Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

FedPACT: A Policy Calibration Framework for Federated Digital Twins in Healthcare Cyber-Physical Systems

Patrick Sello
Univesity of the Western Cape
Antoine Bagula
University of the Western Cape
Olasupo Ajayi
olasupoajayi@gmail.com
University of the Western Cape Faculty of Natural Science https://orcid.org/0000-0001-6583-3749
Ferdinand Kahenga
University of the Western Cape
Landry Mbale
University of the Western Cape
Jennifer Chipps
University of the Western Cape
Damian Clarke
University of KwaZulu-Natal
Jovita Mateus
Namibia University of Science and Technology

Research Article

Keywords: Cyber-Physical Systems, Digital Shadows, Federated Policy Calibration, Healthcare Digital Twins, Policy Coordination, Readmission Prediction

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11553661/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

FedPACT: A Policy Calibration Framework for Federated Digital Twins in Healthcare Cyber-Physical Systems

Patrick Sello[1*], Antoine Bagula[1,2], Olasupo Ajayi[1,3*], Ferdinand Kahenga[1], Landry Mbale[1,2], Jennifer Chipps[4], Damian Clarke[5], Jovita Mateus[6]

[1*]ISAT Laboratory, University of the Western Cape, Robert Sobukwe Road, Bellville, 7535, Western Cape, South Africa.

[2*]Université Nouveaux Horizons, Route Kasapa, Quartier Gambela, Lubumbashi, DRC.

[3]Electrical and Computer Engineering Department, Queen's University, Union Street, Kingston, K7L3N9, Ontario, Canada.

[4]Community Health Sciences, University of the Western Cape, Voortrekker Road, Bellville, 7535, Western Cape, South Africa.

[5]Department of Surgery, University of KwaZulu-Natal, Umbilo Road, Congella, 4001, Durban, South Africa.

[6]Computer Science Department, Namibia University of Science and Technology, Jackson Kaujeua Street, 13388, Windhoek, Namibia.

*Corresponding author(s). E-mail(s): 3919362@myuwc.ac.za; o.ajayi@queensu.ca;
Contributing authors: abagula@uwc.ac.za; ferdinandkahenga@esisalama.org; 6509774@myuwc.ac.za; jchipps@uwc.ac.za; damianclar@gmail.com; jmateus@nust.na;

Abstract

Healthcare institutions increasingly use locally deployed predictive models, yet collaboration across institutions typically focuses on model training rather than the operational policies governing decisions based on model outputs. This paper introduces FedPACT, a federated policy-calibration framework that coordinates local decision thresholds across distributed healthcare systems without exchanging patient-level data or predictive-model parameters.

Each hospital derives an initial threshold from its local predicted-risk distribution and simulated escalation capacity. FedPACT iteratively updates these thresholds using a weighted reference and a local-retention parameter, while preserving local data, models, and policy execution. The mechanism is analysed in terms of convergence, stability, threshold bounds, and computational and communication complexity.

The framework is evaluated using precomputed 15- and 30-day readmission-risk estimates for Trauma and General Surgery, partitioned across five simulated hospital nodes with heterogeneous patient volumes, risk distributions, and escalation capacities. Experiments examine static calibration, sensitivity to the local-retention parameter, baseline comparisons, repeated dynamic-capacity simulations, and a virtualised communication microbenchmark.

Across the evaluated configurations, local thresholds converged geometrically towards an invariant weighted reference, reducing inter-hospital threshold dispersion. This behaviour persisted under the tested dynamic-capacity conditions. However, threshold alignment did not equalise the number of patients eligible for escalation, which remained dependent on local patient volumes, risk distributions, and simulated capacities. The communication microbenchmark indicated a small application payload and sub-millisecond local processing overhead in the tested environment.

These results demonstrate the feasibility of federated calibration of capacity-derived decision thresholds without exchanging patient-level records or predictive-model parameters. They do not establish clinical effectiveness, formal privacy guarantees, real-world network performance, or readiness for autonomous clinical decision-making.

Keywords: Cyber-Physical Systems, Digital Shadows, Federated Policy Calibration, Healthcare Digital Twins, Policy Coordination, Readmission Prediction

1 Introduction

Healthcare systems can be viewed as Cyber-Physical Systems (CPS) in which patients, clinicians, medical devices, and hospital infrastructure interact with digital processes for sensing, analysis, and decision support. Digital Twins provide a broader architectural perspective for representing and monitoring such systems. However, healthcare implementations differ in their data flows, feedback capabilities, and degrees of autonomy. In particular, predictive systems that analyse patient data and support decisions without automated feedback should not be treated as fully autonomous Digital Twins.

Hospital readmission prediction provides a relevant setting for examining this distinction. Readmission risk can reflect injury severity, comorbidities, discharge destination, medication complexity, post-discharge complications, and continuity of care [1]. Predictive analytics can help identify patients at elevated risk [2, 3], but predictive models are often deployed as analytical components rather than integrated into systems with explicitly defined feedback and control mechanisms [4, 5]. A maturity-aware architectural perspective helps establish the capabilities and

limitations of such systems without equating predictive performance with autonomous clinical operation [7].

A further challenge arises when predictive systems operate across multiple hospitals. Differences in patient volumes, predicted-risk distributions, and available escalation capacity can lead institutions to adopt different decision thresholds, even when their predictive models are comparable. These thresholds determine which patients become eligible for further review or follow-up. Coordinating operational thresholds is therefore distinct from coordinating the training of the predictive models that generate risk estimates.

Federated learning (FL) has expanded beyond conventional model aggregation to address personalised learning, fairness-aware optimisation, resource-aware training, probability calibration, and uncertainty quantification [10–15]. Nevertheless, these approaches do not directly resolve how hospitals should coordinate capacity-derived operational thresholds when their risk distributions and escalation capacities differ. This study addresses that gap through the federated calibration of decision policies across distributed healthcare systems.

We introduce Federated Policy Alignment and Calibration across Digital Twins (FedPACT), a framework that extends federation from the predictive-model layer to the operational policy layer. The participating hospital nodes are positioned at the Digital Shadow stage of the Digital Twin Maturity Framework (DTMF) [43]: patient data are processed in the analytical environment, and predictions can inform decisions, but automated feedback or control to the represented clinical environment is not implemented. The maturity framework establishes this architectural boundary, while FedPACT coordinates the thresholds applied to locally generated predictions. Each hospital derives an initial threshold from its local predicted-risk distribution and simulated escalation capacity. The hospitals then iteratively calibrate their thresholds using a weighted reference and a local-retention parameter, without exchanging patient-level records or predictive-model parameters. Thresholds identify patients eligible for possible escalation; they do not automatically trigger clinical interventions.

The study analyses the convergence, stability, threshold bounds, and computational and communication complexity of the calibration mechanism. It evaluates four surgical readmission-prediction tasks across five simulated hospital nodes with heterogeneous patient volumes, predicted-risk distributions, and escalation capacities. Experiments examine baseline comparisons, sensitivity to the local-retention parameter, static calibration, repeated dynamic-capacity simulations, and threshold-eligible escalation workloads. The results show that thresholds converge towards an invariant weighted reference under the stated assumptions, while numerical alignment does not necessarily equalise eligible workloads across hospitals.

The study addresses three research questions:

RQ1: How do differences in predicted-risk distributions and institutional capacities affect capacity-derived local escalation thresholds across healthcare Digital Shadows?

RQ2: To what extent can FedPACT achieve stable cross-institutional policy-threshold alignment without exchanging patient-level data or

predictive-model parameters, and how is convergence influenced by the local-retention parameter?

RQ3: How does federated policy-threshold alignment affect the distribution of threshold-eligible escalation workload across hospitals with heterogeneous predicted-risk distributions and institutional capacities?

Contributions. The principal contributions are:

- A federated policy-calibration framework that coordinates capacity-derived decision thresholds across distributed healthcare systems while keeping patient-level data, predictive models, and policy execution local.
- A weighted iterative calibration mechanism with a local-retention parameter, accompanied by analysis of convergence, threshold-bound preservation, dispersion reduction, and consensus stability.
- An analytical and empirical assessment of the computational and communication requirements of scalar threshold calibration, including a virtualised five-node microbenchmark.
- A simulation-based evaluation across four surgical readmission-prediction tasks and five heterogeneous hospital nodes, including baseline comparisons, parameter sensitivity, and repeated dynamic-capacity simulations.
- An analysis showing that threshold alignment does not, by itself, guarantee equivalent threshold-eligible escalation workloads across hospitals.

The remainder of the paper reviews related work, presents the system model and FedPACT architecture, formalises the calibration mechanism and its analytical properties, and describes the experimental setup and results. The discussion examines the implications and limitations of the findings before the paper concludes.

