International Journal of Emerging Research in Engineering and Technology
Pearl Blue Research Group| 9th ICAILLMBA (ALLAM) 2025-Conference Proceedings
ISSN: 3050-922X | https://doi.org/10.63282/3050-922X.ICAILLMBA-101

Original Article

Continuous Model Adaptation in Distributed Healthcare Systems: A MLOps Framework for Federated Learning in Multi-Institutional Cardiovascular Risk Assessment

Arjun Kamisetty
Software Developer, Fannie Mae, Reston, VA 20190, USA.

Abstract *- When deploying machine learning models across multiple hospitals, we face a critical challenge: patient data is scattered everywhere, models degrade with changing populations, and no one wants to centralize sensitive health information. This paper explores how Machine Learning Operations combined with federated learning can help by letting each institution train models on its own data while sharing only algorithmic improvements, not patient records. We reviewed literature on operational model maintenance and privacy-preserving data sharing in healthcare organizations. Our main finding is that this MLOps-federated approach addresses three essential needs simultaneously: maintaining model performance through automatic retraining when it drifts, protecting patient privacy since raw data never leaves hospitals, and catching bias across diverse patient populations. The challenge lies in handling messy, non-independent clinical data across institutions with different computing resources. The real operational headaches emerge when managing distributed computational resources and when hospitals disagree on measuring model performance fairly. Effective governance requires clear protocols for model versioning, regular performance checks with intervention thresholds, and data quality standards respecting each institution's circumstances. Healthcare systems must invest in infrastructure for automated model management before attempting federated approaches, and regulators need frameworks treating these systems appropriately. This work bridges the gap between theoretical federated learning and actual hospital implementation, providing practical guidance for deploying responsible AI across multiple institutions while maintaining patient privacy.*

Keywords *- Mlops, Federated Learning, Healthcare AI, Model Drift Detection, Privacy-Preserving Machine Learning, Cardiovascular Risk Prediction.*

1. Introduction

Healthcare organizations increasingly recognize that machine learning models can improve patient outcomes, particularly in cardiovascular risk assessment where early prediction saves lives. However, the journey from developing a model in a research environment to maintaining it across multiple hospitals presents substantial challenges that existing literature has not adequately addressed [1]. Traditional centralized approaches require aggregating patient data in a single location, which raises privacy concerns, violates data governance policies, and may not even be legally permissible under regulations like HIPAA and GDPR [2].

The problem becomes more complex when we consider that clinical populations change over time. A model trained on data from 2020 may perform poorly in 2024 because patient demographics shift, treatment protocols evolve, and disease patterns change [3]. Without continuous monitoring and adaptation, these models gradually lose accuracy in what researchers call model drift. Meanwhile, each hospital operates with different electronic health record systems, varying levels of computational infrastructure, and distinct patient populations that may not be well-represented in centrally trained models [4].

Federated learning offers a potential solution by enabling collaborative model training without centralizing sensitive data. Instead of moving patient records to a central server, the learning algorithm travels to each institution, trains on local data, and only shares model updates [5]. However, implementing this approach operationally requires more than just the federated learning algorithm itself. Healthcare systems need robust MLOps practices including automated monitoring for drift, version control for models, systematic evaluation pipelines, and governance frameworks that work across institutional boundaries [6].

This paper examines how MLOps frameworks can support federated learning deployment specifically for cardiovascular risk assessment across multiple healthcare institutions. We focus on the practical operational requirements rather than algorithmic improvements, addressing questions that hospital IT departments and clinical leadership actually face when implementing these systems. Our contribution lies in synthesizing insights from distributed systems, healthcare informatics, and machine learning operations to provide actionable guidance for multi-institutional model deployment.

2. Literature Review

The intersection of MLOps and federated learning in healthcare represents an emerging research area with contributions from several distinct communities. Understanding the current state requires examining work on federated learning foundations, MLOps practices, and healthcare-specific considerations.

