FedMAC: Tackling Partial-Modality Missing in Federated Learning with Cross-Modal Aggregation and Contrastive Regularization

Manh Duong Nguyen[1], Trung Thanh Nguyen[2], Huy Hieu Pham[3], Trong Nghia Hoang[4], Phi Le Nguyen[1], Thanh Trung Huynh[5]

[1]Hanoi University of Science and Technology, Hanoi, Vietnam

[2]Nagoya University, Nagoya, Japan

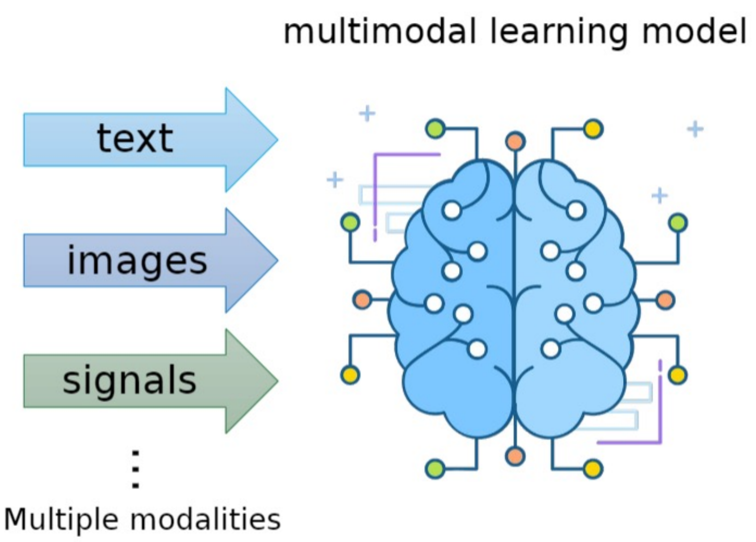
[3]VinUniveristy, Hanoi, Vietnam

[4]Washington State University, State of Washington, United States,

[5]Swiss Federal Institute of Technology Lausanne, Vaud, Switzerland

The 22nd International Symposium on Network Computing and Applications (NCA 2024)

Introduction

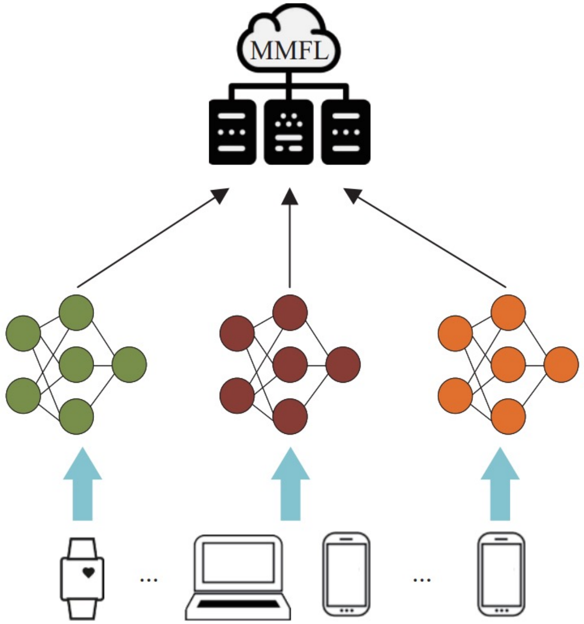


Centralized Multimodal Learning:

- Gather feature from multi-modality data to increase knowledge for model
- Missing modalities: Incomplete dataset, i.e., some of training/inference samples may miss some modalities

Federated Multimodal Learning:

- Collaboratively train with clients having different modality set
- Missing modalities: Deal with diverse set of modality, each set corresponding to a client



Missing modalities in those contexts is different

Problem Definition

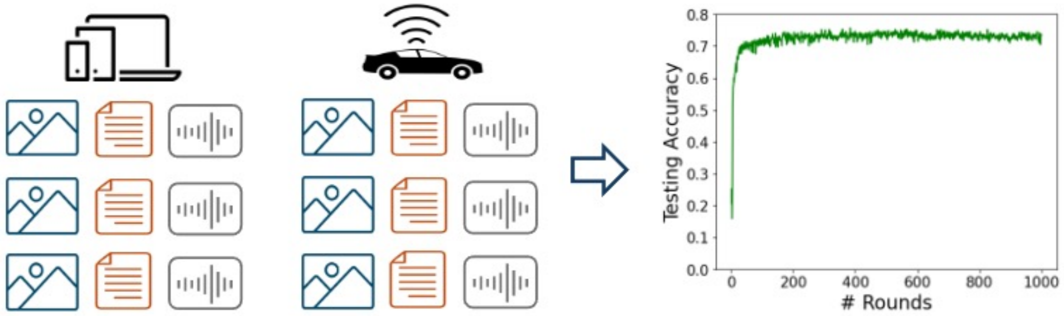
Missing Modalities in FL:

- Separate modality subset per client **(complete missing)**
- In practice, each client may have incomplete local dataset **(partial missing)**

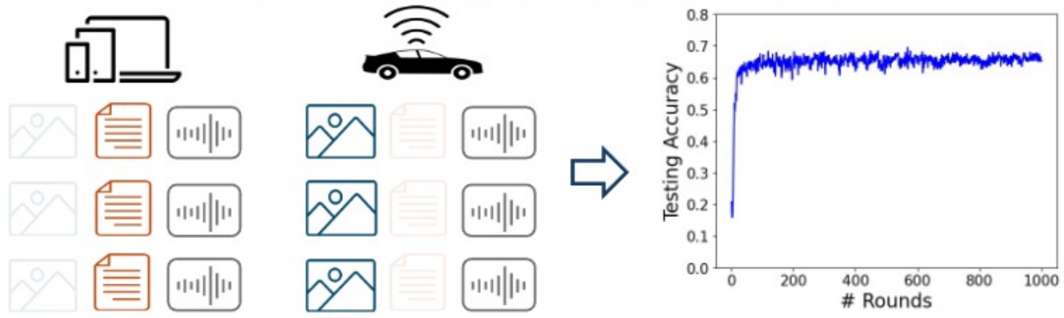
E.g. Top-tier hospitals register that they have 5 modern sensors, however, one or some of them may be in mainteinance

Motivation

- How to design robust algorithms ?
- How large the impact of partial missing can be ?



