Modeling and Analyzing Non-IID Data in Federated Learning based ECG Arrhythmia Detection Scenarios

Davide Cantoro[1,†], Angela-Tafadzwa Shumba[1,2,†], Gianluigi Semeraro[1,†], Mattia Cotardo[1,†], Davide Rollo[1,†], Teodoro Montanaro[1,†], Ilaria Sergi[1,†], Massimo De Vittorio[2,†] and Luigi Patrono[1,*,†]

[1] *Dept. of Engineering for Innovation - University of Salento - Lecce, Italy*
[2] *Istituto Italiano di Tecnologia - Center for Biomolecular Nanotechnologies - Arnesano, Lecce, Italy*

Abstract

Federated Learning (FL) enables the training of Artificial Intelligence (AI) models directly on edge devices, such as those used in Internet of Medical Things (IoMT) scenarios, without transferring sensitive patient data to centralized servers, ensuring compliance with healthcare privacy regulations. However, the inherently non-Independent and Identically Distributed (non-IID) nature of data generated by IoMT devices poses significant challenges to effective model training, including delayed or unstable convergence. This study evaluates a methodology for generating realistic non-IID datasets from existing centralized healthcare datasets to support the evaluation of FL strategies under real-world heterogeneity. Using the FedArtML dataset generation tool, we simulate varying degrees of data distribution skew through advanced statistical partitioning techniques. This enables controlled experimentation and benchmarking of FL performance in healthcare scenarios, such as electrocardiogram (ECG) arrhythmia detection. In a use case based on the MIT-BIH Arrhythmia dataset, we assess the effects of different non-IID conditions on model accuracy and computational workload across clients. While accuracy remains stable (with minimal degradation), extreme non-IID settings lead to substantial variability in training times. These findings demonstrate that controlled dataset generation using FedArtML enables realistic FL evaluations and provides insights into the operational challenges of deploying FL in clinical environments.

Keywords

IoT, Federated Learning, Non-IID data, AI, Machine Learning, Arrhythmia

1. Introduction

The transition from Healthcare 4.0 to Healthcare 5.0 marks a paradigm shift toward systems that support hyper-personalized, predictive, and collaborative care models [1]. In this context, the Internet of Things (IoT) continues to serve as a foundational infrastructure, facilitating real-time physiological monitoring, automated clinical decision support, and the integration of heterogeneous health data sources [2, 3]. As patient engagement and autonomy increase in Healthcare 5.0, there is a corresponding emphasis on data sovereignty, privacy-preserving computation, and decentralized intelligence [4]. These priorities are particularly critical in use cases such as electrocardiogram (ECG) arrhythmia detection, where sensitive biometric signals are acquired and analyzed directly on edge devices, often in resource-constrained or privacy-sensitive environments [5, 6].

Traditional data-driven approaches to ECG arrhythmia detection typically involve transmitting raw or preprocessed signal data to centralized computing infrastructures, where computationally intensive machine learning (ML) models are trained and deployed [7, 8]. While this paradigm can deliver high

Ital-IA 2025: 5th National Conference on Artificial Intelligence, organized by CINI, June 23-24, 2025, Trieste, Italy

*Corresponding author.
†These authors contributed equally.
davide.cantoro@unisalento.it (D. Cantoro); angela.shumba@unisalento.it (A. Shumba); gianluigi.semeraro@studenti.unisalento.it (G. Semeraro); mattia.cotardo@studenti.unisalento.it (M. Cotardo); davide.rollo1@studenti.unisalento.it (D. Rollo); teodoro.montanaro@unisalento.it (T. Montanaro); ilaria.sergi@unisalento.it (I. Sergi); massimo.devittorio@iit.it (M. De Vittorio); luigi.patrono@unisalento.it (L. Patrono)
0009-0004-6903-2198 (D. Cantoro); 0000-0002-5888-7180 (A. Shumba); 0009-0006-4672-1278 (G. Semeraro); 0009-0005-2103-9080 (M. Cotardo); 0009-0006-2148-4333 (D. Rollo); 0000-0003-1750-8268 (T. Montanaro); 0000-0002-3797-3039 (I. Sergi); 0000-0003-1601-6392 (M. De Vittorio); 0000-0002-8591-1190 (L. Patrono)