2 Related Work

2.1 Healthcare Digital Twins and Digital Shadows

Digital Twins extend industrial Cyber-Physical Systems (CPS) through digital representations for monitoring, simulation, and optimisation [8, 9, 16]. Digital Models, Digital Shadows, and intelligent (bidirectional) Digital Twins differ in their data flows and feedback capabilities: Digital Shadows receive data from the represented system, whereas bidirectional Twins additionally support feedback or control [19–23].

Healthcare applications include personalised medicine, chronic disease management, clinical workflow optimisation, and hospital operations [24–26]. These systems combine clinical records, physiological measurements, and predictive models to analyse disease progression and treatment response [27, 28]. However, some healthcare implementations lack bidirectional control, highlighting the need to distinguish analytical systems from Twins with validated feedback and adaptive capabilities [29, 30, 35, 36]. The hospital nodes evaluated in this study are Digital Shadows: predictions may inform decisions, but automated clinical feedback and control are not implemented.

2.2 Federated Learning and Predictive Calibration

Readmission prediction and clinical risk stratification are established applications of machine learning, although predictive components are not always integrated into governed CPS architectures [4, 5, 31, 32]. FL enables collaborative model development while keeping datasets local. Beyond Federated Averaging [54], research addresses personalised learning, fairness, resource-aware optimisation, probability calibration, and uncertainty quantification [10–15].

These methods address collaborative prediction and model-output reliability, but do not directly coordinate the operational thresholds that translate local risk estimates into escalation eligibility. This gap is important when hospitals differ in patient volumes, predicted-risk distributions, and escalation capacities.

2.3 Research Gap and Positioning of FedPACT

Digital Twin for Healthcare (DT4H) research focuses on digital representation and maturity, while predictive analytics and FL primarily address risk estimation and collaborative model development. Neither model aggregation nor Digital Shadow classification specifies how operational decision thresholds should be coordinated across institutions.

FedPACT addresses this gap by federating capacity-derived thresholds rather than patient records or predictive-model parameters. Each node retains its data, model, and policy execution locally, exchanging scalar thresholds for iterative weighted calibration.
The evaluation also distinguishes numerical threshold alignment from workload equalisation: heterogeneous patient volumes and risk distributions can produce different threshold-eligible workloads even when local thresholds converge. FedPACT therefore targets policy calibration, not workload optimisation or clinical-outcome improvement.

Table 1 positions FedPACT relative to the principal research streams.

3 Federated Policy Calibration and Theoretical Analysis

3.1 FedPACT Architecture and Workflow

Figure 1 illustrates the architecture of FedPACT and the interaction between local healthcare nodes and the federated policy-calibration layer.

The architecture comprises three functional layers. The local healthcare layer processes patient records using existing predictive models and derives initial decision thresholds from local predicted-risk distributions and configured escalation capacities.
The federated calibration layer collects these thresholds, computes a weighted reference, and updates each local threshold according to the calibration rule. The local policy-execution layer applies the updated thresholds to identify records eligible for further clinical review.

Only the information required for threshold calibration is exchanged by the core mechanism; patient-level records and predictive-model parameters remain local.

Table 1 – Comparison of related research streams with FedPACT.

Research stream	DT maturity	Model federation	Policy federation	Capacity aware policies	Workload effects
DT maturity and Digital Shadow studies [19, 20, 22]	✓	–	–	–	–
DT4H frameworks [24, 30, 36]	✓	–	–	Limited	–
Healthcare predictive and readmission analytics [2–4]	–	–	–	Limited	–
Conventional FL [54]	–	✓	–	–	–
Proposed FedPACT	✓	Conceptual	✓	✓	✓

The calibrated thresholds support local decision-making but do not automatically initiate clinical interventions. Thus, FedPACT coordinates operational policies across Digital Shadows without itself increasing their Digital Twin maturity or implementing bidirectional clinical control.

3.2 Capacity-Derived Local Thresholds

Let $p_i(x) = M_i(x)$ denote the locally generated readmission-risk estimate for patient x at hospital i. Let $q_i \in (0, 1)$ denote the hospital's assumed escalation capacity as a fraction of its local population. The initial threshold is defined by the empirical $(1 - q_i)$-quantile of its local risk estimates:

$$\tau_i^{(0)} = Q_{1-q_i}(\{p_i(x) : x \in D_i\}) \quad . \tag{1}$$

At calibration round r, the local policy determines eligibility for possible escalation according to

$$\Pi_i^{(r)}(x) = \begin{cases} \text{eligible for escalation}, & p_i(x) \geq \tau_i^{(r)}, \\ \text{standard pathway}, & p_i(x) < \tau_i^{(r)}. \end{cases} \tag{2}$$

The capacity parameter determines the initial threshold, whereas FedPACT subsequently calibrates the threshold across hospitals. Because the local risk distributions may differ, the resulting threshold-eligible workloads need not equal the assumed capacities after calibration. Eligibility denotes a simulated policy outcome, not an automatically executed clinical intervention.

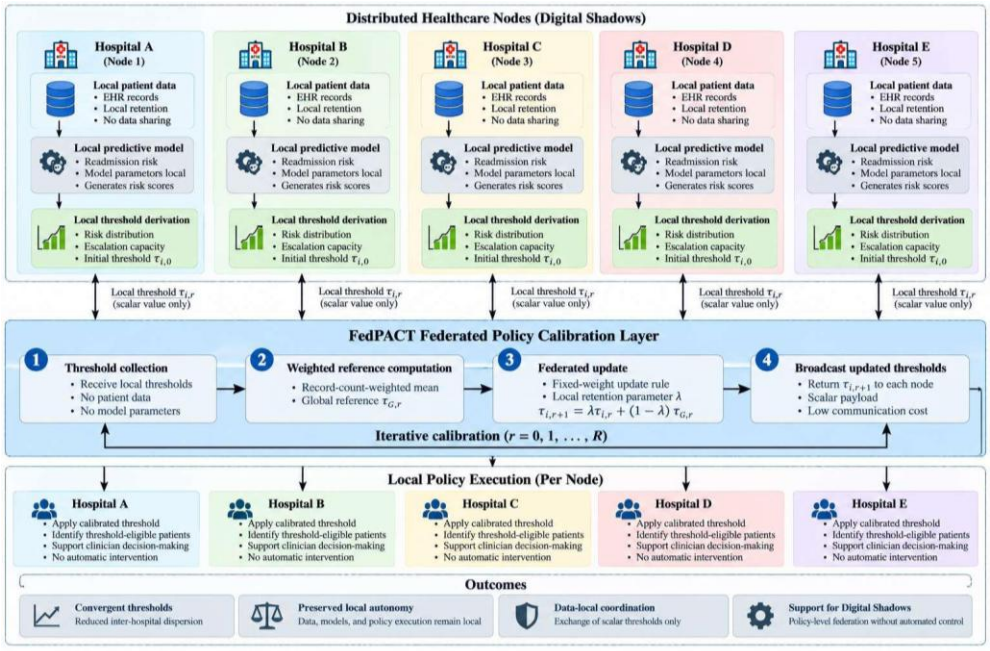


Fig. 1 – FedPACT architecture for federated policy calibration across five simulated healthcare nodes operating as Digital Shadows. Local patient data and predictive models remain within each node. The coordinator aggregates scalar decision thresholds and returns calibrated values. Local policy execution remains separate from the calibration process, with no automated clinical intervention or bidirectional control.

3.3 Weighted Policy-Threshold Calibration

Let $\tau_i^{(r)} \in [0, 1]$ denote the threshold at hospital i during calibration round r. Given fixed non-negative weights w_i satisfying $\sum_{i=1}^{H} w_i = 1$, FedPACT computes the weighted reference threshold

$$\tau_{\text{global}}^{(r)} = \sum_{i=1}^{H} w_i \tau_i^{(r)} . \tag{3}$$

This reference is the minimiser of the weighted squared distance from the local thresholds:

$$\tau_{\text{global}}^{(r)} = \arg\min_{z \in [0,1]} \sum_{i=1}^{H} w_i \left(\tau_i^{(r)} - z\right)^2 . \tag{4}$$

Each hospital updates its threshold according to

$$\tau_i^{(r+1)} = \lambda\tau_i^{(r)} + (1-\lambda)\tau_{\text{global}}^{(r)}, \tag{5}$$

Algorithm 1 FedPACT baseline threshold calibration

Require: Initial thresholds $\{\tau_i^{(0)}\}_{i=1}^{H}$, normalised weights $\{w_i\}_{i=1}^{H}$, parameter $\lambda \in [0, 1]$, and number of rounds R

Ensure: Calibrated thresholds $\{\tau_i^{(R)}\}_{i=1}^{H}$

for r = 0 to R − 1 do

$\tau_{global}^{(r)} \leftarrow \sum_{i=1}^{H} w_i \tau_i^{(r)}$

for all i ∈ H do

$\tau_i^{(r+1)} \leftarrow \lambda \tau_i^{(r)} + (1-\lambda)\tau_{global}^{(r)}$

return $\{\tau_i^{(R)}\}_{i=1}^{H}$

where $\lambda \in [0, 1]$ controls local-threshold retention. For $0 < \lambda < 1$, thresholds move progressively towards the reference. When $\lambda = 0$, every hospital adopts the weighted (global) reference in one round, while $\lambda = 1$ implies that thresholds remain unchanged, i.e., each hospital retains its local threshold without coordination.