(a) Full-modality and its performance.



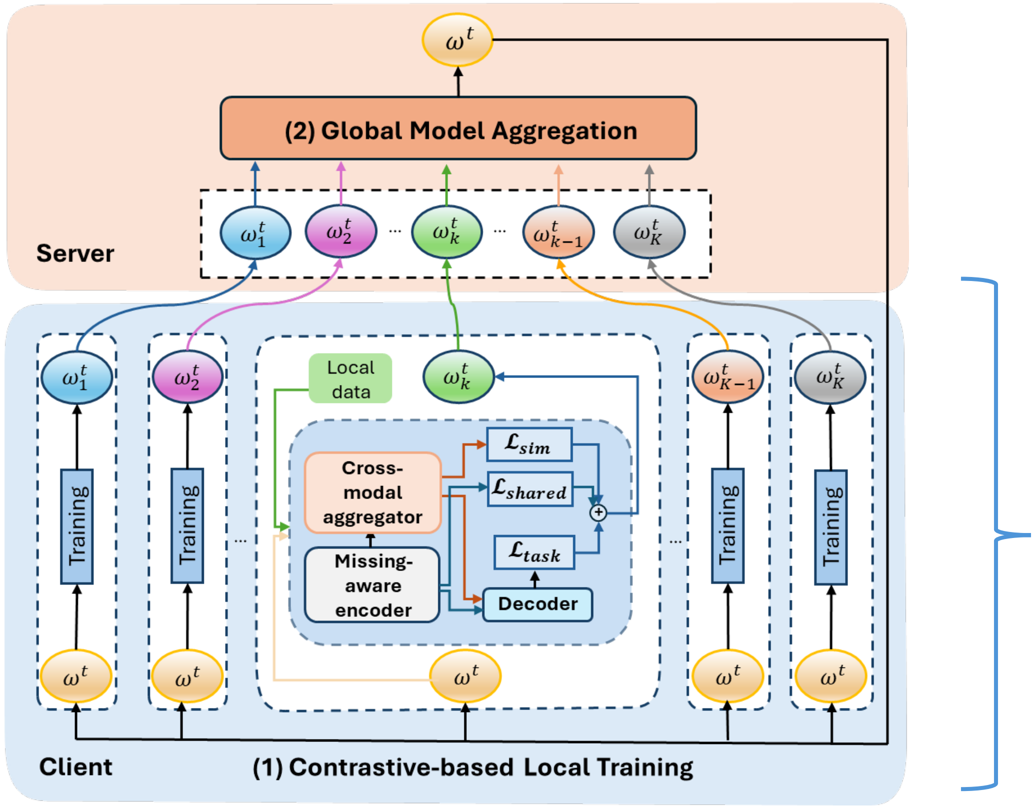
(b) Complete-modality missing and its performance.

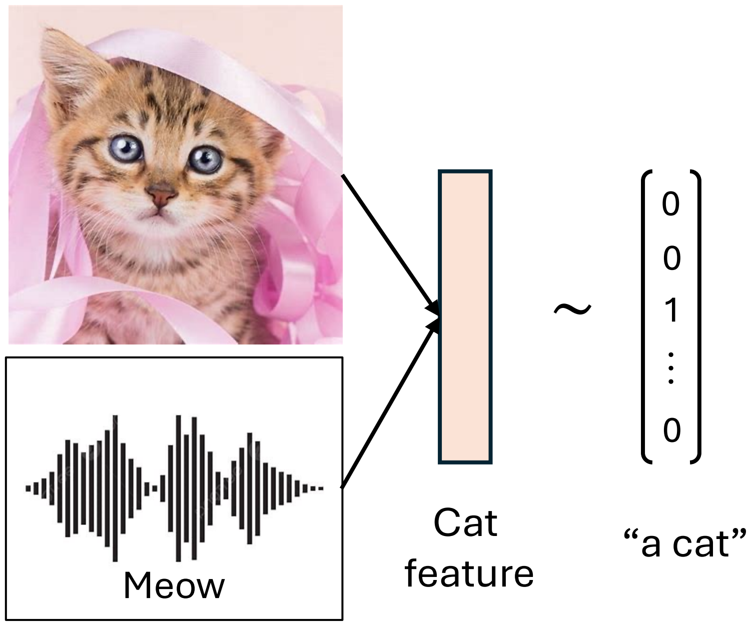
(c) Partial-modality missing and its performance.

