fed})[:, : n^{\mathrm{t}}] - x^{\mathrm{t},al}\right|] \tag{21}$$

$$\nabla_{\theta_{g^{al}}}\left(\mathcal{L}_{distill}\right) = \frac{\partial}{\partial\theta_{g^{al}}}[1 - \frac{E^{\mathrm{t},al}\left(G^{al}_{z'\sim p_{\mathrm{t}^{al}+z}(z')}(z' \mid \hat{y}^{fed})\right)\cdot h^{\mathrm{t}}}{\|E^{\mathrm{t},al}\left(G^{al}_{z'\sim p_{\mathrm{t}^{al}+z}(z')}(z' \mid \hat{y}^{fed})\right)\|\cdot\|h^{\mathrm{t}}\|}] \tag{22}$$

$$\nabla_{\theta_{g^{al}}}\left(\mathcal{L}_{ce}\right) = \frac{\partial}{\partial\theta_{g^{al}}}[-\frac{1}{m^{al}}\sum_{i=1}^{m^{al}}\hat{y}^{fed}\cdot\log(C^{al}\left(E^{\mathrm{t},al}\left(G^{al}_{z'\sim p_{\mathrm{t}^{al}+z}(z')}(z' \mid \hat{y}^{fed})\right)\right)_i)] \tag{23}$$

$$\nabla_{\theta_{g^{al}}}\left(\mathcal{L}^{al}\right) = \nabla_{\theta_{g^{al}}}\left(\mathcal{L}^{al}_G\right) + \nabla_{\theta_{g^{al}}}\left(\mathcal{L}^{al}_{recons}\right) + \nabla_{\theta_{g^{al}}}\left(\mathcal{L}_{distill}\right) + \nabla_{\theta_{g^{al}}}\left(\mathcal{L}_{ce}\right) \tag{24}$$

$$\theta_{g^{al}} \leftarrow \theta_{g^{al}} - \eta_{g^{al}}\nabla_{\theta_{g^{al}}}\left(\mathcal{L}^{al}\right) \tag{25}$$

4.3.3. Private representation augmentation

The purpose of Step 2.2 is to augment the representation of private data. The architecture of Step 2.2 is illustrated in Fig. 7.

Prerequisites: Before formally training the private data $x^{\mathrm{t},nl}$ in the t side, it is important to note that since the federated representation is solely computed from the shared data, and the private data resides within the t side, the knowledge extracted from the federated representation indirectly transfers to the private data. This data transfer is manifested in the sharing of the shared data generator and discriminator with the private data:

$$\theta^s_{v^{nl}\backslash 1} \overset{\text{initial}}{\leftarrow} \theta^s_{v^{al}\backslash 1}, \quad \theta^s_{v^{nl},1} \overset{\text{initial}}{\leftarrow} \theta^s_{v^{al},1,\Theta}, \quad \theta_{v^{nl},1} \subseteq \theta_{v^{al},1},$$
$$\forall v \in \{g, d\}, \quad s = 1, 2, \ldots, T \tag{26}$$

Here, $\theta_{v^{al}\backslash 1}$ and $\theta_{v^{nl}\backslash 1}$ represent the model parameters of $\tilde{x}^{\mathrm{t},al}$ and $\tilde{x}^{\mathrm{t},nl}$ in the generator or discriminator excluding the input layer. $\theta_{v^{al},1}$ and $\theta_{v^{nl},1}$ represent the model parameters of $\tilde{x}^{\mathrm{t},al}$ and $\tilde{x}^{\mathrm{t},nl}$ in the input layer of the generator or discriminator. $\theta_{v^{al},1,\Theta}$ refers to the shared parameters of $\tilde{x}^{\mathrm{t},al}$ in the generator or discriminator's input layer, excluding the parameters of the conditional label $\hat{y}^{fed}$.

Private data $x^{\mathrm{t},nl}$ in t and d sides are unlabeled, hence $\mathcal{L}_{distill}$ and $\mathcal{L}_{ce}$ are not applicable to $x^{\mathrm{t},nl}$. Adversarial loss and reconstruction loss are retained for private data, but they are distinguished from shared data.

Training Objective: t side generates augmented representations for private data $x^{\mathrm{t},nl}$ with federated knowledge while preserving certain personalization. **Learning Federated Data Distribution**: We employ GAN as the unsupervised feature extractor for this data. Unlike traditional GAN, private data and federated representations are not directly related. Hence, discriminator D^{nl} does not require further training. Leveraging the shared discriminator D^{al}, D^{nl} can distinguish between the t-side representations $\tilde{x}^{\mathrm{t},nl}$ generated by G^{nl} and the federated representation $\tilde{x}^{fed}$, while G^{nl} aims to make the generated $\tilde{x}^{\mathrm{t},nl}$ similar to $\tilde{x}^{fed}$ in terms of data distribution.

We define the prior noise variable z', which is a combination of the original shared data distribution $p_{\mathrm{t}^{nl}}(x)$ and the noise distribution $p_z(z)$, resulting in a composite distribution $p_{\mathrm{t}^{nl}+z}(z')$. This z' is used as input to G^{nl}, resulting in the generated representation $\tilde{x}^{\mathrm{t},nl} = G(z')$.

The loss function for G^{nl} as adversarial loss is defined as follows:

$$\mathcal{L}^{nl}_{adversarial} = \mathcal{L}^{nl}_{G} = -\mathbb{E}_{z' \sim p_{\mathrm{t}^{nl}+z}(z')}[\log D^{nl}(G^{nl}(z'))] \tag{27}$$

Through G^{al}, the representation of private data $\tilde{x}^{\mathrm{t},nl}$ can indirectly learn federated knowledge from the data distribution of $\tilde{x}^{fed}$:

$$\tilde{x}^{\mathrm{t},nl} = G^{nl}_{z' \sim p_{\mathrm{t}^{nl}+z}(z')}(z') \tag{28}$$

Retaining Personalization: As the majority of the private data, it has a more direct influence on the personalized nature of the generated representation compared to shared data. Therefore, we still add reconstruction loss to preserve personalization:

$$\mathcal{L}^{nl}_{recons} = \left|\tilde{x}^{\mathrm{t},nl}[:, :n^{\mathrm{t}}] - x^{\mathrm{t},nl}\right| \tag{29}$$

Therefore, for private data, we still employ a multi-objective optimization strategy to augment this part of the representation, integrating multiple loss functions into the private data's loss function as follows:

$$\gamma^{nl}_a + \gamma^{nl}_r = 1 \tag{30}$$

Here, γ^{nl}_a and γ^{nl}_r are the weights for the adversarial loss and reconstruction loss for private data, respectively, used to balance the augmentation of the private data representation. To simultaneously consider both factors, minimizing $\mathcal{L}^{al}$ and ensuring the sum of weights is 1, we employ the Lagrangian multiplier method to address this:

$$\mathcal{L}^{nl}_{Lagrange} = \mathcal{L}^{nl} - \lambda^{nl}\left(\gamma^{nl}_a + \gamma^{nl}_r - 1\right) \tag{31}$$

Here, λ is the Lagrange multiplier for shared data. We take the partial derivative of each weight and the Lagrange multiplier with respect to the corresponding loss function and set it equal to 0.

$$\frac{\partial \mathcal{L}^{nl}_{Lagrange}}{\partial \gamma^{nl}} = 0, \quad \forall \gamma^{nl} \in \left\{\gamma^{nl}_a, \gamma^{nl}_r\right\} \tag{32}$$

$$\frac{\partial \mathcal{L}^{nl}_{Lagrange}}{\partial \lambda^{nl}} = 0 \tag{33}$$

Solving this set of equations yields the gradients for each weight used in private data.

It is worth noting that compared to γ^{al}_a and γ^{al}_r, γ^{nl}_a and γ^{nl}_r are larger, indicating that private data has a greater influence on the personalized augmentation of the overall representation. After each round of private data training, the generator for private data provides feedback to the shared data:

$$\theta^{s+1}_{g^{al}\backslash 1} \overset{\text{initial}}{\leftarrow} \theta^{s}_{g^{nl}\backslash 1}, \quad \theta^{s+1}_{g^{al},1,\Theta} \overset{\text{initial}}{\leftarrow} \theta^{s}_{g^{nl},1}, \quad s = 1, 2, \ldots, T \tag{34}$$

Algorithm 2 Private Representation Augmentation

```
Input: dataset $\mathcal{D}^{\mathrm{t},nl} = \left\{\left(x^{\mathrm{t}}_i\right)\right\}^{m^{\mathrm{t},nl}}_{i=1} \in \mathbb{R}^{\left|m^{\mathrm{t},nl}\right| \times |n^{\mathrm{t}}|}$; generator $G^{nl}$; discriminator $D^{nl}$.
Output: $G^{nl}$.
⇒ Run on the t side.
$\mathcal{B}$ divide $\mathcal{D}^{\mathrm{t},nl}$ into batches of batch size.
for batch in $\mathcal{B}$ do
    $\nabla_{\theta_{g^{nl}}}\left(\mathcal{L}^{nl}\right) = \nabla_{\theta_{g^{nl}}}\left(\mathcal{L}^{nl}_{G}\right) + \nabla_{\theta_{g^{nl}}}\left(\mathcal{L}^{nl}_{recons}\right)$;
    $\theta_{g^{nl}} \leftarrow \theta_{g^{nl}} - \eta_{g^{nl}} \nabla_{\theta_{g^{nl}}}\left(\mathcal{L}^{nl}\right)$.
    $\nabla_{\gamma^{nl}}\left(\mathcal{L}^{nl}_{\text{Lagrange}}\right) = \frac{\partial \mathcal{L}^{nl}}{\partial \gamma^{nl}} - \lambda^{nl}, \quad \forall \gamma^{nl} \in \left\{\gamma^{nl}_a, \gamma^{nl}_r\right\}$;
    $\gamma^{nl} = \gamma^{nl} - \eta_{\gamma^{nl}} \nabla_{\gamma^{nl}}$;
    $\nabla_{\lambda^{nl}}\left(\mathcal{L}^{nl}_{\text{Lagrange}}\right) = \sum_{\gamma^{nl}} \gamma^{nl} - 1$;
    $\lambda^{nl} = \lambda^{nl} - \eta_{\lambda^{nl}} \nabla_{\lambda^{nl}}$.
end for
```

The gradient updates for G^{nl} in the augmentation of private data representation are as follows:

$$\nabla_{\theta_{g^{nl}}}\left(\mathcal{L}^{nl}_{G}\right) = \frac{\partial}{\partial \theta_{g^{nl}}}[\log(D^{nl}(G^{nl}_{z' \sim p_{\mathrm{t}^{al}+z}(z')}(z')))] \tag{35}$$

$$\nabla_{\theta_{g^{nl}}}\left(\mathcal{L}^{nl}_{recons}\right) = \frac{\partial}{\partial \theta_{g^{nl}}}[\left|G^{nl}_{z' \sim p_{\mathrm{t}^{nl}+z}(z')}(z')[:, :n^{\mathrm{t}}] - x^{\mathrm{t},nl}\right|] \tag{36}$$

$$\nabla_{\theta_{g^{nl}}}\left(\mathcal{L}^{nl}\right) = \nabla_{\theta_{g^{nl}}}\left(\mathcal{L}^{nl}_{G}\right) + \nabla_{\theta_{g^{nl}}}\left(\mathcal{L}^{nl}_{recons}\right) \tag{37}$$

$$\theta_{g^{nl}} \leftarrow \theta_{g^{nl}} - \eta_{g^{nl}} \nabla_{\theta_{g^{nl}}}\left(\mathcal{L}^{nl}\right) \tag{38}$$

