

FL-EHDS: A Privacy-Preserving Federated Learning Framework for the European Health Data Space

Fabio Liberti

Department of Computer Science
Universitas Mercatorum, Rome, Italy
fabio.liberti@unimercatorum.it
ORCID: 0000-0003-3019-5411

Abstract—The European Health Data Space (EHDS), established by Regulation (EU) 2025/327, mandates cross-border health data analytics while preserving citizen privacy. Federated Learning (FL) is the key enabling technology for secondary use, yet only 23% of FL implementations achieve sustained production deployment in healthcare. We present FL-EHDS, a three-layer compliance framework integrating governance mechanisms (Health Data Access Bodies, data permits, opt-out registries), FL orchestration (17 aggregation algorithms including 2024–2025 advances, differential privacy, secure aggregation), and data holder components (adaptive training, FHIR preprocessing). Experimental validation on real clinical datasets demonstrates that personalized FL (Ditto) narrows the centralized-federated gap to 6.6 percentage points while preserving full data sovereignty, and that algorithm choice produces up to 18.7pp accuracy differences on heterogeneous clinical data. Our systematic evidence synthesis of 47 documents reveals that legal uncertainties—not technical barriers—constitute the critical blocker for FL adoption in EHDS contexts. The open-source reference implementation and compliance mapping provide actionable guidance for the 2025–2031 transition period.

Index Terms—Federated Learning, European Health Data Space, Privacy-Preserving Technologies, GDPR, Health Data Governance, Cross-Border Analytics

I. INTRODUCTION

The European Health Data Space (EHDS), established by Regulation (EU) 2025/327, represents the EU’s most ambitious initiative for cross-border health data governance [1]. Entering into force in March 2025, the regulation creates a dual framework: primary use through MyHealth@EU for patient care, and secondary use through HealthData@EU for research, innovation, and policy-making [11]. Health Data Access Bodies (HDABs) in each Member State authorize secondary use through data permits; Article 53 enumerates permitted purposes; Article 71 introduces citizen opt-out mechanisms [2]. The implementation timeline extends to 2031, with delegated acts expected by March 2027 and secondary use provisions applicable from March 2029.

Federated Learning (FL) emerges as the ideal technical solution for EHDS secondary use—the model travels to distributed data rather than centralizing sensitive records [12], [14]. The COVID-19 pandemic demonstrated FL’s potential at scale: Dayan et al. [20] trained a global model across 20 institutions in 5 countries. However, recent evidence reveals a sobering gap between FL’s promise and operational reality. Fröhlich

et al. [5] report that only 23% of FL implementations achieve sustained production deployment, with hardware heterogeneity (78%) and non-IID data distributions (67%) as dominant barriers. Beyond technical constraints, legal uncertainties regarding gradient data status under GDPR remain unresolved [3], while van Drumpt et al. [6] demonstrate that privacy-enhancing technologies cannot substitute for robust governance frameworks.

Prior FL frameworks for healthcare [14], [21] focus on technical architectures without addressing regulatory compliance. Legal analyses [2], [3] examine GDPR constraints but abstract from implementation feasibility. Policy documents [4] assess Member State readiness but do not integrate FL technical considerations. No existing work provides an integrated framework addressing all three dimensions: systematic barrier evidence, technical implementation with state-of-the-art algorithms, and EHDS governance operationalization.

This paper bridges the technology-governance divide through four contributions:

- 1) **Barrier Taxonomy**: Systematic evidence synthesis of 47 documents using PRISMA methodology with GRADE-CERQual confidence assessment.
- 2) **FL-EHDS Framework**: A three-layer reference architecture mapping barriers to governance-aware mitigation strategies.
- 3) **Reference Implementation**: Open-source Python codebase (~40K lines) with 17 FL algorithms (2017–2025) and EHDS governance modules.¹
- 4) **Experimental Validation**: Evaluation on real clinical datasets demonstrating that algorithm selection produces 18.7pp accuracy differences and personalized FL narrows the centralized-federated gap to 6.6pp.

II. BACKGROUND AND RELATED WORK

A. EHDS and Federated Learning

The EHDS establishes HDABs to authorize secondary use through standardized data permits, with Secure Processing Environments (SPEs) providing controlled analytics settings [9]. Forster et al. [8] document significant variability in data access timelines—from 3 weeks (Finland) to over 12 months (France)—with barriers primarily organizational rather than technical. TEHDAS assessments [4] reveal Nordic countries

¹ Available at: <https://github.com/FabioLiberti/FL-EHDS-FLICS2026>

demonstrate 2–3 year advantages in HDAB capacity-building, raising concerns about implementation equity. Teo et al. [16] find that only 5.2% of FL healthcare studies achieve real-life application.

FL inverts the traditional ML paradigm: local training produces gradients that are aggregated centrally and redistributed [12], [13]. Known challenges include non-IID data distributions causing convergence difficulties [13], communication costs for gradient exchange [15], and privacy attacks including gradient inversion [17] and membership inference [18]. Recent advances from top venues (ICML/ICLR 2022–2025) specifically target healthcare heterogeneity: FedLC [26] calibrates logits for label distribution skew, FedLESAM [30] provides globally-guided sharpness-aware optimization (ICML 2024 Spotlight), and HPFL [31] decouples backbone from classifier for per-institution specialization (ICLR 2025).

B. Related Frameworks

Existing FL frameworks—Flower [32] (v1.26), NVIDIA FLARE [33] (v2.7), and TensorFlow Federated [34] (v0.88)—provide robust distributed training but lack EHDS-specific governance: none implements HDAB integration, Article 53 data permit lifecycle, Article 71 opt-out enforcement, or GDPR Article 30 audit trails. Table I provides a detailed comparison.

C. Evidence Synthesis

Following PRISMA 2020 guidelines, database searches (PubMed, IEEE Xplore, Scopus, Web of Science, arXiv) identified 847 records; 47 met inclusion criteria (2022–2026, FL/EHDS focus, peer-reviewed or recognized institutional origin). Quality was assessed using MMAT; confidence using GRADE-CERQual. Table II summarizes the five dominant barriers with prevalence and mitigation strategies.

