This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

FedCoRe: Target-Adaptive Completion for Missing Modalities in Healthcare Federated Learning

Holger R. Roth*, Ziyue Xu, and Peter Cnudde

NVIDIA, Santa Clara, CA, USA
hroth@nvidia.com

Abstract. Federated multimodal models often assume every site has every modality, although hospitals differ in access to EHRs, chest radiographs, and ECGs. We study this setting on a MIMIC-derived respiratory deterioration task with simulated FL clients and introduce FedCoRe (Federated Cross-Modal Representation Completion). FedCoRe learns representation- or logit-space corrections rather than generating synthetic ECGs or CXR images. When a client observes a modality that may be missing at deployment, it evaluates the same example with and without that modality to obtain paired supervision. Only clients with such pairs update the completion module, and validation may retain the unchanged prediction. We freeze the trained multimodal predictor during evaluation so that measured differences come only from completion. Hiding ECG reduced AUROC by about 0.085; paired-example FedAvg restored 0.0415 AUROC, or 49.0% of the lost performance. We therefore report two distinct effects: paired-example FedAvg partially recovers the missing-ECG gap, while validation-selected completion is a task-specific classifier-logit correction rather than literal ECG recovery. For CXR, effect-aware completion recovers 52.8% of the loss in a controlled test where CXR is hidden. Paired-example FedAvg transfers part of this effect, but validation keeps the no-completion baseline for deployment cases whose inputs lack CXR. Thus, FedCoRe should be read as a validation-gated completion/correction framework: it can recover missing-modality signal in supported settings, but it should be deployed only when paired examples and validation evidence support that modality.

Keywords: Federated learning · Missing modalities · Multimodal foundation models

1 Introduction

Multimodal clinical prediction increasingly combines structured context, radiology images, physiological signals, and text. Complete-modality evaluation is often overly optimistic for federated learning (FL): some hospitals may support

* Corresponding author.

different modalities, while others may include patients with only a subset of the observed modalities. Figure 1 illustrates this gap between evaluation and deployment.

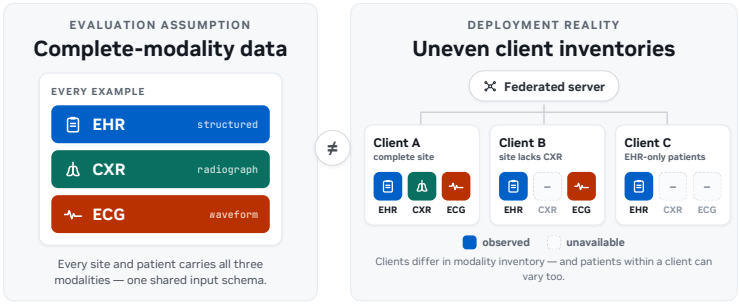

Fig. 1. Complete-modality evaluation contrasts with federated deployment, where sites and patients may have different subsets of EHR, CXR, and ECG.

Our question is whether clients observing a modality can teach a lightweight correction that benefits clients or patients missing it, without centralizing raw records [16]. Here, "completion" means adding a learned residual in representation or classifier-logit space, not generating a synthetic ECG or CXR. We separate two possible sources of improvement: **(1)** robust predictor training and **(2)** the completion operator itself. Accordingly, we ask whether collaborative training through FL helps missing-modality patients and, with that predictor frozen, whether completion adds further signal.

FedCoRe learns task-aware corrections, updates each operator only from clients with valid modality pairs, aggregates those updates, and uses validation to apply the operator or retain the unchanged prediction.

Related Work. FedAvg is the canonical decentralized optimization algorithm [14]; platforms such as Flower [2] or NVIDIA FLARE [17] can coordinate healthcare FL while keeping raw records at their source. Recent reviews identify modality heterogeneity as a central barrier [18] and prior multimodal FL addresses missingness through pseudo-modality generation [20], cross-contrast synthesis [19,3], retrieval [15], contrastive ensembles [21], embedding transfer [13], and prototype/mask completion [1]. FedCoRe instead learns modality-specific corrections from paired outputs without requiring universal modality availability. Raw records remain local, but we do not claim formal privacy guarantees; model updates may leak information without mechanisms such as secure aggregation or differential privacy [12].