Algorithm 3 CrossKT-FRA

```
Input: datasets $\mathcal{D}^{\mathrm{t},al} = \left\{\left(x^{\mathrm{t},al}_i\right)\right\}^{m^{al}}_{i=1} \in \mathbb{R}^{|m^{al}| \times |n^{\mathrm{t}}|}, \mathcal{D}^{\mathrm{d},al} = \left\{\left(x^{\mathrm{d},al}_i\right)\right\}^{m^{al}}_{i=1} \in \mathbb{R}^{|m^{al}| \times |n^{\mathrm{d}}|}$; epoch number $T$.
Output: $\tilde{x}^{\mathrm{t}}$.
⇒ Run on the d side.
$\tilde{x}^{fed} \leftarrow$ FedSVD $\left(\mathcal{D}^{\mathrm{t},al}, \mathcal{D}^{\mathrm{d},al}\right)$ or VFedPCA $\left(\mathcal{D}^{\mathrm{t},al}, \mathcal{D}^{\mathrm{d},al}\right)$.
$E^{\mathrm{d}} \leftarrow$ RunLocal $\left(\tilde{x}^{fed}, y^{fed}\right)$;
Send $(\tilde{x}^{fed}, E^{\mathrm{d}})$ to t side.
⇒ Run on the t side.
$E^{\mathrm{t},al} \overset{\text{initial}}{\leftarrow} E^{\mathrm{d}}$.
for $s$ in $1, 2, ..., T$ do
    $G^{al,s}, D^{al,s} \leftarrow$ Shared Representation Augmentation;
    $\theta^{s}_{v^{nl}\backslash 1} \overset{\text{initial}}{\leftarrow} \theta^{s}_{v^{al}\backslash 1}, \quad \theta^{s}_{v^{nl},1} \overset{\text{initial}}{\leftarrow} \theta^{s}_{v^{al},1,\Theta}, \quad \forall v \in \{g, d\}$;
    $G^{nl,s} \leftarrow$ Private Representation Augmentation;
    $\theta^{s+1}_{g^{al}\backslash 1} \overset{\text{initial}}{\leftarrow} \theta^{s}_{g^{nl}\backslash 1}, \quad \theta^{s+1}_{g^{al},1,\Theta} \overset{\text{initial}}{\leftarrow} \theta^{s}_{g^{nl},1}$.
end for
$\tilde{x}^{\mathrm{t},al} = G^{al}_{z' \sim p_{\mathrm{t}^{al}+z}(z')}(z' \mid \hat{y}^{fed})$;
$\tilde{x}^{\mathrm{t},nl} = G^{nl}_{z' \sim p_{\mathrm{t}^{nl}+z}(z')}(z')$;
$\tilde{x}^{\mathrm{t}} = \left\{\tilde{x}^{\mathrm{t},al}, \tilde{x}^{\mathrm{t},nl}\right\}$.
```

5. Experiment

5.1. Experimental setup

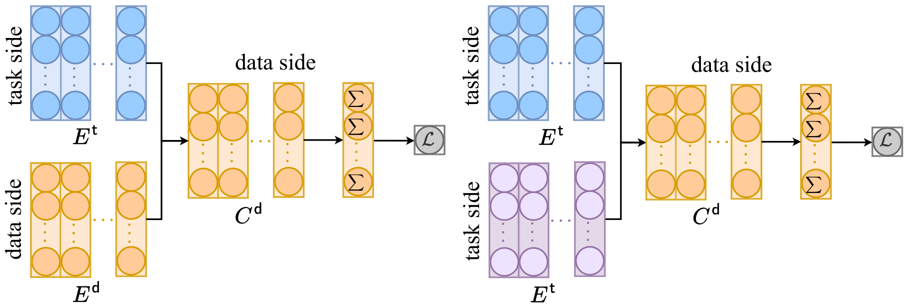
5.1.1. Computational resources

Our experiments utilized an NVIDIA GeForce RTX 4080 GPU, a 13th Gen Intel(R) Core(TM) i7-13700KF processor, 32 GB of RAM, PyTorch 2.0.0, Python 3.9, and CUDA 12.2.

5.1.2. Dataset information

We evaluate our mechanism on the four medical-related datasets:

- *Medical Information Mart for Intensive Care III (MIMIC-III)* dataset, as referenced by Johnson et al. [72], contains clinical data from 58,976 hospitalized patients. It encompasses 26 features covering patients' demographics, diagnoses, treatment processes, laboratory test results, medication usage records, and more. The target variable is Length of Stay (LOS), which ranges from 0 to 3 for prediction.



(a) Split Vertical Federated Learning(SplitVFL).

(b) Center Split Vertical Federated Learning(SplitVFL$_c$).

Fig. 8. Two categories of SplitVFL.

- *Breast* dataset, referenced by Street et al. [73], comprises digital images of 569 breast cell biopsy samples. It encompasses 31 numerical features derived from digital image processing and feature extraction, including characteristics of cell nuclei morphology and structure. The target variable is the diagnosis of breast cell tumors, which is a binary classification (M = malignant, B = benign).

5.1.3. Baseline

Based on the training process of VFL, it can be subdivided into two training modes: Split Vertical Federated Learning (SplitVFL) and Aggregated Vertical Federated Learning (AggVFL).

Split Vertical Federated Learning(SplitVFL): Split learning is an approach in VFL. As illustrated in Fig. 8(a), in SplitVFL, each task side executes a deep learning model adapted to its own dataset. This model serves as the first half of the overall model, ensuring consistency in output format at the final layer, thus enabling aggregation of results from different task sides. Aggregation can take various forms such as simple averaging or weighted averaging, combining outputs from different task sides to generate intermediate results of the overall model. These intermediate results are then passed to the data side, who executes the second half of the overall model to generate the final output.

In the training process of SplitVFL, due to the nature of split learning, the model consists of both the task-side model and the data-side model. The parameters of the task-side model typically require collaborative efforts between the task side and the data side. Local gradients are computed separately by each side, and the parameters are trained using composite derivatives. The gradient calculation for the overall model parameters is as follows:

$$\frac{\partial \mathcal{L}}{\partial \omega_a} = \frac{\partial \mathcal{L}}{\partial o_a}\frac{\partial o_a}{\partial \omega_a} \tag{39}$$

$$\frac{\partial \mathcal{L}}{\partial \omega_p} = \frac{\partial \mathcal{L}}{\partial o_a}\frac{\partial o_a}{\partial o_p}\frac{\partial o_p}{\partial \omega_p} \tag{40}$$

Where $\mathcal{L}$ represents the loss function, ω_a and o_a are the model parameters and model outputs of the data side, respectively, while ω_p and o_p represent the model parameters and model outputs of the task side.

As shown in Fig. 8, a variant of SplitVFL known as SplitVFL$_c$ exists, where the data side's dataset lacks feature information and only contains label information, acting as a centralized entity in such scenarios.

SplitVFL offers the advantage of allowing task sides to learn from the model parameters and results aggregated from the data side's posterior model and the final aggregated model output. This theoretically improves model performance, and different task sides can design structurally different but output-consistent front-end models according to their needs. However, SplitVFL has the disadvantage of requiring the joint participation of task sides and data side in the training process and model deployment. This increases the computational burden on the data side and introduces inconvenience.

Aggregated Vertical Federated Learning(AggVFL): Aggregation learning, another approach in VFL, operates as depicted in Fig. 9(a). In AggVFL, each task side executes a deep learning model tailored to its dataset, generating uniformly formatted model outputs. These outputs are then transmitted to the data side for aggregation, typically through simple averaging or weighted averaging. The aggregated output is used by the data side for loss calculation, completing the algorithm's forward computation.

In the backward training process of AggVFL, since the model only resides on the task side while the model outputs are aggregated by the data side, collaboration between task and data sides is still required for computing local gradients and training model parameters using composite derivatives. The calculation of model parameter gradients is as follows:

$$\frac{\partial \mathcal{L}}{\partial \omega_p} = \frac{\partial \mathcal{L}}{\partial o_a}\frac{\partial o_a}{\partial o_p}\frac{\partial o_p}{\partial \omega_p} \tag{41}$$

Here, $\mathcal{L}$ represents the loss function, o_a denotes the aggregated model output from the data side, and ω_p and o_p are respectively the model parameters and outputs from the task side.

As depicted in Fig. 9, AggVFL also has a variant known as AggVFL$_c$. In this scenario, the data side's dataset lacks feature information and contains only label information, serving as a form of centralized presence.

The advantage of AggVFL lies in the fact that task sides can design models with different structures according to their needs, as long as the output formats remain consistent. Moreover, aggregation only occurs on the data side during the training process, reducing the computational burden on the data side. Additionally, once the model is deployed, task sides can operate independently, enhancing convenience. However, AggVFL's drawback is that during training, learning about the results and experiences of different task sides is limited to the aggregation performed on the data side, which theoretically restricts the model's performance.

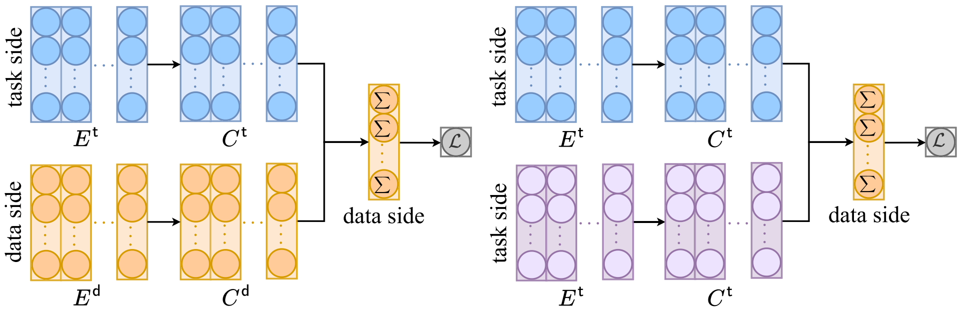
5.1.4. Notation

The used notation can be found in Table 1:

5.1.5. Model setup

In our experiments, we uniformly set up one medical institution as the task side and one medical institution as the data side.

In FRL, we utilized FedSVD and VFedPCA. FedSVD is the default method for the Breast dataset, while VFedPCA is the default for the MIMIC-III dataset. The key parameters for both methods are listed in Table 2.



(a) Aggregated Vertical Federated Learning(AggVFL). (b) Center Aggregated Vertical Federated Learning(AggVFL$_c$).

Fig. 9. Two categories of AggVFL.

Table 1
List of used notions.

Notation	Description
t, d	Task institution and Data institution.
$\mathcal{D}^{\mathrm{t}} = \left\{\left(x_i^{\mathrm{t}}\right)\right\}_{i=1}^{m^{\mathrm{t}}}$	Data from task institution.
$\mathcal{D}^{\mathrm{d}} = \left\{\left(x_i^{\mathrm{d}}, y_i^{\mathrm{d}}\right)\right\}_{i=1}^{m^{\mathrm{d}}}$	Data from data institution.
$\mathrm{p} \in \{\mathrm{t}, \mathrm{d}\}$	Either participating institution of the task or data institution.
$m^{\mathrm{p}}, n^{\mathrm{p}}$	Samples and features of participating institutions.
m^{al}, n^{al}	Samples and features of overlapping data.
m^{nl}, n^{nl}	Samples and features of non-overlapping data.
x^{al}, x^{nl}	Shared and private data of participating institutions.
$\tilde{x}^{fed}$	Federated representation.
G^{al}, G^{nl}	Task-side generator of shared and private data.
D^{al}, D^{nl}	Task-side discriminator of shared and private data.
$E^{\mathrm{p},al}$	Extractor for shared data from institutions.
C^{al}	Classifier for shared data from institutions.
$v \in \{g, d\}$	Either role of generator or discriminator.
$\theta_{v^{al}\backslash 1}\theta_{v^{nl}\backslash 1}$	Parameters of the models for shared and private data of roles, excluding the input layer.
$\theta_{v^{al},1}$	Parameters of the model input layer for shared data of roles.
$\theta_{v^{nl},1,\Theta}$	Parameters of the model input layer for private data of roles, excluding the label condition.
$\mathcal{L}_{adversarial}, \mathcal{L}_G, \gamma_a$	Adversarial loss and its corresponding weight.
$\mathcal{L}_{recons}, \gamma_r$	Reconstruction loss and its corresponding weight.
$\mathcal{L}_{distill}, \gamma_d$	Distillation loss and its corresponding weight.
$\mathcal{L}_{ce}, \gamma_c$	classification loss and its corresponding weight.
$\tilde{x}^{\mathrm{t}}$	Augmented representation of task institution.
$\mathcal{L}^{al}_{\mathrm{Lagrange}}, \mathcal{L}^{al}_{\mathrm{Lagrange}}$	Lagrangian functions for shared and private data.
$\lambda^{al}, \lambda^{nl}$	Lagrangian multipliers for shared and private data.