3.4 Convergence and Threshold Bounds

Theorem 1 (Convergence). Let the aggregation weights be fixed and satisfy $w_i \geq 0$ and $\sum_i w_i = 1$. For $0 \leq \lambda < 1$, the updates in (5) converge to the weighted reference determined by the initial thresholds:

$$\tau^{\star} = \sum_{i=1}^{H} w_i \tau_i^{(0)}. \tag{6}$$

In particular,

$$\tau_i^{(r)} = \tau^{\star} + \lambda^r \left(\tau_i^{(0)} - \tau^{\star}\right), \tag{7}$$

and hence $\lim_{r\to\infty} \tau_i^{(r)} = \tau^{\star}$ for every hospital.

Proof. Multiplying (5) by w_i and summing over hospitals shows that the weighted mean is invariant:

$$\sum_{i=1}^{H} w_i \tau_i^{(r+1)} = \sum_{i=1}^{H} w_i \tau_i^{(r)}.$$

The invariant value is $\tau^{\star}$. Subtracting $\tau^{\star}$ from the update gives

$$\tau_i^{(r+1)} - \tau^{\star} = \lambda\left(\tau_i^{(r)} - \tau^{\star}\right).$$

Applying this relation recursively yields (7). Since $\lambda^r \to 0$ for $0 \leq \lambda < 1$, convergence follows. □

Lemma 1 (Threshold-bound preservation). If $\tau_i^{(0)} \in [0, 1]$ for all hospitals and $\lambda \in [0, 1]$, then $\tau_i^{(r)} \in [0, 1]$ for every round r. This follows because both the weighted reference and the local update are convex combinations of values in [0, 1].

Lemma 2 (Geometric reduction in dispersion). For any hospitals i and j,

$$\tau_i^{(r)} - \tau_i^{(r)} = \lambda^r \left(\tau_i^{(0)} - \tau_i^{(0)} \right). \quad (8)$$

Consequently, the threshold spread $\Delta^{(r)} = \max_i \tau_i^{(r)} - \min_i \tau_i^{(r)}$ satisfies

$$\Delta^{(r)} = \lambda^r \Delta^{(0)}. \quad (9)$$

For the unweighted variance $\sigma_r^2 = H^{-1} \sum_i (\tau_i^{(r)} - \tau^{-(r)})^2$, where $\tau^{-(r)} = H^{-1} \sum_i \tau_i^{(r)}$, the same update gives

$$\sigma_r^2 = \lambda^{2r} \sigma_0^2 . \quad (10)$$

Thus, threshold dispersion decreases geometrically for $0 \leq \lambda < 1$. At $\lambda = 0$, consensus is achieved after one update; at $\lambda = 1$, no dispersion reduction is guaranteed.

Proposition 1 (Consensus stability). If all hospital thresholds equal τ^* at a given round, subsequent updates preserve that value. This follows directly from (3) and (5).

These results establish convergence, preservation of threshold bounds, and geometric reduction in numerical threshold dispersion under fixed weights. They do not establish equal threshold-eligible workloads, equivalent clinical outcomes, or clinical validity across hospitals.

4 Case Study and Experimental Methodology

4.1 Readmission-Prediction Case Study

The study uses de-identified registry data from Grey's Hospital, KwaZulu-Natal, South Africa, and previously developed prediction pipelines [42, 44]. The Trauma dataset contains 22,704 records from 21,250 unique patients, while General Surgery contains 33,637 records from 25,237 patients. Four prediction tasks are considered: Trauma and General Surgery at 15- and 30-day horizons (TM15, TM30, GS15, and GS30).

Each task is partitioned across five simulated hospital nodes (A–E), yielding heterogeneous record counts and predicted-risk distributions. As shown in Table 2, mean predicted risk ranges from 0.69% to 2.19% for TM15 and from 1.77% to 3.20% for GS30. These are model-predicted probabilities, not observed readmission rates or direct measures of predictive accuracy. Differences may reflect case mix, data characteristics, or model calibration; temporal consistency and calibration against observed outcomes require separate validation.

4.2 Capacity-Derived Local Policies

Each node is assigned a hypothetical escalation-capacity fraction q_i, with $q_A = 0.20$, $q_B = 0.13$, $q_C = 0.15$, $q_D = 0.10$, and $q_E = 0.12$. These settings were not estimated from measured staffing or follow-up resources. For local records D_i with predicted risk $p(x)$, the initial threshold is the empirical $(1 - q_i)$-quantile:

$$\tau_i^{(0)} = Q_{1-q_i}(\{p(x) : x \in D_i\}) . \quad (11)$$

Table 2 – Predicted readmission-risk statistics across the five simulated hospital nodes.

Hospital	Task	Count	Mean (%)	Median (%)	P75 (%)	Maximum (%)
A	TM15	4555	2.19	2.10	2.32	4.47
A	TM30	4555	2.48	2.37	2.58	4.99
A	GS15	6805	2.50	2.43	2.47	7.01
A	GS30	6805	3.20	3.12	3.16	7.46
B	TM15	4539	1.41	1.34	1.51	3.94
B	TM30	4539	1.78	1.70	1.89	3.90
B	GS15	6556	1.87	1.81	1.83	6.15
B	GS30	6556	2.60	2.52	2.55	6.80
C	TM15	4549	1.19	1.12	1.30	2.77
C	TM30	4549	1.04	1.03	1.03	5.62
C	GS15	6840	2.09	1.43	2.23	6.10
C	GS30	6840	2.12	2.07	2.09	6.59
D	TM15	4423	1.01	0.72	0.82	5.47
D	TM30	4423	1.24	0.90	1.06	5.64
D	GS15	6628	1.02	1.00	1.00	5.66
D	GS30	6628	1.77	1.74	1.76	5.10
E	TM15	4638	0.69	0.68	0.68	5.43
E	TM30	4638	1.28	0.94	1.09	5.69
E	GS15	6808	1.19	1.17	1.17	5.69
E	GS30	6808	2.20	2.15	2.17	6.40

A record is eligible for simulated escalation when

$$\Pi_i(x) = \begin{cases} \text{Escalation eligible}, & p(x) \geq \tau_i, \\ \text{Standard-care category}, & p(x) < \tau_i. \end{cases} \quad (12)$$

These categories represent eligibility only; they do not imply that interventions were delivered or clinical decisions automated.

4.3 Execution Environment

The five hospital partitions were processed in separate virtual machines (VM) on a physical host (AMD Ryzen 7 5700G, 64 GB RAM, Windows 11). Each VM used four virtual CPUs, 4 GB RAM, a 32 GB virtual disk, and Debian GNU/Linux 13. The software stack included Python 3.13.5, NumPy, pandas, scikit-learn, imbalanced-learn, SciPy, and XGBoost. CodeCarbon estimated energy use and emissions from software-observed resource utilisation; these were not physical power measurements.
The setup does not establish performance on physical edge devices or independent hospital networks.

4.4 Policy-Calibration Strategies

Four strategies were compared over eight update rounds.

Local-only:

$$\tau_{i,\text{local}}^{(r)} = \tau_i^{(0)}. \quad (13)$$

Direct weighted adoption: nodes adopt the initial weighted reference

$$\tau_{\text{global}}^{(0)} = \sum_{i=1}^{H} w_i \tau_i^{(0)}, \qquad \sum_{i=1}^{H} w_i = 1. \tag{14}$$

Unweighted gradual coordination: nodes update towards the arithmetic mean:

$$\bar{\tau}^{(r)} = \frac{1}{H} \sum_{i=1}^{H} \tau_i^{(r)}, \qquad \tau_i^{(r+1)} = \lambda \tau_i^{(r)} + (1 - \lambda)\, \bar{\tau}^{(r)}. \tag{15}$$

Weighted gradual coordination (FedPACT): at each round,

$$\begin{aligned} \tau_{\text{global}}^{(r)} &= \sum_{i=1}^{H} w_i \tau_i^{(r)}, \\ \tau_i^{(r+1)} &= \lambda \tau_i^{(r)} + (1 - \lambda) \tau_{\text{global}}^{(r)}. \end{aligned} \tag{16}$$

The baseline uses $\lambda = 0.5$; sensitivity is evaluated for $\lambda \in \{0.1, 0.3, 0.5, 0.8\}$.

