Personalized Federated Graph Learning for Heterogeneous Incomplete EHRs

Liping Chen[1], Guangqing Bai[2], Zhuoyang Han[2], Jing Ren[1], Mujie Liu[3], Tao Tang[4(✉)], Shuo Yu[2], and Ivan Lee[4]

[1] RMIT University, Melbourne, Australia
[2] Dalian University of Technology, Dalian, China
[3] Federation University Australia, Ballarat, Australia
[4] University of South Australia, Adelaide, Australia
{lp.chen, zhuoyang.han, jing.ren}@ieee.org
{mujie.liu, tao.tang, shuo.yu}@ieee.org
baigq@mail.dlut.edu.cn, Ivan.Lee@unisa.edu.au

Abstract. Real-world electronic health records (EHRs) are decentralized and typically exhibit incomplete information across both modalities and features. This heterogeneous missingness poses a major challenge for federated graph learning, which aims to collaboratively model distributed medical data while preserving privacy. Existing methods often assume uniform data completeness or apply imputation strategies, which struggle to maintain consistent predictive performance under non-uniform and severe data incompleteness. To address this issue, we propose MissPCL, a personalized federated graph contrastive learning framework tailored to heterogeneous client-level missingness. Each client estimates its missingness rate and constructs a bipartite patient-modality graph from observed features. Clients with low missingness participate in global training using contrastive learning on local graphs, while high-missingness clients are excluded from aggregation and instead perform personalized fine-tuning. Experiments on two real-world EHR datasets under diverse missingness scenarios demonstrate that MissPCL consistently outperforms state-of-the-art baselines in classification accuracy and robustness. Notably, our framework achieves stable performance across varying degrees of data incompleteness, showing its practical utility in realistic federated clinical settings. Code implementations and the supplementary materials are available at https://github.com/TutaResearch/MISSPCL.

Keywords: Personalized Federated Learning · Graph Contrastive Learning · Heterogeneous Missingness

1 Introduction

Graph learning offers a powerful paradigm for capturing complex relationships across data entities and has demonstrated strong performance in domains such as healthcare, transportation, and social networks [5,12]. In the healthcare domain, multimodal graph learning integrates heterogeneous information from structured

electronic health records (EHRs), imaging, and lab tests, improving patient representation learning for tasks such as disease prediction and biomarker discovery [14,22,29]. However, multimodal clinical data are typically fragmented across institutions due to non-standardized collection protocols and strict privacy regulations [3], motivating the use of federated learning as a privacy-preserving paradigm for collaborative modeling without centralized data sharing. Recently, federated multimodal graph learning has gained attention as a promising approach to leverage distributed, complementary medical data across clients [7].

Beyond data decentralization, a major challenge lies in pervasive and non-random data incompleteness, which occurs at both the modality and feature levels [11]. Such fine-grained missingness arises from clinical heterogeneity, varying resource availability, and differences in patient characteristics, and is further amplified in federated settings where each client may exhibit distinct and sparse patterns [15]. For example, while one hospital may record full lab and imaging data, another may only collect basic vitals due to operational constraints. This heterogeneity severely limits global model optimization, particularly when clients with richer data dominate training [13].

To address missing data, prior studies have proposed modality imputation [17] and modality-invariant representation learning [30]. However, these methods often assume consistent modality availability or rely on reconstruction, which may be infeasible or privacy-sensitive in federated environments [4]. In contrast, graph contrastive learning (GCL) has recently emerged as a powerful alternative [18], offering robust representation learning without requiring full data reconstruction. Recent studies demonstrate its ability to effectively mitigate modality collapse and extract discriminative patient features even under severe missingness [24,27]. Current GCL-based methods are typically designed for centralized settings and do not account for federated heterogeneity or client-level missingness variation.

To bridge this gap, we propose MissPCL, a missingness-aware personalized federated graph contrastive learning framework for decentralized healthcare. Each client estimates its feature-level missingness and builds a bipartite graph connecting patients and available modalities. Clients with low missingness collaborate in global model training using a mutual-consistent contrastive learning objective that combines label supervision and graph augmentation. High-missingness clients are excluded from global updates and instead fine-tune the global model locally. This personalized framework jointly addresses client heterogeneity and missingness variation, yielding robust, consistent performance across diverse real-world scenarios. We summarize our key contributions as follows:

1) We propose MissPCL, a novel personalized federated graph contrastive learning framework tailored to heterogeneous missingness in decentralized multimodal healthcare data.