Federated learning emerged as a solution for training models across decentralized data sources while preserving privacy. McMahan et al. introduced the foundational Federated Averaging algorithm, demonstrating that averaging model weights from distributed training can approximate centralized learning [7]. Subsequent work has addressed communication efficiency, security against adversarial participants, and convergence guarantees under non-identical data distributions [8]. However, most theoretical work assumes relatively stable data distributions and does not address operational concerns about long-term model maintenance.

Recent healthcare applications have demonstrated federated learning's viability for clinical prediction tasks. Rieke et al. surveyed federated learning applications in medical imaging, showing that models trained across multiple hospitals can match or exceed the performance of models trained at individual institutions [9]. For cardiovascular risk specifically, Liu et al. demonstrated federated learning for predicting heart failure using electronic health records from three hospitals, though their study focused on initial model development rather than ongoing operational maintenance [10].

The MLOps community has developed practices for managing machine learning systems in production environments. Paleyes et al. provided a comprehensive overview of challenges in deploying and maintaining ML systems, emphasizing the need for monitoring, automated retraining pipelines, and clear governance [1]. Drift detection has received particular attention, with researchers proposing various methods for identifying when model performance degrades due to changing data distributions [11]. However, most MLOps literature assumes centralized data access and does not address the distributed setting of federated learning.

Some recent work has begun bridging these areas. Mothukuri et al. discussed operational considerations for federated learning systems, including resource management and fault tolerance [12]. Zhang et al. proposed a framework for continuous learning in federated settings, though their focus was primarily on algorithmic efficiency rather than operational governance [13]. What remains underexplored is how healthcare institutions can actually implement and govern these systems given real-world constraints around computational resources, data quality variations, and institutional autonomy.

The cardiovascular risk assessment domain presents specific challenges relevant to our work. Traditional risk calculators like the Framingham Risk Score and Pooled Cohort Equations rely on demographic factors and laboratory values [14]. Machine learning approaches can incorporate additional variables from electronic health records, but they require careful validation across diverse populations to avoid perpetuating healthcare disparities [15]. The need for ongoing monitoring is particularly acute because cardiovascular disease patterns change with evolving treatment standards and population health trends.

3. Methodology

Our methodology combines systematic literature analysis with framework development informed by practical requirements from healthcare IT systems. Rather than implementing a complete system, we focused on identifying the essential components and governance structures needed for sustainable operation. We analyzed existing MLOps platforms and federated learning frameworks to understand their capabilities and limitations in healthcare contexts. This included examining tools like Kubeflow, MLflow, and TensorFlow Federated alongside healthcare-specific platforms. We evaluated each against requirements derived from cardiovascular risk assessment use cases, including the need to handle heterogeneous data formats, support asynchronous training across institutions with varying computational resources, and provide audit trails for regulatory compliance.

The proposed framework architecture consists of three layers addressing different operational concerns, as illustrated by the visual description provided in Figure 1. The institutional layer operates within each hospital's data center and includes local data preprocessing, model training capabilities, and drift detection monitors. Each institution maintains sovereignty over its computational resources and can set policies about when and how frequently to participate in federated training rounds [6]. The federation layer coordinates training across institutions through a central orchestrator that manages model aggregation, version control, and communication protocols. Importantly, this layer never accesses raw patient data, receiving only encrypted model updates from participating institutions. The governance layer spans both institutional and federation levels, establishing protocols for model validation, performance monitoring, and intervention thresholds.