4.5 Evaluation Metrics

Threshold alignment is measured by the inter-node spread

$$\Delta^{(r)} = \max_i \tau_i^{(r)} - \min_i \tau_i^{(r)}, \tag{17}$$

threshold variance, the rounds required to reduce initial spread by at least 95%,

$$r_{95} = \min\{r : \Delta^{(r)} \le 0.05 \Delta^{(0)}\}, \tag{18}$$

and the maximum single-round adjustment,

$$A_{\max} = \max_{i,r} \left| \tau_i^{(r+1)} - \tau_i^{(r)} \right|. \tag{19}$$

If the 95% target was not reached within eight rounds, it was recorded as not reached.

Threshold-eligible workload is evaluated separately:

$$W_i(\tau) = \sum_{x \in D_i} I\{p(x) \ge \tau\}, \tag{20}$$

where $I\{\cdot\}$ is the indicator function. Workload variation across the five nodes is summarised by its mean, variance, and coefficient of variation,

$$CV_W = \frac{\sigma_W}{\mu_W}. \tag{21}$$

Table 3 – Local threshold policies across hospital nodes

Hospital	Cohort	Records	q_i	15-day Threshold (%)	30-day Threshold (%)	Escalations per Horizon
Hospital A	GS	6805	0.20	2.4824	3.1839	1361
Hospital A	TM	4555	0.20	2.3862	2.6510	911
Hospital B	GS	6556	0.13	1.8747	2.6377	853
Hospital B	TM	4539	0.13	1.6691	2.0942	591
Hospital C	GS	6840	0.15	3.4027	2.1258	1026
Hospital C	TM	4549	0.15	1.4196	1.0339	683
Hospital D	GS	6628	0.10	1.0227	1.8171	663
Hospital D	TM	4423	0.10	1.5158	2.0027	443
Hospital E	GS	6808	0.12	1.1878	2.2458	817
Hospital E	TM	4638	0.12	0.6813	1.6942	557

These metrics count records eligible for escalation, not observed clinical workload or completed interventions.

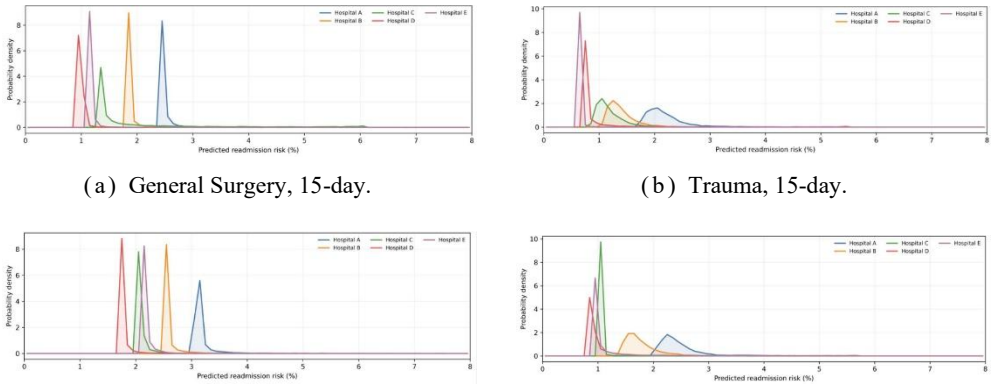
Additional experiments examine the specified stochastic dynamic-capacity model and an in-process communication microbenchmark. The benchmark measures serialised application payloads and local execution time, not network transmission, packet loss, authentication, encryption, or retransmission. Consequently, findings about dynamic behaviour apply only to the simulated capacity and disturbance conditions.

5 Simulation Results and Analysis

FedPACT was evaluated on four readmission-prediction tasksTrauma and General Surgery at 15- and 30-day horizons (TM15, TM30, GS15, GS30)using five simulated hospital nodes. The analysis examines heterogeneous predicted-risk distributions, local capacity-derived thresholds, static and dynamic policy calibration, threshold-eligible workload, and computational overhead. The risk estimates were obtained from prior work [42, 44].

5.1 Risk heterogeneity and local policy derivation

The hospital-level risk distributions were non-IID across sites, cohorts, and prediction horizons. As shown in Table 3, hospital A had the highest mean predicted risk for both General Surgery (2.48% at 15 days and 3.18% at 30 days) and Trauma (2.38% and 2.65%, respectively). Hospital D had the lowest mean General Surgery risk at both horizons (1.02% and 1.82%), while Hospital E had the lowest 15-day Trauma mean (0.68%). These differences imply that a common numerical threshold may select different proportions of records at different hospitals. The estimates are model-predicted probabilities, not observed readmission rates, and do not establish differences in clinical outcomes.



(a) General Surgery, 15-day. (b) Trauma, 15-day.

(c) General Surgery, 30-day. (d) Trauma, 30-day.

Fig. 2 – Predicted readmission-risk distributions across the five simulated hospital nodes.

Each hospital derived a task-specific initial threshold from its local predicted-risk distribution and simulated capacity fraction q_i, using the $(1-q_i)$-quantile. Thresholds differ because the risk distributions and capacity assumptions differ. Ties and quantile-boundary effects were handled through deterministic ranking to approximate the intended local allocation.

5.2 Static federated policy calibration

For each task t, the federated reference threshold was the record-count-weighted mean of the local thresholds:

$$\tau_{G,t} = \frac{\sum_{i=1}^{K} n_{i,t}\tau_{i,t}}{\sum_{i=1}^{K} n_{i,t}}, \tag{22}$$

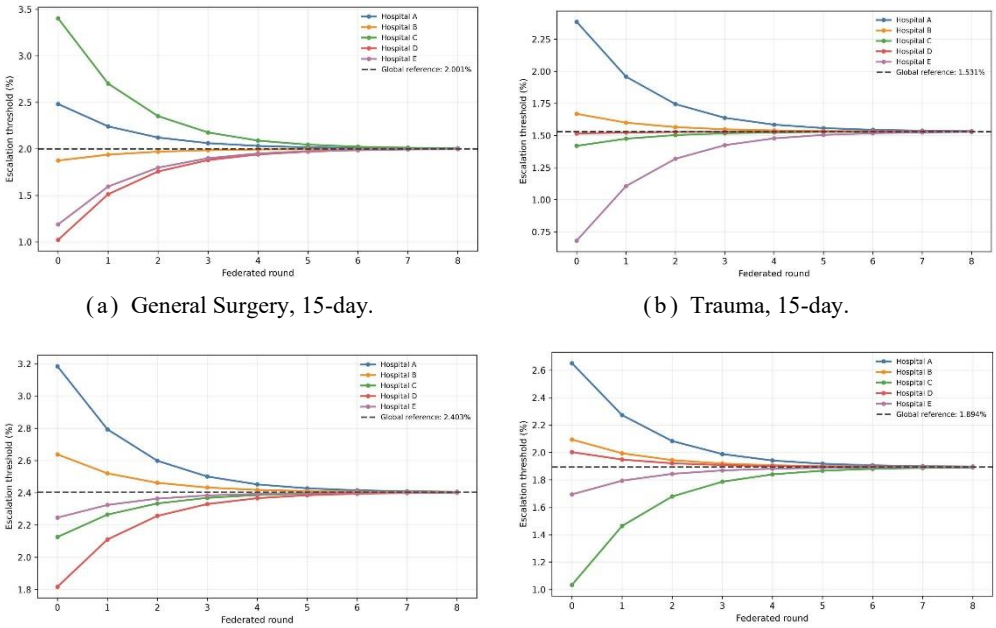
where $n_{i,t}$ is the number of prediction records at node i. Each local threshold was updated over eight rounds using $\lambda = 0.5$:

$$\tau_{i,t}^{(r+1)} = \lambda\tau_{i,t}^{(r)} + (1 - \lambda)\tau_{G,t}. \tag{23}$$

By so doing, patient-level records and predictive-model parameters remained local, while the coordination mechanism exchanged only the aggregate policy information.

Figure 3 shows monotonic movement towards the task-specific reference without overshooting. After eight updates, threshold spreads decreased by approximately 99.61%, consistent with the contraction factor $\lambda^8 = 0.5^8$. The final threshold ranges were 1.5280–1.5347% for TM15, 1.8906–1.8969% for TM30, 1.9976–2.0069% for GS15, and 2.4008–2.4061% for GS30.