2 Method

Intuition. Suppose one client has EHR, CXR, and ECG, while another lacks ECG. The first client compares predictions with and without ECG to train an ECG-specific completion operator. The server aggregates operator updates only from clients that can form such pairs; the second client applies the operator only when validation shows that it helps. Thus, FedCoRe creates paired views, trains correction operators, aggregates only valid updates, and validates deployment.

Paired views. Let client k contain examples $\mathcal{D}_k = \{(x_i^{M_i}, y_i)\}_{i=1}^{n_k}$, where y_i is the clinical outcome and $M_i \subseteq \mathcal{M}$ is the observed modality set; $\mathcal{M} = \{\text{EHR}, \text{CXR}, \text{ECG}\}$. A fixed encoder f_ϕ and prediction head g_ψ map the modalities available for an example to an internal representation $h_i(M_i)$ and classifier logits $\ell_i(M_i)$:

$$h_i(M_i) = f_\phi(x_i^{M_i}), \qquad \ell_i(M_i) = g_\psi(h_i(M_i)). \tag{1}$$

Let t denote the input modality we want to complete when it is missing at deployment. Paired supervision exists only if $t \in M_i$ and $M_i \setminus \{t\} \neq \emptyset$. We evaluate each such example twice to form a pair:

$$h_i^{+t} = h_i(M_i), \qquad h_i^{-t} = h_i(M_i \setminus \{t\}), \tag{2}$$

Here, h_i^{+t} is the representation with t observed and h_i^{-t} is the representation after removing t. The corrected output introduced below is the *completed* representation or logit. Examples without t, or with only t, cannot supervise its completion.

Completion operator. FedCoRe learns the change induced by modality t. For pooled hidden states, a modality-specific operator C_{θ_t} predicts a correction to the t-removed representation:

$$\tilde{h}_i^t = h_i^{-t} + \alpha_t C_{\theta_t}(h_i^{-t}), \tag{3}$$

The scalar α_t controls how strongly that correction is applied; a zero value leaves the original representation unchanged. For logit-delta completion, the operator instead corrects the frozen classifier output directly:

$$\tilde{\ell}_i^t = \ell_i^{-t} + \alpha_t C_{\theta_t}(h_i^{-t}). \tag{4}$$

Thus, C_{θ_t} can operate either in representation space or in classifier-logit space: it predicts a hidden-state residual in Eq. (3), or an additive logit residual in Eq. (4). In both cases, the paired t-present pass provides the training reference: h_i^{+t} for representation completion and $\ell_i(M_i)$ for logit completion. The formulation is backbone-agnostic: it requires only paired t-present and t-removed representations or logits.

For CXR, we use an effect-aware logit interface. Let

$$d_i^{\text{CXR}} = \ell_i(M_i) - \ell_i(M_i \setminus \{\text{CXR}\}), \qquad \Delta_i^{\text{CXR}} = \gamma_i r_\theta(h_i^{-\text{CXR}}),$$

where d_i^{CXR} is the observed two-logit CXR effect and $\gamma_i \in [0,1]$ is an instance gate. Deployment uses

$$\tilde{\ell}_i^{\mathrm{CXR}} = \ell_i^{-\mathrm{CXR}} + \alpha_{\mathrm{CXR}} \Delta_i^{\mathrm{CXR}}.$$

The CXR loss adds Smooth-L1 effect matching, probability consistency, gate supervision, and gate sparsity to the task/logit terms. The implementation uses λ_d, λ_p, λ_g, and λ_s for effect matching, probability consistency, gate supervision, and gate sparsity, respectively.

Local objective. Each client learns the correction from its valid pairs using a task-aware objective:

$$\mathcal{L}_i^t = \lambda_{\mathrm{task}} \mathrm{CE}(g_\psi(\tilde{h}_i^t), y_i) + \lambda_{\mathrm{logit}} \|g_\psi(\tilde{h}_i^t) - g_\psi(h_i^{+t})\|_2^2 + \lambda_{\mathrm{rep}} \|\tilde{h}_i^t - h_i^{+t}\|_2^2. \quad (5)$$

The three terms preserve label prediction, align the completed and t-present logits, and regularize the completed representation, respectively. The main missing-ECG result retains the task and logit terms and sets $\lambda_{\mathrm{rep}} = 0$ after validation.