2) MissPCL estimates client-specific missingness rates and dynamically selects participants for global training, enabling personalized model adaptation for high-missingness clients.

3) Extensive experiments on real-world EHR datasets demonstrate that MissPCL achieves superior generalization and personalization under diverse missingness patterns.

2 Related Work

2.1 Multimodal Graph Learning with Missing Modalities

Multimodal graph learning enables integration of structured healthcare data such as imaging, clinical notes, and lab tests. Existing methods addressing missing modalities include imputation-based models, which reconstruct missing data via latent or clustering strategies, and representation learning approaches that aim for modality-invariant embeddings [28,17]. For instance, M3Care [28] uses modality-specific graphs for latent reconstruction, while Cafe [17] clusters clients by missingness for feature completion. Recent fusion-based models (e.g., ViLT [10], HGMF [1]) bypass reconstruction by learning from observed modalities, though they may overlook modality-specific cues. Contrastive frameworks like MUSE [24] and FuseMoE [6] improve flexibility but often ignore feature-level informativeness. However, these methods often treat modalities as atomic units and do not explicitly quantify feature-level informativeness under heterogeneous missingness. In contrast, our approach uses graph-based contrastive learning to selectively propagate informative signals, improving robustness to both modality and feature-level incompleteness.

2.2 Personalized Federated Learning on Heterogeneous Data

Personalized federated learning (PFL) mitigates Non-IID issues by adapting models to client-specific distributions. Existing methods include global model personalization (e.g., fine-tuning [21], reweighted aggregation [25,2]) and client-specific models using generative adversarial networks [23], prompt tuning [26], or adapters [19]. While effective for data heterogeneity, most existing methods assume uniform modality availability across clients-an unrealistic assumption in federated healthcare due to inconsistent data acquisition protocols. In contrast, our method explicitly models heterogeneous missingness and adopts selective client participation and fine-tuning, enhancing global model robustness and local adaptation in the presence of incomplete and non-uniform data.

3 Proposed Method

In this paper, we propose MissPCL, a missingness-aware personalized federated graph contrastive learning framework, designed to address heterogeneous data incompleteness in decentralized healthcare environments. The framework operates in three key stages: (1) each client quantifies its local feature-level missingness to guide personalized participation; (2) low-missingness clients collaboratively

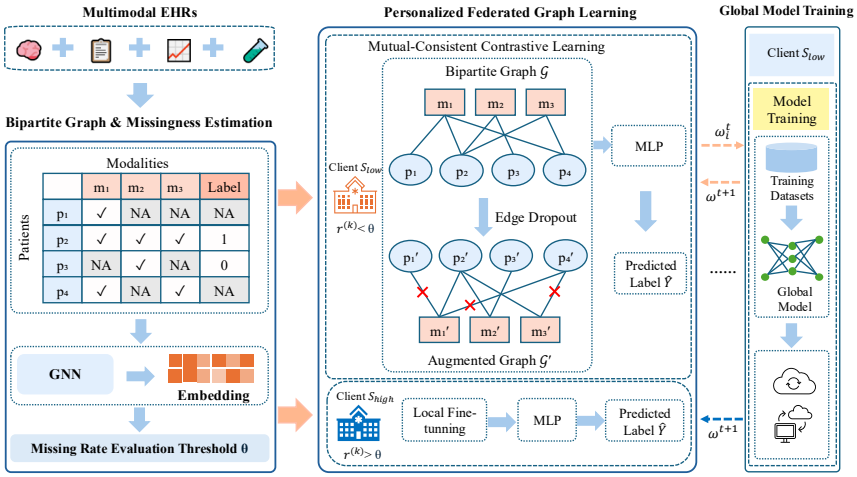

Fig. 1: Overview of the proposed MissPCL framework.

train a global model via federated optimization on locally constructed bipartite patient-modality graphs using contrastive learning; and (3) high-missingness clients personalize the global model via fine-tuning on their incomplete data. The method integrates mutual-consistent contrastive learning to enhance robustness and a personalized client selection mechanism to improve adaptability under diverse missingness conditions. This design enables more reliable disease prediction across clients with varying levels of data quality. The overall framework of the proposed MissPCL is shown in Figure 1.