Overview of FedMAC

Hypothesis:

- Modality ~ Augmented version of a core hidden feature





If the core feature is extracted completely, our model won't be affected by the modality subset of each instance

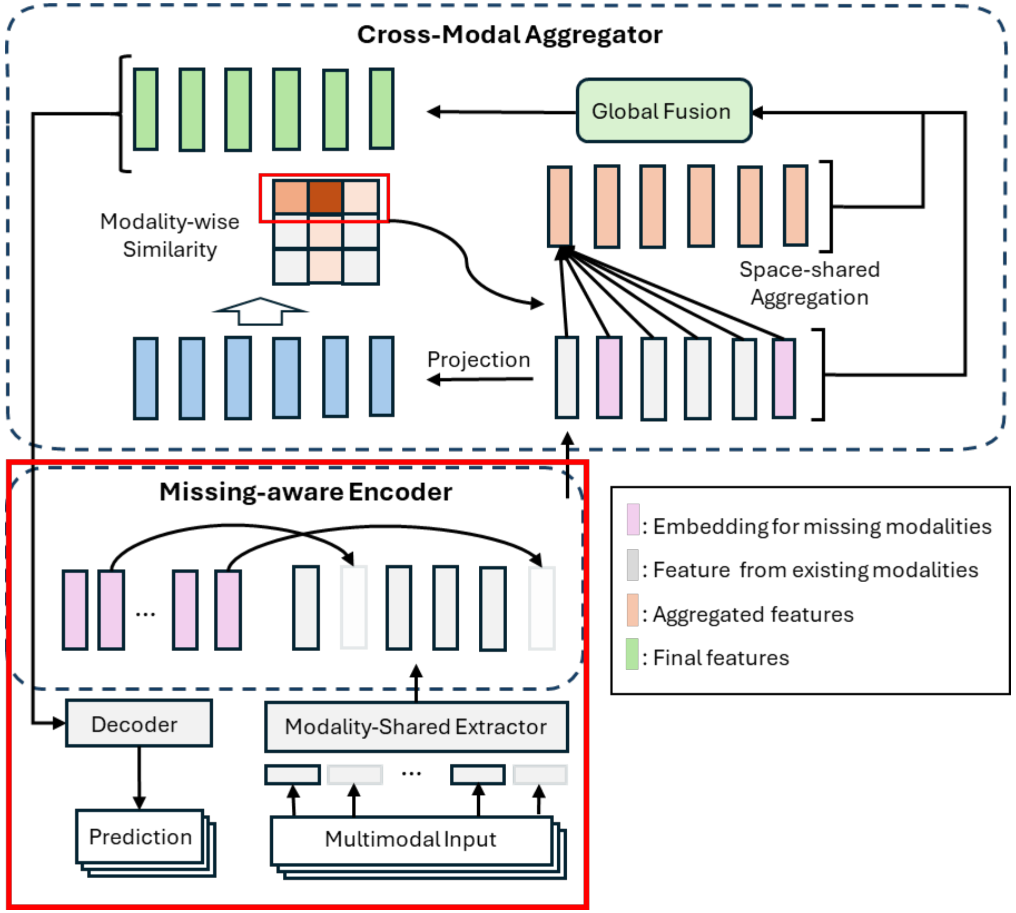
FedMAC: Client's Architecture

Missing-Aware Encoder

- Modality-Shared Extractor: Trying to extract shared core feature from different modalities
- Missed features are replaced by corresponding embeddings

However...

- Imputation embeddings are shared among instances
- Thus, these embeddings can not represent the core features of each instance completely



Necessity of **reformulating missing-modality features from existing ones, allowing for instance-level specifications**