Aggregation by available supervision. Clients share completion modules without assuming that every client can supervise every modality. For modality t, the valid paired set at client k is

$$\mathcal{P}_{k,t} = \{i \in \mathcal{D}_k : t \in M_i,\ M_i \setminus \{t\} \neq \emptyset\} \quad (6)$$

Clients with $|\mathcal{P}_{k,t}| = 0$ return no completion weights and receive zero aggregation weight. The server averages the remaining updates in proportion to the number of valid pairs:

$$\theta_t^{r+1} = \sum_{k=1}^{K} \frac{|\mathcal{P}_{k,t}|}{\sum_j |\mathcal{P}_{j,t}|} \theta_{k,t}^{r+1}. \quad (7)$$

Consequently, a client that cannot supervise completion for t cannot dilute or corrupt its global completion operator.

Validation-based deployment. Finally, validation selects a candidate $c = (s, \alpha, \tau) \in \mathcal{C}_t$: an operator source s, completion strength α, and optional gate threshold τ. The no-completion candidate c_0 has $\alpha = 0$ and leaves predictions unchanged. For deployment, $V_{\mathrm{miss}}^t(c)$ is computed on cases whose observed inputs exclude t; for a frozen-predictor completion test, it uses complete cases after deliberately removing t. $V_{\mathrm{safe}}(c)$ optionally measures a protected validation view, such as complete-modality cases. We select

$$c_t^\star = \arg\max_{c \in \mathcal{C}_t} V_{\mathrm{miss}}^t(c) \quad \text{s.t.} \quad V_{\mathrm{miss}}^t(c) \geq V_{\mathrm{miss}}^t(c_0),\ V_{\mathrm{safe}}(c) \geq V_{\mathrm{safe}}(c_0) - \epsilon. \quad (8)$$

The second constraint is omitted when no separate safety view is used. If no completion candidate qualifies, c_0 leaves the original prediction unchanged. Figure 2 summarizes the complete workflow.

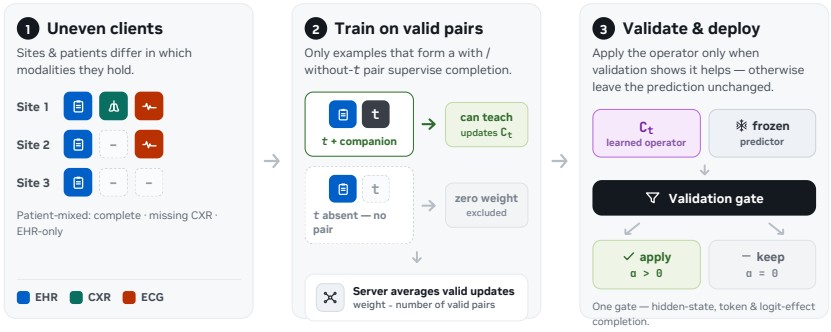

Fig. 2. FedCoRe workflow: paired observations train the completion operator, sites without valid pairs receive zero weight, and validation applies completion or leaves predictions unchanged.

3 Task and Federated Setup

Cohort and label. We construct a MIMIC-derived benchmark using public datasets from PhysioNet[1] [4]: MIMIC-IV [6,9], MIMIC-CXR [8,10], MIMIC-CXR-JPG [7,11], and MIMIC-IV-ECG [5]. The benchmark links structured EHR context, a CXR image, and a diagnostic ECG record. The task is binary respiratory deterioration within 48 hours of the later ECG–CXR "index" timestamp, operationalized as future invasive ventilation, ICU transfer, or all-cause in-hospital death. To reduce leakage, structured context is pre-index, outcome event times are strictly post-index, ICU-transfer labels use post-index transfer times, and encounters in which the patient was already in an ICU at the index time are excluded. The ventilation endpoint counts only new post-index invasive ventilation events. ECG and CXR observations are matched within 6 hours. The cohort contains 13,914 encounters from 11,331 patients, split patient-disjoint into 9,777 train, 2,056 validation, and 2,081 test examples; the positive class is 304/13,914 (2.2%).

FL modality heterogeneity. Let $M_i \subseteq \{\text{EHR}, \text{CXR}, \text{ECG}\}$ be the observed modalities for example i. We simulate client-level missingness and patient-mixed missingness, where a client contains both complete and incomplete examples. Splits are patient-disjoint across train/validation/test and FL clients. Table 1 summarizes the main templates used for the missing-ECG frozen-predictor completion test.