3.1 Missingness Estimation and Bipartite Graph Construction

Missingness Estimation. To quantify the degree of data incompleteness on each client, we compute a missingness rate based on the proportion of missing feature values. Specifically, for client $k \in (0, \ldots, K)$, let $N^{(k)}$ be the number of patients and $\mathcal{F}$ the set of features. We define the missingness rate $r^{(k)}$ as the number of missing entries in its local dataset, divided by the total number of entries on client k, denoted as $N^{(k)}|\mathcal{F}|$. The number of missing entries is counted by scanning all feature values for all patients and checking whether a value is missing (e.g., equals NA or null). This count is then divided by the total number of entries.

Bipartite Graph Construction. After quantifying missingness rates, each client constructs an undirected bipartite patient-modality graph $\mathcal{G}^{(k)}$, following [24], to explicitly model the relationship between patients and modalities under the federated learning setting. For clarity, we omit the superscript k in subsequent descriptions.

Formally, each bipartite graph is defined as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ denotes the set of nodes and $\mathcal{E}$ the set of edges. The node set $\mathcal{V}$ consists of two disjoint

subsets: the set of patient nodes $\mathcal{V}_P = \{v_1, v_2, \ldots, v_N\}$ and the set of modality nodes $\mathcal{V}_M = \{v_{N+1}, v_{N+2}, \ldots, v_{N+M}\}$, where N and M denote the number of patients and modalities, respectively. The edge set $\mathcal{E}$ is determined by the modality presence matrix $\mathbf{M} \in \{0,1\}^{N \times M}$, where an edge e_{v_i,v_j} exists if patient node $v_i \in \mathcal{V}_P$ has records corresponding to modality node $v_j \in \mathcal{V}_M$.

To capture semantic relationships between patients and modalities, we initialize edge embeddings based on raw modality features.

Specifically, for an edge e_{v_i,v_j}, its embedding is computed by applying a modality-specific encoder to the corresponding raw input by Equation (1).

$$\mathbf{e}_{v_i,v_j} = \text{Encoder}^{(j)}(\mathbf{x}_{i,j}), \tag{1}$$

where $\mathbf{x}_{i,j}$ denotes the raw feature vector of modality j for patient i, and $\text{Encoder}^{(j)}(\cdot)$ is a learnable encoder dedicated to modality j.

3.2 Personalized Federated Graph Learning

Representation Learning on Bipartite Graphs. After constructing the bipartite patient-modality graph $\mathcal{G}$ locally at each client, we utilize a multi-layer Graph Neural Network (GNN) to iteratively aggregate and update node embeddings.

Specifically, we adopt distinct initialization strategies depending on node type. Each modality node v_j is initialized as $\mathbf{h}_{v_j}^{(0)} = \text{Embedding}\,(j)$, while each patient node v_i is initialized with a constant vector of ones, i.e., $\mathbf{h}_{v_i}^{(0)} = 1 \in \mathbb{R}^d$, where the superscript (0) denotes the initial layer, and d represents the embedding dimension.

Therewith, the message from modality node v_j to patient node v_i at the l-th propagation layer is computed by Equation (2).

$$\mathbf{m}_{v_j \to v_i}^{(l)} = \text{GNN}^{(l)}\left(\left[\mathbf{h}_{v_j}^{(l-1)} \| \mathbf{h}_{v_i}^{(l-1)} \| \mathbf{e}_{v_j \to v_i}^{(l-1)}\right]\right), \tag{2}$$

where $\mathbf{h}_{v_j}^{(l-1)}$ and $\mathbf{h}_{v_i}^{(l-1)}$ represent embeddings of the source and target nodes at layer $(l-1)$ respectively, $\mathbf{e}_{v_j \to v_i}^{(l-1)}$ denotes the embedding of the edge from v_j to v_i, and $\|$ denotes vector concatenation.