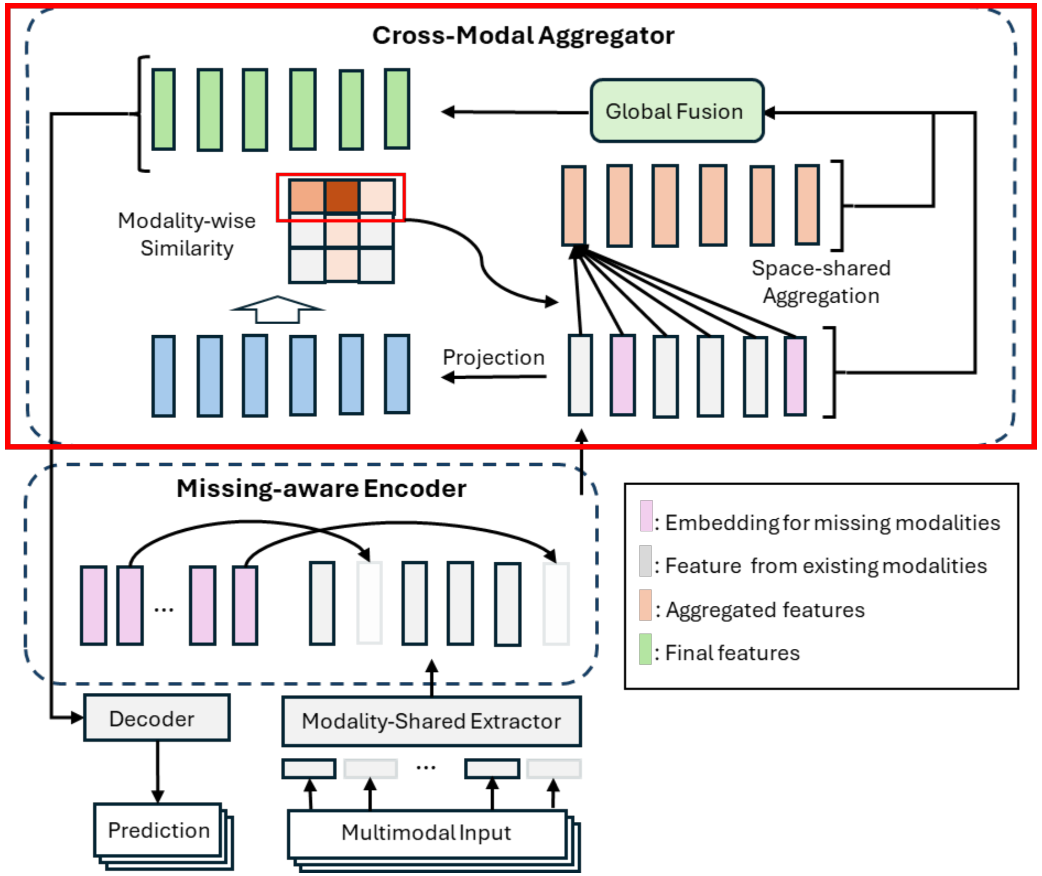
FedMAC: Client's Architecture

Cross-Modal Aggregator

- Maps space-shared features into a higher-level space to learn modality-wise similarities
- These similarities then guide a weighted aggregation back into the shared space

Global Fusion

- Benefits of aggregated features for existing-modality features might be less than missing ones
- A Global Fusion module dynamically fuses features before and after the aggregation enabling adaptive integration across modalities

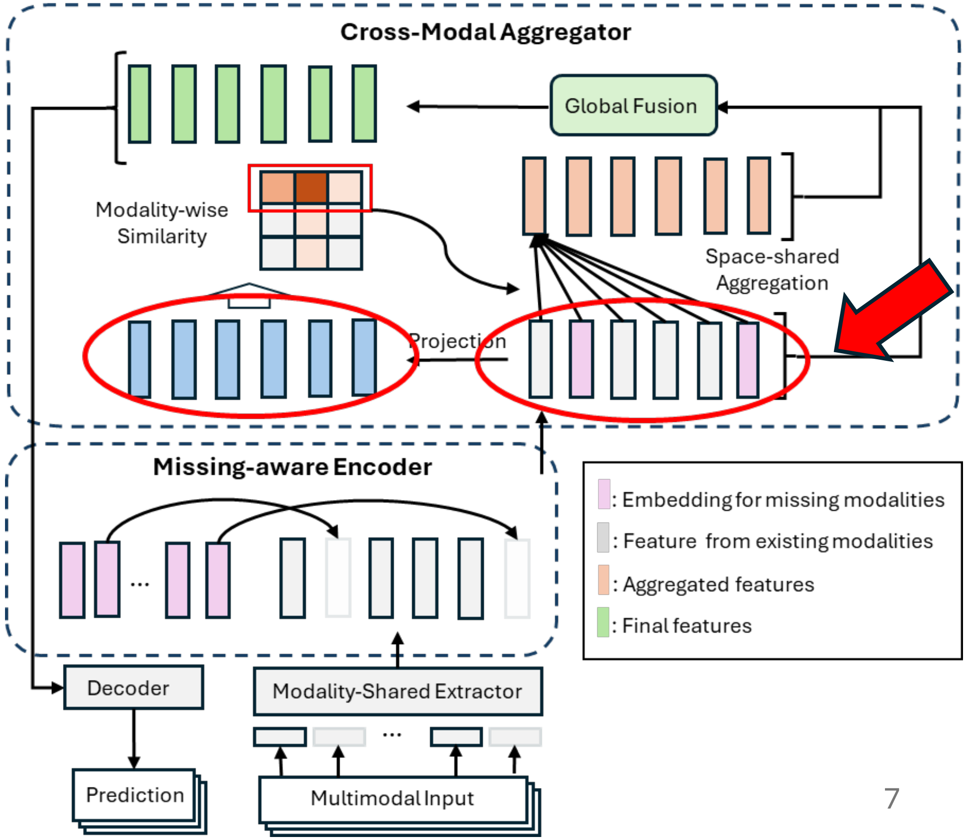


FedMAC: Contrastive-based Regularization

Space-shared Loss

- Ensure all modality-specific features belonging to a single instance are closely aligned in the shared space
- Positive pairs: modalities of the same instance, negative pairs for otherwise

$$\mathcal{L}_{shared} = -\sum_{h \in H} \sum_{h' \in H^{+}} \log \frac{\exp(R_H^{h,h'}/\tau)}{\sum_{h'' \in H/\{h\}} \exp(R_H^{h,h''}/\tau)},$$