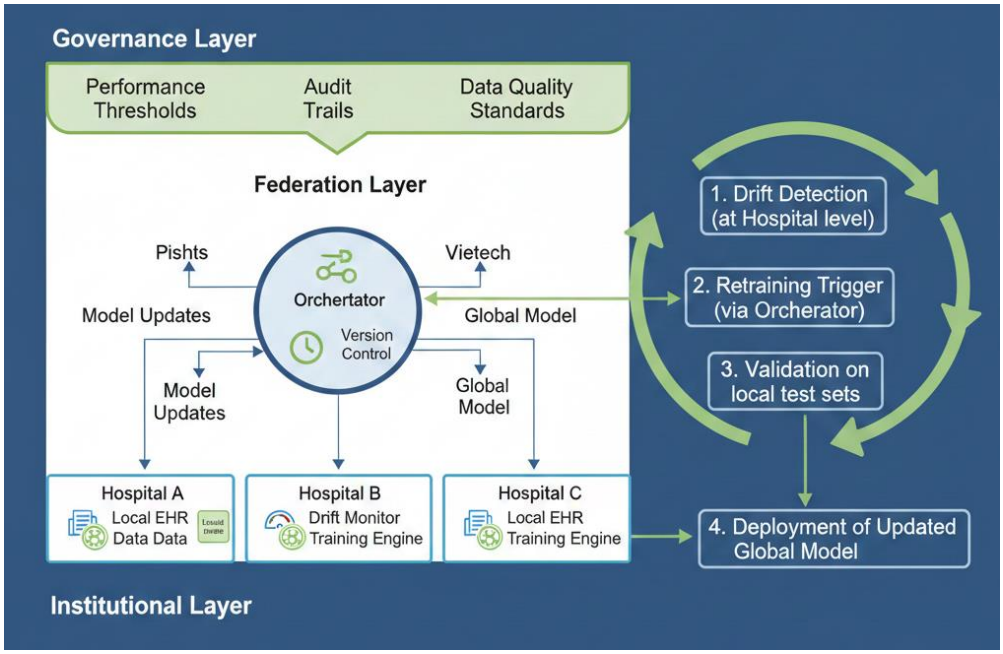

Fig 1: Mlops-Federated Learning Framework for Cardiovascular Risk Assessment

For cardiovascular risk assessment specifically, we designed the workflow to align with clinical validation requirements. Initial model development occurs through federated training across participating institutions, with each site using its historical data to contribute to a global model. Once deployed, each institution runs continuous monitoring to detect when local model performance degrades below acceptable thresholds, measured through metrics like area under the ROC curve for discrimination and calibration slopes [14]. When drift is detected at multiple institutions or when the federation orchestrator identifies systematic performance degradation, an automated retraining cycle initiates.

The retraining process follows a structured protocol addressing practical operational concerns. First, participating institutions verify their local data quality against predefined standards, checking for completeness, consistency, and recency. Second, each institution trains on its recent data using the current global model as initialization, which improves convergence compared to training from scratch [13]. Third, the federation orchestrator aggregates updates using weighted averaging based on each institution's sample size and data quality metrics. Fourth, the candidate updated model undergoes validation at each institution using held-out test sets before deployment. This validation step is critical because it allows each hospital to verify that the updated model performs acceptably for its specific patient population.

We established governance requirements through analysis of healthcare regulatory frameworks and institutional review board protocols. Model versioning must maintain complete audit trails showing which data contributed to each model version and which institutions participated in training. Performance monitoring requires standardized metrics that account for different base rates of cardiovascular events across institutions with varying patient populations. Data quality standards must be flexible enough to accommodate different electronic health record systems while maintaining minimum thresholds for completeness and accuracy [4].

4. Results and Discussion

The proposed framework addresses several critical operational requirements that previous work has not fully resolved. Most significantly, it provides a practical path for healthcare institutions to maintain model performance over time without sacrificing data privacy or institutional autonomy.

For model drift detection, our approach combines institution-level and federation-level monitoring to balance sensitivity with computational efficiency. Each hospital runs lightweight monitoring on a weekly basis, checking performance metrics on recent patient cohorts. When an institution detects drift exceeding its threshold, it reports this to the federation orchestrator without sharing patient-level data. The orchestrator aggregates these signals across institutions to distinguish true drift affecting multiple sites from local anomalies [11]. This two-level approach prevents unnecessary retraining triggered by temporary local variations while ensuring timely response to genuine performance degradation affecting the broader healthcare network.

Privacy preservation emerges as a multifaceted concern beyond the basic guarantee that raw data never leaves institutional boundaries. The framework must also protect against inference attacks where adversaries might attempt to extract patient

information from model updates. We recommend that institutions apply differential privacy techniques to their local model updates before sharing them with the federation orchestrator, adding calibrated noise that preserves privacy while maintaining model utility [8]. The privacy-utility trade off requires careful tuning because excessive noise degrades model performance, but insufficient protection could theoretically allow reconstruction of training data characteristics.

