聯邦時間序列補值異質性 MNAR 缺失之研究
Federated Clinical Time-Series Imputation under Heterogeneous MNAR (Missing Not At Random)

專題成員：曾令城, C-17
指導教授：莊坤達
National Cheng Kung University

專題背景和動機

- **想解決什麼問題？**

  如何使用聯邦學習，提升每個客戶（醫療場所）的缺失值（生命體徵）補值準確度

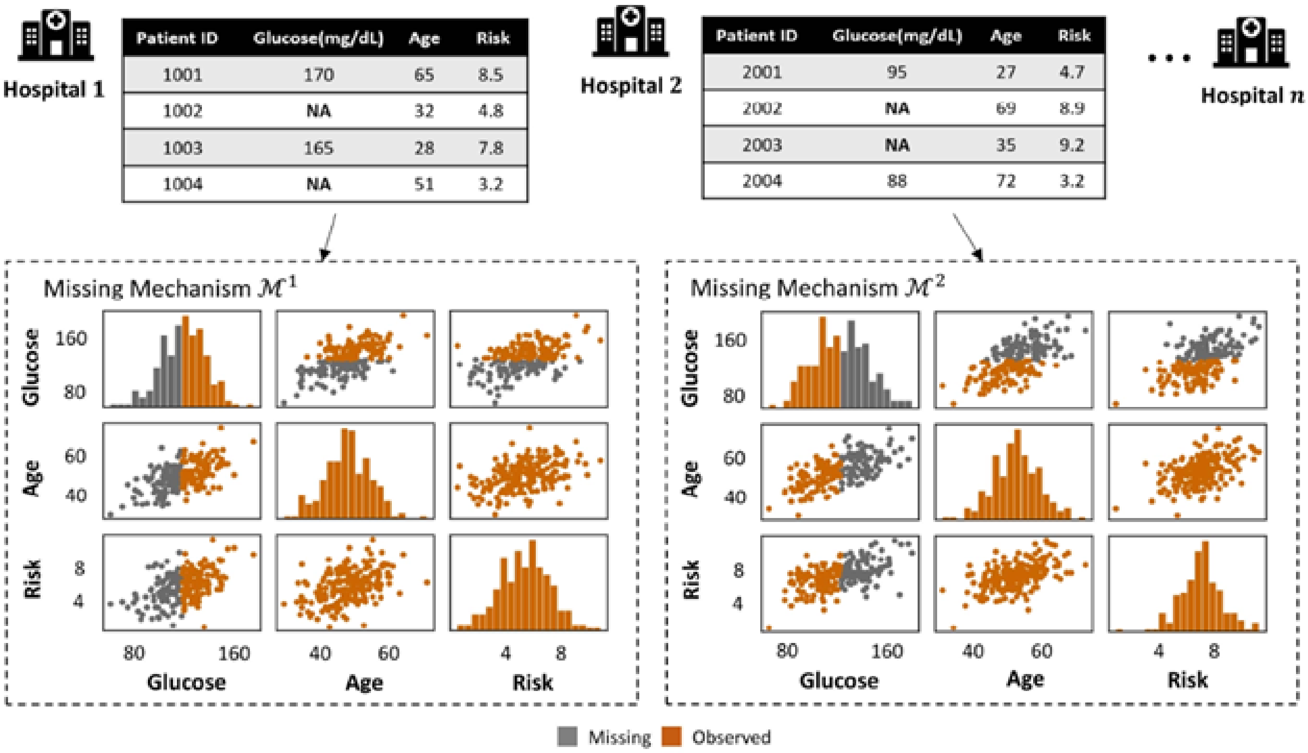
  。

- **為什麼？**

  1. 生命徵象是臨床決策與病患監測的核心依據
  2. 資料缺失影響後續分析、機器學習效能

ID	Time	HR	RR	Temp	SBP	...
0	00:00	NaN	17.2	36.5	NaN	...
1	08:00	94.5	17.4	NaN	124.5	...
2	16:00	96.5	NaN	NaN	124	...
⋮	⋮	⋮	⋮	⋮	⋮	...