© 2025 Creative Commons License Attribution 4.0 International (CC BY 4.0)

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073
published 2025-12-07

CEUR-WS.org/Vol-4121/Ital-IA_2025_paper_56.pdf

predictive accuracy, it is subject to several critical limitations. The reliance on stable, high-bandwidth communication links for continuous data offloading imposes scalability constraints, particularly in bandwidth-limited or latency-sensitive environments. Moreover, the centralized aggregation of sensitive physiological data introduces significant privacy and security risks, which are increasingly unacceptable in modern, patient-centric healthcare systems.

Federated Learning (FL) has emerged as an alternative innovative technology to address these limitations. It enables model training directly on edge devices, aggregating model updates, rather than raw data, into a global model that is periodically redistributed for continued local training and inference [9]. This approach enhances privacy, scalability, and system responsiveness, aligning well with the operational and ethical goals of Healthcare 5.0. In traditional distributed ML, data is typically assumed to be independent and identically distributed (IID) across clients. However, in real-world Internet of Medical Things (IoMT) deployments, such as ECG-based clinical decision support systems, this assumption rarely holds. Clients typically hold data from individual patients, resulting in non-IID, imbalanced, and heterogeneous datasets, which can significantly degrade FL performance, particularly in terms of model convergence and training stability [10].

To investigate these challenges, this study adopts FedArtML [11], a recently introduced open-source tool for generating synthetic non-IID datasets from centralized sources with systematic control over distributional heterogeneity. We applied FedArtML to multiple publicly available healthcare datasets to explore its versatility in simulating diverse non-IID scenarios. We particularly focused on the MIT-BIH Arrhythmia dataset to assess the impact of different levels and types of data skew on model accuracy and computational workload distribution across clients in an FL scenario. This work demonstrates the general applicability of FedArtML for structured dataset preparation and highlights its practical value in developing and benchmarking FL strategies for IoT-enabled, privacy-preserving clinical environments.

2. Related Works

Due to the highly sensitive and inherently decentralized nature of data in ECG-based arrhythmia detection IoMT scenarios, many researchers have begun investigating the application of Federated Learning in such scenarios [12, 13, 14, 15]. However, the challenges posed by employing distributed non-IID datasets in FL scenarios have been addressed in a few studies [16, 17].

One of the major challenges is that of non-Independent and Identically Distributed (non-IID) data, which significantly impacts communication efficiency, model training, convergence, and overall performance [18]. The literature categorizes the various types of data heterogeneity into various classes, including, feature distribution skew, label distribution skew, feature space skew, label space skew, and quantity skew, reflecting how different clients may exhibit distinct feature or label distributions, or even entirely disjoint sets of features or labels [19].

Frameworks, such as FedCurv and asynchronous FL methods [20, 17] have been proposed to address the Non-IID challenges in Federated Learning. FedCurv [20] applies regularization penalties to models whose updates deviate excessively from earlier parameter values that previously led to performance improvements. In contrast, Asynchronous Federated Learning techniques are used to tackle training speed and convergence challenges in Non-IID scenarios by allowing clients to update the global model without waiting for all participants to synchronize [17].

Given the limited attention to distributed non-IID ECG data in IoMT-based arrhythmia detection, this work employs FedArtML to systematically evaluate how varying degrees of data heterogeneity affect model accuracy and computational load distribution in an IoT edge Federated Learning scenario.

3. Experimental Setup

The experimental platform, shown in Fig. 1, is a fully containerized system that ensures reproducibility and scalability across IoT edge devices. All services run in Docker containers and are orchestrated via Kubernetes. The codebase is Python-based, leveraging Flower [21] for the Federated Learning workflow, GitHub (with GitHub Actions) for CI/CD, and Mlflow [22] for experiment tracking. FedArtML handles Non-IID dataset generation [11].