Table 2
Default key parameters in FRL.

FRL	Parameter	Default	Description
FedSVD	*num_party*	2	The number of participants.
	block_size	100	Build fx-size block in orthogonal matrix generation.
VFedPCA	*party_party*	2	The number of participants.
	iter_num	100	The number of local power iteration.
	period_num	10	The number of communication period.
	warm_start	True	Use the previous global aggregation vector.

The components of the representation augmentation module, $\mathcal{L}adversarial$, $\mathcal{L}distill$, and $\mathcal{L}_{ce}$, correspond to the (generator + discriminator), extractor, and classifier, respectively. The generator and discriminator employ a sequence of multiple linear layers and non-linear activation functions. The extractor and classifier adopt Convolutional Neural Network (CNN). The key parameters for these components are shown in Table 3.

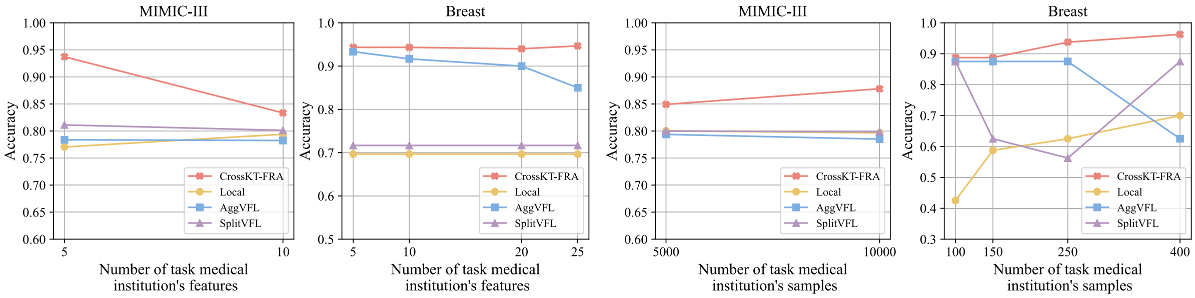
5.2. Experimental analysis

We designed three sets of experiments.

1. **Independent Application**: To assess the standalone performance of CrossKT-FRA, wherein augmented local representations are directly applied to training and prediction in local medical models. This approach aims to evaluate the generalizability across medical data with different heterogeneous distributions, and to explore the impact of personalization and federated knowledge on model performance. Experimentally, we adjust the level of personalized knowledge contained in the private data of the task medical institution by varying the feature quantity and sample size. This is done to observe the model's performance under different levels of personalization. We also adjust the size of global knowledge contained in the shared data by varying the feature quantity and sample size between the task medical institution and the data medical institution. Finally, we adjust the size of indirect global knowledge contained in the private data of the data medical institution by varying its feature quantity and sample size.
2. **Enhancement Optimization**: To validate the enhancement optimization performance of CrossKT-FRA, wherein CrossKT-FRA replaces the feature extraction module of local models or VFL models, and the augmented representation replaces the original module output. This method optimizes the generalizability of different local models or VFL models applied in medical scenarios. Experimentally, the local model is trained at the task medical institution, while the VFL model is jointly trained at both the task and data medical institutions. By varying the feature quantity and sample size of the task data and shared

Table 3
Default key parameters in the representation augmentation module.

Component	Parameter	Default	Description
Discriminator and Generator	*d_depth*	4	The depth of discriminator.
	g_depth	4	The depth of generator.
	activation	LeakyReLU	The activation function
	d_activation_last	Sigmoid	The activation function for the last layer of the discriminator.
	g_activation_last	Tanh	The activation function for the last layer of the generator.
	negative_slope	0.2	The angle of the negative slope.
	learning_rate	0.0002	The learning rate.
	batch_size	100	Batch size.
	epochs	20	Number of iterations.
Extractor	*Conv_depth*	2	Number of convolution layers.
	Conv_kernel_size	3	Convolution kernel size.
	Conv_stride	1	Convolutional step.
	Conv_padding	1	Convolution filling.
	Conv_kernel_size	2	Pooling kernel size.
	Conv_stride	2	Pooling step.
	Conv_padding	1	Pooling filling.
Classifier	*depth*	2	Number of linear layers.
	fc1_output	512	First linear layer output.
	fc2_output	num_classes	Second linear layer output (total number of categories).
	activation	LeakyReLU	The activation function.
	activation_last	Softmax	The activation function for the last layer.



(a) Varying features of task medical institution.

(b) Varying samples of task medical institution.

Fig. 10. Predictions accuracy by varying the task medical institution's data.

data, we observe the performance impact of the augmented representation optimizing the VFL model.

3. **Ablation Study**: By adding and removing the components proposed by CrossKT-FRA, namely $\mathcal{L}_{adversarial}$, $\mathcal{L}_{recons}$, $\mathcal{L}_{distill}$, $\mathcal{L}_{ce}$, and incorporating private data into the knowledge transfer process, we discuss the reasons behind the performance improvement for each component.

5.2.1. *Independent application*

Fig. 10 illustrates the impact of varying the number of features and samples from task medical institutions on the predictive performance for the MIMIC-III and Breast datasets. The results demonstrate that CrossKT-FRA consistently outperforms the baselines, maintaining high diagnostic accuracy across different levels of data personalization. The feature and sample numbers of medical institutions directly reflect the amount of private data, which is intrinsically linked to the degree of data personalization. After the federated knowledge helps the task medical institution find the globally optimal solution range, personalized knowledge can better assist in searching for the optimal solution that suits its specific needs within that range. The experiments also confirm the effectiveness of the knowledge transfer mechanism between shared and private data in task medical institutions.

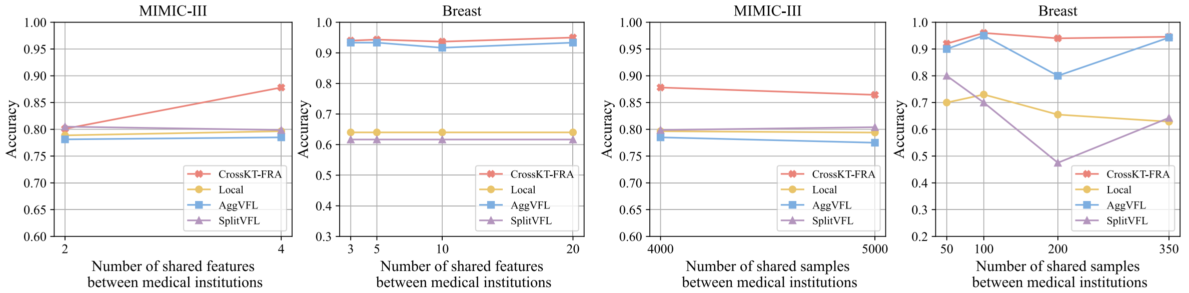
Fig. 11 illustrates the impact of varying the number of shared data features and samples between task and data medical institutions on the predictive performance for the MIMIC-III and Breast datasets. The results demonstrate that CrossKT-FRA consistently outperforms the baselines, maintaining high diagnostic accuracy across different volumes of federated knowledge. The number of shared data features and samples among data medical institutions directly reflects the volume of shared data, which is intrinsically linked to the amount of federated knowledge. Federated knowledge helps to determine the range of optimal solutions. The experiments also confirm the effectiveness of the federated knowledge transfer mechanism between data and task medical institutions.

Fig. 12 illustrates the impact of varying the number of features and samples of the data medical institution on the predictive performance of MIMIC-III and Breast datasets. The results demonstrate that CrossKT-FRA consistently outperforms the baseline. On the smaller Breast dataset, it maintains high diagnostic accuracy, while on the larger MIMIC-III dataset, accuracy significantly increases with the growth in the number of features and samples. The volume of data from medical institutions is crucial for the capability of federated knowledge representation extraction, laying the groundwork for the federated knowledge transfer mechanism. The experiments also confirm that the representation extraction capability effectively supports the knowledge transfer mechanism.

In each group of simulated split data environments, the data from each medical institution and the shared data for comparison between CrossKT-FRA independent application and the base model are shown in Table 4. Each model epoch is set to 50.

5.2.2. *Enhanced optimization: Local optimization*

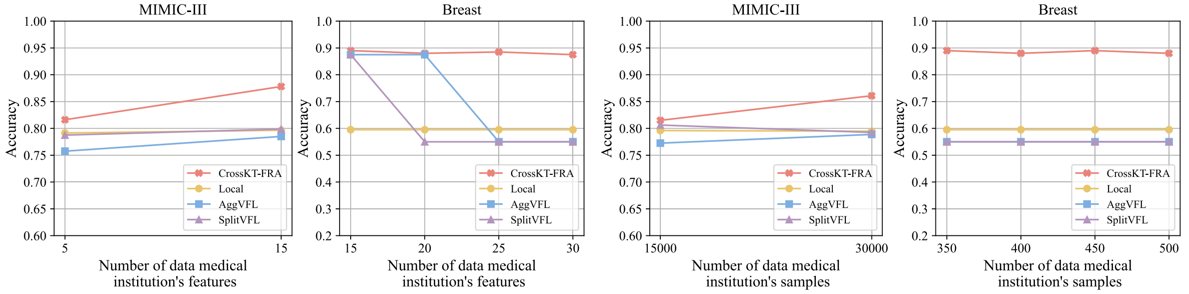
In medical downstream tasks, healthcare institutions can directly optimize their local medical models using the representations augmented by CrossKT-FRA. Tables 5 and 6 display the accuracy of different machine learning(ML) algorithms on the MIMIC-III and Breast datasets, respectively, before and after applying CrossKT-FRA augmented representations. These tables illustrate results under various



(a) Varying shared features between medical institutions.

(b) Varying shared samples between medical institutions.

Fig. 11. Predictions accuracy by varying the shared data between medical institutions.



(a) Varying features of data medical institution.

(b) Varying samples of data medical institution.

Fig. 12. Predictions accuracy by varying the data medical institution's data.

Table 4
The default samples and features held by each medical institution's data and the shared data in the independent application experiment.