研究方向：醫院間的缺失資料有互補性



Hospital 1

Patient ID	Glucose(mg/dL)	Age	Risk
1001	170	65	8.5
1002	NA	32	4.8
1003	165	28	7.8
1004	NA	51	3.2

Hospital 2

Patient ID	Glucose(mg/dL)	Age	Risk
2001	95	27	4.7
2002	NA	69	8.9
2003	NA	35	9.2
2004	88	72	3.2

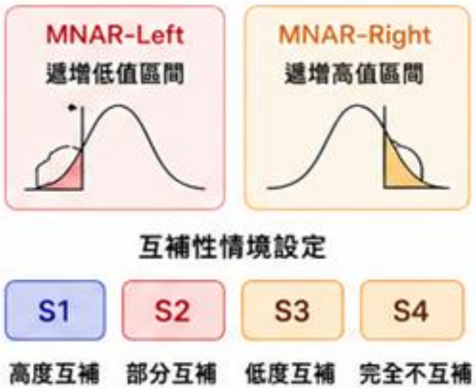


Fig. 1. Min, S., Wang, X., Asif, H., & Vaidya, J. (2025).CAFE: Improved Federated Data Imputation by Leveraging Missing Data Heterogeneity. IEEE TKDE

研究方法-計算互補性流程

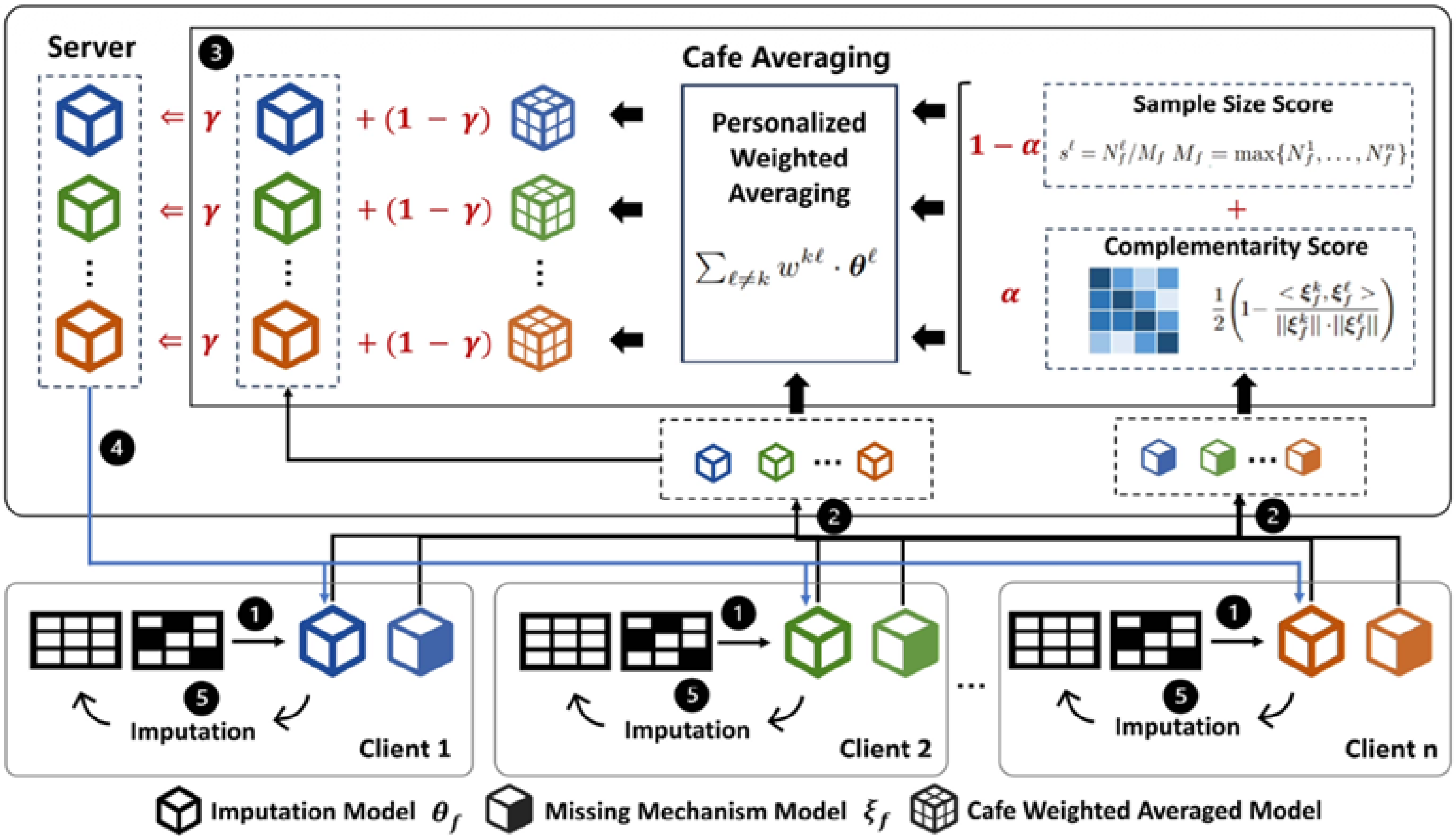
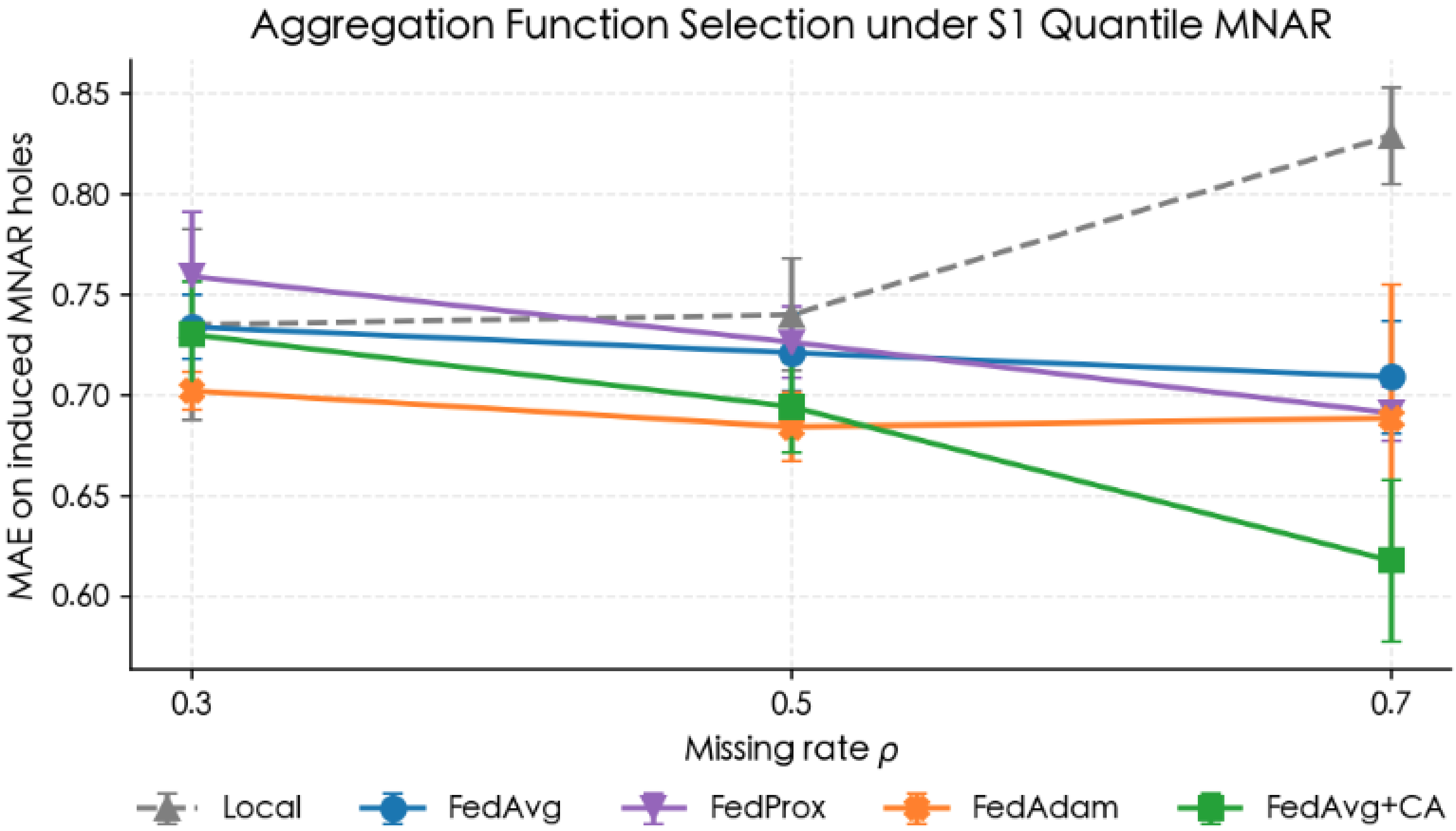