4 Experiments

Baselines and controls. We compare no completion, local-only completion, paired-example FedAvg completion, and validation-selected completion; full-

[1] https://physionet.org/

Table 1. Core FL heterogeneity templates for the missing-ECG frozen-predictor completion test. Missing-ECG mass is expected test-set patient mass under the client mask; paired sources are examples in which ECG is observed and can supervise completion.

Template	Clients	Full-modality clients	Missing-ECG mass	Paired-example structure	Purpose
No full-modality client	4	0	66.3%	ECG-present but no complete site	Negative control
Patient-mixed any-missing	4	Mixed	30.0%	Complete and ECG-present patients within clients	Main heterogeneous setting
Patient-mixed heavy	4	Mixed	36.3%	Heavier missing-ECG patient mix	Stress setting

modality runs provide a reference, not a centralized pooled or directly reimplemented prior-method comparison. CXR controls include server-selected and model-soup sources, scalar logit recalibration, pseudo-CXR tokens, and forced fixed-scale completion. In Table 2, "Before" is NoCompletion, FedAvg uses the aggregated operator with a validation-selected strength, and "Val-selected" additionally chooses among local and FedAvg sources. FedCoRe operates on previously trained global multimodal predictors, independent of whether predictor training is centralized or federated. During controlled completion experiments, each predictor is held fixed so that performance changes are attributable to the target-specific completion operators trained and aggregated across clients. The ECG predictor was obtained through prior modality-heterogeneous federated training, whereas the CXR reference predictors were trained separately on the complete EHR+CXR training view.

CXR configuration. Missing-CXR experiments use a frozen Qwen3-VL-8B-Instruct predictor with LoRA rank 8 adapters, pre-index EHR context, and CXR images. The effect-aware gated MLP uses hidden size 4096, dropout 0.25, learning rate 5×10^{-5}, 80 local operator-training epochs, batch size 128, uncapped completion train/validation sets, and balanced task weighting. Its loss weights are $\lambda_{\text{task}} = 1$, $\lambda_{\text{logit}} = 8$, and $\lambda_{\text{rep}} = 0$, with Smooth-L1 effect matching, $\lambda_p = 0.5$, $\lambda_g = 1.0$, and $\lambda_s = 0.005$. Validation considers $\alpha \in \{0, .1, .25, .5, .75, 1, 1.25, 1.5, 2, 3, 4, 6, 8\}$ and gate thresholds $\{0, .1, .25, .5, .75\}$.

ECG/EHR configuration. Supporting ECG/EHR rows use an earlier Qwen-family backend with lead-II ECG patch tokens. The main ECG operator is a two-layer logit-delta MLP with hidden size 2048, dropout 0.3, learning rate 10^{-4}, batch size 128, 50 local epochs, and one FedAvg completion round across four simulated clients, with $\lambda_{\text{task}} = 1$, $\lambda_{\text{logit}} = 4$, and $\lambda_{\text{rep}} = 0$.

Selection protocol. Candidate completion operators are selected using validation data only, and every strength grid includes $\alpha = 0$ (no completion). For ECG, each source selects $\alpha \in \{0, 2, 4, 8, 12\}$ by missing-ECG validation AUROC subject to nonnegative gain; "Val-selected" then chooses the source with the greatest validation gain. The no-full-modality, any-missing, and heavy settings evaluate 3, 11, and 8 sources, respectively. All positive reported rows have unique validation maxima; any tie is resolved by the smallest $|\alpha|$ and a deterministic source identifier. For CXR, we evaluate the fixed operator over the 13 strengths and five gate thresholds above. A CXR candidate must improve remove-CXR validation AUROC, have at least 0.02 full-minus-removed validation headroom, and reduce full-view validation AUROC by no more than $\epsilon = 0.005$. The validation/test manifests contain 2,056/2,081 examples with 44/42 positives. For the

ECG-present reference, the same frozen predictor receives CXR, EHR, and ECG; the missing view removes only ECG on the same manifest. Table 2 changes the available completion sources and paired-example pattern, not the frozen predictor. Test predictions were saved during replay but summarized only after validation fixed the source, strength, gate threshold, or no-completion choice; test metrics were never used for selection.