The setup is based on a client-server architecture, with a Dell PowerEdge T350 used as a server and Four Raspberry Pi 4 devices acting as IoT-edge clients. The Federated Learning configuration employs FedAVG [10], a well-known Federated Learning Algorithm, over 10 communication rounds.

An important step in the experimental setup is the data preprocessing phase, which transforms an originally IID dataset into multiple Non-IID partitions. This work uses a publicly available tabular ECG dataset on Kaggle [23]. The dataset contains data extracted from various well-known datasets such as the MIT-BIH Supraventricular Arrhythmia Database [24] and the MIT-BIH Arrhythmia Database [25], for both Lead II and Lead V.

To balance class proportions and reduce computational load on IoT edge devices, each of the lead-specific datasets was undersampled. Subsequently, FedArtML [11] was employed to induce and regulate varying degrees of data heterogeneity. Three partitioning schemes (IID, Moderate Non-IID, and Strong Non-IID) were generated for each dataset, producing four client-specific subsets per scheme. Data skew was controlled via the setup parameter α, and inter-client divergence was quantified using the Hellinger Distance (HD), in (1).

4. Results and Discussion

The tabulated Lead II and Lead V features from the MIT-BIH Supraventricular Arrhythmia Database [24] and the MIT-BIH Arrhythmia Databases [25] were used to convert the datasets into Non-IID partitions divisible among four clients, leveraging FedArtML [11]. Three Non-IID configurations were generated for each of the source datasets: IID, Moderate Non-IID, and Extreme Non-IID. These configurations are summarized in Table 1. The FL experiments were carried out using the Lead II partitions from the MIT-BIH Supraventricular database, which exhibited a clear, monotonic heterogeneity gradient. As α decreases from 1000 to 10 to 3.162, HD increases smoothly from 0.01 (IID) to 0.18 (Moderate Non-IID) to 0.34 (Strong Non-IID). Other leads exhibit abrupt HD changes (e.g., Lead V's HD jumps from 0.13 to 0.41) or display a narrower range of HD values (e.g., Arrhythmia Lead V spans only 0.01→0.13→0.23). Such patterns can obscure intermediate effects.

Following FL implementation, the IID scenario achieved the highest accuracy at 90.3663%, compared to 90.0606% and 90.2165% observed in the Moderate Non-IID and Strongly Non-IID configurations, respectively (see Figure 2). Although the Strongly Non-IID configuration gradually converges toward the IID baseline over time, Figure 2 highlights that its elevated data heterogeneity induces irregular training dynamics, with pronounced fluctuations in per-round accuracy.

More notable patterns were observed in training time variations as shown in Table 2. In the Strongly Non-IID scenario, high data heterogeneity produces marked per-client execution time differences of approximately 50% (from 33.6 to 66.0 minutes). The Moderate Non-IID scenario exhibits reduced variability of approximately 15.5% (34.2 to 53.3 minutes), while the IID configuration shows minimal variation, with client execution times tightly varying between 44.7 and 45.9 minutes.

5. Conclusion

This study demonstrates how controlled generation of Non-IID datasets from initially IID ECG collections is achievable for Federated Learning on real IoT-edge devices, specifically in arrhythmia detection scenarios. Leveraging FedArtML enables the creation of reproducible, quantifiable Non-IID partitions to systematically evaluate the system behavior under different heterogeneity levels. Performance metrics reveal that Non-IID conditions drop only by 0.31 % accuracy compared to IID, highlighting the robustness of the FL approach. However, extreme Non-IID settings exhibit local divergence that needs several federation rounds to stabilize, indicating a need for stabilization or personalization strategies. Additionally, strong data heterogeneity causes pronounced disparities in per-client execution times, up to 50 % (33.6 min vs. 66.0 min), compared to approximately 15.5 % variability in moderate Non-IID and

2.7 % in the IID scenario. Significant execution time variability suggests the need for more adaptive aggregation and load-balancing schemes; future work will investigate adaptive methods designed for heterogeneous computational loads.