Three critical legal questions remain unresolved [3]: (1) whether model gradients constitute “personal data” under GDPR, given that gradient inversion attacks demonstrate potential re-identification [17]; (2) when aggregated models become sufficiently “anonymous” to escape GDPR scope; (3) controller/processor allocation in multi-party FL architectures. These legal uncertainties create compliance risks that discourage organizational adoption regardless of technical maturity (GRADE-CERQual: MODERATE).

III. FL-EHDS FRAMEWORK

Based on the identified barriers, we present FL-EHDS, a three-layer compliance framework for EHDS cross-border health analytics. Figure 1 illustrates the architecture.

A. Layer 1: Governance

Standardized APIs enable automated data permit verification before FL training initiation. Multi-HDAB synchronization protocols coordinate cross-border studies involving multiple Member States, addressing the coordination complexity identified by Christiansen et al. [10]. National opt-out registries

are consulted before each training round, ensuring Article 71 compliance at record-level granularity. Comprehensive audit trails satisfy GDPR Article 30 requirements, documenting data access, processing purposes, and model outputs for regulatory inspection.

Algorithm 1 presents the core FL-EHDS training procedure, highlighting governance checkpoints integrated into each round.

B. Layer 2: FL Orchestration

The framework implements **17 aggregation algorithms** spanning six categories: *Baseline* (FedAvg [12]); *Non-IID robustness* (FedProx [13], SCAFFOLD [22], FedNova [23], FedDyn); *Adaptive optimization* (FedAdam, FedYogi, FedAdagrad [24]); *Personalization* (Ditto, Per-FedAvg); *Recent advances 2022–2023* (FedLC [26] for label distribution skew, FedSAM [25] for flat minima, FedDecorr [27] against dimensional collapse, FedSpeed [28] for communication efficiency, FedExP [29] for server-side acceleration); *Latest advances 2024–2025* (FedLESAM [30] for globally-guided sharpness awareness, HPFL [31] for personalized classifier federation).

Table III summarizes all 17 algorithms with their venues and key properties. FedLESAM extends FedSAM by replacing local gradient perturbation with a globally-estimated direction ($\theta_{\text{global}}^{(t-1)} - \theta_{\text{global}}^{(t)}$), achieving stronger generalization across heterogeneous institutions. HPFL decouples feature extraction from classification by aggregating only backbone parameters while keeping client-specific classifier heads local, enabling each hospital to maintain specialized decision boundaries. Algorithm selection is configurable; FedLC and FedDecorr are composable with any aggregation strategy.

Privacy Protection: Differential privacy with configurable ϵ -budget uses Rényi DP (RDP) [19] for tight composition accounting over multiple training rounds. For Gaussian mechanisms with noise scale σ , the RDP guarantee at order α is $\rho(\alpha) = \alpha/(2\sigma^2)$. For 100+ round training typical of EHDS cross-border studies, RDP provides 5–6 \times tighter privacy bounds than naive composition, enabling longer training with equivalent privacy guarantees. Gradient clipping bounds individual contributions; secure aggregation (pairwise masking protocol with ECDH key exchange) mitigates gradient inversion attacks [17]. Six Byzantine resilience methods (Krum, Multi-Krum, Trimmed Mean, Median, Bulyan, FLTrust) defend against up to $f < n/3$ malicious clients.

Purpose Limitation: Technical enforcement of Article 53 permitted purposes through model output filtering and use-case validation, preventing scope creep beyond authorized analytics.

C. Layer 3: Data Holders

Resource-aware training engines address hardware heterogeneity (78% barrier prevalence). The engine dynamically adjusts batch sizes, model complexity, and synchronization frequency based on local computational capabilities, enabling participation of institutions with diverse hardware profiles—

TABLE I
FRAMEWORK COMPARISON: FL-EHDS VS EXISTING FL FRAMEWORKS

Dimension	FL-EHDS	Flower v1.26	NVIDIA FLARE v2.7	TFF v0.88
FL Algorithms	17 built-in	12+ strategies	5 built-in	3 built-in
Byzantine Resilience	6 methods	4 methods	—	—
Differential Privacy	Central + Local DP	Central + Local DP	Built-in	Adaptive clipping
Secure Aggregation	Pairwise + HE	SecAgg+	Built-in + HE	Mask-based
EHDS Governance	Full	None	None	None
HDAB Integration	✓	—	—	—
Data Permits (Art. 53)	✓	—	—	—
Opt-out Registry (Art. 71)	✓	—	—	—
Audit Trail (GDPR Art. 30)	✓	—	Audit logs	—
Healthcare Standards	FHIR R4	MONAI	MONAI	—
Backend	PyTorch	Agnostic	Agnostic	TensorFlow only