Subsequently, each patient node v_i updates its embedding by aggregating messages from its neighboring modality nodes using mean pooling:

$$\mathbf{h}_{v_i}^{(l)} = \text{Aggregate}\left(\left\{\mathbf{m}_{v_j \to v_i}^{(l-1)} \mid v_j \in \mathcal{N}(v_i)\right\}\right), \tag{3}$$

where $\mathcal{N}(v_i) \subseteq \mathcal{V}_M$ denotes the set of modality nodes connected to patient node v_i.

After L propagation layers, we obtain the final patient node embeddings $\mathbf{h}_{v_i}^{(L)}$, which are then used for the contrastive learning.

Personalized Client Selection. To mitigate performance degradation caused by clients with high missingness rates, we introduce a personalized federated framework that dynamically selects participating clients based on their local missingness scores.

Given a tunable threshold θ, the client set is partitioned into:

- **Low-missingness clients** $\mathcal{S}_{\text{low}} = \{i \mid r^{(i)} < \theta\}$, who participate in global model training.
- **High-missingness clients** $\mathcal{S}_{\text{high}} = \{i \mid r^{(i)} \geq \theta\}$, who receive the trained global model and perform local fine-tuning.

The global training objective over $\mathcal{S}_{\text{low}}$ is defined as:

$$\min_{\mathbf{w}} \ \mathcal{L}_{\text{global}}(\mathbf{w}) = \sum_{i \in \mathcal{S}_{\text{low}}} \frac{n_i}{n_{\text{low}}} \mathcal{L}_i(\mathbf{w}), \tag{4}$$

where n_i is the number of samples on client S_i, $n_{\text{low}} = \sum_{i \in \mathcal{S}_{\text{low}}} n_i$, and $\mathcal{L}_i(\mathbf{w})$ is the local loss computed on client i.

Mutual-Consistent Contrastive Learning. To enhance robustness and representation consistency, we adopt Mutual-Consistent Contrastive Learning (MCCL) [24]. Specifically, we generate an augmented bipartite graph $\mathcal{G}'$ by randomly dropping edges from the original graph $\mathcal{G}$. We then construct contrastive pairs based on patient identity, encouraging embeddings of different modalities from the same patient (positive pairs) to be close, while pushing apart embeddings from different patients (negative pairs).

Local Training Objectives. On each selected client $S_i \in \mathcal{S}_{\text{low}}$, the local training objective integrates three components: unsupervised contrastive loss, supervised contrastive loss, and classification loss. These losses are computed based on patient node embeddings obtained from the local bipartite graph. We adopt NT-Xent loss [27] for contrastive objective and standard cross-entropy loss for local supervision. It is to be noted that the details of the constrastive loss, the supervised loss, and classification loss are introduced in the supplementary materials.

Personalized Fine-Tuning for High-Missingness Clients. After global training, the server distributes the optimized global model $\mathbf{w}^*$ to high-missingness clients $S_i \in \mathcal{S}_{\text{high}}$. Each client then performs local fine-tuning to better adapt to its heterogeneous and incomplete data distribution. The fine-tuning objective is formulated in Equation (5).

$$\min_{\mathbf{w}_i} \ \mathcal{L}_i(\mathbf{w}_i) + \frac{\alpha}{2} \|\mathbf{w}_i - \mathbf{w}^*\|^2, \tag{5}$$

where $\mathcal{L}_i(\mathbf{w}_i)$ denotes the local training loss on client S_i, and α controls the strength of regularization towards the global model.

4 Experiments

4.1 Experimental Setup

Datasets. We evaluate our method on two real-world EHR datasets: Alzheimer's Disease Neuroimaging Initiative (ADNI) [8] and MIMIC-IV [9]. For ADNI, we use the curated TADPOLE challenge dataset [16], which includes structural MRI, FDG-PET, diffusion tensor imaging (DTI), CSF biomarkers, and clinical assessments, covering 1,737 patients and 12,741 visits. For MIMIC-IV, we extract structured data modalities including diagnoses (ICD codes), procedures, prescriptions, and laboratory tests, following prior categorization [20]. Each dataset is split into training, validation, and test sets with a 70%:10%:20% ratio.