Acknowledgments

This study was partially funded by the Italian Ministry of Health, Italian Health Operational Plan (Cohesion and Development Fund 2014-2020), trajectory 2 "eHealth, advanced diagnostics, medical device and mini invasiveness", project "Sistema di Monitoraggio ed Analisi basato su intelligenza aRTificiale per pazienti affetti da scompenso CARdiaco cronico con dispositivi medici mini-invasivi e indossabili Evoluti – SMART CARE" (CUP F83C22001380006), by the European Union - Next Generation EU, PRIN 2022 PNRR call, under the project "Interactive digital twin solutions for cardiovascular disease Management, PReventiOn and treatment leVeraging the internet of things and Edge intelligence paradigms - IMPROVE (CUP F53D23009250001), and by the project FAIR - Future AI Research (PE00000013), under the NRRP MUR program funded by the NextGenerationEU, and specifically under the project "A feDerAted learning aPproach To ImproVe Edge-based AI in iot-aware pervasive environments - ADAPTIVE-AI" (CUP H97G22000210007). This research is part of the Net Service Project "CLEAR - Cybersecurity Laboratory for EdgeComputing and Advanced Research", funded by the European Union, NextGenerationEU, PE00000014, CUP J33C22002810001 with Project SPOKE 8 – SERICS "SECURITY AND RIGHTS IN THE CYBERSPACE" - "RISK MANAGEMENT AND GOVERNANCE". Furthermore, the authors wish to thank the "MUR, Ministero dell'Universit'a e della Ricerca" for its financial support to the project "IDAS – Innovazione Digitale in Ambito Salute" through the Italian Ministerial Decree "Decreto Ministeriale n. 737 - 25-06-2021 - Criteri di riparto e utilizzazione del Fondo per la promozione e lo sviluppo delle politiche del Programma Nazionale per la Ricerca (PNR)" - CUP: F84D22000270001. Finally, authors would like to thank NetService for its funding support of Davide Cantoro.

Declaration on Generative AI

During the preparation of this work, the author(s) used ChatGPT, Grammarly in order to: Grammar and spelling check, Paraphrase and reword. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] J. Basulo-Ribeiro, L. Teixeira, The future of healthcare with industry 5.0: Preliminary interview-based qualitative analysis, Future Internet 16 (2024) 68. doi:10.3390/fi16030068.

[2] R. S. Delsi, B. Amutha, IoT revolutionizing healthcare: A survey of smart healthcare system architectures, in: 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), IEEE, 2023. doi:10.1109/rmkmate59243.2023.10369980.

[3] A. T. Shumba, S. M. Demir, V. M. Mastronardi, F. Rizzi, G. De Marzo, L. Fachechi, P. M. Ros, D. Demarchi, L. Patrono, M. De Vittorio, Monitoring cardiovascular physiology using bio-compatible aln piezoelectric skin sensors, IEEE Access 12 (2024) 16951–16962. doi:10.1109/ACCESS.2024.3359058.

[4] D. Saraswat, P. Bhattacharya, A. Verma, V. K. Prasad, S. Tanwar, G. Sharma, P. N. Bokoro, R. Sharma, Explainable ai for healthcare 5.0: Opportunities and challenges, IEEE Access 10 (2022) 84486–84517. doi:10.1109/ACCESS.2022.3197671.

[5] B. Binny, N. Praneel, A. Niranjan, A. B. Mathew, A. Vijay Vasanth, Iot-enabled arrhythmia detection using rnns for ecg analysis, in: 2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS), 2024, p. 388 – 395. doi:10.1109/ICICNIS64247.2024.10823369.