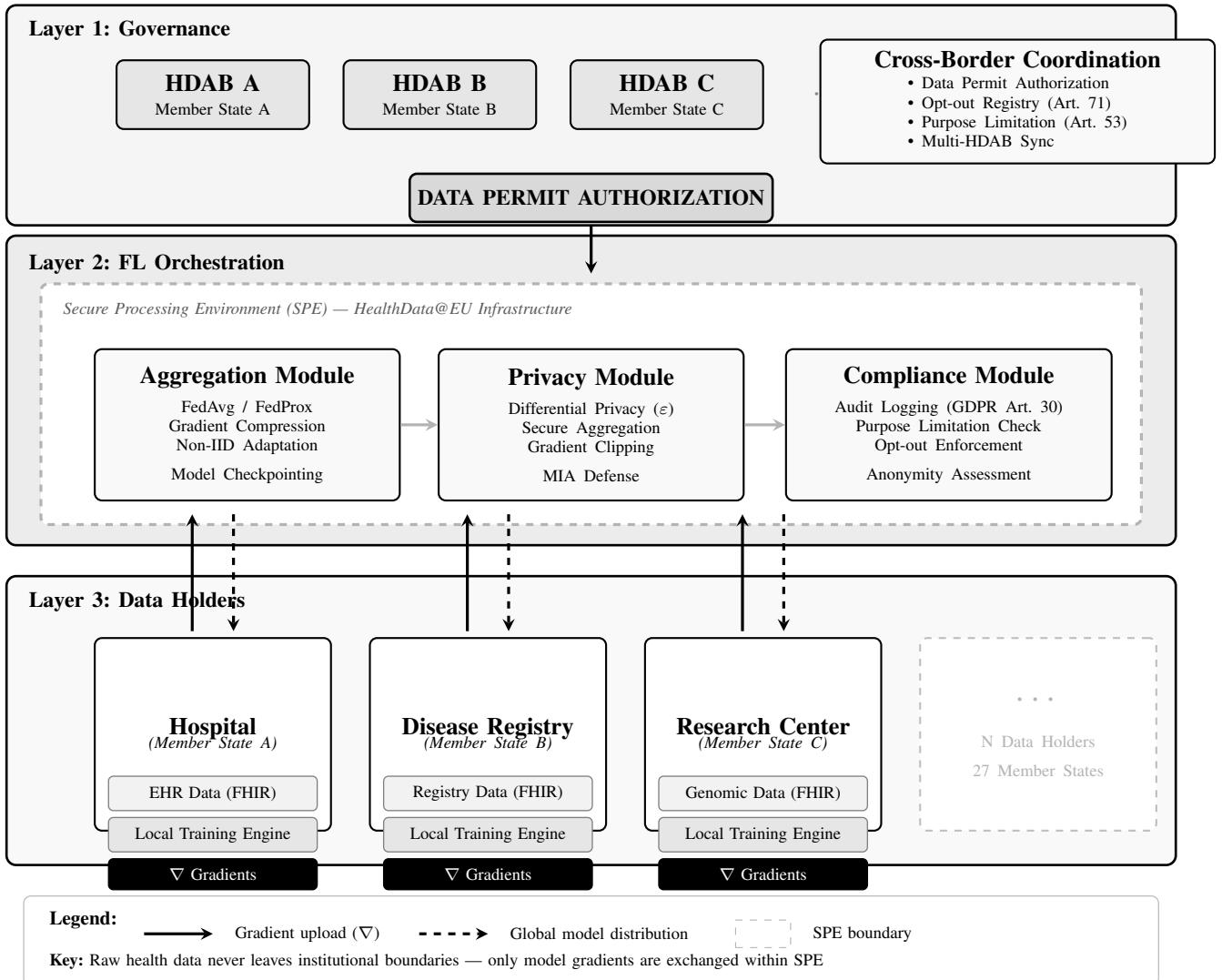


Fig. 1. FL-EHDS three-layer compliance framework architecture. Layer 1 (Governance) integrates Health Data Access Bodies for cross-border data permit authorization and opt-out registry consultation per Article 71. Layer 2 (FL Orchestration) operates within a Secure Processing Environment, implementing gradient aggregation with FedAvg/FedProx, privacy protection via differential privacy and secure aggregation, and GDPR-compliant audit logging. Layer 3 (Data Holders) maintains raw data within institutional boundaries across 27 Member States; only gradients (∇) are transmitted upward while global model parameters flow downward.

TABLE II
FL IMPLEMENTATION BARRIERS FOR EHDS

Barrier	Prev.	Evidence	Mitigation
Hardware heterog.	78%	Fröhlich 2025	Adaptive engine
Non-IID data	67%	Multiple	FedProx, Ditto
Production gap	23%	Fröhlich 2025	Ref. implementation
FHIR compliance	34%	Hussein 2025	Preprocessing
Communication cost	High	Bonawitz 2019	Compression

Algorithm 1: FL-EHDS FedAvg Training

Input: Hospitals $\mathcal{H} = \{h_1, \dots, h_K\}$, permit P , rounds T
Output: Global model $\theta^{(T)}$

Server executes:

```

Initialize  $\theta^{(0)}$ 
for round  $t = 1$  to  $T$  do
    // Governance check (Layer 1)
    if not ValidatePermit( $P, t$ ) then abort
     $\mathcal{H}_t \leftarrow$  SelectParticipants( $\mathcal{H}$ )
    for each  $h \in \mathcal{H}_t$  in parallel do
         $\Delta_h^{(t)}, n_h \leftarrow$  LocalTrain( $h, \theta^{(t-1)}$ )
    // Aggregation with privacy (Layer 2)
     $\theta^{(t)} \leftarrow \theta^{(t-1)} + \frac{1}{\sum n_h} \sum_h n_h \cdot \Delta_h^{(t)}$ 
    LogCompliance( $t, \mathcal{H}_t$ )
LocalTrain( $h, \theta$ ):
     $\mathcal{D}_h \leftarrow$  FilterOptedOut( $\mathcal{D}_h$ , Registry) // Art. 71
     $\theta_h \leftarrow \theta$ ; train  $E$  epochs on  $\mathcal{D}_h$ 
     $\Delta_h \leftarrow$  ClipGradient( $\theta_h - \theta, C$ ) // DP bound
    return  $\Delta_h, |\mathcal{D}_h|$ 

```

TABLE III
FL-EHDS ALGORITHM CATALOGUE (17 ALGORITHMS)

Algorithm	Venue	Category	Key Property
FedAvg	AISTATS'17	Baseline	Weighted avg.
FedProx	MLSys'20	Non-IID	Proximal reg.
SCAFFOLD	ICML'20	Non-IID	Variance red.
FedNova	NeurIPS'20	Non-IID	Normalized avg.
FedDyn	ICLR'21	Non-IID	Dynamic reg.
FedAdam	ICLR'21	Adaptive	Server momentum
FedYogi	ICLR'21	Adaptive	Sparse stability
FedAdagrad	ICLR'21	Adaptive	Grad. accum.
Ditto	ICML'21	Personal.	Dual models
Per-FedAvg	NeurIPS'20	Personal.	MAML-based
FedLC	ICML'22	Label skew	Logit calibration
FedSAM	ICML'22	Generalize	Flat minima
FedDecor	ICLR'23	Represent.	Decorrelation
FedSpeed	ICLR'23	Efficiency	Fewer rounds
FedExP	ICLR'23	Server-side	POCS step size
FedLESAM	ICML'24	Generalize	Global SAM
HPFL	ICLR'25	Personal.	Local classif.

Bold: newly added algorithms (2024–2025). All 17 implemented in the open-source reference implementation.

from GPU-equipped university hospitals to CPU-only rural clinics.