Figure 4 compares the locally derived thresholds with the corresponding federated reference. Like the results in Table 3, the figures reveal substantial inter-hospital variation, reflecting the combined influence of non-IID predicted-risk distributions and hospital-specific escalation capacities.



(a) General Surgery, 15-day. (b) Trauma, 15-day.

(c) General Surgery, 30-day. (d) Trauma, 30-day.

Fig. 3 – Convergence of hospital thresholds across federated rounds.

The variance plot, Figure 5, provides a second view of the same convergence process. To reduce redundancy, it is retained as a single representative panel rather than repeating four task-specific variance plots.

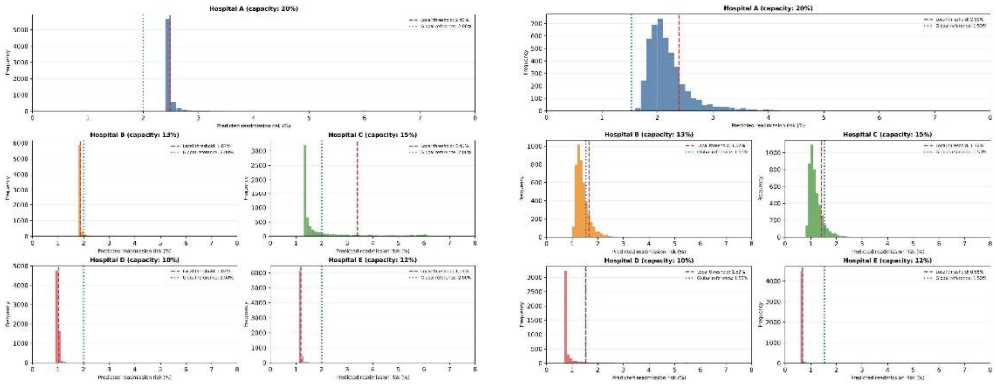
5.3 Threshold-eligible workload and strategy comparison

Threshold convergence does not imply workload equalisation. For node i, task t, and round r, the number of threshold-eligible records is

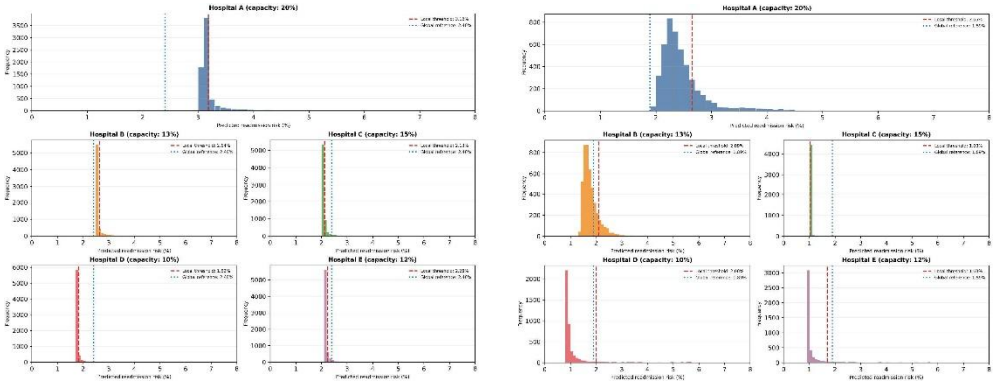
$$W_{i,t}^{(r)} = \sum_{k=1}^{N_{i,t}} I\left(R_{i,t,k} \geq \tau_{i,t}^{(r)}\right), \tag{24}$$

where $R_{i,t,k}$ is the predicted risk for record k. As the local thresholds approach a shared value, non-IID risk distributions can still produce different eligible proportions and workloads.

Table 4 compares local-only, direct weighted adoption, unweighted gradual coordination, and weighted gradual FedPACT after eight rounds. Direct adoption aligned thresholds in one round but produced the largest single-round adjustments. Both gradual strategies reduced initial threshold spread by at least 95% within



(a) General Surgery, 15-day. (b) Trauma, 15-day.



(c) General Surgery, 30-day. (d) Trauma, 30-day.

Fig. 4 – Local and federated thresholds across hospital nodes and prediction horizons.

five rounds, with smaller maximum adjustments. Weighted and unweighted gradual coordination performed similarly because node record counts were relatively balanced.

The workload coefficient of variation increased from approximately 24.68–25.28% under local-only policies to 113.06–140.84% under weighted gradual FedPACT. Total threshold-eligible records also increased substantially, and in five hospital–task combinations every record became eligible because the calibrated threshold fell below the local minimum score. Thus, the method aligns numerical thresholds but does not preserve local capacity constraints or guarantee workload balance.

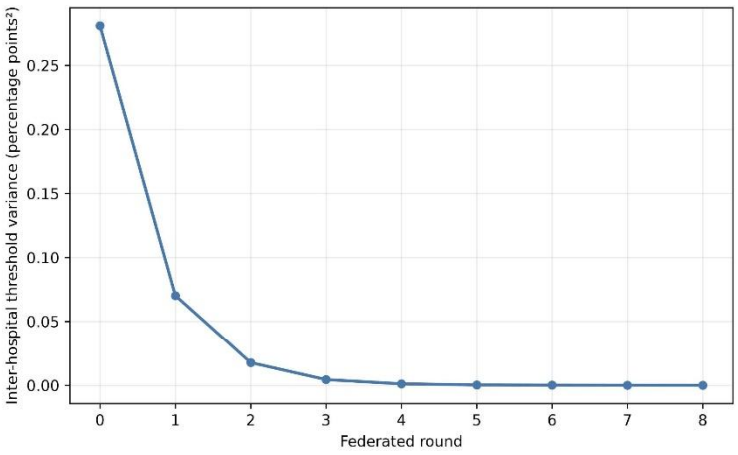


Fig. 5 – Reduction in inter-hospital threshold variance across federated rounds.

Table 4 – Comparison of policy-coordination strategies after eight rounds

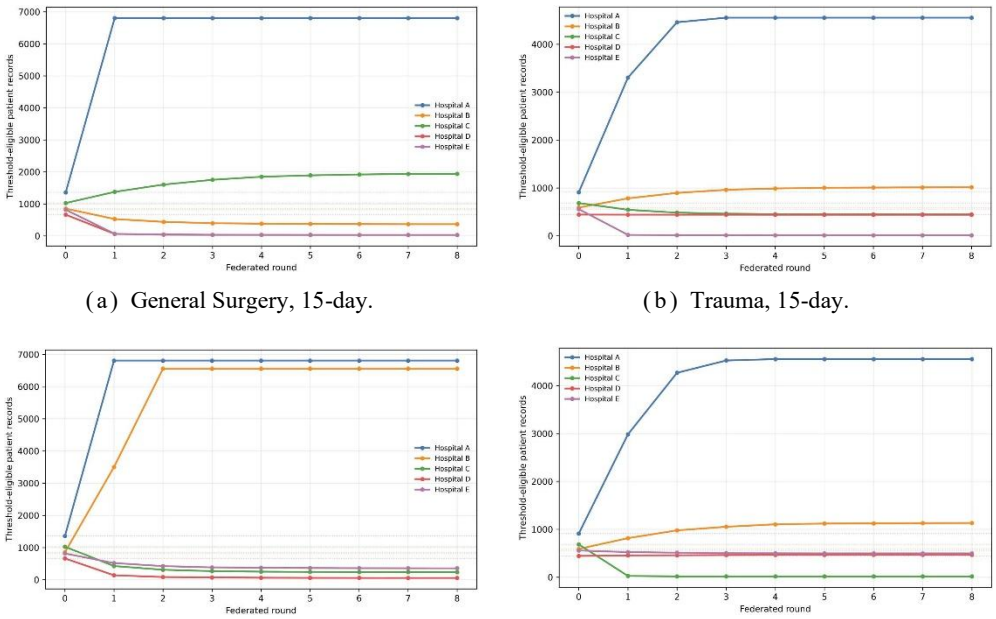
Task	Strategy	$\Delta^{(8)}$ (pp)	$\mathrm{Var}(\tau^{(8)})$	r_{95}	A_{max} (pp)	CV_W	W_{tot}
GS15	Local	2.379940	0.766084	–	0.000000	0.2528	4,721
GS15	Direct-W	0.000000	0.000000	1	1.401221	1.4083	9,174
GS15	Grad-U	0.009297	1.17×10^{-5}	5	0.704305	1.4046	9,191
GS15	FedPACT-W	0.009297	1.17×10^{-5}	5	0.700611	1.4084	9,172
GS30	Local	1.366820	0.221944	–	0.000000	0.2522	4,721
GS30	Direct-W	0.000000	0.000000	1	0.780823	1.1306	14,013
GS30	Grad-U	0.005339	3.39×10^{-6}	5	0.390928	1.1302	14,016
GS30	FedPACT-W	0.005339	3.39×10^{-6}	5	0.390412	1.1306	14,013
TM15	Local	1.704964	0.297026	–	0.000000	0.2468	3,185
TM15	Direct-W	0.000000	0.000000	1	0.854901	1.2843	6,469
TM15	Grad-U	0.006660	4.53×10^{-6}	5	0.426569	1.2889	6,450
TM15	FedPACT-W	0.006660	4.53×10^{-6}	5	0.427450	1.2850	6,466
TM30	Local	1.617120	0.280934	–	0.000000	0.2468	3,185
TM30	Direct-W	0.000000	0.000000	1	0.860016	1.2358	6,673
TM30	Grad-U	0.006317	4.29×10^{-6}	5	0.430644	1.2383	6,661
TM30	FedPACT-W	0.006317	4.29×10^{-6}	5	0.430008	1.2371	6,667