Metrics. Our primary endpoint is the absolute change in AUROC among cases missing the modality to be completed. Secondary measures are relative lift and, when available, aggregate deployment AUROC change. We report positive-row counts and selected strength; main ECG/CXR analyses also include AUPRC and patient-level paired bootstrap 95% CIs. The intervals condition on the validation-selected operator and exclude source, strength, and gate selection uncertainty. With 2.2% prevalence and 42 test positives, AUROC does not establish precision or clinical-threshold utility; results are model-development evidence.

5 Results & Discussion

ECG is the most reliably completed modality. Across the exploratory matrix, missing ECG improves in 69/70 rows, with mean +0.059 AUROC and +9.2% relative lift. Missing EHR improves more modestly in 20/20 rows (mean +0.027), while missing CXR depends more strongly on the backbone and completion interface. Because the modalities use different source inputs, backbones, and completion interfaces, these numbers are not a ranking of clinical importance. Figure 3 summarizes the main decisions, with detailed ECG and CXR results in Tables 2 and 3.

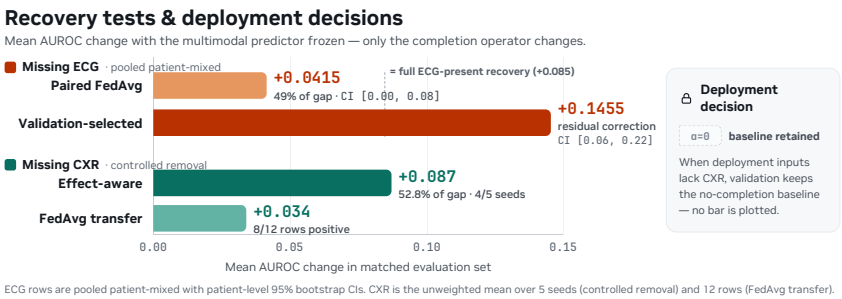

Fig. 3. Observed gains with the multimodal predictor frozen. Bars report matched target-removal/recovery tests rather than deployment gains; the box separately reports CXR deployment, where validation selected $\alpha = 0$ and retained the baseline.

Paired-example FedAvg recovers about half of the missing-ECG gap. With ECG present, the frozen predictor reaches 0.6627 AUROC; removing ECG reduces it to 0.5781. Across the six patient-mixed rows, FedAvg raises

AUROC to 0.6195, adding +0.0415 [0.0023, 0.0778] and recovering 49.0% of the lost performance. The predictor and patients are unchanged; only the completion operator differs. In the no-full-modality-client control, validation leaves the prediction unchanged because none of the available corrections helps.

Validation-selected ECG completion acts as residual correction. It reaches 0.7235 AUROC, an improvement of +0.1455 [0.0617, 0.2213] that exceeds the ECG-present reference. We therefore interpret this result as classifier-logit correction beyond modality recovery, not literal ECG reconstruction. AUPRC changes remain unstable with 42 positive test examples; selected-source validation/test deltas are +0.120–0.136/+0.104–0.169.

Table 2. Frozen-predictor missing-ECG test. Both methods select scale on validation; "Val-selected" also selects source. CIs use shared patient-level resamples across the six matched patient-mixed predictions. Full ref is ECG-present replay; reference-normalized gain is (After − Before)/(Full ref − Before), where values above 100% denote correction beyond the reference.

Scenario	Completion	Seeds	Before	After	Full ref	Ref.-norm.	Δ AUROC [95% CI]	AUPRC	Positive
No full-modality client	FedAvg	7,17,42	0.5781	0.5781	0.6627	0.0%	+0.0000 [0.0000, 0.0000]	0.0375 / 0.0375	0/3
No full-modality client	Val-selected	7,17,42	0.5781	0.5781	0.6627	0.0%	+0.0000 [0.0000, 0.0000]	0.0375 / 0.0375	0/3
Patient-mixed combined	**FedAvg**	**6 rows**	**0.5781**	**0.6195**	**0.6627**	**49.0%**	**+0.0415 [0.0023, 0.0778]**	**0.0375 / 0.0368**	**6/6**
Patient-mixed combined	**Val-selected**	**6 rows**	**0.5781**	**0.7235**	**0.6627**	**171.8%**	**+0.1455 [0.0617, 0.2213]**	**0.0375 / 0.0495**	**6/6**

Additional modality tests. A single missing-ECG scenario gives +0.095 mean AUROC for FedAvg and +0.116 for validation-selected completion across six seeds; missing-EHR is weaker but non-harmful (+0.014, three seeds).