Computational resource heterogeneity across institutions presents practical challenges that pure federated learning algorithms do not address. Large academic medical centers may have substantial GPU clusters available for model training, while smaller community hospitals might only have CPU-based servers. Our framework accommodates this through asynchronous training where institutions contribute updates when their resources allow rather than requiring synchronous participation [12]. The federation orchestrator weights contributions not just by sample size but also by recency, giving more influence to institutions that train on the most current data. This approach prevents larger institutions from dominating the global model while ensuring that all participants can contribute meaningfully.

Bias detection and mitigation become more tractable in federated settings compared to centralized approaches, somewhat counterintuitively. Because each institution evaluates model performance on its own patient population, systematic biases affecting specific demographic groups become visible at the institutional level. For cardiovascular risk assessment, this means hospitals serving predominantly Hispanic populations can identify if the model underestimates risk for their patients, while institutions serving older populations can detect age-related bias [15]. The federation orchestrator can aggregate these bias reports to identify systematic issues requiring algorithmic intervention, such as reweighting training contributions or incorporating demographic factors differently.

Governance structures must balance standardization with institutional flexibility. We propose a tiered governance model where core requirements around data quality, privacy protection, and performance monitoring apply to all participants, while institutions retain flexibility in their local implementation details. For example, all hospitals must monitor model calibration, but each can choose whether to use their own validation cohorts or rely on the federated validation approach. This flexibility acknowledges that institutions have different resources and priorities while maintaining sufficient standardization for the federation to function effectively [6].

The operational reality involves numerous practical considerations that research papers often overlook. IT departments need clear documentation about computational requirements and integration points with existing electronic health record systems. Clinical leadership requires evidence that the operational complexity provides meaningful benefit over simpler alternatives. Regulatory affairs teams want assurance that the system complies with patient privacy laws and maintains appropriate audit trails. Legal departments seek clarity on liability when models trained across multiple institutions make predictions affecting patient care. Our framework provides structure for addressing these concerns but does not eliminate them entirely.

Implementation challenges will inevitably arise. Data quality variations across institutions may be larger than anticipated, requiring more sophisticated preprocessing or quality-weighted aggregation. Communication infrastructure between institutions and the federation orchestrator may face network limitations or security requirements that slow training cycles. Institutional review boards may have questions about the federated approach that require education and documentation. Clinical champions at each hospital need ongoing support to maintain engagement as the initial excitement of a new technology fades into routine operation.

5. Conclusion and Further Research

This work demonstrates that combining MLOps practices with federated learning provides a viable path forward for deploying and maintaining machine learning models across multiple healthcare institutions. The framework we propose addresses the practical operational requirements that determine whether these systems succeed or fail in real-world healthcare environments. Rather than focusing solely on algorithmic improvements, we emphasize governance structures, monitoring protocols, and institutional coordination mechanisms that enable sustainable operation.

Several limitations of this work point toward important directions for further research. We have not implemented the complete framework in a real multi-institutional setting, so questions remain about practical challenges that will emerge during actual deployment. The computational and communication costs of federated training may prove prohibitive for some institutions, suggesting the need for more efficient algorithms or tiered participation models. Our focus on cardiovascular risk assessment provides useful specificity but limits the generalizability of some recommendations to other clinical prediction tasks.

Future research should pursue several specific directions. First, empirical studies implementing federated MLOps frameworks across real healthcare institutions would provide invaluable insights about practical barriers and solutions. Second, developing standardized evaluation protocols for multi-institutional model validation would help institutions assess whether

federated approaches provide sufficient benefit to justify their complexity. Third, investigating how to effectively involve patients and community stakeholders in governance of federated health AI systems would improve trust and accountability. Fourth, exploring federated learning for rare conditions where even multi-institutional collaboration provides limited data could extend benefits to underserved patient populations.