FedMAC: Contrastive-based Regularization

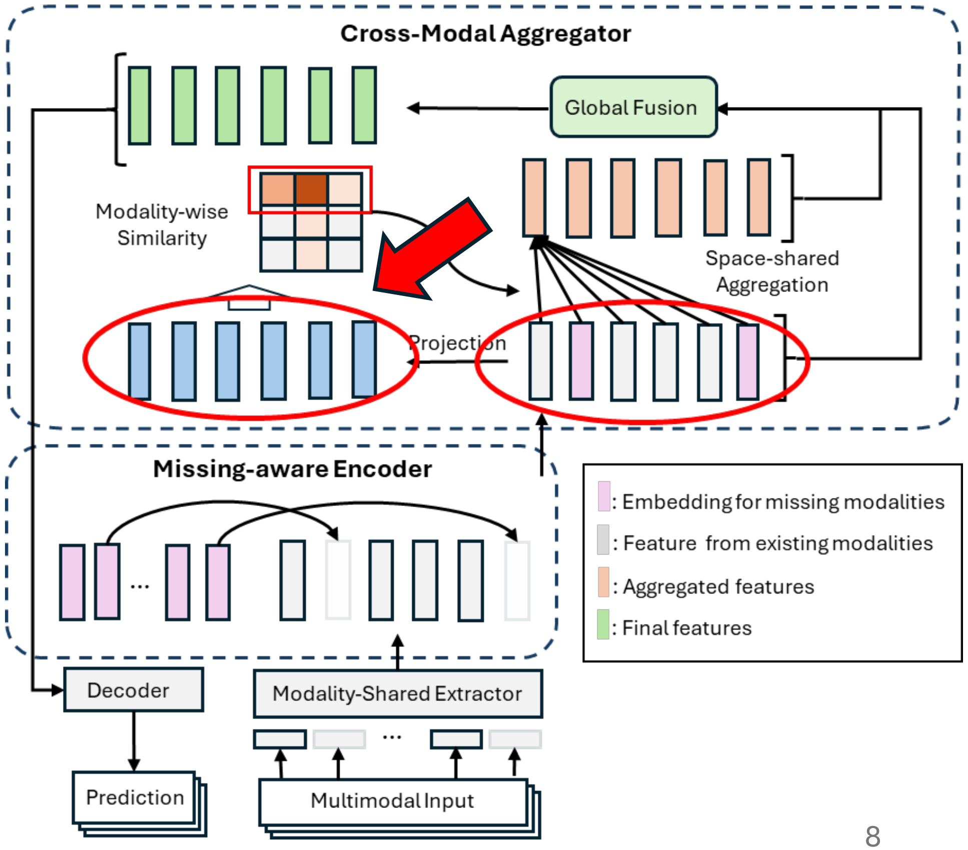
Modality-wise Similarity Loss

- Preserve accurate cross-modal relationships within the shared representation space
- Positive and negative pairs are defined similarly to space-shared loss, but in modality-wise similarity space

$$\mathcal{L}_{sim} = -\sum_{z \in Z} \sum_{z' \in Z^+} \log \frac{\exp(S_Z^{z,z'}/\tau)}{\sum_{z'' \in Z/\{z\}} \exp(S_Z^{z,z''}/\tau)},$$

$$\boxed{\mathcal{L} = \mathcal{L}_{task} + \lambda(\mathcal{L}_{shared} + \mathcal{L}_{sim}),}$$

Our final loss function



Experiment Settings

Dataset and Preparation

- A subset of PTB-XL dataset[1] with 3,963 clinical samples is used. Each sample has 12 modalities (electrocardiogram recordings) and is assigned to a single class

Partial Missing-Modality Data Simulation

- The missing pattern $(\mathrm{p}_\mathrm{m}^\mathrm{k}, \mathrm{p}_\mathrm{s}^\mathrm{k})$ is assigned for a client k with a dataset D_k and represented by a missing matrix:
  - $\mathrm{b}_\mathrm{i}^\mathrm{m} \in \{0, 1\}$ indicates whether modality m is missing or available for sample i
  - $\mathrm{M}_\mathrm{k}, \mathrm{N}_\mathrm{k}$ are the number of modalities and samples, respectively

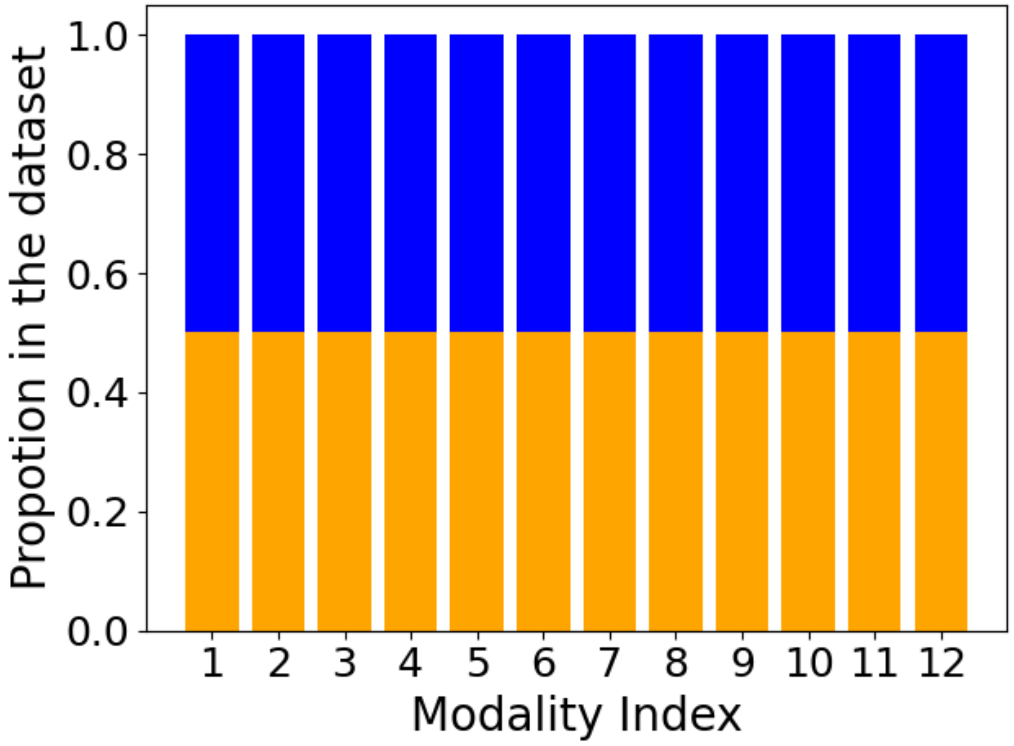
$$\psi(D_k) = \begin{bmatrix} b_1^1 & \dots & b_1^{M_k} \\ \vdots & \ddots & \vdots \\ b_{N_k}^1 & \dots & b_{N_k}^{M_k} \end{bmatrix}$$

- The partial missing-modality instances then can be obtained via following equation:

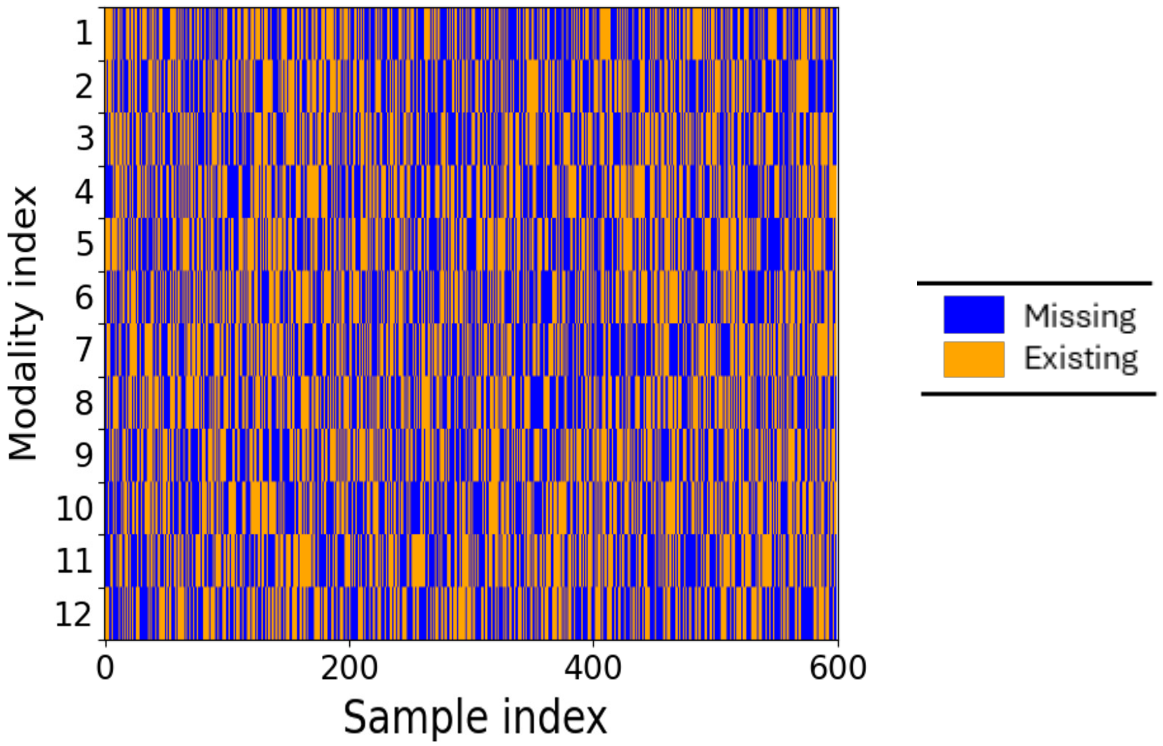
$$\hat{x}_i = [x_i^1, \dots, x_i^{M_k}] \odot [b_i^1, \dots, b_i^{M_k}],$$

[1] P. Wagner, N. Strodthoff, R.-D. Bousseljot, D. Kreiseler, F. I. Lunze,W. Samek, and T. Schaeffter, "PTB-XL, a large publicly available electrocardiography dataset," Scientific Data, vol. 7, no. 1, p. 154, 2020.

Experiment Settings



The modality proportion in the dataset



The modality distribution in each sample

Quantitative Results

TABLE I: Comparison of FedMAC with other methods under various missing statistics in the IID and Non-IID settings, where the missing statistics of the clients and server are *similar*. The best and second-best results are highlighted in **bold** and underlined, respectively.

Model	Missing statistics (p_m/p_s) in the IID setting						Missing statistics (p_m/p_s) in the Non-IID setting					
	1.0/0.1	1.0/0.5	1.0/0.8	0.5/0.5	0.8/0.5	0.8/0.8	1.0/0.1	1.0/0.5	1.0/0.8	0.5/0.5	0.8/0.5	0.8/0.8
FedProx	76.67	72.51	56.99	64.31	63.56	46.53	75.16	60.53	37.96	56.24	35.06	34.43
FedMSplit	72.51	68.10	51.32	56.62	60.40	35.56	71.43	58.13	36.47	57.88	57.38	36.70
MIFL	76.29	72.13	55.23	65.19	64.31	37.45	73.01	64.43	38.97	62.04	58.51	36.95
FedInMM	70.37	56.66	28.12	43.25	53.59	35.18	67.47	55.36	41.36	59.64	56.37	47.16
FedMAC	**77.42**	**76.16**	**66.96**	**75.03**	**75.28**	**72.27**	**76.29**	**74.40**	**64.82**	**72.38**	**73.39**	**67.59**

Quantitative Results

TABLE I: Comparison of FedMAC with other methods under various missing statistics in the IID and Non-IID settings, where the missing statistics of the clients and server are *similar*. The best and second-best results are highlighted in **bold** and underlined, respectively.