Task Settings. We consider three prediction tasks: (i) mortality prediction and (ii) hospital readmission prediction, both based on MIMIC-IV, and (iii) Alzheimer's disease progression prediction based on ADNI.

Evaluation Metrics. For the mortality prediction task, we report the Area Under the ROC Curve (AUC-ROC) and the Area Under the Precision–Recall Curve (AUC-PRC). For the Alzheimer's disease progression (ADP) task, we report Accuracy and AUC-ROC.

Baselines. We compare our model with sota methods including **Cafe** [17], **PEARL** [21], and **M3Care** [28]. Additionally, we examine the effect of differential privacy (DP) by applying DP-SGD to global updates to assess robustness under stricter privacy constraints.

Data Pre-processing. Since ADNI and MIMIC-IV are centralized datasets, we simulate a federated setting by partitioning each into 8 clients with roughly equal patient counts. To capture real-world heterogeneity, we design three synthetic missingness configurations by assigning different missing rates to clients. Empirically, both datasets exhibit ∼30% feature-level missingness. Based on this, we simulate client-wise missingness ranging from 0.3 to 1.0, where 1.0 represents fully missing input and is used to test robustness under extreme incompleteness.

Specifically, we define three settings:

- **Case 1 (increasing missingness):** (0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0).
- **Case 2 (mixed missingness):** (0.3, 0.3, 0.5, 0.5, 0.7, 0.7, 0.9, 0.9).
- **Case 3 (grouped high missingness):** (0.6, 0.6, 0.7, 0.7, 0.8, 0.8, 0.9, 0.9).

For each client, missingness is simulated by randomly masking features to match the assigned rate. This setup enables a systematic evaluation of robustness under varying degrees and patterns of modality incompleteness. Further details of the experimental setup are provided in the supplementary material.

4.2 Results Analysis

As shown in Table 1, our method **MissPCL** achieves the best overall performance across all three tasks under increasing missingness (Case 1). On ADNI (ADP task), it obtains the highest AUC-ROC of 0.6427, outperforming all baselines. On MIMIC-IV, MissPCL also delivers strong results, reaching the highest accuracy for both mortality and readmission prediction tasks. While **Cafe** shows competitive AUC-ROC for the mortality task, it underperforms on the ADP task. **M3Care** attains good accuracy but degrades with severe missingness, and **PEARL** consistently lags due to limited ability to handle incomplete modalities.

MissPCL's advantage stems from its missingness-aware client selection. We set $\theta = 0.8$ for ADNI and $\theta = 0.9$ for MIMIC-IV. For example, in ADNI Case 1, clients with missingness below 0.8 (i.e., 0.3–0.7) join global training, while higher-missingness clients skip and fine-tune locally. This design lets the global model learn from more complete data while still enabling adaptation for sparse clients.

Table 1: Baseline results with Case 1 client settings. AUC represents AUROC, ACC represents Accuracy. Best values are in bold.

Baseline	ADP (AUC)	Mortality AUC	Mortality ACC	Readmission AUC	Readmission ACC
Cafe	0.5509	**0.8923**	0.8880	**0.6834**	0.7634
PEARL	0.4683	0.6026	0.4918	0.5330	0.6105
M3Care	0.5963	0.6318	0.9127	0.5665	0.7615
MissPCL (ours)	**0.6427**	0.8210	**0.9132**	0.6491	**0.7675**

Table 2: Baseline results with Case 2 client settings.

Baseline	ADP (AUC)	Mortality AUC	Mortality ACC	Readmission AUC	Readmission ACC
Cafe	0.5489	**0.8914**	0.8906	**0.6882**	0.7637
PEARL	0.6166	0.6922	0.8724	0.5261	0.5814
M3Care	0.6018	0.6388	0.9127	0.5714	0.7615
MissPCL (ours)	**0.6621**	0.8252	**0.9137**	0.6485	**0.7670**

As shown in Table 2, **MissPCL** delivers strong results under mixed missingness (Case 2) across all tasks. On ADNI-ADP, it achieves the highest AUC-ROC of 0.6621, surpassing all baselines. For MIMIC-IV, it attains the best accuracy on both mortality (91.37%) and readmission (76.70%) prediction, reflecting superior generalization. Although **Cafe** yields the top AUC-ROC on MIMIC-IV tasks, its ADP performance is poor, likely due to reliance on imputation that benefits binary outcomes but falters under complex missingness. **PEARL** and **M3Care** also underperform, struggling with heterogeneous missingness. In this

case, we set $\theta = 0.6$ for ADNI and $\theta = 0.8$ for MIMIC-IV, enabling clients below the threshold to join collaborative training while others fine-tune locally. This strategy lets **MissPCL** exploit reliable data globally while adapting to sparse clients, enhancing robustness.