Group	Dataset	Feature			Sample		
		Task	Shared	Data	Task	Shared	Data
Task feature	MIMIC-III	**5,10**	4	15	10 000	4000	15 000
	Breast	**5,10,20,25**	3	30	400	300	500
Task sample	MIMIC-III	10	4	15	**5000,10000**	4000	15 000
	Breast	20	15	30	**100,150,250,400**	80	500
Shared feature	MIMIC-III	10	**2,4**	15	10 000	4000	15 000
	Breast	25	**3,5,10,20**	30	400	300	500
Shared sample	MIMIC-III	10	4	15	1000	**4000,5000**	15 000
	Breast	25	10	30	400	**50,100,200,350**	500
Data feature	MIMIC-III	10	4	**5,15**	10 000	4000	15 000
	Breast	12	10	**15,20,25,30**	300	200	500
Data sample	MIMIC-III	10	4	15	10 000	4000	**15000,30000**
	Breast	12	10	30	300	200	**350,400,450,500**

feature and sample sizes shared between different medical institutions and the task-specific institution (considering the negligible impact of the private data volume from the data-owning institution on the local model, we did not separately control for the feature and sample size of the data-owning institution). The results indicate that after applying CrossKT-FRA for representation optimization, the performance of different ML algorithms improves. Notably, this augmentation is more significant on the MIMIC-III dataset, while only marginal improvements are observed on the Breast dataset.

We believe that the differences in these results can be attributed to the following factors:

1. **Dataset**: The MIMIC-III dataset is relatively large, containing more samples and features, which poses a greater challenge for models in finding optimal solutions. In contrast, the smaller Breast dataset allows models to find optimal solutions more easily. Additionally, the complexity of the MIMIC-III dataset is higher than that of the Breast dataset. CrossKT-FRA augments representation learning, enabling the extraction of more meaningful features that adapt to complex data characteristics, thus improving model performance on large complex datasets.
2. **ML Algorithms**: Traditional ML algorithms, when trained directly on raw data, achieve an average accuracy of only around 30%. This indicates that the feature representations extracted by traditional ML algorithms from raw data are insufficient to support high-performance predictive models. In contrast, CrossKT-FRA provides superior feature representations, enabling traditional ML algorithms to more easily find optimal solutions during training.
3. **CrossKT-FRA's Representation Augmentation Mechanism**: CrossKT-FRA combines global knowledge from different data-owning institutions with local personalized knowledge, offering an effective strategy for knowledge transfer and sharing.

Therefore, the experiments demonstrate that CrossKT-FRA can effectively improve the performance of local ML algorithms, particularly when handling large-scale and complex datasets. It can extract

Table 5
Accuracy of downstream tasks on MIMIC-III after Local optimization.

Method	Feature		Sample		Accuracy(%)	
	Task	Shared	Task	Shared	Original	Ours
Neural Network [74]	5	3	1×10^4	5×10^3	31.20	**88.54**
	10	5	1×10^4	5×10^3	29.65	**73.15**
	5	3	1.5×10^4	8×10^3	30.26	**94.93**
	10	5	1.5×10^4	8×10^3	29.76	**71.56**
Random Forest [75]	5	3	1×10^4	5×10^3	31.50	**95.45**
	10	5	1×10^4	5×10^3	30.50	**70.20**
	5	3	1.5×10^4	8×10^3	30.36	**91.56**
	10	5	1.5×10^4	8×10^3	32.26	**83.56**
K Nearest Neighbors [76]	5	3	1×10^4	5×10^3	29.20	**80.65**
	10	5	1×10^4	5×10^3	29.20	**85.90**
	5	3	1.5×10^4	8×10^3	29.26	**80.40**
	10	5	1.5×10^4	8×10^3	28.96	**62.93**
Adaptive Boosting [77]	5	3	1×10^4	5×10^3	30.95	**77.35**
	10	5	1×10^4	5×10^3	30.05	**64.75**
	5	3	1.5×10^4	8×10^3	32.03	**57.03**
	10	5	1.5×10^4	8×10^3	31.83	**55.33**
eXtreme Gradient Boosting [78]	5	3	1×10^4	5×10^3	29.95	**97.20**
	10	5	1×10^4	5×10^3	30.15	**86.25**
	5	3	1.5×10^4	8×10^3	30.36	**95.50**
	10	5	1.5×10^4	8×10^3	30.53	**93.16**

Table 6
Accuracy of downstream tasks on Breast after Local optimization.

Method	Feature		Sample		Accuracy(%)	
	Task	Shared	Task	Shared	Original	Ours
Neural Network [74]	5	2	350	300	62.85	**64.28**
	10	4	350	300	60.00	**64.28**
	5	2	400	350	65.00	**67.50**
	10	4	400	350	61.25	**71.25**
Random Forest [75]	5	2	350	300	58.57	**71.42**
	10	4	350	300	65.71	**71.42**
	5	2	400	350	58.75	**71.25**
	10	4	400	350	65.00	**66.25**
K Nearest Neighbors [76]	5	2	350	300	60.00	**67.14**
	10	4	350	300	60.00	**64.28**
	5	2	400	350	62.50	**63.75**
	10	4	400	350	65.00	**66.25**
Adaptive Boosting [77]	5	2	350	300	61.42	**71.42**
	10	4	350	300	62.85	**64.28**
	5	2	400	350	58.75	**65.00**
	10	4	400	350	61.25	**68.75**
eXtreme Gradient Boosting [78]	5	2	350	300	61.42	**64.28**
	10	4	350	300	65.71	**70.00**
	5	2	400	350	62.50	**63.75**
	10	4	400	350	65.00	**67.50**

more meaningful features, reducing the search difficulty during model training and significantly boosting model performance. CrossKT-FRA's representation augmentation mechanism proves to be applicable and effective across different datasets and scenarios.

We utilize various machine learning algorithms to train models for downstream tasks. In our experiments, we use Neural Network [74], Random Forest [75], K Nearest Neighbors [76], Adaptive Boosting [77], and eXtreme Gradient Boosting [78] for robustness assessment. The parameters of the downstream models are summarized in Table 7.

5.2.3. Enhancement optimization: VFL optimization

Traditional VFL requires participating healthcare institutions to find valuable information in overlapping data (as discussed in Section 1). In Figs. 13 and 14, we control for different amounts of shared data (in terms of features and samples) to optimize AggVFL and SplitVFL prediction performance on the MIMIC-III and Breast datasets, respectively. The results demonstrate that with the augmented representations provided by CrossKT-FRA, the optimized VFL algorithms consistently outperform the original VFL baselines. This indicates that CrossKT-FRA effectively compensates for the lack of shared data features by transferring knowledge from private data to shared data, thus improving prediction accuracy. In addition to performance improvements, the VFL algorithms using CrossKT-FRA augmented representations also exhibit significant advantages in computation time. This is because CrossKT-FRA optimizes the representations during the learning stage, reducing the complexity of subsequent model training and thereby increasing computational efficiency. Consequently, CrossKT-FRA not only improves prediction accuracy across various datasets and scenarios but also significantly reduces computation time, demonstrating its superiority and practicality for large-scale complex datasets.

In each simulated split data environment, when comparing the CrossKT-FRA optimized VFL model with the original VFL model, the data from each medical institution is consistent, and efforts are made to maximize the reduction of the impact caused by this part of the data, as shown in Table 8. Each model epoch is set to 50.

5.2.4. Ablation study

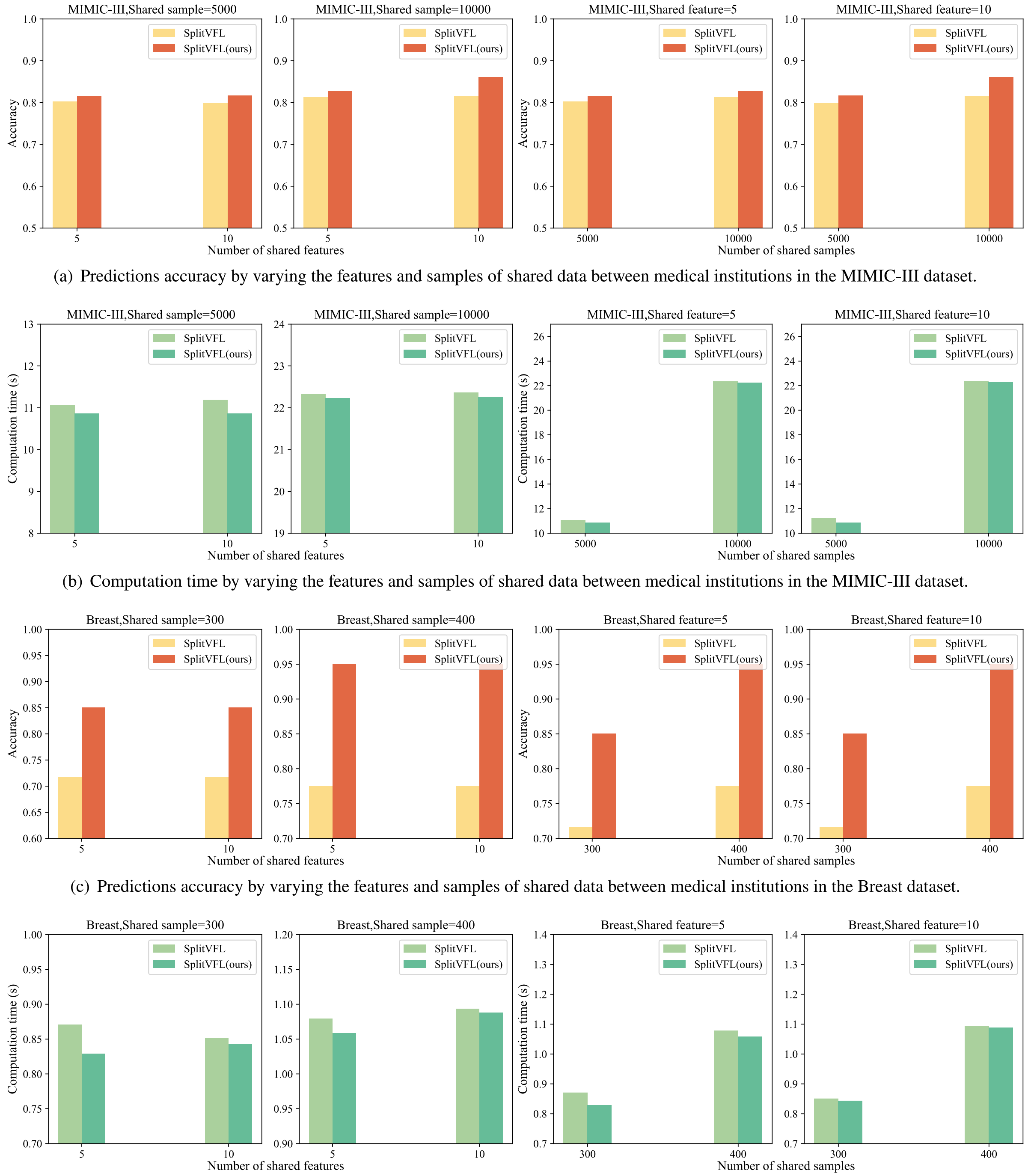
Figs. 15 and 16 illustrate the impact on predictive performance for the MIMIC-III and Breast datasets when using the shared data representation augmentation module. The analysis considers scenarios with and without $\mathcal{L}_{adversarial}$(adversarial loss), $\mathcal{L}_{distill}$(distillation loss), $\mathcal{L}_{recons}$(reconstruction loss) and $\mathcal{L}_{ce}$(classification loss). The performance is evaluated by varying the number of features and samples of the task medical institution, as well as by varying features and samples of the shared data between the task and data medical institutions. The results demonstrate that incorporating all these losses leads to the highest prediction accuracy. Incorporating $\mathcal{L}_{adversarial}$ helps extract global knowledge from federated representations. The inclusion of $\mathcal{L}_{distill}$ effectively distills critical information from the generated representations. $\mathcal{L}_{recons}$ plays a crucial role in preserving the personalized knowledge of the original data in the data distribution. $\mathcal{L}_{ce}$ ensures consistency in the labels of the augmented representations. By combining these losses, our method improves prediction performance across various datasets and data distribution scenarios, highlighting the effectiveness of our knowledge transfer mechanism.