Model	Missing statistics (p_m/p_s) in the IID setting						Missing statistics (p_m/p_s) in the Non-IID setting					
	1.0/0.1	1.0/0.5	1.0/0.8	0.5/0.5	0.8/0.5	0.8/0.8	1.0/0.1	1.0/0.5	1.0/0.8	0.5/0.5	0.8/0.5	0.8/0.8
FedProx	76.67	72.51	56.99	64.31	63.56	46.53	75.16	60.53	37.96	56.24	35.06	34.43
FedMSplit	72.51	68.10	51.32	56.62	60.40	35.56	71.43	58.13	36.47	57.88	57.38	36.70
MIFL	76.29	72.13	55.23	65.19	64.31	37.45	73.01	64.43	38.97	62.04	58.51	36.95
FedInMM	70.37	56.66	28.12	43.25	53.59	35.18	67.47	55.36	41.36	59.64	56.37	47.16
FedMAC	**77.42**	**76.16**	**66.96**	**75.03**	**75.28**	**72.27**	**76.29**	**74.40**	**64.82**	**72.38**	**73.39**	**67.59**

FedMAC consistently outperforms other baselines in all tested missing scenarios, especially when the missing statistics increases, e.g., in 1.0/0.8

Quantitative Results

TABLE I: Comparison of FedMAC with other methods under various missing statistics in the IID and Non-IID settings, where the missing statistics of the clients and server are *similar*. The best and second-best results are highlighted in **bold** and underlined, respectively.

Model	Missing statistics (p_m/p_s) in the IID setting						Missing statistics (p_m/p_s) in the Non-IID setting					
	1.0/0.1	1.0/0.5	1.0/0.8	0.5/0.5	0.8/0.5	0.8/0.8	1.0/0.1	1.0/0.5	1.0/0.8	0.5/0.5	0.8/0.5	0.8/0.8
FedProx	76.67	72.51	56.99	64.31	63.56	46.53	75.16	60.53	37.96	56.24	35.06	34.43
FedMSplit	72.51	68.10	51.32	56.62	60.40	35.56	71.43	58.13	36.47	57.88	57.38	36.70
MIFL	76.29	72.13	55.23	65.19	64.31	37.45	73.01	64.43	38.97	62.04	58.51	36.95
FedInMM	70.37	56.66	28.12	43.25	53.59	35.18	67.47	55.36	41.36	59.64	56.37	47.16
FedMAC	**77.42**	**76.16**	**66.96**	**75.03**	**75.28**	**72.27**	**76.29**	**74.40**	**64.82**	**72.38**	**73.39**	**67.59**

While other methods' performance significantly degrades, FedMAC remains robust when missing statistic increases

Quantitative Results

TABLE II: Comparison of FedMAC with other methods under various missing statistics in the IID setting, where the missing statistics of the clients and server are *different*. The best and second-best results are highlighted in **bold** and underlined, respectively.

Client's missing statistics (p_m/p_s)	Model	Server's missing statistics (p_m/p_s)					
		1.0/0.1	1.0/0.5	1.0/0.8	0.5/0.95	0.8/0.95	0.1/0.95
1.0/0.1	FedProx	76.54	68.72	51.83	70.49	51.83	76.54
1.0/0.1	FedMSplit	72.76	66.33	52.33	69.99	59.27	72.89
1.0/0.1	MIFL	76.67	71.00	**53.85**	72.01	61.79	76.04
1.0/0.1	FedInMM	68.73	57.63	43.76	64.94	55.11	71.75
1.0/0.1	**FedMAC**	**77.93**	**72.51**	52.46	**75.53**	**65.83**	**77.30**
1.0/0.5	FedProx	**75.53**	72.13	60.40	72.76	65.83	75.03
1.0/0.5	FedMSplit	72.13	67.59	59.39	69.10	64.44	71.88
1.0/0.5	MIFL	74.65	71.63	60.78	71.25	66.08	72.76
1.0/0.5	FedInMM	63.05	57.50	45.77	57.88	54.48	64.56
1.0/0.5	**FedMAC**	75.16	**76.16**	**66.71**	**74.15**	**73.27**	**77.30**
1.0/0.8	FedProx	69.98	66.46	56.12	69.23	64.19	69.36
1.0/0.8	FedMSplit	65.32	60.91	49.31	65.32	57.63	66.83
1.0/0.8	MIFL	68.35	64.69	54.22	66.71	57.00	67.59
1.0/0.8	FedInMM	31.40	25.98	25.98	38.46	26.10	31.02
1.0/0.8	**FedMAC**	**75.28**	**74.02**	**66.96**	**74.91**	**74.27**	**75.28**

Quantitative Results

Baselines perform relatively well with lower missing statistics but struggle as the server's missing statistic increases

In contrast, FedMAC maintains strong performance, consistently achieving the best results across nearly all experiments

TABLE II: Comparison of FedMAC with other methods under various missing statistics in the IID setting, where the missing statistics of the clients and server are *different*. The best and second-best results are highlighted in **bold** and underlined, respectively.

Client's missing statistics (p_m/p_s)	Model	Server's missing statistics (p_m/p_s) 1.0/0.1	1.0/0.5	1.0/0.8	0.5/0.95	0.8/0.95	0.1/0.95
1.0/0.1	FedProx	76.54	68.72	51.83	70.49	51.83	76.54
	FedMSplit	72.76	66.33	52.33	69.99	59.27	72.89
	MIFL	76.67	71.00	**53.85**	72.01	61.79	76.04
	FedInMM	68.73	57.63	43.76	64.94	55.11	71.75
	FedMAC	**77.93**	**72.51**	52.46	**75.53**	**65.83**	**77.30**
1.0/0.5	FedProx	**75.53**	72.13	60.40	72.76	65.83	75.03
	FedMSplit	72.13	67.59	59.39	69.10	64.44	71.88
	MIFL	74.65	71.63	60.78	71.25	66.08	72.76
	FedInMM	63.05	57.50	45.77	57.88	54.48	64.56
	FedMAC	75.16	**76.16**	**66.71**	**74.15**	**73.27**	**77.30**
1.0/0.8	FedProx	69.98	66.46	56.12	69.23	64.19	69.36
	FedMSplit	65.32	60.91	49.31	65.32	57.63	66.83
	MIFL	68.35	64.69	54.22	66.71	57.00	67.59
	FedInMM	31.40	25.98	25.98	38.46	26.10	31.02
	FedMAC	**75.28**	**74.02**	**66.96**	**74.91**	**74.27**	**75.28**