Table 3: MissPCL performance under different θ thresholds across three client cases. PRC represents AUPRC.

(a) Case 1

θ	ADP (AUC)	Mortality-AUC	Mortality-PRC	Readmission-AUC	Readmission-PRC
0.4	0.6224	0.7956	0.3001	0.6323	0.3678
0.5	0.6244	0.8087	0.3167	0.6417	0.3776
0.6	0.6216	0.8141	0.3262	0.6448	0.3831
0.7	0.6304	0.8175	0.3429	0.6459	0.3844
0.8	**0.6427**	0.8194	0.3494	0.6470	0.3861
0.9	0.6390	**0.8210**	0.3513	0.6491	**0.3874**
1.0	0.6362	0.8195	**0.3543**	**0.6506**	0.3868

(b) Case 2

θ	ADP (AUC)	Mortality-AUC	Mortality-PRC	Readmission-AUC	Readmission-PRC
0.4	0.6531	0.8129	0.3228	0.6395	0.3769
0.6	**0.6621**	0.8199	0.3499	0.6452	0.3845
0.8	0.6586	**0.8252**	**0.3657**	**0.6485**	**0.3877**
1.0	0.6547	0.8185	0.3594	0.6431	0.3862

(c) Case 3

θ	ADP (AUC)	Mortality-AUC	Mortality-PRC	Readmission-AUC	Readmission-PRC
0.7	0.6123	0.8056	0.3110	0.6364	0.3737
0.8	0.6155	0.8145	**0.3351**	0.6120	0.3815
0.9	**0.6202**	0.8191	0.3464	**0.6457**	**0.3848**
1.0	0.6176	**0.8215**	**0.3551**	0.6419	0.3814

4.3 Hyperparameter Analysis

Effect of Threshold θ. We analyze the sensitivity of the threshold parameter θ across three client configurations to evaluate its impact on performance. As shown in Table 3, higher θ generally improves **AUC-ROC** and **AUC-PRC**. The best results occur at $\theta = 0.8$ for ADNI and $\theta = 0.9$ for MIMIC-IV, suggesting that filtering out clients with severe missingness improves generalization by reducing noisy updates.

Beyond these points (e.g., $\theta = 1.0$), performance plateaus or slightly declines, indicating that including highly sparse clients may degrade the global model. Accordingly, we set $\theta = 0.8$ for ADNI and $\theta = 0.9$ for MIMIC-IV.

Case 2 (mixed missingness) shows MissPCL performing best on ADNI at $\theta = 0.6$ and on MIMIC-IV at $\theta = 0.8$, underscoring the trade-off between data quality and participation. Case 3 (high missingness) yields lower but stable performance, with optimal results again at $\theta = 0.9$, indicating robustness under challenging conditions.

Structured Modality-level Missingness. To assess robustness under more realistic conditions, we simulate missingness patterns based on empirical modality availability observed in the original datasets. In ADNI, cognitive assessments, MRI, and PET/CSF biomarkers are available in roughly 50%, 35%, and 15% of cases, respectively. In MIMIC-IV, discharge notes and lab vectors follow a 0.72-to-0.28 distribution. These ratios guide modality-wise masking, yielding client data distributions that reflect real-world clinical heterogeneity—such as limited access to PET scans or missing labs in smaller hospitals.

Under this setting (aligned with Case 1), **MissPCL** achieves 0.7836 AUC-ROC on ADNI, and 0.6233 AUC-ROC with 91.26% accuracy on MIMIC-IV mortality prediction. For readmission, it obtains 0.5767 AUC-ROC and 76.20% accuracy. These results highlight MissPCL's robustness under structured, non-uniform missingness.