Figs. 17 and 18 illustrate the impact on prediction performance for the MIMIC-III and Breast datasets when using the shared data representation augmentation module. The analysis considers scenarios with and without $\mathcal{L}_{adversarial}$(adversarial loss) and $\mathcal{L}_{recons}$(reconstruction loss). The performance is evaluated by varying the number of features and samples of the task medical institution, as well as by varying features and samples of the shared data between the task and data medical institutions. The experiments demonstrate that incorporating both $\mathcal{L}_{adversarial}$ and $\mathcal{L}_{recons}$ results in the highest prediction accuracy. $\mathcal{L}_{adversarial}$ aids in extracting global knowledge from federated representations, while $\mathcal{L}_{recons}$ preserves the personalized knowledge of the original data in the distribution of a large volume of private data. Therefore, the combination of these losses improves prediction performance across different datasets and data distribution scenarios, showcasing the effectiveness of our knowledge transfer mechanism.

In each simulated data split scenario, the distribution of data among the medical institutions and shared data when components are ablated in the private data part of CrossKT-FRA can be found in Table 9. The number of epochs for each model is set to 50.

To empirically validate the benefits of utilizing private data in the semi-supervised representation augmentation module, we conducted an ablation study. This study involved varying the feature and sample of both task data and shared data to indirectly control private data, allowing us to compare model performance with and without private data participation in knowledge transfer. Table 10 summarizes the experimental results, showing the prediction accuracy on test sets comprising both shared and private data across different configurations.

It is evident that when private data participates in knowledge transfer, the augmented representations exhibit improved performance,



(a) Predictions accuracy by varying the features and samples of shared data between medical institutions in the MIMIC-III dataset.

(b) Computation time by varying the features and samples of shared data between medical institutions in the MIMIC-III dataset.

(c) Predictions accuracy by varying the features and samples of shared data between medical institutions in the Breast dataset.

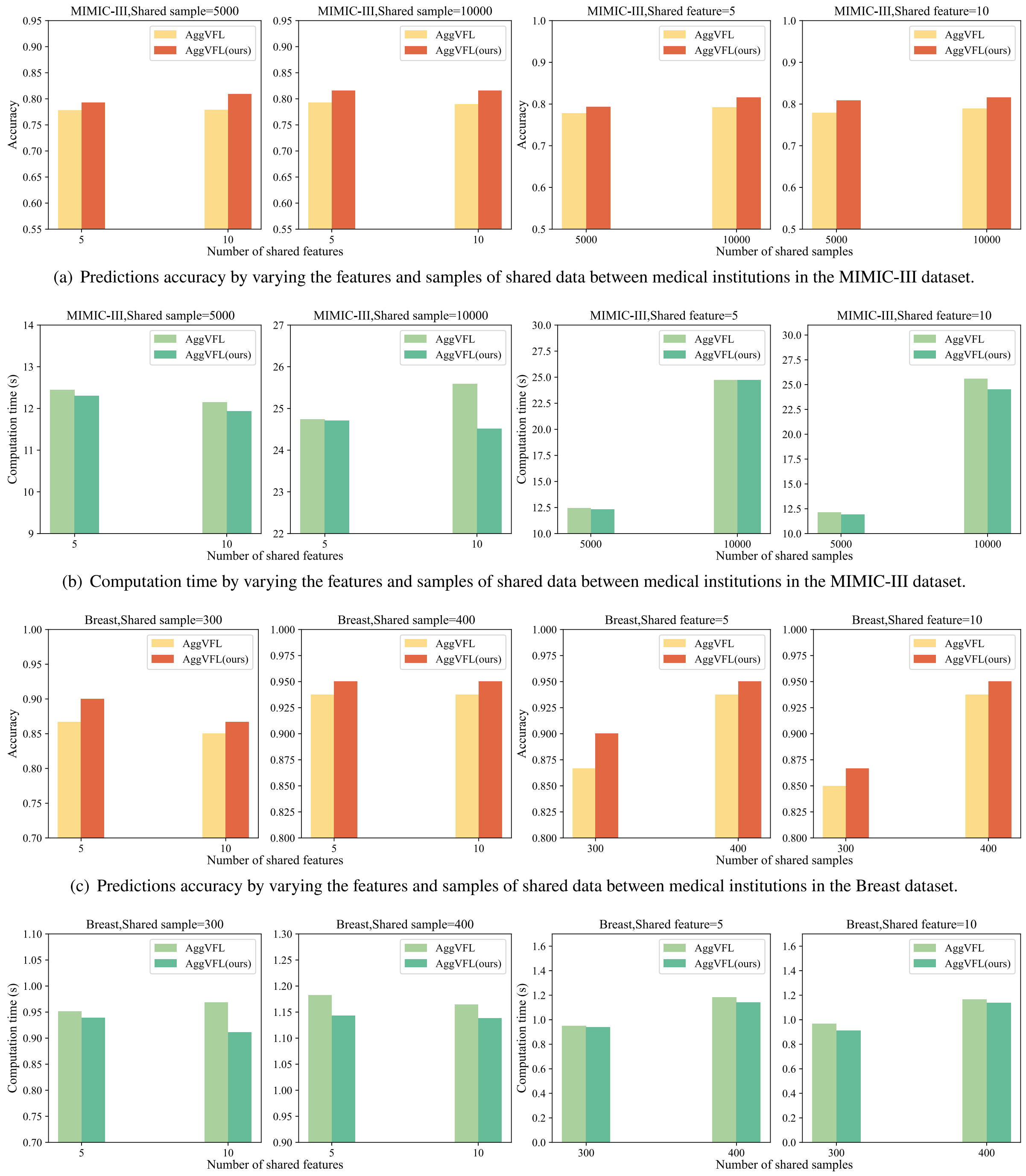
(d) Computation time by varying the features and samples of shared data between medical institutions in the Breast dataset.

Fig. 13. Predictions accuracy and computation time of SplitVFL by varying the features of shared data between medical institutions.

benefiting from more comprehensive and personalized information. Additionally, this process equips the model with the capability to be applied to new, unexplored medical datasets. By incorporating private data into the knowledge transfer process, we can better balance the extraction of global federated knowledge and the retention of local personalized information. In the MIMIC-III dataset, due to its larger data volume, more private data can be simulated, resulting in a more significant performance improvement. Conversely, in the Breast dataset, the smaller data volume limits the amount of private data that can be simulated, leading to less noticeable improvement. This indicates that the availability of private data constrains the extent of representation augmentation.

To empirically verify the versatility of the FRL module, including FedSVD and VFedPCA, we conducted ablation experiments. These experiments primarily varied the feature and sample of both task data and shared data to compare the performance of different FRL methods in knowledge transfer. We varied the shared data because FRL is used to extract the federated representation of the shared data. Adjusting the task data indirectly changes the private data, reflecting the impact of federated representation on knowledge transfer to the private data. Table 11 summarizes these experimental results, showing the prediction accuracy of FedSVD and VFedPCA under different configurations.

It is evident that both methods significantly contribute to algorithm optimization by extracting the federated representation of shared data.



(a) Predictions accuracy by varying the features and samples of shared data between medical institutions in the MIMIC-III dataset.

(b) Computation time by varying the features and samples of shared data between medical institutions in the MIMIC-III dataset.

(c) Predictions accuracy by varying the features and samples of shared data between medical institutions in the Breast dataset.

(d) Computation time by varying the features and samples of shared data between medical institutions in the Breast dataset.

Fig. 14. Predictions accuracy and computation time of AggVFL by varying the features of shared data between medical institutions.

In these experiments, the federated representation extracted by both methods was used to augment the representation of the shared data for the task institution, which in turn was used to augment the representation of the private data, yielding commendable results. Therefore, both methods played a vital role in knowledge transfer. Specifically, for the larger-scale MIMIC-III dataset, VFedPCA provided greater optimization, indicating that VFedPCA can more effectively extract the federated representation of shared data when dealing with large-scale datasets, thereby augmenting the task institution's representation and improving model performance. Conversely, for the smaller-scale Breast dataset, FedSVD provided greater optimization. This suggests that FedSVD is more effective on smaller datasets, better capturing and utilizing the representation of shared data to augment the task institution's representation.

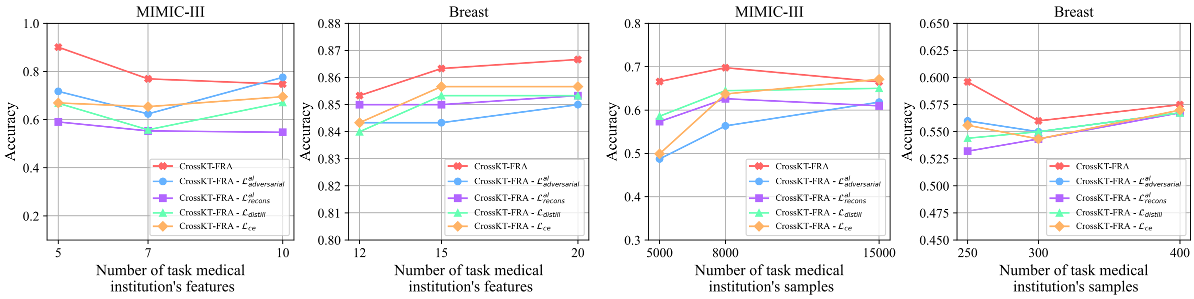
6. Discussion

6.1. Security and privacy

Security and privacy are critical factors in the design of FL mechanisms. Although our proposed framework primarily focuses on enhancing model performance through semi-supervised representation learning and knowledge transfer, privacy protection remains an integral part of our approach.

Table 7
Default key parameters in th downstream medical models.

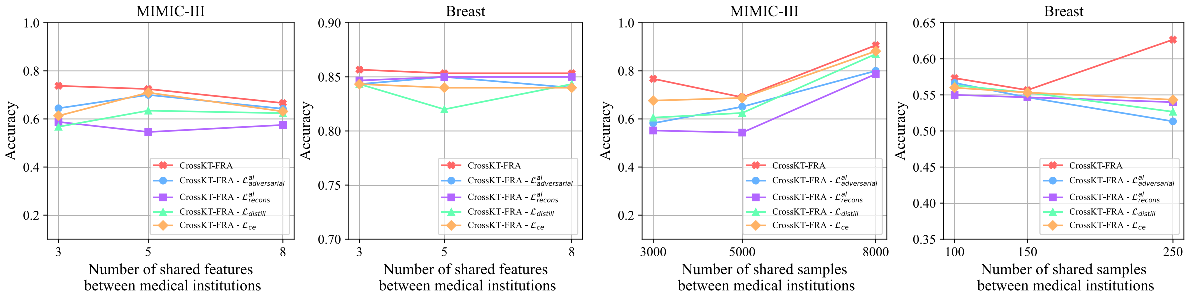
Model	Parameter	Default	Description
Neural Network [74]	*hidden_layer_sizes*	(100,100,50)	The number of units in hidden layers.
	α	0.01	Weight of the L2 regularization term.
	max_iter	400	Maximum of iterations.
	activation	relu	Activation function for the hidden layer.
Random Forest [75]	*n_estimators*	200	The number of the trees.
	max_depth	10	The maximum depth of the tree.
K Nearest Neighbors [76]	*n_neighbors*	8	Number of neighbors.
Adaptive Boosting [77]	*max_depth*	3	DecisionTreeClassifer's maximum depth.
	n_estimators	100	The maximum number of estimators.
	learning_rate	0.5	Each classifier's weight at each iteration.
eXtreme Gradient Boosting [78]	*max_depth*	7	The maximum depth of a tree.
	learning_rate	0.01	Weight at each iteration.



(a) Varying features of task medical institution.

(b) Varying samples of task medical institution.