Fig. 2. Min, S., Wang, X., Asif, H., & Vaidya, J. (2025).CAFE: Improved Federated Data Imputation by Leveraging Missing Data Heterogeneity. IEEE TKDE

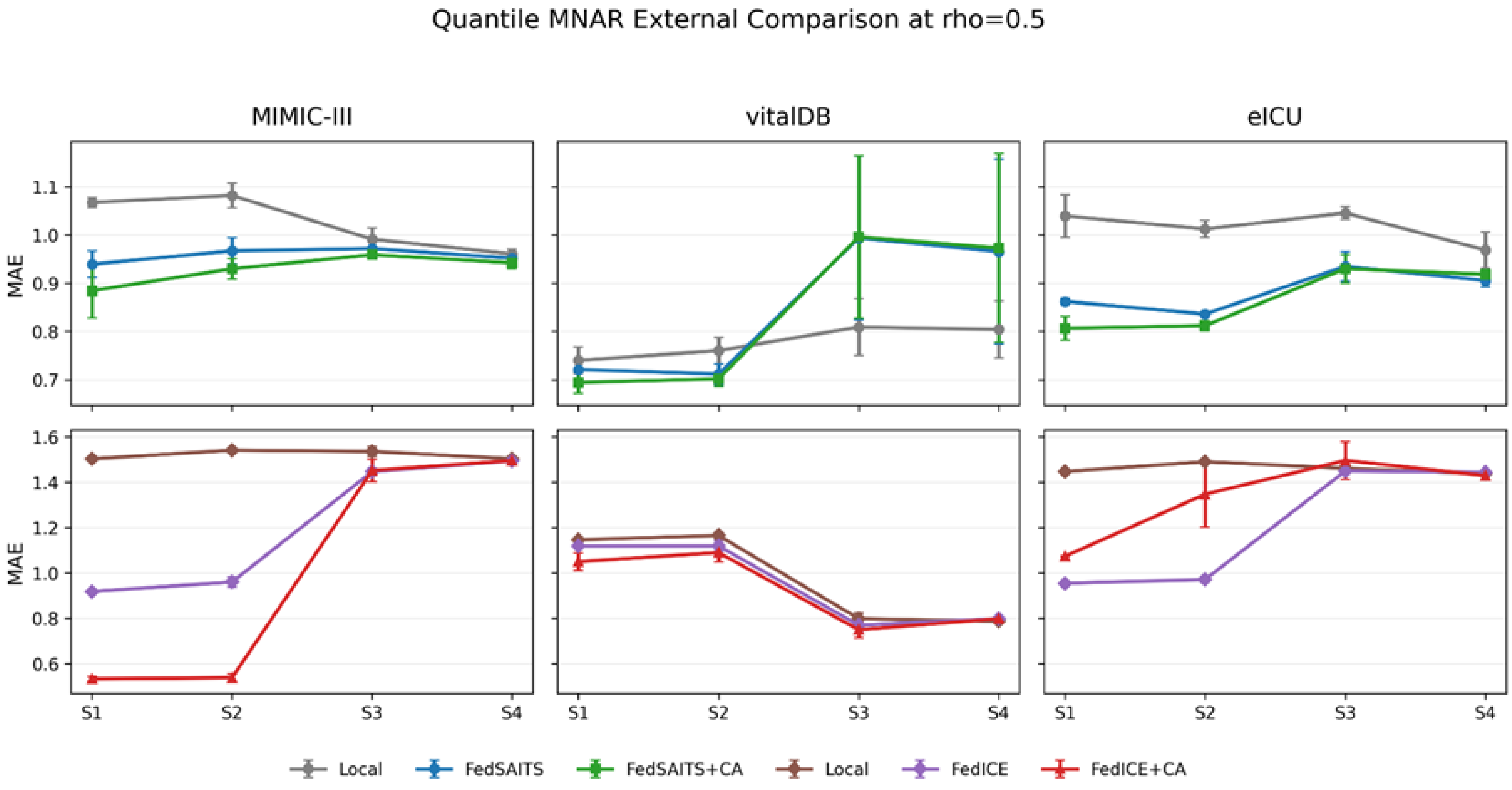
實驗設定：

1. 資料來源：VitalDB (首爾大學醫院, 1 min/step, 5 clients), MIMIC-III (ICU, 1 hr/step, 3 clients), eICU-demo (PhysioNet 2.0.1, 5 min/step, 5 hospital clusters)
2. 模型: 統計 ICE (Iterative Conditional Expectation), 深度學習：SAITS (Self-Attention-based Imputation for Time Series)
3. 缺失比率：$\rho \in \{0.3, 0.5, 0.7\}$
4. 互補情況：MNAR scenarios(S1, S2, S3, S4)
5. 訓練方法：Local training (baseline) vs. federated aggregation (FedAvg, FedProx, FedAdam, FedAvg+CA)
6. 評估指標: MAE (Mean Absolute Error) , F1-score, AUROC