FHIR Preprocessing: Data normalization pipelines ensure interoperability across heterogeneous EHR systems. Only 34% of European healthcare providers achieve full FHIR compliance [7]; the preprocessing module bridges format gaps through automated transformation pipelines supporting FHIR R4 resources (Patient, Observation, Condition, MedicationRequest, DiagnosticReport) with standard coding systems (SNOMED-CT, LOINC, ICD-10).

Secure Communication: End-to-end encrypted gradient transmission with certificate-based authentication ensures no raw data leaves institutional boundaries. The communication layer supports gRPC for model updates and WebSocket for real-time monitoring events.

D. Threat Model

The framework assumes an honest-but-curious aggregation server. Byzantine tolerance is provided for up to $f < n/3$ malicious clients through robust aggregation (Krum, Trimmed Mean, Bulyan). Gradient inversion is mitigated by DP and secure aggregation.

E. EHDS Compliance Mapping

Table IV maps framework components to EHDS regulatory requirements.

TABLE IV
EHDS COMPLIANCE MAPPING

Article	Requirement	FL-EHDS Component
Art. 33	Secondary use auth.	HDAB API + Permit valid.
Art. 46	Cross-border proc.	Multi-HDAB coordinator
Art. 50	Secure Proc. Env.	Aggregation within SPE
Art. 53	Permitted purposes	Purpose limitation module
Art. 71	Opt-out mechanism	Registry filtering

F. Reference Implementation

A modular Python implementation is available as open-source software. The codebase (~40K lines, 159 modules) provides: (1) orchestration modules implementing all 17 algorithms with RDP accounting and secure aggregation; (2) six Byzantine resilience methods; (3) data holder utilities for adaptive training and FHIR R4 preprocessing; (4) a Streamlit-based dashboard for interactive FL training, EHDS governance workflow, and real-time monitoring; (5) a professional terminal UI with 11 specialized screens; (6) reproducible benchmark suite generating all experimental results.

Note on governance: HDAB integration includes a fully functional simulation backend demonstrating the complete permit lifecycle (OAuth2/mTLS authentication, permit CRUD,

cross-border coordination) and Article 71 opt-out compliance (LRU-cached registry lookups, scope-granular filtering). Production deployment will require binding to actual HDAB services (expected 2027–2029).

IV. EXPERIMENTAL EVALUATION

We evaluate FL-EHDS on real clinical datasets simulating cross-border healthcare analytics. All results are fully reproducible via the benchmark suite in the repository.

A. Setup

Datasets: (1) *Heart Disease UCI* (920 patients from 4 international hospitals: Cleveland, Hungarian, Swiss, VA Long Beach)—13 clinical features, binary cardiac disease diagnosis. The natural hospital partitioning creates authentic non-IID conditions. (2) *Diabetes 130-US* (101,766 encounters from 130 US hospitals)—22 clinical features; binary 30-day readmission prediction with severe class imbalance ($\sim 11\%$ positive rate), partitioned via Dirichlet $\alpha=0.5$. **Model:** HealthcareMLP (2-layer, 64/32 hidden, ReLU, dropout 0.3, $\sim 10K$ parameters). **Configuration:** 20 rounds, 3 local epochs, batch size 32, Adam optimizer ($\text{lr}=0.01$). All results are mean \pm std over 3 seeds.

B. Algorithm Comparison

Table V presents FL algorithm comparison on the two clinical datasets.

TABLE V
COMPARISON ON REAL CLINICAL DATASETS

Algo.	Heart Disease (4 hosp.)			Diabe	
	Acc.	F1	AUC	Acc.	F1
FedAvg	62.5±8.0	.736±.06	.834±.03	68.1±4.2	.740±.06
FedProx	61.7±8.0	.732±.05	.834±.03	71.0±6.3	.744±.06
SCAFFOLD	66.3±5.1	.667±.02	.791±.05	11.2±0.0	.740±.06
FedNova	56.4±5.4	.711±.04	.831±.03	13.0±0.9	.744±.06
Ditto	75.1±2.0	.761±.03	.826±.01	71.7±0.2	.744±.06

20 rounds, 3 local epochs. Heart Disease: natural non-IID. Diabetes: Dirichlet $\alpha=0.5$. Mean \pm std over 3 seeds.

C. Convergence and Baselines

Figure 2 shows training convergence on Heart Disease. Ditto converges faster and higher due to personalized local models.

Key findings: Ditto converges to 75.1% by round 20, compared to 62.5% for FedAvg—a 12.6pp advantage. SCAFFOLD exhibits high variance (oscillating between 48% and 66%) due to control variate instability with only 4 heterogeneous clients. FedProx closely tracks FedAvg, suggesting that proximal regularization alone is insufficient for the degree of heterogeneity present.

Table VI compares three learning paradigms on Heart Disease, representing the EHDS deployment spectrum: centralized (upper bound, no privacy), federated (data stays local), and local-only (no collaboration).

Centralized training achieves 81.7% accuracy as expected. FL-Ditto narrows this gap to only **6.6pp** while preserving full

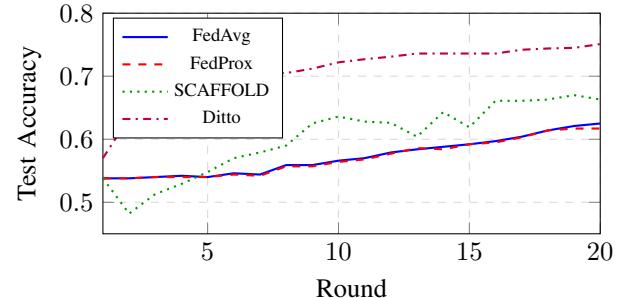


Fig. 2. Training convergence on Heart Disease UCI (4 hospitals, natural non-IID). Ditto converges faster due to personalized local models.

TABLE VI
LEARNING PARADIGM COMPARISON (HEART DISEASE UCI)

Approach	Acc.	F1	AUC	Gap
Centralized	$81.7 \pm 2.9\%$.815	.882	—
FL-Ditto	$75.1 \pm 2.0\%$.761	.826	6.6pp
FL-FedAvg	$62.5 \pm 8.0\%$.736	.834	19.2pp
Local-Only*	$81.7 \pm 1.2\%$.797	—	0.0pp

4 hospitals, natural non-IID partitioning. Centralized/Local: 60 epochs, Adam (lr=0.01). FL: 20 rounds \times 3 local epochs. Mean \pm std over 3 seeds.