GS15: General Surgery 15-day; GS30: General Surgery 30-day; TM15: Trauma 15-day; TM30: Trauma 30-day. Local: local-only policy; Direct-W: direct weighted-global adoption; Grad-U: unweighted gradual coordination; FedPACT-W: weighted gradual FedPACT coordination. $\Delta^{(8)}$ and A_{max} are in percentage points. W_{tot} is the total number of threshold-eligible records across the five nodes. r_{95} is the number of rounds required to reduce initial spread by 95%; – means this criterion was not reached.

5.4 Dynamic-capacity evaluation

Dynamic capacity was modelled as a bounded periodic trajectory with stochastic perturbations:

$$\tilde{q}_i(t) = q_i^{base} + A\,[\sin(\omega t + \varphi_i) - \sin(\varphi_i)] + \epsilon_i(t),$$
$$q_i(t) = \min\{q_{max}, \max[q_{min}, \tilde{q}_i(t)]\}, \quad (25)$$



(a) General Surgery, 15-day. (b) Trauma, 15-day.

(c) General Surgery, 30-day. (d) Trauma, 30-day.

Fig. 6 – Threshold-eligible escalation demand during federated policy calibration. Demand stabilises as thresholds converge but remains heterogeneous across nodes.

Table 5 – Threshold-eligible escalation demand before and after federated policy calibration.

Prediction task	Before federation (r = 0)			After federation (r = 8)		
	Mean	Variance	CV (%)	Mean	Variance	CV (%)
GS15	944.2	56,963.76	25.28	1,834.4	6,675,208.24	140.84
GS30	944.2	56,684.56	25.22	2,802.6	10,040,781.84	113.06
TM15	637.0	24,706.00	24.68	1,293.2	2,761,408.16	128.50
TM30	637.0	24,706.00	24.68	1,333.4	2,720,813.04	123.71

with $A = 0.02$, $\omega = 2\pi/10$, $T = 20$, $\varphi_i = 2\pi i/H$, $H = 5$, $\epsilon_i(t) \sim N(0, 0.005^2)$, and $0.05 \leq q_i(t) \leq 0.25$. The disturbance was zero at $t = 0$, so all runs began at baseline capacities. For each of ten seeds (42–51), local quantile thresholds were recalculated at every time step and calibrated over eight rounds with $\lambda = 0.5$. Each time step was an independent calibration snapshot; the preceding time step's final thresholds were not carried forward.

Across the ten realisations, final threshold variance and spread remained small after eight updates. Mean total threshold-eligible workloads varied over time from 6,277.5–6,724.2 for TM15, 6,618.1–7,293.8 for TM30, 8,907.3–9,259.7 for GS15, and 13,996.9–14,017.1 for GS30. Workload coefficients of variation remained high

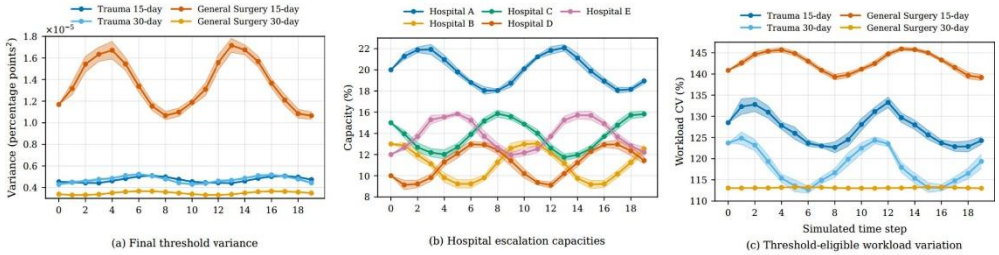
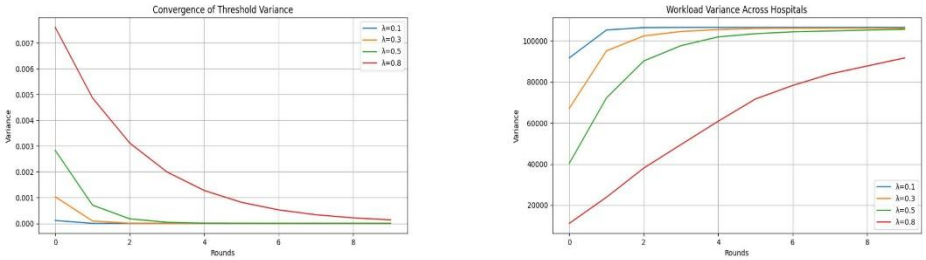


Fig. 7 – Dynamic-capacity results across ten stochastic realisations (seeds 42–51). Solid lines show cross-seed means and shaded regions show 95% confidence intervals: (a) final threshold variance after eight updates; (b) hospital capacity trajectories; and (c) workload coefficient of variation.



(a) Threshold variance for different λ values. (b) Workload variation for different λ values.

Fig. 8 – Sensitivity of threshold alignment and threshold-eligible workload to λ.

(approximately 1.13–1.46 across tasks and time steps). The dynamic experiment therefore supports numerical alignment under the tested periodic-capacity model, but not workload equalisation or robustness to abrupt disruptions and real-world hospital conditions.

5.5 Sensitivity to the coordination parameter

The local-retention parameter was varied over $\lambda \in \{0.1, 0.3, 0.5, 0.8\}$. Smaller values accelerated threshold convergence, while larger values produced more gradual changes in threshold-eligible workload. This smoothness does not imply reduced workload disparity.

5.6 Computational and communication performance

The communication benchmark represented five hospital nodes and a coordinator within one Python process. It used 100 warm-up executions and 2,000 measured repetitions of the four tasks over eight rounds, producing 64,000 task-rounds per execution. The complete local processing cycle remained below one millisecond in

Table 6 – Hospital-level computational performance

Measurement	Hospital A	Hospital B	Hospital C	Hospital D	Hospital E
Real time (s)	512.305	260.001	310.800	290.226	246.837
User CPU time (s)	772.107	641.867	800.952	761.627	605.131
System CPU time (s)	37.955	58.724	60.892	44.595	48.824
Total energy (kWh)	0.00320	0.00253	0.00281	0.00292	0.00223
CPU energy (kWh)	0.00178	0.00182	0.00195	0.00213	0.00156
RAM energy (kWh)	0.00142	0.000709	0.000851	0.000795	0.000676
Emissions (gCO_2e)	2.262	1.788	1.985	2.069	1.579

all rounds. These timings exclude network transmission, authentication, encryption, persistent storage, and model inference; they are not estimates of inter-hospital network latency and do not demonstrate scaling across larger networks.

The local prediction pipeline was also profiled in the five VMs. Table 6 reports runtime and software-estimated resource and environmental measures. The wall-clock runtime ranged from 246.837 s at Hospital E to 512.305 s at Hospital A.

5.7 Summary of results

Across the four tasks, FedPACT consistently reduced inter-hospital threshold dispersion under the tested static and periodic dynamic-capacity conditions. Gradual calibration reduced the size of individual threshold changes relative to direct adoption, while the weighted and unweighted gradual strategies behaved similarly in this relatively balanced five-node setup. Crucially, threshold consensus did not preserve capacity-derived allocations or equalise threshold-eligible workload. The computational benchmark indicates modest local processing but does not establish network performance, formal privacy guarantees, clinical effectiveness, or readiness for deployment.

6 Discussion

6.1 Principal Findings and Implications

FedPACT coordinates operational decision thresholds across five simulated hospital nodes without exchanging patient-level records or predictive-model parameters. Under the evaluated fixed-weight update rule, threshold dispersion decreased as predicted by the convergence analysis; smaller values of the retention parameter λ accelerated alignment, while larger values retained more of each local threshold.