Finally, we also evaluate MissPCL under varying values of θ and missingness rates r for Case 2 clients, and report detailed **AUC-PRC** results in the supplementary materials. To address privacy concerns in federated learning, we additionally extend MissPCL with differential privacy mechanisms; results and analysis of this extension are likewise included in the supplementary materials.

5 Conclusion

We propose MissPCL, a missingness-aware personalized federated graph contrastive learning framework designed for decentralized healthcare data with heterogeneous incompleteness. By estimating client-level missingness and dynamically selecting participants, MissPCL balances data quality and participation, while enabling personalized adaptation for highly incomplete clients. Leveraging bipartite graph modeling and mutual-consistent contrastive learning, it effectively captures patient-disease relationships under varying missingness conditions. Extensive experiments on real-world EHR datasets show that MissPCL consistently outperforms state-of-the-art baselines across diverse missingness patterns.

In future work, we aim to explore more fine-grained missingness modeling beyond scalar rates, integrate uncertainty-aware representations, and deploy the framework in real-world clinical federated settings to further evaluate its scalability and practical utility.

References

1. Chen, J., Zhang, A.: Hgmf: Heterogeneous graph-based fusion for multimodal data with incompleteness. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1295–1305 (2020)
2. Chen, Y., Lu, W., Qin, X., Wang, J., Xie, X.: Metafed: Federated learning among federations with cyclic knowledge distillation for personalized healthcare. IEEE Transactions on Neural Networks and Learning Systems **35**(11), 16671–16682 (2024)
3. Esmaeilzadeh, P.: Challenges and strategies for wide-scale artificial intelligence (ai) deployment in healthcare practices: A perspective for healthcare organizations. Artificial Intelligence in Medicine **151**, 102861 (2024)
4. Guan, H., Yap, P.T., Bozoki, A., Liu, M.: Federated learning for medical image analysis: A survey. Pattern Recognition p. 110424 (2024)
5. Han, H., Zhao, T., Yang, C., Zhang, H., Liu, Y., Wang, X., Shi, C.: Openhgnn: an open source toolkit for heterogeneous graph neural network. In: Proceedings of the 31st ACM International Conference on Information & Knowledge Management. pp. 3993–3997 (2022)
6. Han, X., Nguyen, H., Harris, C.W., Ho, N., Saria, S.: Fusemoe: Mixture-of-experts transformers for fleximodal fusion. In: The Thirty-eighth Annual Conference on Neural Information Processing Systems (2024)
7. Huang, W., Wang, D., Ouyang, X., Wan, J., Liu, J., Li, T.: Multimodal federated learning: Concept, methods, applications and future directions. Information Fusion **112**, 102576 (2024)
8. Jack Jr, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., L. Whitwell, J., Ward, C., et al.: The alzheimer's disease neuroimaging initiative (adni): Mri methods. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine **27**(4), 685–691 (2008)
9. Johnson, A.E., Bulgarelli, L., Shen, L., Gayles, A., Shammout, A., Horng, S., Pollard, T.J., Hao, S., Moody, B., Gow, B., et al.: Mimic-iv, a freely accessible electronic health record dataset. Scientific Data **10**(1) (2023)
10. Kim, W., Son, B., Kim, I.: Vilt: Vision-and-language transformer without convolution or region supervision. In: International conference on machine learning. pp. 5583–5594. PMLR (2021)
11. Koo, H.: Overcoming uncertain incompleteness for robust multimodal sequential diagnosis prediction via curriculum data erasing guided knowledge distillation. In: 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (2025)
12. Lablack, M., Yu, S., Xu, S., Shen, Y.: Long-sequence model for traffic forecasting in suboptimal situation. In: Proceedings of the 18th Workshop on Mobility in the Evolving Internet Architecture. pp. 25–30 (2023)
13. Li, Y., Hu, X., Yu, S., Ding, W., Pedrycz, W., Kiat, Y.C., Liu, Z.: A vertical federated multi-view fuzzy clustering method for incomplete data. IEEE Transactions on Fuzzy Systems (2025)
14. Liu, Q., Hu, J., Xiao, Y., Zhao, X., Gao, J., Wang, W., Li, Q., Tang, J.: Multimodal recommender systems: A survey. ACM Comput. Surv. **57**(2) (2024)
15. Liu, Y., Wang, C., Yuan, X.: Fedmobile: Enabling knowledge contribution-aware multi-modal federated learning with incomplete modalities. In: Proceedings of the ACM on Web Conference 2025. pp. 2775–2786 (2025)