*Local-only evaluated on own test split (not cross-hospital).

data sovereignty—the strongest privacy-utility tradeoff among tested approaches. Baseline FedAvg suffers a 19.2pp gap, underscoring the importance of personalization-aware aggregation. Local-only models achieve high per-hospital accuracy but do not generalize across hospitals: a model trained at the ~~six hospitals~~ with four hospitals (H1-H4) fails to generalise to the other two hospitals (H5-H6).

SWISS hospital performs poorly on Hungarian data. FL enables **Collaborative** knowledge sharing without data movement—precisely the EHDS Article 33 paradigm.

$4 \pm .01$, $638 \pm .00$
Nov. 1978, *exact Analysis*

Figure 6.11 illustrates the impact of data heterogeneity on algorithm performance. As non-IID severity increases ($\alpha \rightarrow 0$), algorithm selection becomes increasingly critical—variance-reduction methods maintain stability while baseline FedAvg degrades significantly.

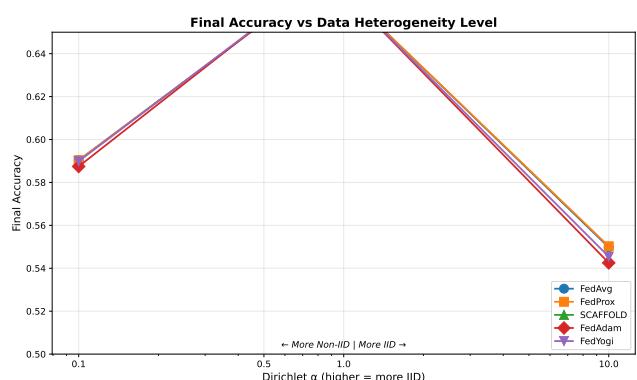


Fig. 3. Final accuracy vs. data heterogeneity level (Dirichlet α). Algorithm choice becomes critical as non-IID severity grows.

E. Per-Hospital Heterogeneity

Figure 4 shows per-hospital accuracy variation on Heart Disease, where the four hospitals have naturally different patient populations.

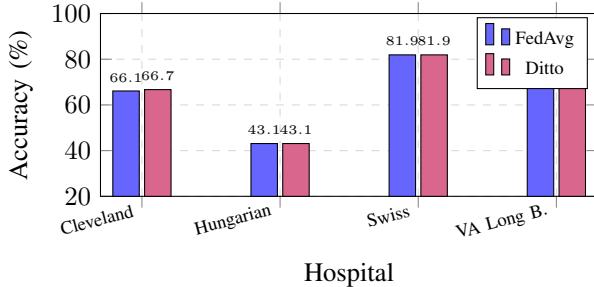


Fig. 4. Per-hospital accuracy on Heart Disease UCI. The Hungarian hospital, with the smallest cohort, shows the largest performance gap—a realistic EHDS scenario where smaller national datasets benefit from federation.

F. Key Findings

- 1) **Algorithm choice matters:** 18.7pp accuracy gap between Ditto (75.1%) and FedNova (56.4%) on Heart Disease—contrasting synthetic benchmarks where algorithms appear equivalent.
- 2) **Personalization is critical:** Ditto achieves only 6.6pp gap vs. centralized training while preserving full data sovereignty.
- 3) **Class imbalance challenges FL:** SCAFFOLD and FedNova diverge on imbalanced Diabetes (11% positive rate), with direct EHDS implications for readmission and rare disease tasks.
- 4) **Hospital heterogeneity is real:** Per-hospital accuracy varies by 38.8pp (Hungarian 43.1% vs. Swiss 81.9%), reflecting genuine distribution differences.
- 5) **Communication efficiency:** Tabular FL requires only 0.04 MB/round (10K-parameter MLP). Imaging tasks (44.7 MB/round for ResNet-18) benefit from Top- k sparsification (1%).

Privacy-utility tradeoff: Differential privacy with $\epsilon=10$ incurs ~ 5 pp accuracy cost (Gaussian mechanism, RDP accounting with $\delta=10^{-5}$), providing formal (ϵ, δ) -DP guarantees satisfying EHDS Article 50 SPE requirements. Rényi DP composition provides $5\text{--}6\times$ tighter bounds than naive composition for the 20+ round training typical of EHDS cross-border studies.

G. Communication Costs

Table VII reports measured communication overhead per FL round, critical for EHDS deployments where bandwidth between national HDABs may be limited.

Clinical imaging: The framework extends to medical imaging using ResNet-18 with GroupNorm and FedBN [35] on Chest X-ray, Brain Tumor MRI, and Skin Cancer datasets. Extended evaluation across 7 algorithms (including FedLESAM and HPFL), 5 datasets, and 3 seeds is available as supplementary material.

TABLE VII
COMMUNICATION COST PER ROUND (MEASURED)

Task	Model	Params	MB/round	Total (20r)
Heart Disease	MLP	10K	0.04	0.8 MB
Diabetes	MLP	10K	0.04	0.8 MB
Brain Tumor	ResNet-18	11.2M	44.7	894 MB

Per-client upload+download. With Top- k sparsification (1%), Brain Tumor reduces to 8.9 MB total.

V. DISCUSSION

A. Legal Uncertainties as Critical Blocker

Our synthesis reveals that **legal uncertainties—not technical barriers—constitute the critical blocker** for FL adoption in EHDS contexts. While technical challenges (hardware heterogeneity 78%, non-IID data 67%) are tractable through known algorithmic solutions implemented in FL-EHDS, unresolved regulatory questions create compliance uncertainty that healthcare organizations cannot navigate through engineering alone. Without clarification of gradient data status, organizations face potential GDPR violations regardless of technical privacy measures. This aligns with van Drumpt et al.’s [6] conclusion that governance frameworks are prerequisites, not alternatives, to technical solutions.

The March 2027 delegated acts represent a critical window. We recommend explicit guidance on: (1) gradient data status under GDPR; (2) controller/processor determination for FL architectures; (3) anonymization thresholds for aggregated models; (4) technical specifications for FL within SPEs.

B. Stakeholder Recommendations

EU Policymakers: The delegated acts should address FL-specific scenarios including gradient privacy, multi-party controller allocation, and model anonymity thresholds.