CXR is partially recoverable when removed. With Qwen3-VL, we evaluate complete cases with CXR present, deliberately remove CXR, and apply the effect-aware logit operator. Across five seeds, completion adds +0.087 AUROC on average (4/5 positive) and recovers 52.8% of the full-vs-missing gap. The effect varies: seed 23 leaves the prediction unchanged, and only two per-seed intervals exclude zero; we therefore report paired CIs per seed rather than treat seeds as independent clinical replications. AUPRC improves by +0.010 on average but is unstable with 42 positives. Scalar recalibration gives zero rank-metric lift. Pseudo-CXR tokens are a negative ablation: validation rejects them, while forced completion reduces AUROC by 0.082. This is evidence that some CXR-induced logit signal is recoverable from the remaining inputs, not that deployment without CXR is solved.

CXR completion transfers through FedAvg but is not deployed when unsupported. Paired-example FedAvg improves the remove-CXR test in 8/12 rows (mean +0.034 AUROC); server-selected sources are weaker (+0.017, 2/6 positive), and source-subset operators are negative (0/12). For deployment cases whose observed inputs lack CXR, validation selects $\alpha = 0$. FedCoRe therefore retains the no-completion baseline instead of applying an unsupported CXR correction.

Table 3. Frozen-predictor missing-CXR completion test. Validation selects scale and gate threshold; only scale is shown. Delta intervals are paired patient-level bootstrap 95% CIs. Means and gap recovery are unweighted across seeds.

Seed	Full AUROC	Missing	Completed	Δ AUROC [95% CI]	Gap rec.	AUPRC before/after	α
7	0.6673	0.5237	0.6454	+0.1217 [0.0005, 0.2330]	84.7%	0.0204 / 0.0409	0.50
17	0.6969	0.5933	0.6228	+0.0295 [-0.0520, 0.1104]	28.5%	0.0402 / 0.0331	0.75
23	0.7303	0.6187	0.6187	+0.0000 [0.0000, 0.0000]	0.0%	0.0297 / 0.0297	0.00
31	0.7072	0.5188	0.6594	+0.1405 [0.0301, 0.2464]	74.6%	0.0255 / 0.0345	0.10
42	0.7192	0.5307	0.6739	+0.1431 [-0.0037, 0.2826]	75.9%	0.0327 / 0.0611	0.10
Mean	0.7042	0.5570	0.6440	+0.0870	52.8%	0.0297 / 0.0399	–

6 Conclusion

FedCoRe learns modality-specific operators from valid pairs, excludes uninformative clients from aggregation, and can retain the unchanged prediction. Missing-ECG benefits from completion; missing-CXR shows recoverable signal in remove-CXR tests but not unsupported deployment cases. Controlled four-client masks impose modality availability rather than reflecting natural missingness or institutional covariate shift; external multi-site validation remains future work.

Code and AI use disclosure. Experiments used NVIDIA FLARE[2]. Codex and ChatGPT assisted with experimentation and writing, and Claude Design with figures; the authors verified all content and remain responsible.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bao, G., Zhang, Q., Miao, D., Gong, Z., Hu, L., Liu, K., Liu, Y., Shi, C.: Multimodal federated learning with missing modality via prototype mask and contrast (2023), arXiv:2312.13508
2. Beutel, D.J., Topal, T., Mathur, A., Qiu, X., Fernandez-Marques, J., Gao, Y., et al.: Flower: A friendly federated learning research framework. arXiv preprint arXiv:2007.14390 (2020)
3. Dalmaz, O., Mirza, M.U., Elmas, G., Özbey, M., Dar, S.U.H., Ceyani, E., Avestimehr, S., Çukur, T.: One model to unite them all: Personalized federated learning of multi-contrast MRI synthesis. Medical Image Analysis **94**, 103121 (2024). https://doi.org/10.1016/j.media.2024.103121
4. Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., et al.: Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. Circulation **101**(23), e215–e220 (2000). https://doi.org/10.1161/01.CIR.101.23.e215