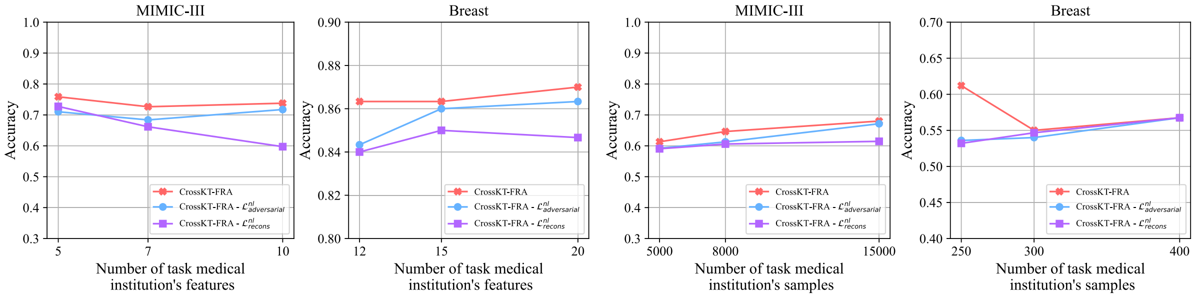
Fig. 15. Predictions accuracy by varying the features and samples of the task medical institution's data while including or excluding $\mathcal{L}^{al}_{adversarial}$, $\mathcal{L}^{al}_{recons}$, $\mathcal{L}_{distill}$, and $\mathcal{L}_{ce}$ in the representation augmentation module for shared data.



(a) Varying shared features between medical institutions.

(b) Varying shared samples between medical institutions.

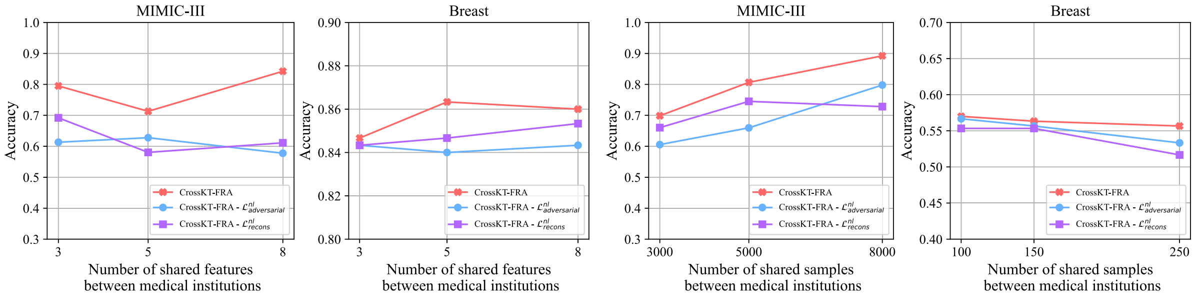
Fig. 16. Predictions accuracy by varying the features and samples between the task and data medical institution's data while including or excluding $\mathcal{L}^{al}_{adversarial}$, $\mathcal{L}^{al}_{recons}$, $\mathcal{L}_{distill}$, and $\mathcal{L}_{ce}$ in the representation augmentation module for shared data.



(a) Varying features of task medical institution.

(b) Varying samples of task medical institution.

Fig. 17. Predictions accuracy by varying the features and samples of the task medical institution's data while including or excluding $\mathcal{L}^{nl}_{adversarial}$ and $\mathcal{L}^{nl}_{recons}$ in the representation augmentation module for private data.



(a) Varying shared features between medical institutions.

(b) Varying shared samples between medical institutions.

Fig. 18. Predictions accuracy by varying the features and samples between the task and data medical institution's data while including or excluding $\mathcal{L}^{nl}_{adversarial}$ and $\mathcal{L}^{nl}_{recons}$ in the representation augmentation module for private data.

Table 8
The default samples and features held by each medical institution's data in the enhanced optimization experiment.

Dataset	Feature		Sample	
	Task	Data	Task	Data
MIMIC-III	10	15	15 000	30 000
Breast	25	30	400	500

Our framework incorporates existing mature VFL algorithms in the FRL module, including FedSVD [67] and VFedPCA [68]. These algorithms provide robust privacy protection mechanisms. FedSVD uses two random orthogonal matrices to mask the original data, ensuring that the third-party server can only obtain SVD results using masked data. VFedPCA allows the third-party server to perform a weighted summation of the eigenvectors and eigenvalues of both parties' data. These methods prevent non-owners from directly using the data, thereby protecting privacy.

However, the novelty of our method lies in the knowledge transfer mechanism following the FRL module process. Specifically, the knowledge obtained from the federated representation is first transferred to shared data, then from shared data to private data, and finally back to shared data from private data. This iterative knowledge transfer augments representations while maintaining privacy.

Given the limitations of our study, a detailed analysis of robustness against privacy attacks is beyond our current research scope. For specific security and privacy analyses, we refer to the original works on FedSVD [67] and VFedPCA [68].

6.2. Future work and extensions

In addition to the specific medical datasets examined in this paper, our research has broad application potential. While our current study primarily focuses on semi-supervised representation augmentation for tabular data in the medical field, the augmentation characteristics at the representation level theoretically make this method applicable to other types of data, such as image data.

Specifically, the FedSVD and VFedPCA algorithms involved in the FRL module of CrossKT-FRA have been proven effective on image datasets, and the $\mathcal{L}_{adversarial}$ component using GANs has demonstrated its effectiveness across multiple image datasets. Furthermore, the $\mathcal{L}_{distill}$ and $\mathcal{L}_{ce}$ components calculate losses from representations and labels, respectively, without requiring any changes. For the $\mathcal{L}_{recons}$ component, although cosine similarity is used for tabular data, it can be replaced with other suitable metrics such as KL divergence for image datasets. Therefore, the CrossKT-FRA method can theoretically be applied to image datasets. Moreover, the current algorithm is theoretically applicable to other privacy-preserving FL fields, such as education and government. In these fields, data can also benefit from semi-supervised representation augmentation to improve model performance and protect data privacy.

In future work, we plan to further explore the application of CrossKT-FRA in these fields and investigate its applicability to different types of data to fully realize its value in FL. We also intend to design more targeted knowledge transfer strategies in federated transfer learning (FTL) [79] environments with medical data and address more

Table 9
Default samples and features held by each medical institution's data and the shared data in the component ablation experiment.

Module	Group	Dataset	Feature			Sample		
			Task	Shared	Data	Task	Shared	Data
Shared	Task feature	MIMIC-III	**5,7,10**	4	15	5000	4000	30 000
		Breast	**12,15,20**	10	30	300	300	500
	Task sample	MIMIC-III	10	4	15	**5000,8000,15000**	4000	30 000
		Breast	10	10	30	**250,300,400**	200	500
	Shared feature	MIMIC-III	10	**3,5,8**	15	5000	4000	30 000
		Breast	10	**3,5,8**	30	300	300	500
	Shared sample	MIMIC-III	10	4	15	10 000	**3000,5000,8000**	30 000
		Breast	10	10	30	300	**100,150,250**	500
Private	Task feature	MIMIC-III	**5,7,10**	4	15	5000	4000	30 000
		Breast	**12,15,20**	10	30	300	300	500
	Task sample	MIMIC-III	10	4	15	**5000,8000,15000**	4000	30 000
		Breast	10	10	30	**250,300,400**	200	500
	Shared feature	MIMIC-III	10	**3,5,8**	15	5000	4000	30 000
		Breast	10	**3,5,8**	30	300	300	500
	Shared sample	MIMIC-III	10	4	15	10 000	**3000,5000,8000**	30 000
		Breast	10	10	30	300	**100,150,250**	500

Table 10
Predicting accuracy by varying the feature and sample of private data from the task medical institution, with and without knowledge transfer participation ("Ours-*nl*" represents no participation, "Ours" represents participation).

Dataset	Feature		Sample		Private data		Accuracy(%)	
Task	Shared	Task	Shared	$n^{t,nl}$	$m^{t,nl}$	Ours-*nl*	Ours	
MIMIC-III	5	4	10 000	4000	1	6000	77.21	**80.82**
	10	4	10 000	4000	6	6000	57.82	**77.24**
	10	4	5000	4000	6	1000	58.32	**74.58**
Breast	10	4	400	300	6	100	57.75	**58.25**
	20	4	400	300	16	100	58.00	**59.25**
	10	4	350	300	6	50	56.85	**57.42**

Table 11
Predicting accuracy by varying the feature and sample of both task institution data and shared data during knowledge transfer with either FedSVD or VFedPCA in the FRL module. ("FedSVD" and "VFedPCA" respectively represent the scenarios where the corresponding FRL methods were involved in knowledge transfer.)

Dataset	Feature		Sample		Accuracy(%)	
	Task	Shared	Task	Shared	FedSVD	VPedVCA
MIMIC-III	5	4	15 000	4000	84.22	**89.53**
	10	4	15 000	4000	72.31	**74.87**
	10	4	10 000	4000	**70.56**	67.75
Breast	10	4	400	300	**58.50**	**58.50**
	20	4	400	300	**60.00**	58.25
	10	4	350	300	**57.99**	56.85

federated medical issues [38,80,81] to tackle the challenges of non-overlapping distributed data scenarios in FL [82]. Additionally, we aim to explore and integrate more advanced privacy protection technologies and conduct extensive experiments to evaluate our framework's resilience against various data leakage attempts.

7. Conclusion

We propose a collaborative framework for federated healthcare based on a semi-supervised representation augmentation mechanism with cross-institutional knowledge transfer (CrossKT-FRA), aiming to improve the efficiency of resource utilization in cross-institutional collaboration scenarios within the medical field. CrossKT-FRA achieves federated knowledge transfer between medical institutions and mutual knowledge transfer between local shared and private data. Extensive experiments conducted on medical datasets validate the superiority of CrossKT-FRA in independent operation settings, the universality of the enhanced optimization process and the effectiveness of each model component.

CRediT authorship contribution statement

Zilong Yin: Writing – review & editing, Methodology, Resources, Validation, Investigation, Formal analysis, Data curation. **Haoyu Wang:** Writing – review & editing, Validation, Software, Formal analysis. **Bin Chen:** Validation, Supervision, Resources, Methodology. **Xin Zhang:** Conceptualization. **Xiaogang Lin:** Writing – review & editing, Visualization, Validation. **Hangling Sun:** Data curation. **Anji Li:** Conceptualization. **Chenyu Zhou:** Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] Ezekiel J. Emanuel, Govind Persad, Ross Upshur, Beatriz Thome, Michael Parker, Aaron Glickman, Cathy Zhang, Connor Boyle, Maxwell Smith, James P. Phillips, Fair allocation of scarce medical resources in the time of Covid-19, N. Engl. J. Med. 382 (21) (2020) 2049–2055.

[2] Guoguang Rong, Arnaldo Mendez, Elie Bou Assi, Bo Zhao, Mohamad Sawan, Artificial intelligence in healthcare: Review and prediction case studies, Engineering 6 (3) (2020) 291–301.

[3] Georgios A. Kaissis, Marcus R. Makowski, Daniel Rückert, Rickmer F. Braren, Secure, privacy-preserving and federated machine learning in medical imaging, Nat. Mach. Intell. 2 (6) (2020) 305–311.

[4] Sanket S. Dhruva, Joseph S. Ross, Joseph G. Akar, Brittany Caldwell, Karla Childers, Wing Chow, Laura Ciaccio, Paul Coplan, Jun Dong, Hayley J. Dykhoff, Stephen Johnston, Todd Kellogg, Cynthia Long, Peter A. Noseworthy, Kurt Roberts, Anindita Saha, Andrew Yoo, Nilay D. Shah, Aggregating multiple real-world data sources using a patient-centered health-data-sharing platform, npj Digit. Med. 3 (1) (2020) 60.