Threshold alignment did not equalise threshold-eligible workloads. Eligible records remained dependent on local patient volumes and predicted-risk distributions, and in some cases a calibrated threshold fell below the minimum local score, making all records eligible. Numerical consensus and workload coordination are therefore distinct objectives; workload balancing would require an explicit capacity- or workload-aware objective and separate evaluation.

The communication benchmark showed low local-processing times in the tested five-node virtualised configuration. Because it ran within one Python process, however, it characterises local coordination costs rather than end-to-end performance across independent hospital networks.

The evaluated configuration is a Digital Shadow under DTMF [43]: it supports analytical processing but does not implement bidirectional feedback or automated clinical control. FedPACT describes cross-institutional policy coordination, not Intelligent Digital Twin maturity. Local prediction, policy calibration, and decision execution remain separate; calibrated thresholds do not automatically initiate clinical actions.

Local data retention and recorded threshold updates support inspectability but do not establish formal privacy, accountability, or clinical adequacy. These require separate operational assessment informed by relevant digital-health and trustworthy-AI guidance [53, 55–57]. Progression towards an Intelligent Digital Twin would require real-time integration, governed feedback, validated clinical decision support, prospective safety assessment, and appropriate human oversight; these capabilities were not evaluated here.

6.2 Answers to the Research Questions

RQ1: Effects of risk distributions and capacity on local thresholds. Capacity-derived thresholds depend on each node's predicted-risk distribution and configured escalation-capacity fraction. Heterogeneous distributions mean that the same capacity fraction need not produce the same numerical threshold. Higher configured capacity generally corresponds to a lower quantile threshold and a larger eligible fraction. The dynamic-capacity experiment likewise showed threshold changes as simulated capacity varied.

RQ2: Convergence and stability. FedPACT reduced threshold dispersion in the evaluated static configurations and repeated dynamic-capacity simulations. Under fixed, non-negative, normalised weights and $0 \leq \lambda < 1$, the analysis establishes convergence to an invariant weighted reference. Smaller λ values produce faster alignment. These results do not establish formal privacy protection or performance over independent hospital networks.

RQ3: Effects on threshold-eligible workload. Threshold convergence did not equalise eligible workloads. Their magnitude and distribution remained dependent on local patient volumes and risk distributions, and workload variation increased under common-reference calibration in several configurations. FedPACT therefore addresses numerical policy alignment, not workload equalisation; capacity preservation or workload fairness requires an additional objective and evaluation.

6.3 Limitations and Future Work

The findings constitute a simulation-based proof of concept with the following limitations:

1. Single-source data and simulated hospitals. Five nodes were formed by partitioning data from Grey's Hospital Trauma Unit in KwaZulu-Natal, South

Africa, rather than using independent data from five hospitals. The partitions cannot represent the full clinical, organisational, demographic, and technical variation across institutions.

2. Precomputed predictions. The study used precomputed readmission-risk estimates; it did not evaluate federated model training, continuous recalibration, or prospective prediction. Predicted probabilities are not observed readmission rates and do not establish predictive accuracy or clinical outcomes.
3. Simplified policy representation. FedPACT calibrated a scalar threshold derived from a local risk distribution and configured capacity fraction. The policy design was informed by an interpretive application of WHO-INTEGRATE [53], not a validated operationalisation. Real escalation decisions may also depend on staffing, bed availability, comorbidities, care pathways, patient preferences, and changing demand.
4. Restricted dynamic conditions. The dynamic experiment used bounded periodic capacity with Gaussian disturbances over 20 time steps and ten seeds. Abrupt demand surges, sustained capacity loss, delayed updates, node failure, and communication interruption were not tested; stability claims are limited to the evaluated conditions.
5. Coordination assumptions. The convergence analysis assumes fixed, non-negative, normalised weights within a calibration episode and availability of local thresholds. Coordinator failure, incomplete participation, asynchronous updates, and malicious or erroneous contributions were not evaluated. The proof does not cover arbitrary changes in weights.
6. Virtualised infrastructure and communication. All five VMs ran on one physical host. The benchmark measured application payload and local processing in one Python process, not network transmission, authentication, encryption, packet loss, retransmission, or performance across independent physical edge devices and hospital networks.
7. Estimated environmental impact. CodeCarbon estimates, rather than physical power-meter measurements, were used for energy and emissions. They do not reproduce the thermal, processor, input/output, or power characteristics of physical single-board computers.
8. Clinical validity, privacy, and deployment. No automated intervention, closed-loop feedback, or prospective clinical use was implemented. Data locality alone is not a formal privacy guarantee. The results do not establish clinical effectiveness, workload fairness, regulatory compliance, or deployment readiness.

Future work should use independently collected multi-institutional data and physically distributed infrastructure; investigate capacity- and workload-aware objectives, changing weights, asynchronous coordination, and failure conditions; and assess richer but interpretable policies. Any progression towards feedback-driven Intelligent Digital Twins would require real-time integration, prospective clinical validation, governance, and human oversight.

7 Conclusion

FedPACT was formulated and evaluated as a policy-level federation mechanism that coordinates operational decision thresholds rather than patient-level records or predictive-model parameters. Under fixed, non-negative, normalised weights and $0 \leq \lambda < 1$, the analysis establishes convergence to an invariant weighted reference, preservation of threshold bounds, and geometric reduction in threshold dispersion.

Experiments across four surgical readmission-prediction tasks and five simulated hospital nodes showed threshold alignment under the tested static and dynamic-capacity conditions. The retention parameter λ controlled convergence speed, but threshold alignment did not equalise threshold-eligible workloads, which remained dependent on local patient volumes and risk distributions. Workload balancing must therefore be treated as a separate coordination objective.

The scalar mechanism had a linear communication requirement in the number of participating hospitals, and the in-process benchmark showed low local-processing overhead in the tested configuration. These measurements do not establish end-to-end network performance, formal privacy guarantees, or clinical effectiveness. The implementation remains a Digital Shadow without automated clinical intervention or bidirectional control. Multi-institutional validation, distributed deployment testing, appropriate governance, and human oversight are necessary before operational use.

Statements and Declarations

- Funding: The research reported in this article was supported by the South African Medical Research Council (SAMRC) with funds received from the Department of Science and Innovation. The content and findings reported and/or illustrated are the sole deduction, view, and responsibility of the researcher and do not reflect the official position and sentiments of SAMRC or the Department of Science and Innovation.
- Conflict of interest/Competing interests: Not applicable
- Ethics approval and consent to participate: Not applicable
- Consent for publication: All authors gave their consent.
- Data availability: Data used are available in Grey's Hospital private repository.
- Materials availability: Not applicable
- Code availability: all codes are available on the first author's github repository.
- Author contribution: PS: Conceptualization, Investigation, Writing - Original draft preparation; AB: Supervision, Validation, Writing - Review & Editing; OA: Methodology, Visualisation, Writing - Original draft preparation; FK: Formal analysis, Validation; LM: Methodology, Software; JC: Resources, Funding & grant acquisition, Writing - Review & Editing; DC: Data Curation, Resources; JM: Validation, Writing - Review & Editing

References

[1] B. Savitsky et al., “Risk factors for rehospitalisation after trauma: a follow-up study,” European J. Trauma and Emergency Surgery, 2025.

[2] D. Bertsimas et al., “From predictions to prescriptions in healthcare: recent advances in predictive analytics,” Nature Med., vol. 30, 2024.

[3] A. Rajkomar et al., “Machine learning in medicine: opportunities and challenges for predictive healthcare,” Nature Med., 2024.

[4] E. Topol, “The convergence of human and artificial intelligence in medicine,” Nature Med., vol. 25, no. 1, pp. 44056, 2019.

[5] C. Krittanawong et al., “Bridging predictive AI and clinical implementation in healthcare systems,” npj Digital Med., 2025.

[6] B. Tudor et al., “A scoping review of human digital twins in healthcare applications and usage patterns,” npj Digital Med., vol. 8, no. 1, 2025.

[7] E. Katsoulakis et al., “Digital twins for health: a scoping review". NPJ digital medicine, vol. 7, no. 1, pp. 77, 2024

[8] H. Kopetz and W. Steiner, “Temporal Consistency of Data and Information in Cyber-Physical Systems,” arXiv preprint arXiv:2409.19309, 2024.

[9] G. Vidyalakshmi et al., “Digital Twins and Cyber-Physical Systems: A New Frontier in Computer Modeling,” Comp. Modeling in Engineering & Sci., 2025.

[10] B. Liu et al., “Recent advances on federated learning: A systematic survey,” Neurocomputing, vol. 597, Art. no. 128019, 2024.

[11] M. Li et al., “From challenges and pitfalls to recommendations and opportunities: Implementing federated learning in healthcare,” Medical Image Analysis, vol. 101, Art. no. 103497, 2025.