Ablation Studies

We evaluate two key components of FedMAC in the IID setting, where the missing statistics of clients and server are similar

- **FedC:** FedMAC without Cross-**M**odal **A**ggregator
- **FedMA:** FedMAC without **C**ontrastive Regularization. We exclude $\mathcal{L}_{sim}$ and $\mathcal{L}_{shared}$ from the training loss

TABLE III: Ablation study on different components of FedMAC under various missing statistics in the IID setting, where the missing statistics of the clients and server are *similar*. The best and second-best results are highlighted in **bold** and underlined, respectively.

Model	Missing statistics (p_m/p_s)					
	1.0/0.1	1.0/0.5	1.0/0.8	0.5/0.5	0.8/0.5	0.8/0.8
FedC	**77.80**	<u>76.08</u>	**67.09**	<u>73.26</u>	<u>74.40</u>	<u>68.22</u>
FedMA	73.52	72.76	66.82	61.69	66.96	61.64
FedMAC	<u>77.42</u>	**76.16**	<u>66.96</u>	**75.03**	**75.28**	**72.26**

Ablation Studies

We evaluate two key components of FedMAC in the IID setting, where the missing statistics of clients and server are similar

- **FedC:** FedMAC without Cross-**M**odal **A**ggregator
- **FedMA:** FedMAC without **C**ontrastive Regularization. We exclude $\mathcal{L}_{sim}$ and $\mathcal{L}_{shared}$ from the training loss

TABLE III: Ablation study on different components of FedMAC under various missing statistics in the IID setting, where the missing statistics of the clients and server are *similar*. The best and second-best results are highlighted in **bold** and underlined, respectively.

Model	Missing statistics (p_m/p_s)					
	1.0/0.1	1.0/0.5	1.0/0.8	0.5/0.5	0.8/0.5	0.8/0.8
FedC	**77.80**	<u>76.08</u>	**67.09**	<u>73.26</u>	<u>74.40</u>	<u>68.22</u>
FedMA	73.52	72.76	66.82	61.69	66.96	61.64
FedMAC	<u>77.42</u>	**76.16**	<u>66.96</u>	**75.03**	**75.28**	**72.26**

FedC performs adequately when the missing modality pattern is uniform across clients ($p_m = 1$) but struggles when the pattern varies ($p_m < 1$)

Ablation Studies

We evaluate two key components of FedMAC in the IID setting, where the missing statistics of clients and server are similar

- **FedC:** FedMAC without Cross-**M**odal **A**ggregator
- **FedMA:** FedMAC without **C**ontrastive Regularization. We exclude $\mathcal{L}_{sim}$ and $\mathcal{L}_{shared}$ from the training loss

TABLE III: Ablation study on different components of FedMAC under various missing statistics in the IID setting, where the missing statistics of the clients and server are *similar*. The best and second-best results are highlighted in **bold** and underlined, respectively.

Model	Missing statistics (p_m/p_s)					
	1.0/0.1	1.0/0.5	1.0/0.8	0.5/0.5	0.8/0.5	0.8/0.8
FedC	**77.80**	76.08	**67.09**	73.26	74.40	68.22
FedMA	73.52	72.76	66.82	61.69	66.96	61.64
FedMAC	<u>77.42</u>	**76.16**	<u>66.96</u>	**75.03**	**75.28**	**72.26**

FedMA shows notable performance declines, particularly in scenarios with high client variation ($p_m < 1$), underscoring the role of contrastive mechanisms

Ablation Studies

We evaluate two key components of FedMAC in the IID setting, where the missing statistics of clients and server are similar

- **FedC:** FedMAC without Cross-**M**odal **A**ggregator
- **FedMA:** FedMAC without **C**ontrastive Regularization. We exclude $\mathcal{L}_{sim}$ and $\mathcal{L}_{shared}$ from the training loss

TABLE III: Ablation study on different components of FedMAC under various missing statistics in the IID setting, where the missing statistics of the clients and server are *similar*. The best and second-best results are highlighted in **bold** and underlined, respectively.

Model	Missing statistics (p_m/p_s)					
	1.0/0.1	1.0/0.5	1.0/0.8	0.5/0.5	0.8/0.5	0.8/0.8
FedC	**77.80**	<u>76.08</u>	**67.09**	<u>73.26</u>	<u>74.40</u>	<u>68.22</u>
FedMA	73.52	72.76	66.82	61.69	66.96	61.64
FedMAC	<u>77.42</u>	**76.16**	<u>66.96</u>	**75.03**	**75.28**	**72.26**

FedMAC outperforms FedC and FedMA in most cases. Overall, the both Cross-Modal Aggregation and Contrastive Regularization in FedMAC is crucial for robust and consistent performance

Conclusions

Our contributions:

- The first to thoroughly investigate the problem of partially missing modalities in FL
- Introduce FedMAC, an innovative client-side architecture designed to handle complete and partial modality missing issues

Results:

- Consistently outperforms baselines with a large margin in every testing scenarios when the missing statistics of server and clients are similar
- Outpeforms existing methods in most cases when the missing statistics of server and clients are different

Future works:

- Conducting more extensive experiments including different datasets and varying missing scenarios
- Design an appropriate server aggregation algorithm which can handle partial missing modality under Non-IID settings