[5] Smadar Shilo, Hagai Rossman, Eran Segal, Axes of a revolution: challenges and promises of big data in healthcare, Nat. Med. 26 (1) (2020) 29–38.

[6] Mohammad Fattahi, Esmaeil Keyvanshokooh, Devika Kannan, Kannan Govindan, Resource planning strategies for healthcare systems during a pandemic, European J. Oper. Res. 304 (1) (2023) 192–206, The role of Operational Research in future epidemics/ pandemics.

[7] Yujie Feng, Jiangtao Wang, Yasha Wang, Sumi Helal, Completing missing prevalence rates for multiple chronic diseases by jointly leveraging both intra- and inter-disease population health data correlations, in: Proceedings of the Web Conference 2021, WWW '21, Association for Computing Machinery, New York, NY, USA, 2021, pp. 183–193.

[8] Quande Liu, Cheng Chen, Jing Qin, Qi Dou, Pheng-Ann Heng, FedDG: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2021, pp. 1013–1023.

[9] Jafar A. Alzubi, Omar A. Alzubi, Ashish Singh, Manikandan Ramachandran, Cloud-IIoT-based electronic health record privacy-preserving by CNN and blockchain-enabled federated learning, IEEE Trans. Ind. Inform. 19 (1) (2023) 1080–1087.

[10] Meng Shen, Junxian Duan, Liehuang Zhu, Jie Zhang, Xiaojiang Du, Mohsen Guizani, Blockchain-based incentives for secure and collaborative data sharing in multiple clouds, IEEE J. Sel. Areas Commun. 38 (6) (2020) 1229–1241.

[11] Rodolfo Stoffel Antunes, Cristiano André da Costa, Arne Küderle, Imrana Abdullahi Yari, Björn Eskofier, Federated learning for healthcare: Systematic review and architecture proposal, ACM Trans. Intell. Syst. Technol. 13 (4) (2022).

[12] Chandra Thapa, Seyit Camtepe, Precision health data: Requirements, challenges and existing techniques for data security and privacy, Comput. Biol. Med. 129 (2021) 104130.

[13] Nazish Khalid, Adnan Qayyum, Muhammad Bilal, Ala Al-Fuqaha, Junaid Qadir, Privacy-preserving artificial intelligence in healthcare: Techniques and applications, Comput. Biol. Med. 158 (2023) 106848.

[14] Sotirios Messinis, Nikos Temenos, Nicholas E. Protonotarios, Ioannis Rallis, Dimitrios Kalogeras, Nikolaos Doulamis, Enhancing internet of medical things security with artificial intelligence: A comprehensive review, Comput. Biol. Med. 170 (2024) 108036.

[15] Qiang Yang, Yang Liu, Tianjian Chen, Yongxin Tong, Federated machine learning: Concept and applications, ACM Trans. Intell. Syst. Technol. 10 (2) (2019).

[16] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, Blaise Aguera Arcas, Communication-efficient learning of deep networks from decentralized data, in: Artificial Intelligence and Statistics, PMLR, 2017, pp. 1273–1282.

[17] Taisa Kushner, Amit Sharma, Bursts of activity: Temporal patterns of help-seeking and support in online mental health forums, in: Proceedings of the Web Conference 2020, WWW '20, Association for Computing Machinery, New York, NY, USA, 2020, pp. 2906–2912.

[18] Jing Ma, Qiuchen Zhang, Jian Lou, Li Xiong, Joyce C. Ho, Communication efficient federated generalized tensor factorization for collaborative health data analytics, in: Proceedings of the Web Conference 2021, WWW '21, Association for Computing Machinery, New York, NY, USA, 2021, pp. 171–182.

[19] China Knowledge Centre for Engineering Sciences and Technology, Aminer database, 2016-2023, https://www.aminer.org.

[20] Jinpeng Hou, Mang Su, Anmin Fu, Yan Yu, Verifiable privacy-preserving scheme based on vertical federated random forest, IEEE Internet Things J. 9 (22) (2022) 22158–22172.

[21] Abhishek Hazra, Mainak Adhikari, Sudarshan Nandy, Khushbu Doulani, Varun G Menon, Federated-learning-aided next-generation edge networks for intelligent services, IEEE Netw. 36 (3) (2022) 56–64.

[22] Yang Liu, Yan Kang, Tianyuan Zou, Yanhong Pu, Yuanqin He, Xiaozhou Ye, Ye Ouyang, Ya-Qin Zhang, Qiang Yang, Vertical federated learning: Concepts, advances, and challenges, IEEE Trans. Knowl. Data Eng. (2024) 1–20.

[23] Peter Kairouz, H. McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Bhagoji, Kallista Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, Rafael D'Oliveira, Hubert Eichner, Salim El Rouayheb, David Evans, Josh Gardner, Zachary Garrett, Adrià Gascón, Badih Ghazi, Phillip Gibbons, Sen Zhao, Advances and Open Problems in Federated Learning, Foundations and Trends® in Machine Learning, 2021.

[24] Xuefei Yin, Yanming Zhu, Jiankun Hu, A comprehensive survey of privacy-preserving federated learning: A taxonomy, review, and future directions, ACM Comput. Surv. 54 (6) (2021).

[25] Lingxiao Huang, Zhize Li, Jialin Sun, Haoyu Zhao, Coresets for vertical federated learning: regularized linear regression and k-means clustering, in: Proceedings of the 36th International Conference on Neural Information Processing Systems, NIPS '22, Curran Associates Inc., Red Hook, NY, USA, 2024.

[26] Yan Kang, Yang Liu, Tianjian Chen, FedMVT: Semi-supervised vertical federated learning with MultiView training, 2020, CoRR abs/2008.10838.

[27] Yan Kang, Yang Liu, Xinle Liang, FedCVT: Semi-supervised vertical federated learning with cross-view training, ACM Trans. Intell. Syst. Technol. 13 (2022).

[28] Quim Zaldo-Aubanell, Isabel Serra, Albert Bach, Pablo Knobel, Ferran Campillo i López, Jordina Belmonte, Pepus Daunis i Estadella, Roser Maneja, Environmental heterogeneity in human health studies. A compositional methodology for land use and land cover data, Sci. Total Environ. 806 (2022) 150308.

[29] Juexiao Zhou, Longxi Zhou, Di Wang, Xiaopeng Xu, Haoyang Li, Yuetan Chu, Wenkai Han, Xin Gao, Personalized and privacy-preserving federated heterogeneous medical image analysis with PPPML-HMI, Comput. Biol. Med. 169 (2024) 107861.

[30] Sabyasachi Dash, Sushil Kumar Shakyawar, Mohit Sharma, Sandeep Kaushik, Big data in healthcare: management, analysis and future prospects, J. Big Data 6 (1) (2019) 54.

[31] Subrato Bharati, M. Rubaiyat Hossain Mondal, Prajoy Podder, V.B. Surya Prasath, Federated learning: Applications, challenges and future directions, Int. J. Hybrid Intell. Syst. 18 (1–2) (2022) 19–35.

[32] Pushpa Devi, Kishori Lal Bansal, Data science in healthcare: Techniques, challenges and opportunities, Health Technol. 14 (4) (2024) 623–634.

[33] Xinzhi Zhang, Eliseo Pérez-Stable, Philip Bourne, Emmanuel Peprah, Obidiugwu Duru, Nancy Breen, David Berrigan, Fred Wood, James Jackson, David Wong, Joshua Denny, Big data science: Opportunities and challenges to address minority health and health disparities in the 21st century, Ethnicity Dis. 27 (2017) 95.

[34] Andre Esteva, Alexandre Robicquet, Bharath Ramsundar, Volodymyr Kuleshov, Mark DePristo, Katherine Chou, Claire Cui, Greg Corrado, Sebastian Thrun, Jeff Dean, A guide to deep learning in healthcare, Nat. Med. 25 (1) (2019) 24–29.

[35] Leslie Lenert, Brooke Yeager McSwain, Balancing health privacy, health information exchange, and research in the context of the COVID-19 pandemic, J. Am. Med. Inform. Assoc. 27 (6) (2020) 963–966.

[36] Georgios A. Kaissis, Marcus R. Makowski, Daniel Rückert, Rickmer F. Braren, Secure, privacy-preserving and federated machine learning in medical imaging, Nat. Mach. Intell. 2 (6) (2020) 305–311.

[37] Jie Xu, Benjamin S. Glicksberg, Chang Su, Peter Walker, Jiang Bian, Fei Wang, Federated learning for healthcare informatics, J. Healthc. Inform. Res. 5 (1) (2021) 1–19.

[38] Jie Xu, Benjamin S. Glicksberg, Chang Su, Peter Walker, Jiang Bian, Fei Wang, Federated learning for healthcare informatics, J. Healthc. Inform. Res. 5 (1) (2021) 1–19.

[39] Ittai Dayan, Holger R. Roth, Aoxiao Zhong, Ahmed Harouni, Amilcare Gentili, Anas Z. Abidin, Andrew Liu, Anthony Beardsworth Costa, Bradford J. Wood, Chien-Sung Tsai, Chih-Hung Wang, Chun-Nan Hsu, C.K. Lee, Peiying Ruan, Daguang Xu, Dufan Wu, Eddie Huang, Felipe Campos Kitamura, Griffin Lacey, Gustavo César de Antônio Corradi, Gustavo Nino, Hao-Hsin Shin, Hirofumi Obinata, Hui Ren, Jason C. Crane, Jesse Tetreault, Jiahui Guan, John W. Garrett, Joshua D. Kaggie, Jung Gil Park, Keith Dreyer, Krishna Juluru, Kristopher Kersten, Marcio Aloisio Bezerra Cavalcanti Rockenbach, Marius George Linguraru, Masoom A. Haider, Meena AbdelMaseeh, Nicola Rieke, Pablo F. Damasceno, Pedro Mario Cruz e Silva, Pochuan Wang, Sheng Xu, Shuichi Kawano, Sira Sriswasdi, Soo Young Park, Thomas M. Grist, Varun Buch, Watsamon Jantarabenjakul, Weichung Wang, Won Young Tak, Xiang Li, Xihong Lin, Young Joon Kwon, Abood Quraini, Andrew Feng, Andrew N. Priest, Baris Turkbey, Benjamin Glicksberg, Bernardo Bizzo, Byung Seok Kim, Carlos Tor-Díez, Chia-Cheng Lee, Chia-Jung Hsu, Chin Lin, Chiu-Ling Lai, Christopher P. Hess, Colin Compas, Deepeksha Bhatia, Eric K. Oermann, Evan Leibovitz, Hisashi Sasaki, Hitoshi Mori, Isaac Yang, Jae Ho Sohn, Krishna Nand Keshava Murthy, Li-Chen Fu, Matheus Ribeiro Furtado de Mendonça, Mike Fralick, Min Kyu Kang, Mohammad Adil, Natalie Gangai, Peerapon Vateekul, Pierre Elnajjar, Sarah Hickman, Sharmila Majumdar, Shelley L. McLeod, Sheridan Reed, Stefan Gräf, Stephanie Harmon, Tatsuya Kodama, Thanyawee Puthanakit, Tony Mazzulli, Vitor Lima de Lavor, Yothin Rakvongthai, Yu Rim Lee, Yuhong Wen, Fiona J. Gilbert, Mona G. Flores, Quanzheng Li, Federated learning for predicting clinical outcomes in patients with COVID-19, Nat. Med. 27 (10) (2021) 1735–1743.

[40] Zelei Liu, Yuanyuan Chen, Yansong Zhao, Han Yu, Yang Liu, Renyi Bao, Jinpeng Jiang, Zaiqing Nie, Qian Xu, Qiang Yang, Contribution-aware federated learning for smart healthcare, in: Proceedings of the AAAI Conference on Artificial Intelligence, 36, vol. No. 11: IAAI-22, EAAI-22, AAAI-22 Special Programs and Special Track, Student Papers and Demonstrations, 2022.