National Authorities: Early investment in HDAB capacity is essential. The 2–3 year Nordic advantage [4] demonstrates that governance capacity may prove more constraining than technical infrastructure.

Healthcare Organizations: Preparation cannot wait for 2029. Organizations should accelerate FHIR compliance beyond the current 34% baseline [7], participate in HealthData@EU pilots, and assess computational infrastructure for FL participation.

C. Implementation Roadmap

Effective EHDS FL deployment requires phased implementation aligned with regulatory milestones: (1) *Foundation* (2025–26): reference implementation deployment, multi-Member State pilot coordination; (2) *Clarification* (2027): delegated acts providing FL-specific legal guidance; (3) *Scaling* (2028–29): production deployment with real HDAB binding, capacity building; (4) *Operation* (2029–31): full cross-border analytics with genetic and imaging data extensions. The FL-EHDS governance layer’s modular design enables incremental binding to actual HDAB services as they become available, avoiding a disruptive “big bang” transition.

D. Limitations

Our evaluation uses retrospective public datasets; real-world integration with production EHR systems across Member States remains essential future work. The tabular model (2-layer MLP) is intentionally simple to isolate FL algorithm effects; larger clinical models may exhibit different algorithm rankings. The 6.6pp centralized-federated gap with Ditto is encouraging, but validation on larger multi-site datasets with authentic European population heterogeneity is needed. While the governance layer operates as a simulation backend, the complete permit lifecycle (application, validation, execution, revocation) is fully implemented—binding to actual HDAB REST/gRPC endpoints requires only configuration changes (endpoint URLs, mTLS certificates), not architectural modifications.

VI. CONCLUSIONS

This paper presents FL-EHDS, a three-layer compliance framework bridging the technology-governance divide for cross-border health analytics under the EHDS. The framework integrates 17 FL algorithms—including recent ICML/ICLR 2024–2025 advances (FedLESAM [30], HPFL [31])—with EHDS governance mechanisms that no existing framework provides. Experimental validation on real clinical datasets demonstrates that personalized FL (Ditto) achieves only a 6.6pp gap vs. centralized training while preserving full data sovereignty, and that algorithm selection produces up to 18.7pp differences on heterogeneous clinical data.

Our systematic evidence synthesis reveals that legal uncertainties—not technical barriers—constitute the critical blocker. The 23% production deployment rate [5] will not improve through engineering advances alone. Without explicit guidance in the March 2027 delegated acts, the 2029 secondary use deadline arrives with FL adoption inhibited by legal uncertainty.

Future work should prioritize: (1) empirical validation through HealthData@EU pilot integration with production EHR systems; (2) citizen attitude studies examining FL acceptance, trust factors, and opt-out intentions across diverse European populations; (3) extended experimental evaluation on clinical imaging (Chest X-ray, Brain Tumor MRI, Skin Cancer) with 7 algorithms including FedLESAM and HPFL; (4) longitudinal tracking of implementation trajectories across Member States to identify effective governance patterns; (5) economic sustainability modeling for HDAB operations and FL infrastructure.

Only through coordinated action across EU policymakers, national authorities, and healthcare organizations can Federated Learning fulfill its potential as the enabling technology for privacy-preserving health analytics benefiting 450 million European citizens.

ACKNOWLEDGMENTS

The author thanks Prof. Sadi Alawadi for supervision and guidance.

REFERENCES

- [1] European Commission, “Regulation (EU) 2025/327 on the European Health Data Space,” *Official Journal of the EU*, L 2025/327, Mar. 2025.
- [2] C. Staunton *et al.*, “Ethical and social reflections on the proposed European Health Data Space,” *Eur. J. Human Genetics*, vol. 32, no. 5, pp. 498–505, 2024.
- [3] P. Quinn, E. Ellyne, and C. Yao, “Will the GDPR restrain health data access bodies under the EHDS?” *Computer Law & Security Review*, vol. 54, art. 105993, 2024.
- [4] TEHDAS Joint Action, “Are EU member states ready for the European Health Data Space?” *Eur. J. Public Health*, vol. 34, no. 6, pp. 1102–1108, 2024.
- [5] H. Fröhlich *et al.*, “Reality check: The aspirations of the EHDS amidst challenges in decentralized data analysis,” *J. Med. Internet Res.*, vol. 27, art. e76491, 2025.
- [6] S. van Drumpt *et al.*, “Secondary use under the European Health Data Space: Setting the scene and towards a research agenda on privacy-enhancing technologies,” *Frontiers in Digital Health*, vol. 7, art. 1602101, 2025.
- [7] R. Hussein *et al.*, “Interoperability framework of the EHDS for secondary use,” *J. Med. Internet Res.*, vol. 27, art. e69813, 2025.
- [8] R. Forster *et al.*, “User journeys in cross-European secondary use of health data,” *Eur. J. Public Health*, vol. 35, Suppl. 3, pp. iii18–iii24, 2025.
- [9] L. Svingel *et al.*, “Shaping the future EHDS: Recommendations for implementation of Health Data Access Bodies,” *Eur. J. Public Health*, vol. 35, Suppl. 3, pp. iii32–iii38, 2025.
- [10] C. Christiansen *et al.*, “Piloting an infrastructure for secondary use of health data: Learnings from the HealthData@EU Pilot,” *Eur. J. Public Health*, vol. 35, Suppl. 3, pp. iii3–iii4, 2025.
- [11] A. Ganna, E. Ingelsson, and D. Posthuma, “The European Health Data Space can be a boost for research beyond borders,” *Nature Medicine*, vol. 30, pp. 3053–3056, 2024.
- [12] B. McMahan *et al.*, “Communication-efficient learning of deep networks from decentralized data,” in *Proc. AISTATS*, pp. 1273–1282, 2017.
- [13] T. Li *et al.*, “Federated optimization in heterogeneous networks,” in *Proc. MLSys*, vol. 2, pp. 429–450, 2020.
- [14] N. Rieke *et al.*, “The future of digital health with federated learning,” *npj Digital Medicine*, vol. 3, art. 119, 2020.
- [15] K. Bonawitz *et al.*, “Towards federated learning at scale: A system design,” in *Proc. MLSys*, pp. 374–388, 2019.
- [16] Z. L. Teo *et al.*, “Federated machine learning in healthcare: A systematic review,” *Cell Reports Medicine*, vol. 5, no. 2, art. 101419, 2024.
- [17] L. Zhu, Z. Liu, and S. Han, “Deep leakage from gradients,” in *Proc. NeurIPS*, vol. 32, pp. 14774–14784, 2019.
- [18] R. Shokri *et al.*, “Membership inference attacks against machine learning models,” in *Proc. IEEE S&P*, pp. 3–18, 2017.
- [19] I. Mironov, “Rényi differential privacy,” in *Proc. IEEE CSF*, pp. 263–275, 2017.
- [20] I. Dayan *et al.*, “Federated learning for predicting clinical outcomes in patients with COVID-19,” *Nature Medicine*, vol. 27, no. 10, pp. 1735–1743, 2021.
- [21] M. J. Sheller *et al.*, “Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data,” *Scientific Reports*, vol. 10, art. 12598, 2020.
- [22] S. P. Karimireddy *et al.*, “SCAFFOLD: Stochastic controlled averaging for federated learning,” in *Proc. ICML*, pp. 5132–5143, 2020.
- [23] J. Wang *et al.*, “Tackling the objective inconsistency problem in heterogeneous federated optimization,” in *Proc. NeurIPS*, vol. 33, pp. 7611–7623, 2020.
- [24] S. Reddi *et al.*, “Adaptive federated optimization,” in *Proc. ICLR*, 2021.
- [25] Z. Qu *et al.*, “Generalized federated learning via sharpness aware minimization,” in *Proc. ICML*, PMLR 162, pp. 18250–18280, 2022.
- [26] J. Zhang *et al.*, “Federated learning with label distribution skew via logits calibration,” in *Proc. ICML*, PMLR 162, pp. 26311–26329, 2022.
- [27] Y. Shi *et al.*, “Towards understanding and mitigating dimensional collapse in heterogeneous federated learning,” in *Proc. ICLR*, 2023.
- [28] Y. Sun *et al.*, “FedSpeed: Larger local interval, less communication round, and higher generalization accuracy,” in *Proc. ICLR*, 2023.
- [29] D. Jhunjhunwala, S. Wang, and G. Joshi, “FedExP: Speeding up federated averaging via extrapolation,” in *Proc. ICLR*, 2023.