[6] A.-T. Shumba, T. Montanaro, I. Sergi, A. Bramanti, M. Ciccarelli, A. Rispoli, A. Carrizzo, M. De Vittorio, L. Patrono, Wearable technologies and ai at the far edge for chronic heart failure prevention and management: A systematic review and prospects, Sensors 23 (2023). doi:10.3390/s23156896.
[7] X. Chen, J. Ji, K. Loparo, P. Li, Real-time personalized cardiac arrhythmia detection and diagnosis: A cloud computing architecture, in: 2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), 2017, pp. 201–204. doi:10.1109/BHI.2017.7897240.
[8] O. Cheikhrouhou, R. Mahmud, R. Zouari, M. Ibrahim, A. Zaguia, T. N. Gia, One-dimensional cnn approach for ecg arrhythmia analysis in fog-cloud environments, IEEE Access 9 (2021) 103513–103523. doi:10.1109/ACCESS.2021.3097751.
[9] J. Wen, Z. Zhang, Y. Lan, Z. Cui, J. Cai, W. Zhang, A survey on federated learning: challenges and applications, International Journal of Machine Learning and Cybernetics 14 (2023) 513–535. doi:10.1007/s13042-022-01647-y.
[10] B. McMahan, E. Moore, D. Ramage, S. Hampson, B. A. y Arcas, Communication-efficient learning of deep networks from decentralized data, in: Artificial Intelligence and Statistics, 2017, pp. 1273–1282. arXiv:1602.05629.
[11] D. M. J. Gutierrez, A. Anagnostopoulos, I. Chatzigiannakis, A. Vitaletti, Fedartml: A tool to facilitate the generation of non-iid datasets in a controlled way to support federated learning research, IEEE Access 12 (2024) 81004–81016. doi:10.1109/ACCESS.2024.3410026.
[12] R. Rb-Silva, X. Ribeiro, F. Almeida, C. Ameijeiras-Rodriguez, J. Souza, L. Conceição, T. Taveira-Gomes, G. Marreiros, A. Freitas, Secur-e-health project: Towards federated learning for smart pediatric care, Stud Health Technol Inform 302 (2023) 516–520. doi:10.3233/SHTI230196.
[13] R. N. Asif, A. Ditta, H. Alquhayz, S. Abbas, M. A. Khan, T. M. Ghazal, S.-W. Lee, Detecting electrocardiogram arrhythmia empowered with weighted federated learning, IEEE Access 12 (2024) 1909–1926. doi:10.1109/ACCESS.2023.3347610.
[14] S. Sakib, M. M. Fouda, Z. Md Fadlullah, K. Abualsaud, E. Yaacoub, M. Guizani, Asynchronous federated learning-based ecg analysis for arrhythmia detection, in: 2021 IEEE International Mediterranean Conference on Communications and Networking (MeditCom), 2021, pp. 277–282. doi:10.1109/MeditCom49071.2021.9647636.
[15] A. Manocha, S. Sood, M. Bhatia, Federated learning-inspired smart ecg classification: an explainable artificial intelligence approach, Multimedia Tools and Applications (2024) 1–24. doi:10.1007/s11042-024-20084-3.
[16] D. M. Jimenez Gutierrez, H. M. Hassan, L. Landi, A. Vitaletti, I. Chatzigiannakis, Application of federated learning techniques for arrhythmia classification using 12-lead ecg signals, in: I. Chatzigiannakis, I. Karydis (Eds.), Algorithmic Aspects of Cloud Computing, Springer Nature Switzerland, Cham, 2024, pp. 38–65.
[17] Y. Chen, Y. Ning, M. Slawski, H. Rangwala, Asynchronous online federated learning for edge devices with non-iid data, in: 2020 IEEE International Conference on Big Data (Big Data), 2020, pp. 15–24. doi:10.1109/BigData50022.2020.9378161.
[18] Z. Lu, H. Pan, Y. Dai, X. Si, Y. Zhang, Federated learning with non-iid data: A survey, IEEE Internet of Things Journal 11 (2024) 19188–19209. doi:10.1109/JIOT.2024.3376548.
[19] X. Ma, J. Zhu, Z. Lin, S. Chen, Y. Qin, A state-of-the-art survey on solving non-iid data in federated learning, Future Generation Computer Systems 135 (2022) 244–258. doi:https://doi.org/10.1016/j.future.2022.05.003.
[20] N. Shoham, T. Avidor, A. Keren, N. Israel, D. Benditkis, L. Mor-Yosef, I. Zeitak, Overcoming forgetting in federated learning on non-iid data, 2019. URL: https://arxiv.org/abs/1910.07796. arXiv:1910.07796.
[21] D. J. Beutel, T. Topal, A. Mathur, X. Qiu, J. Fernandez-Marques, Y. Gao, L. Sani, K. H. Li, T. Parcollet, P. P. B. de Gusmão, N. D. Lane, Flower: A friendly federated learning research framework, 2022. URL: https://arxiv.org/abs/2007.14390. arXiv:2007.14390.
[22] Mlflow: An open source platform for the complete machine learning lifecycle, https://mlflow.org/, 2025. Accessed: March 11, 2025.
[23] S. Sakib, Ecg arrhythmia classification dataset, https://www.kaggle.com/datasets/sadmansakib7/