[41] Qian Yang, Jianyi Zhang, Weituo Hao, Gregory P. Spell, Lawrence Carin, FLOP: Federated learning on medical datasets using partial networks, in: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, KDD '21, Association for Computing Machinery, New York, NY, USA, 2021, pp. 3845–3853.

[42] Cosmin I. Bercea, Benedikt Wiestler, Daniel Rueckert, Shadi Albarqouni, Federated disentangled representation learning for unsupervised brain anomaly detection, Nat. Mach. Intell. 4 (8) (2022) 685–695.

[43] Isaac Adjei-Mensah, Xiaoling Zhang, Isaac Osei Agyemang, Sophyani Banaamwini Yussif, Adu Asare Baffour, Bernard Mawuli Cobbinah, Collins Sey, Linda Delali Fiasam, Ijeoma Amuche Chikwendu, Joseph Roger Arhin, Cov-fed: Federated learning-based framework for COVID-19 diagnosis using chest X-ray scans, Eng. Appl. Artif. Intell. 128 (2024) 107448.

[44] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, Blaise Aguera Arcas, Communication-efficient learning of deep networks from decentralized data, in: Artificial Intelligence and Statistics, PMLR, 2017, pp. 1273–1282.

[45] Bin Gu, An Xu, Zhouyuan Huo, Cheng Deng, Heng Huang, Privacy-preserving asynchronous vertical federated learning algorithms for multiparty collaborative learning, IEEE Trans. Neural Netw. Learn. Syst. 33 (11) (2022) 6103–6115.

[46] Kewei Cheng, Tao Fan, Yilun Jin, Yang Liu, Tianjian Chen, Dimitrios Papadopoulos, Qiang Yang, SecureBoost: A lossless federated learning framework, IEEE Intell. Syst. 36 (6) (2021) 87–98.

[47] Yuncheng Wu, Shaofeng Cai, Xiaokui Xiao, Gang Chen, Beng Chin Ooi, Privacy preserving vertical federated learning for tree-based models, Proc. VLDB Endow. 13 (12) (2020) 2090–2103.

[48] Qiang Yang, Yang Liu, Tianjian Chen, Yongxin Tong, Federated machine learning: Concept and applications, ACM Trans. Intell. Syst. Technol. 10 (2) (2019).

[49] Yaochen Hu, Di Niu, Jianming Yang, Shengping Zhou, FDML: A collaborative machine learning framework for distributed features, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '19, Association for Computing Machinery, New York, NY, USA, 2019, pp. 2232–2240.

[50] Yaochen Hu, Di Niu, Jianming Yang, Shengping Zhou, FDML: A collaborative machine learning framework for distributed features, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '19, Association for Computing Machinery, New York, NY, USA, 2019, pp. 2232–2240.

[51] Jinbo Wang, Xikai Pei, Ruijin Wang, Fengli Zhang, Ting Chen, Federated semi-supervised learning with tolerant guidance and powerful classifier in edge scenarios, Inform. Sci. 662 (2024) 120201.

[52] Zhengyi Zhong, Ji Wang, Weidong Bao, Jingxuan Zhou, Xiaomin Zhu, Xiongtao Zhang, Semi-HFL: semi-supervised federated learning for heterogeneous devices, Complex Intell. Syst. 9 (2) (2023) 1995–2017.

[53] Cobbinah B. Mawuli, Jay Kumar, Ebenezer Nanor, Shangxuan Fu, Liangxu Pan, Qinli Yang, Wei Zhang, Junming Shao, Semi-supervised federated learning on evolving data streams, Inform. Sci. 643 (2023) 119235.

[54] Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, Virginia Smith, Federated optimization in heterogeneous networks, in: I. Dhillon, D. Papailiopoulos, V. Sze (Eds.), in: Proceedings of Machine Learning and Systems, vol. 2, 2020, pp. 429–450.

[55] Jiqiang Gao, Baolei Zhang, Xiaojie Guo, Thar Baker, Min Li, Zheli Liu, Secure partial aggregation: Making federated learning more robust for industry 4.0 applications, IEEE Trans. Ind. Inform. 18 (9) (2022) 6340–6348.

[56] Mengkai Song, Zhibo Wang, Zhifei Zhang, Yang Song, Qian Wang, Ju Ren, Hairong Qi, Analyzing user-level privacy attack against federated learning, IEEE J. Sel. Areas Commun. 38 (10) (2020) 2430–2444.

[57] Xiao Jin, Pin-Yu Chen, Chia-Yi Hsu, Chia-Mu Yu, Tianyi Chen, CAFE: Catastrophic data leakage in vertical federated learning, in: M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, J. Wortman Vaughan (Eds.), in: Advances in Neural Information Processing Systems, vol. 34, Curran Associates, Inc., 2021, pp. 994–1006.

[58] Jiqiang Gao, Boyu Hou, Xiaojie Guo, Zheli Liu, Ying Zhang, Kai Chen, Jin Li, Secure aggregation is insecure: Category inference attack on federated learning, IEEE Trans. Dependable Secure Comput. 20 (1) (2023) 147–160.

[59] Chong Fu, Xuhong Zhang, Shouling Ji, Jinyin Chen, Jingzheng Wu, Shanqing Guo, Jun Zhou, Alex X. Liu, Ting Wang, Label inference attacks against vertical federated learning, in: Kevin R.B. Butler, Kurt Thomas (Eds.), 31st USENIX Security Symposium, USENIX Security 2022, Boston, MA, USA, August 10-12, 2022, USENIX Association, 2022, pp. 1397–1414.

[60] Andrew C. Yao, Protocols for secure computations, in: 23rd Annual Symposium on Foundations of Computer Science, (Sfcs 1982), 1982, pp. 160–164.

[61] Cynthia Dwork, Aaron Roth, The algorithmic foundations of differential privacy, Found. Trends Theor. Comput. Sci. 9 (3–4) (2014) 211–407.

[62] Martin Abadi, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, Li Zhang, Deep learning with differential privacy, in: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, CCS '16, Association for Computing Machinery, New York, NY, USA, 2016, pp. 308–318.

[63] Keith Bonawitz, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H. Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, Karn Seth, Practical secure aggregation for privacy-preserving machine learning, in: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, CCS '17, Association for Computing Machinery, New York, NY, USA, 2017, pp. 1175–1191.

[64] Kang Wei, Jun Li, Ming Ding, Chuan Ma, Howard H. Yang, Farhad Farokhi, Shi Jin, Tony Q.S. Quek, H. Vincent Poor, Federated learning with differential privacy: Algorithms and performance analysis, IEEE Trans. Inf. Forensics Secur. 15 (2020) 3454–3469.

[65] Chung-ju Huang, Leye Wang, Xiao Han, Vertical federated knowledge transfer via representation distillation for healthcare collaboration networks, in: Proceedings of the ACM Web Conference 2023, WWW '23, Association for Computing Machinery, New York, NY, USA, 2023, pp. 4188–4199.

[66] Richard Nock, Stephen Hardy, Wilko Henecka, Hamish Ivey-Law, Giorgio Patrini, Guillaume Smith, Brian Thorne, Entity resolution and federated learning get a federated resolution, 2018, CoRR abs/1803.04035.

[67] Di Chai, Leye Wang, Junxue Zhang, Liu Yang, Shuowei Cai, Kai Chen, Qiang Yang, Practical lossless federated singular vector decomposition over billion-scale data, in: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD '22, Association for Computing Machinery, New York, NY, USA, 2022, pp. 46–55.

[68] Yiu-ming Cheung, Jian Lou, Feng Yu, Vertical federated principal component analysis on feature-wise distributed data, in: Wenjie Zhang, Lei Zou, Zakaria Maamar, Lu Chen (Eds.), Web Information Systems Engineering – WISE 2021, Springer International Publishing, Cham, 2021, pp. 173–188.

[69] Michal Kosinski, David Stillwell, Thore Graepel, Private traits and attributes are predictable from digital records of human behavior, Proc. Natl. Acad. Sci. USA 110 (15) (2013) 5802–5805.

[70] Lihua Yin, Jiyuan Feng, Hao Xun, Zhe Sun, Xiaochun Cheng, A privacy-preserving federated learning for multiparty data sharing in social IoTs, IEEE Trans. Netw. Sci. Eng. 8 (3) (2021) 2706–2718.

[71] Zhou Zhou, Youliang Tian, Jinbo Xiong, Jianfeng Ma, Changgen Peng, Blockchain-enabled secure and trusted federated data sharing in IIoT, IEEE Trans. Ind. Inform. 19 (5) (2023) 6669–6681.

[72] Alistair E.W. Johnson, Tom J. Pollard, Lu Shen, Li-wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, Roger G. Mark, MIMIC-III, a freely accessible critical care database, Sci. Data 3 (1) (2016) 160035.

[73] Kaggle, Breast cancer wisconsin (diagnostic) data set, 2022, Retrieved 2022-06-08 from https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data.

[74] Emilly M Lima, Antônio H Ribeiro, Gabriela MM Paixão, Manoel Horta Ribeiro, Marcelo M Pinto Filho, Paulo R Gomes, Derick M Oliveira, Ester C Sabino, Bruce B Duncan, Luana Giatti, Sandhi M Barreto, Jr Wagner Meira, Thomas B Schön, Antonio Luiz P Ribeiro, Deep neural network estimated electrocardiographic-age as a mortality predictor, 2021, http://dx.doi.org/10.1101/2021.02.19.21251232, medRxiv.

[75] Celestine Iwendi, Ali Kashif Bashir, Atharva Peshkar, R. Sujatha, Jyotir Moy Chatterjee, Swetha Pasupuleti, Rishita Mishra, Sofia Pillai, Ohyun Jo, COVID-19 patient health prediction using boosted random forest algorithm, Front. Public Health 8 (2020).

[76] Wenchao Xing, Yilin Bei, Medical health big data classification based on KNN classification algorithm, IEEE Access 8 (2020) 28808–28819.

[77] Fusheng Li, Wanqi Yang, Qian Ma, Huizhu Cheng, Xin Lu, Yanchun Zhao, X-ray fluorescence spectroscopic analysis of trace elements in soil with an adaboost back propagation neural network and multivariate-partial least squares regression, Meas. Sci. Technol. 32 (10) (2021) 105501.

[78] Amir Bahador Parsa, Ali Movahedi, Homa Taghipour, Sybil Derrible, Abolfazl (Kouros) Mohammadian, Toward safer highways, application of xgboost and SHAP for real-time accident detection and feature analysis, Accid. Anal. Prev. 136 (2020) 105405.

[79] Yang Liu, Yan Kang, Chaoping Xing, Tianjian Chen, Qiang Yang, A secure federated transfer learning framework, IEEE Intell. Syst. 35 (4) (2020) 70–82.

[80] Rodolfo Stoffel Antunes, Cristiano André da Costa, Arne Küderle, Imrana Abdullahi Yari, Björn Eskofier, Federated learning for healthcare: Systematic review and architecture proposal, ACM Trans. Intell. Syst. Technol. 13 (4) (2022).

[81] Qiong Wu, Xu Chen, Zhi Zhou, Junshan Zhang, FedHome: Cloud-edge based personalized federated learning for in-home health monitoring, IEEE Trans. Mob. Comput. 21 (8) (2022) 2818–2832.

[82] Kevin I-Kai Wang, Xiaokang Zhou, Wei Liang, Zheng Yan, Jinhua She, Federated transfer learning based cross-domain prediction for smart manufacturing, IEEE Trans. Ind. Inform. 18 (6) (2022) 4088–4096.