[12] F. Sabah et al., “Model optimization techniques in personalized federated learning: A survey,” Expert Systems with Applications, vol. 243, Art. no. 122874, 2024.

[13] M. Luo et al., “No fear of heterogeneity: Classifier calibration for federated learning with non-IID data,” in Advances in Neural Info. Processing Sys., vol. 34, pp. 5972–5984, 2021.

[14] C. Lu et al., “Federated conformal predictors for distributed uncertainty quantification,” in Proc. 40th Int. Conf. on Machine Learning, ser. Proceedings of Machine Learning Research, vol. 202, pp. 22942–22964, 2023.

[15] S. Maddock, G. Cormode, and C. Maple, “Private federated multiclass post-hoc calibration,” arXiv preprint arXiv:2510.01987, 2025.

[16] M. Singh et al., “Digital Twin: Origin to Future,” Applied Sys. Innov., vol. 4, no. 2, pp. 36, 2021.

[17] F. Fornari et al., “Digital Twins of Business Processes: A Research Manifesto,” Internet of Things, vol. 30, p. 101477, 2025.

[18] K. Iranshahi et al., “Digital Twins: Recent Advances and Future Directions in Engineering Fields,” Intelligent Systems with Applications, vol. 26, p. 200516, 2025.

[19] S. Sepasgozar, “Differentiating Digital Twin from Digital Shadow,” Buildings, vol. 11, no. 4, p. 151, 2021.

[20] J. Ehrhardt and C. Hoffmann, “The Digital Shadow: Developing a universal model for the automated optimization of cyber-physical production systems based on real-time data,” Procedia CIRP, no. 93, pp. 304-310, 2020.

[21] M. Liebenberg and M. Jarke, “Information systems engineering with Digital Shadows: Concept and use cases in the Internet of Production,” Info. Sys, vo. 114, no. 102182, 2023.

[22] F. Flammini, “Digital Twins as Run-Time Predictive Models for CPS,” Philosophical Trans. of the R.S. A, vol. 379, no. 2207, p. 20200369, 2021.

[23] A. Ricci et al., “Pervasive and connected digital twinsa vision for digital health,” IEEE Internet Computing, vol. 25, no. 5, pp. 26-32, 2021.

[24] E. Katsoulakis et al., “Digital Twins for Health: A Scoping Review,” npj Digital Med., vol. 7, no. 1, p. 77, 2024.

[25] P. Armeni et al., “Digital twins in healthcare: is it the beginning of a new era of evidence-based medicine? A critical review,” J. Personalized Med., vol. 12, no. 8, p. 1255, 2022.

[26] S. Elkefi and O. Asan, “Digital twins for managing health care systems: rapid literature review,” J. Medical Internet Research, vol. 24, no. 8, e37641, 2022.

[27] M. Cellina et al., “Digital Twins: The New Frontier for Personalized Medicine?,” Applied Sciences, vol. 13, no. 13, p. 7940, 2023.

[28] K. Zhang et al., “Concepts and Applications of Digital Twins in Healthcare and Medicine,” Patterns, vol. 5, no. 8, 2024.

[29] A. Vallée, “Digital Twin for Healthcare Systems,” Frontiers in Digital Health, vol. 5, p. 1253050, 2023.

[30] G. Pellegrino et al., “A Conceptual Framework for Digital Twin in Healthcare: evidence from a systematic meta-review,” Info. Sys. Frontiers, vol. 27, no. 1, pp. 7-32, 2024.

[31] C. Rebello et al., “Digital Twin Framework for Optimal Decision-Making in CPS,” arXiv, 2023.

[32] K. P. Venkatesh et al., “Health digital twins as tools for precision medicine: Considerations for computation, implementation, and regulation,” npj Digital Med., vol. 5, no. 1, p. 150, 2022.

[33] A. Parnianifard et al., “Digital Twins Towards Cyber-Physical Systems Systems,” In Foundations of Digital Twins, pp. 63-75, 2022.

[34] A. Carbonaro et al., “Corrigendum: CONNECTED: leveraging digital twins and personal knowledge graphs in healthcare digitalization,” Frontiers in Digital Health, vol. 6, p. 1416390, 2024.

[35] H. Suresh et al., “A Framework for Accountability in Healthcare AI,” Communications of the ACM, 2020.

[36] G. Appuhamilage et al., “DT4CCU: A Digital Twin Framework for Critical Care Units,” Research Square, 2024.

[37] K. Rudsari et al., “Digital twins in healthcare: a comprehensive review and future directions,” Frontiers in Digital Health, vol. 7, p. 1633539, 2025.

[38] Z. Elgammal, M. T. Albrijawi, and R. Alhajj, “Digital twins in healthcare: a review of AI-powered practical applications across health domains,” J. Big Data, vol. 12, no. 1, p. 234, 2025.

[39] F. Wilson et al., “Foundations for Digital Twins,” arXiv preprint arXiv:2405.00960, 2024.

[40] R. Walton, F. W. Ciarallo, and L. Champagne, “A unified digital twin approach incorporating virtual, physical, and prescriptive analytical components to support adaptive real-time decision-making,” Comp. & Industrial Engineering, vol. 193, p. 110241, 2024.

[41] B. Tudor et al., “A scoping review of human digital twins in healthcare applications and usage patterns,” npj Digital Med., vol. 8, no. 1, 2025.

[42] A. Bagula et al. "When 99.98% is too Good to be True: Preventing Rule-Induced Overfitting in Embodied Clinical AI for Surgical Readmission Prediction." 2026 ITU Kaleidoscope - AI and Frontier Technologies for Good (ITU K), Geneva, Switzerland, 2026, pp. 1-8, doi: 10.23919/ITUK70038.2026.11632583.

[43] P. Sello et al., “From Digital Models to Intelligent Twins: A Maturity-Driven CPS Framework for Trauma Readmission in Healthcare,” Int. Conf. on ICTs for eHealth (ICTS4eHealth 2026), In Press.

[44] L. Mbale, et al., “A Hybrid Regex-Transformer Model for Enhanced Readmission Risk Detection from Clinical Text,” Int. Conf. on ICTs for eHealth (ICTS4eHealth 2026), In Press.

[45] V. Schmidt, et al., “CodeCarbon: Estimate and Track Carbon Emissions from Machine Learning Computing,” 2021, https://github.com/mlco2/codecarbon, doi:10.5281/zenodo.4658424.

[46] M. Görtz et al., “Digital twins for personalized treatment in uro-oncology in the era of artificial intelligence,” Nature Reviews Urology, vol. 23, no. 1, pp. 29-39, 2026.

[47] M. Xames et al., “Evaluating CFIR 2.0 in identifying digital twin implementation challenges in healthcare: bridging the dichotomy between engineering and healthcare communities,” Frontiers in Digital Health, vol. 7, p. 1611225, 2025.

[48] Z. Zhou et al., “Synthetic Anatomy: Deep Learning Models for Virtual Population GenerationA Review,” medRxiv, 2025-10, 2025.

[49] M. Tortora et al., “Medical Digital Twin: A Review on Technical Principles and Clinical Applications,” J. Clinical Med., vol. 14, no. 2, p. 324, 2025.

[50] R. Pan et al., “Human Digital Twin: Data, Models, Applications, and Challenges,” arXiv preprint arXiv:2508.13138, 2025.

[51] P. Gökta and A. Grzybowski, “Shaping the Future of Healthcare: Ethical Clinical Challenges and Pathways to Trustworthy AI,” J. Clinical Med., vol. 14, no. 5, p. 1605, 2025.

[52] R. Shah, K. Bozic, and P. Jayakumar, “Artificial Intelligence in Value-Based Health Care,” HSS J., vol. 21, no. 3, pp. 307-313, 2025.

[53] E. Rehfuess, et al., “The WHO-INTEGRATE evidence to decision framework version 1.0: integrating WHO norms and values and a complexity perspective,” BMJ Global Health, vol. 4, no. 1, e000844, 2019.

[54] B. McMahan, et al., “Communication-efficient learning of deep networks from decentralized data,” in Proc. 20th Int. Conf. on Artificial Intelligence and Statistics, vol. 54, pp. 1273–1282, PMLR, 2017.

[55] World Health Organisation, “Global Strategy on Digital Health 2020–2025,” WHO, Geneva, Switzerland, 2021.

[56] World Health Organisation, “Ethics and Governance of Artificial Intelligence for Health: large multi-modal models,” WHO, Geneva, Switzerland, 2024.

[57] Int. Telecommunication Union, “AI Standards for Global Impact: From Governance to Action,” ITU, Geneva, Switzerland, 2025.