The broader implication of this work is that operational considerations must guide AI deployment in healthcare, not just algorithmic performance. Healthcare institutions need practical frameworks that acknowledge their constraints around data privacy, computational resources, and institutional autonomy while providing pathways to improve patient care through advanced analytics. The MLOps-federated learning combination we propose represents one such framework, but the field needs continued development of operational practices that bridge the gap between research innovations and healthcare delivery realities.

References

[1] A. Paleyes, R. G. Urma, and N. D. Lawrence, "Challenges in deploying machine learning: a survey of case studies," *ACM Computing Surveys*, vol. 55, no. 6, pp. 1-29, 2022.

[2] D. S. W. Ting, L. R. Pasquale, L. Peng, J. P. Campbell, A. Y. Lee, R. Raman, G. S. W. Tan, L. Schmetterer, P. A. Keane, and T. Y. Wong, "Artificial intelligence and deep learning in ophthalmology," *British Journal of Ophthalmology*, vol. 103, no. 2, pp. 167-175, 2019.

[3] S. Rabanser, S. Günnemann, and Z. Lipton, "Failing loudly: An empirical study of methods for detecting dataset shift," *Advances in Neural Information Processing Systems*, vol. 32, pp. 1396-1408, 2019.

[4] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities," *IEEE Access*, vol. 5, pp. 8869-8879, 2017.

[5] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50-60, 2020.

[6] N. Rieke, J. Hancox, W. Li, F. Milletari, H. R. Roth, S. Albarqouni, S. Bakas, M. N. Galtier, B. A. Landman, K. Maier-Hein, S. Ourselin, M. Sheller, R. M. Summers, A. Trask, D. Xu, M. Baust, and M. J. Cardoso, "The future of digital health with federated learning," *NPJ Digital Medicine*, vol. 3, no. 1, pp. 1-7, 2020.

[7] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Arcas, "Communication-efficient learning of deep networks from decentralized data," *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, pp. 1273-1282, 2017.

[8] R. C. Geyer, T. Klein, and M. Nabi, "Differentially private federated learning: A client level perspective," *arXiv preprint arXiv:1712.07557*, 2017.

[9] N. Rieke, J. Hancox, W. Li, F. Milletarì, H. R. Roth, S. Albarqouni, S. Bakas, M. N. Galtier, B. A. Landman, K. Maier-Hein, S. Ourselin, M. Sheller, R. M. Summers, A. Trask, D. Xu, M. Baust, and M. J. Cardoso, "The future of digital health with federated learning," *NPJ Digital Medicine*, vol. 3, art. 119, 2020.

[10] Y. Liu, T. Fan, T. Chen, Q. Xu, and Q. Yang, "FATE: An industrial grade platform for collaborative learning with data protection," *Journal of Machine Learning Research*, vol. 22, no. 226, pp. 1-6, 2021.

[11] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, and G. Zhang, "Learning under concept drift: A review," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 12, pp. 2346-2363, 2018.

[12] V. Mothukuri, R. M. Parizi, S. Pouriyeh, Y. Huang, A. Dehghantanha, and G. Srivastava, "A survey on security and privacy of federated learning," *Future Generation Computer Systems*, vol. 115, pp. 619-640, 2021.

[13] C. Zhang, Y. Xie, H. Bai, B. Yu, W. Li, and Y. Gao, "A survey on federated learning," *Knowledge-Based Systems*, vol. 216, art. 106775, 2021.

[14] D. C. Goff Jr., D. M. Lloyd-Jones, G. Bennett, S. Coady, R. B. D'Agostino, R. Gibbons, P. Greenland, D. T. Lackland, D. M. Levy, C. J. O'Donnell, J. G. Robinson, J. S. Schwartz, S. T. Shero, S. C. Smith Jr., P. Sorlie, N. J. Stone, and P. W. Wilson, "2013 ACC/AHA guideline on the assessment of cardiovascular risk," *Circulation*, vol. 129, no. 25 suppl 2, pp. S49-S73, 2014.

[15] I. Y. Chen, P. Szolovits, and M. Ghassemi, "Can AI help reduce disparities in general medical and mental health care?," *AMA Journal of Ethics*, vol. 21, no. 2, pp. 167-179, 2019.