[2] https://github.com/NVIDIA/NVFlare

5. Gow, B., Pollard, T., Nathanson, L.A., Johnson, A., Moody, B., Fernandes, C., et al.: MIMIC-IV-ECG: Diagnostic electrocardiogram matched subset. PhysioNet (sep 2023). https://doi.org/10.13026/4nqg-sb35, version 1.0
6. Johnson, A., Bulgarelli, L., Pollard, T., Gow, B., Moody, B., Horng, S., Celi, L.A., Mark, R.: MIMIC-IV. PhysioNet (oct 2024). https://doi.org/10.13026/kpb9-mt58, version 3.1
7. Johnson, A., Lungren, M., Peng, Y., Lu, Z., Mark, R., Berkowitz, S., Horng, S.: MIMIC-CXR-JPG - chest radiographs with structured labels. PhysioNet (mar 2024). https://doi.org/10.13026/jsn5-t979, version 2.1.0
8. Johnson, A., Pollard, T., Mark, R., Berkowitz, S., Horng, S.: MIMIC-CXR Database. PhysioNet (sep 2019). https://doi.org/10.13026/C2JT1Q, version 2.0.0
9. Johnson, A.E.W., Bulgarelli, L., Shen, L., Gayles, A., Shammout, A., Horng, S., et al.: MIMIC-IV, a freely accessible electronic health record dataset. Scientific Data **10**(1), 1 (2023). https://doi.org/10.1038/s41597-022-01899-x
10. Johnson, A.E.W., Pollard, T.J., Berkowitz, S.J., Greenbaum, N.R., Lungren, M.P., Deng, C.y., et al.: MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. Scientific Data **6**(1), 317 (2019). https://doi.org/10.1038/s41597-019-0322-0
11. Johnson, A.E.W., Pollard, T.J., Greenbaum, N.R., Lungren, M.P., Deng, C.y., Peng, Y., et al.: MIMIC-CXR-JPG, a large publicly available database of labeled chest radiographs (2019), arXiv:1901.07042
12. Kairouz, P., et al.: Advances and open problems in federated learning. Foundations and Trends in Machine Learning **14**(1-2), 1–210 (2021). https://doi.org/10.1561/2200000083
13. Le, H.Q., Nguyen, M.N.H., Thwal, C.M., Qiao, Y., Zhang, C., Hong, C.S.: FedMEKT: Distillation-based embedding knowledge transfer for multimodal federated learning. Neural Networks **183**, 107017 (2025). https://doi.org/10.1016/j.neunet.2024.107017
14. McMahan, B., Moore, E., Ramage, D., Hampson, S., Agüera y Arcas, B.: Communication-efficient learning of deep networks from decentralized data. In: Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (2017)
15. Poudel, P., Shrestha, P., Amgain, S., Shrestha, Y.R., Gyawali, P., Bhattarai, B.: CAR-MFL: Cross-modal augmentation by retrieval for multimodal federated learning with missing modalities. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. pp. 102–112. Springer Nature Switzerland (2024). https://doi.org/10.1007/978-3-031-72117-5_10
16. Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H.R., Albarqouni, S., et al.: The future of digital health with federated learning. NPJ digital medicine **3**(1), 119 (2020)
17. Roth, H.R., Cheng, Y., Wen, Y., Yang, I., Xu, Z., Hsieh, Y.T., et al.: NVIDIA FLARE: Federated learning from simulation to real-world (2022), arXiv:2210.13291
18. Thrasher, J., Devkota, A., Siwakoti, P., Chivukula, R., Poudel, P., Hu, C., et al.: Multimodal federated learning in healthcare: A review. Journal of Healthcare Informatics Research (2025). https://doi.org/10.1007/s41666-025-00226-4
19. Wang, J., Xie, G., Huang, Y., Lyu, J., Zheng, F., Zheng, Y., Jin, Y.: FedMed-GAN: Federated domain translation on unsupervised cross-modality brain image synthesis. Neurocomputing **546**, 126282 (2023). https://doi.org/10.1016/j.neucom.2023.126282

20. Yan, Y., Feng, C.M., Li, Y., Li, P., Goh, R.S.M., Lei, B., et al.: Federated pseudo modality generation for incomplete multi-modal MRI reconstruction. IEEE Journal of Biomedical and Health Informatics **29**(8), 5849–5861 (2025). https://doi.org/10.1109/JBHI.2025.3566217
21. Yu, Q., Liu, Y., Wang, Y., Xu, K., Liu, J.: Multimodal federated learning via contrastive representation ensemble. In: International Conference on Learning Representations (2023)