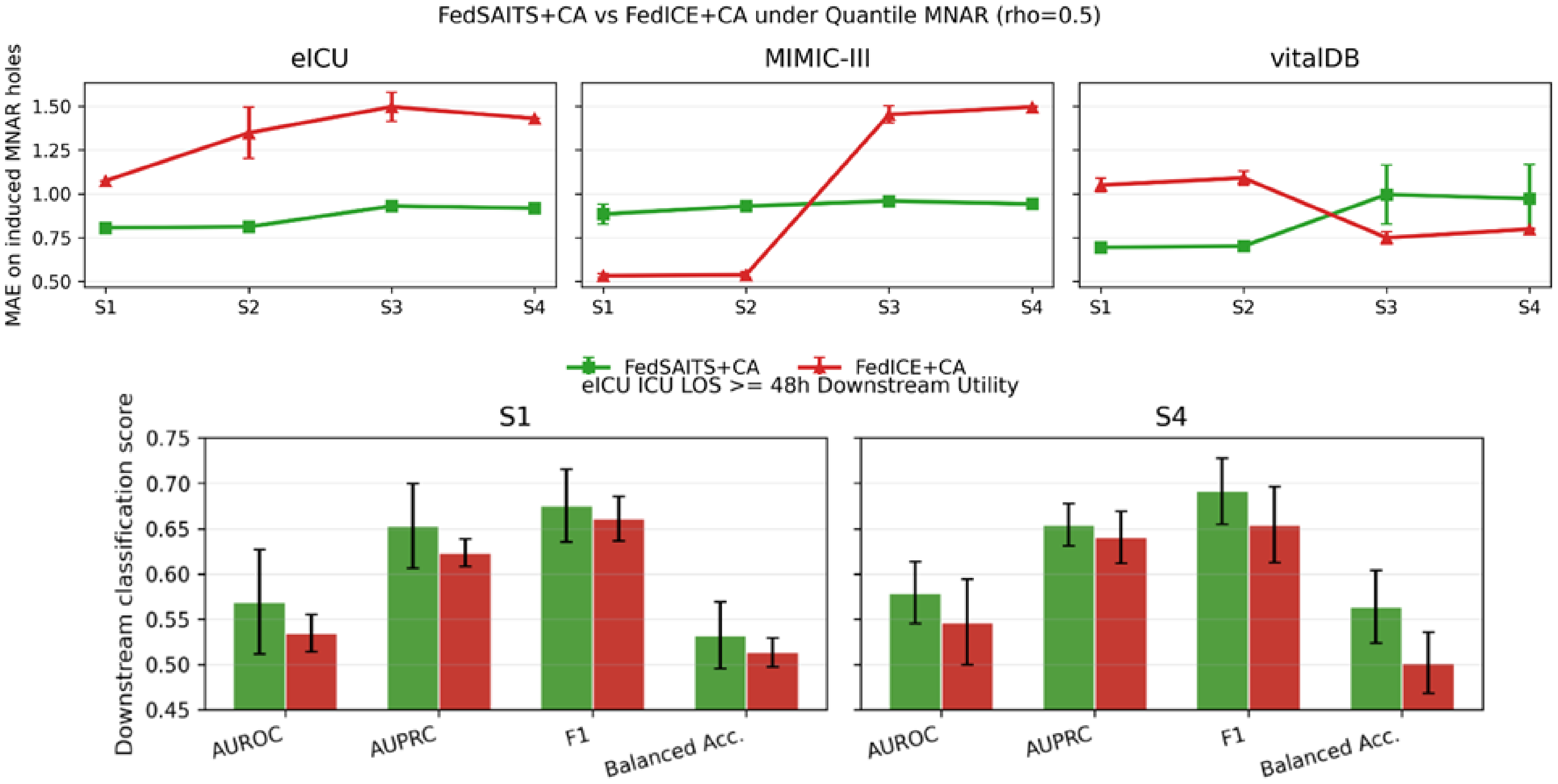
結果一: 聯邦學習可有效降低補值誤差，且缺失率越高效益越明顯



結果二：缺失互補性可作為有效聯邦學習聚合訊號



結果三：異質性MNAR 下不存在Universally Best Imputation Backbone



結論與未來展望

- **結論：**
  在異質性MNAR場景下，互補性聚合可提升補值準確度，沒有單一最佳的聯邦時間序列補值模型

- **未來展望：**
  1)發展更適合SAITS時間序列補值聯邦聚合方法
  2)建立具時間感知能力的 Missingness Mechanism

參考資料

1. Tsai, W.-C., Chung, Y.-T., Tseng, L.-C., Chen, P.-F., Chen, P.-L., Yeh, C.-Y., Ting, L. P.-Y., & Chuang, K.-T. (2025). 基於深度神經網路的臨床生命徵象時間序列補值. UHIMA2025 & TLCMA2025, NSYSU.
2. Du, W., Côté, D., & Liu, Y. (2023). SAITS: Self-Attention-based Imputation for Time Series. Expert Systems with Applications, 219, 119619.
3. Du, W., et al. (2025). TSI-Bench: Benchmarking Time Series Imputation. ICLR 2025 Workshop / Benchmark Paper.
4. Rubin, D. B. (1976). Inference and missing data. Biometrika, 63(3), 581–592.
5. Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. Journal of Statistical Software, 45(3), 1–67.
6. Min, S., Wang, X., Asif, H., & Vaidya, J. (2025). Cafe: Improved federated data imputation by leveraging missing data heterogeneity. IEEE Transactions on Knowledge and Data Engineering.
7. Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L.-W. H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L. A., & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. Scientific Data, 3, 160035.
8. Pollard, T. J., Johnson, A. E. W., Raffa, J. D., Celi, L. A., Mark, R. G., & Badawi, O. (2018). The eICU Collaborative Research Database, a freely available multi-center database for critical care research. Scientific Data, 5, 180178.
9. Lee, H.-C., & Jung, C.-W. (2018). VitalDB: A high-fidelity multi-parameter vital signs database in surgical patients. Scientific Data, 5, 180077.

謝謝聆聽