16. Marinescu, R.V., Oxtoby, N.P., Young, A.L., Bron, E.E., Toga, A.W., Weiner, M.W., Barkhof, F., Fox, N.C., Golland, P., Klein, S., et al.: Tadpole challenge: Accurate alzheimer's disease prediction through crowdsourced forecasting of future data. In: Predictive Intelligence in Medicine: Second International Workshop. pp. 1–10. Springer (2019)
17. Min, S., Asif, H., Wang, X., Vaidya, J.: Cafe: Improved federated data imputation by leveraging missing data heterogeneity. IEEE Transactions on Knowledge and Data Engineering **37**(5), 2266–2281 (2025)
18. Ren, J., Hou, M., Liu, Z., Bai, X.: Eagle: Contrastive learning for efficient graph anomaly detection. IEEE Intelligent Systems **38**(2), 55–63 (2022)
19. Saha, P., Mishra, D., Wagner, F., Kamnitsas, K., Noble, J.A.: Fedpia – permuting and integrating adapters leveraging wasserstein barycenters for finetuning foundation models in multi-modal federated learning. Proceedings of the AAAI Conference on Artificial Intelligence **39**(19) (2025)
20. Sung, M., Hahn, S., Han, C.H., Lee, J.M., Lee, J., Yoo, J., Heo, J., Kim, Y.S., Chung, K.S.: Event prediction model considering time and input error using electronic medical records in the intensive care unit: Retrospective study. JMIR Medical Informatics **9**(11), e26426 (2021)
21. Tang, T., Han, Z., Cai, Z., Yu, S., Zhou, X., Oseni, T., Das, S.K.: Personalized federated graph learning on non-iid electronic health records. IEEE Transactions on Neural Networks and Learning Systems **35**(9), 11843–11856 (2024)
22. Tang, T., Han, Z., Yu, S., Bagirov, A., Zhang, Q.: Fuzzy multiview graph learning on sparse electronic health records. IEEE Transactions on Fuzzy Systems **32**(10), 5520–5532 (2024)
23. Wijesinghe, A., Zhang, S., Ding, Z.: Ps-fedgan: An efficient federated learning framework with strong data privacy. IEEE Internet of Things Journal **11**(16), 27584–27596 (2024)
24. Wu, Z., Dadu, A., Tustison, N., Avants, B., Nalls, M., Sun, J., Faghri, F.: Multimodal patient representation learning with missing modalities and labels. In: The Twelfth International Conference on Learning Representations (2024)
25. Wu, Z., Sun, S., Wang, Y., Liu, M., Xu, K., Wang, W., Jiang, X., Gao, B., Lu, J.: Fedcache: A knowledge cache-driven federated learning architecture for personalized edge intelligence. IEEE Transactions on Mobile Computing **23**(10), 9368–9382 (2024)
26. Yang, F.E., Wang, C.Y., Wang, Y.C.F.: Efficient model personalization in federated learning via client-specific prompt generation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 19159–19168 (2023)
27. You, Y., Chen, T., Sui, Y., Chen, T., Wang, Z., Shen, Y.: Graph contrastive learning with augmentations. In: Advances in Neural Information Processing Systems. vol. 33, pp. 5812–5823 (2020)
28. Zhang, C., Chu, X., Ma, L., Zhu, Y., Wang, Y., Wang, J., Zhao, J.: M3care: Learning with missing modalities in multimodal healthcare data. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. pp. 2418–2428 (2022)
29. Zheng, S., Zhu, Z., Liu, Z., Guo, Z., Liu, Y., Yang, Y., Zhao, Y.: Multi-modal graph learning for disease prediction. IEEE Transactions on Medical Imaging **41**(9), 2207–2216 (2022)
30. Zuo, H., Liu, R., Zhao, J., Gao, G., Li, H.: Exploiting modality-invariant feature for robust multimodal emotion recognition with missing modalities. In: 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (2023)