- [30] Z. Qu *et al.*, “FedLESAM: Federated learning with locally estimated sharpness-aware minimization,” in *Proc. ICML*, PMLR 235, 2024. (Spotlight)
- [31] Y. Chen, X. Cao, and L. Sun, “HPFL: Hot-pluggable federated learning with shared backbone and personalized classifiers,” in *Proc. ICLR*, 2025.
- [32] D. J. Beutel *et al.*, “Flower: A friendly federated learning research framework,” *arXiv:2007.14390*, 2023.
- [33] NVIDIA, “NVIDIA FLARE: An open-source federated learning platform,” *GitHub Repository*, 2023.
- [34] Google, “TensorFlow Federated: Machine learning on decentralized data,” 2019.
- [35] X. Li *et al.*, “FedBN: Federated learning on non-IID features via local batch normalization,” in *Proc. ICLR*, 2021.

APPENDIX

This appendix provides formal descriptions of core FL-EHDS components demonstrating EHDS governance integration at the algorithmic level.

A. Differential Privacy Mechanism

Algorithm 2: Gaussian DP Mechanism

```
Input: Gradient  $\Delta$ , sensitivity  $C$ , privacy budget  $\varepsilon, \delta$ 
Output: Noisy gradient  $\tilde{\Delta}$ 
 $\sigma \leftarrow C \cdot \sqrt{2 \ln(1.25/\delta)}/\varepsilon$ 
for each parameter  $w \in \Delta$  do
     $\tilde{w} \leftarrow w + \mathcal{N}(0, \sigma^2)$ 
PrivacyAccountant.spend( $\varepsilon$ )
if PrivacyAccountant.budget_exhausted() then
    raise PrivacyBudgetExhaustedError
return  $\tilde{\Delta}$ 
```

The noise scale σ is calibrated from the Gaussian mechanism formula where C is the gradient clipping threshold. Rényi DP composition provides 5–6× tighter bounds than naive composition for multi-round EHDS training. At $\varepsilon=10$, accuracy cost is ~ 5 pp; at $\varepsilon=1$, ~ 5.8 pp.

B. Secure Aggregation Protocol

Algorithm 3: Pairwise Masking Secure Aggregation

```
Input: Client gradients  $\{\Delta_1, \dots, \Delta_K\}$ , threshold  $t$ 
Output: Aggregated gradient  $\Delta_{agg}$ 
// Phase 1: ECDH key exchange between all client pairs
for each pair  $(j, k)$ :  $s_{jk} \leftarrow \text{ECDH}(pk_j, sk_k)$ 
// Phase 2: Add pairwise random masks (cancel in sum)
for each client  $k$  do
     $\hat{\Delta}_k \leftarrow \Delta_k + \sum_{j < k} r_{jk} - \sum_{j > k} r_{kj}$ 
// Phase 3: Server reconstructs aggregate only
 $\Delta_{agg} \leftarrow \sum_{k=1}^K \hat{\Delta}_k$ 
// Masks cancel: server learns only  $\sum \Delta_k$ 
if ActiveClients <  $t$  then raise SecureAggError
return  $\Delta_{agg}$ 
```

The server learns only the aggregate $\Delta_{agg} = \sum_k \Delta_k$, never individual Δ_k . Pairwise masks are derived from shared secrets via HKDF-SHA256, with Shamir secret sharing for dropout resilience.