ecg-arrhythmia-classification-dataset, 2021. Accessed: March 11, 2025.
[24] Mit-bih supraventricular arrhythmia database, https://doi.org/10.13026/C2V30W, 2021. doi:10.13026/C2V30W, accessed: March 11, 2025.
[25] Mit-bih arrhythmia database, https://doi.org/10.13026/C2F305, 2021. doi:10.13026/C2F305, accessed: March 11, 2025.

A. Appendix A: Figures, Tables, and Equations

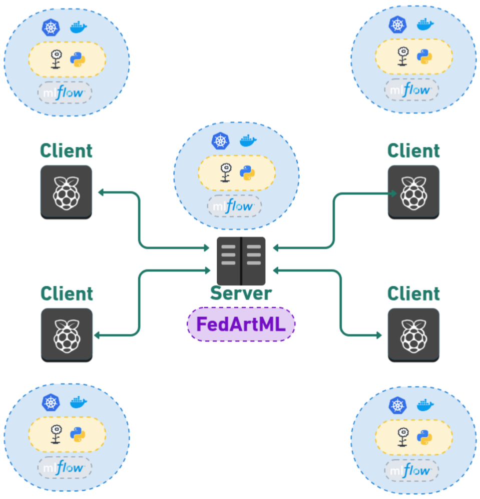

Figure 1: System Architecture

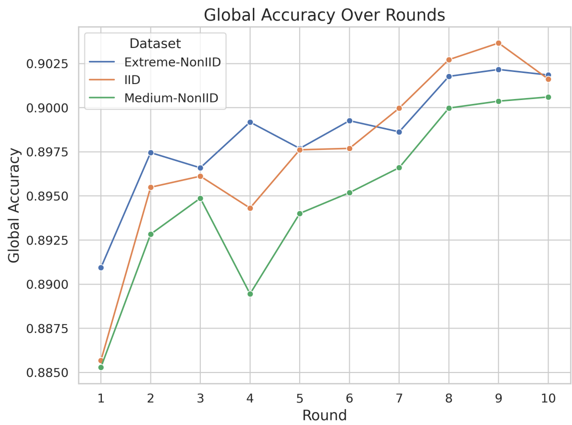

Figure 2: Global Model Accuracy Over Rounds

Table 1
Configurations for generating Non-IID datasets. (1): MIT-BIH Supraventricular Lead II; (2): MIT-BIH Supraventricular Lead V; (3): MIT-BIH Arrhythmia Lead II; (4): MIT-BIH Arrhythmia Lead V.

Partitioning	Dataset 1		Dataset 2		Dataset 3		Dataset 4	
	α_1	HD_1	α_2	HD_2	α_3	HD_3	α_4	HD_4
IID	1000	0.01	1000	0.01	1000	0.01	1000	0.01
Moderate Non-IID	10	0.18	7.934	0.13	12.59	0.20	6.31	0.13
Extreme Non-IID	3.162	0.34	2.512	0.41	3.981	0.33	5.012	0.23

Table 2
Training Time Summary

Dataset	Min Training Time	Max Training Time
Extreme Non-IID	33.6min	66.0min
Moderate Non-IID	34.2min	53.5min
IID	44.7min	45.9min

$$\text{HD}(P_{Y_1}(y), P_{Y_2}(y)) = \frac{1}{\sqrt{2}}\sqrt{\sum_{y\in Y}\left(\sqrt{P_{Y_1}(y)} - \sqrt{P_{Y_2}(y)}\right)^2} \tag{1}$$