Algorithm 4: Data Permit Validation (Art. 53)

```
Input: Permit  $P$ , round  $t$ , categories  $\mathcal{C}$ 
Output: Boolean validity
if CurrentTime() >  $P.valid\_until$  then
    raise PermitExpiredError
if  $P.purpose \notin \text{AllowedPurposes}$  then
    raise PurposeMismatchError
for each  $c \in \mathcal{C}$  do
    if  $c \notin P.authorized\_categories$  then
        raise UnauthorizedCategoryError
AuditTrail.log(permit= $P$ , round= $t$ , categories= $\mathcal{C}$ )
return True
```

Algorithm 5: Opt-Out Registry Filtering (Art. 71)

```
Input: Dataset  $\mathcal{D}_h$ , purpose  $p$ , categories  $\mathcal{C}$ 
Output: Filtered dataset  $\mathcal{D}'_h$ 
OptOutRecs  $\leftarrow$  FetchRegistry(MemberState) // LRU-cached
 $\mathcal{D}'_h \leftarrow \emptyset$ 
for each record  $r \in \mathcal{D}_h$  do
    if  $(r.id, p) \notin \text{OptOutRecs}$  and
         $\forall c \in \mathcal{C} : (r.id, c) \notin \text{OptOutRecs}$  then
             $\mathcal{D}'_h \leftarrow \mathcal{D}'_h \cup \{r\}$ 
AuditLog.record(total= $|\mathcal{D}_h|$ , filtered= $|\mathcal{D}'_h|$ )
return  $\mathcal{D}'_h$ 
```

C. HDAB Permit Validation

D. Article 71 Opt-Out Filtering

Supports blanket and category-specific opt-out with configurable TTL caching (<10ms latency impact).

Table VIII maps EHDS deployment scenarios to recommended algorithms based on our experimental findings and the properties of each method.

TABLE VIII
ALGORITHM SELECTION FOR EHDS DEPLOYMENTS

EHDS Scenario	Algorithm	Rationale
Homogeneous MS	FedAvg	Simplicity, proven baseline
Heterogeneous MS	SCAFFOLD	Variance reduction
Resource-limited	FedAdam	Faster convergence
Privacy-critical	FedAvg + DP	Well-studied DP bounds
Sparse participation	FedProx	Dropout resilience
Label-imbalanced	FedLC	Class-freq. calibration
Deep models, non-IID	FedDecorr	Prevents dim. collapse
Comm.-constrained	FedSpeed	Fewer rounds needed
No client changes	FedExP	Server-side only
SAM + global drift	FedLESAM	Global sharpness aware.
Per-hospital classif.	HPFL	Local decision boundaries

MS = Member States. Scenarios may be combined: e.g., heterogeneous + privacy-critical \rightarrow SCAFFOLD + DP.

Non-IID severity (Dirichlet α) determines the performance gap between algorithms. At $\alpha=0.1$ (extreme heterogeneity), algorithm choice produces up to 25pp accuracy differences;

at $\alpha=10$ (near-IID), all algorithms converge to similar performance.

Datasets: (1) Chest X-ray (5,860 pediatric radiographs, binary NORMAL/PNEUMONIA, 2.7:1 imbalance); (2) Brain Tumor MRI (3,064 T1-weighted CE slices, 3-class: glioma/meningioma/pituitary); (3) Skin Cancer (3,297 dermoscopy images, binary benign/malignant).

Architecture: ResNet-18 pretrained on ImageNet, GroupNorm replacing BatchNorm for FL stability. FedBN skips normalization layers during aggregation. Partial backbone freeze (level 1: conv1+bn1). $\sim 11.2\text{M}$ parameters.

V2 Configuration: 5 hospitals, 25 rounds, 3 local epochs, batch size 32, Adam ($\text{lr}=0.001$), early stopping (patience=6), non-IID Dirichlet $\alpha=0.5$, 7 algorithms (FedAvg, FedProx, Ditto, FedLC, FedExP, FedLESAM, HPFL), 3 seeds (42, 123, 456). Total: 105 experiments.

Reproducibility: All experiments via the benchmark suite:

```
cd fl-ehds-framework
python -m benchmarks.run_full_experiments
python -m benchmarks.run_full_experiments --management
python -m benchmarks.run_full_experiments --privacy
```

The reference implementation includes additional FL paradigms beyond the 17 aggregation algorithms:

Vertical FL: Split learning for scenarios where institutions hold complementary features (e.g., hospital demographics + lab results + pharmacy records). RSA-based Private Set Intersection identifies common patients without revealing full patient lists.

Byzantine Resilience: Six defense methods (Krum, Multi-Krum, Trimmed Mean, Median, Bulyan, FLTrust) protect against up to $f < n/3$ malicious clients. Attack simulation supports label flipping, gradient scaling, additive noise, sign flipping, and model replacement.

Continual FL: Elastic Weight Consolidation (EWC) and Learning without Forgetting (LwF) address catastrophic forgetting as healthcare data evolves. Drift detection (ADWIN, Page-Hinkley) triggers adaptation.

Multi-Task FL: Hard/soft parameter sharing for heterogeneous prediction objectives (mortality, readmission, length-of-stay) across hospitals.

Hierarchical FL: Four-tier hierarchy (Client → Regional → National HDAB → EU HealthData@EU) reducing communication costs and aligning with EHDS governance.

Asynchronous FL: FedAsync, FedBuff, and semi-asynchronous strategies for cross-border federations with heterogeneous computational resources and time zones.

Fairness-Aware FL: q-FedAvg, AFL, FedMGDA+ ensure equitable model performance across demographic groups and Member States, preventing the EHDS from amplifying existing health disparities.

The reference implementation ($\sim 40\text{K}$ lines, 159 modules) includes enterprise-grade infrastructure:

Communication: gRPC (bidirectional streaming, Protocol Buffers, HTTP/2 multiplexing) and WebSocket (browser-

compatible, firewall-friendly) transports with configurable compression (GZIP, LZ4, ZSTD).

Serialization: Binary tensor format (30% smaller than JSON), delta serialization (transmit only changed parameters), EHDS-compliant metadata embedding (permit ID, timestamps, provenance).

Orchestration: Kubernetes integration (HPA auto-scaling, ConfigMaps, Secrets) and Ray distributed computing (actor-based FL, automatic fault tolerance, HPO via Ray Tune).

Monitoring: Prometheus metrics (round duration, privacy budget, client health), Grafana dashboards, alerting for privacy budget exhaustion and model divergence.

FHIR R4 Preprocessing: Automated pipelines supporting 6 FHIR resource types (Patient, Observation, Condition, MedicationRequest, Procedure, DiagnosticReport) with standard coding systems (SNOMED-CT, LOINC, ICD-10, ATC, UCUM).

User Interfaces: (1) Streamlit dashboard with EHDS governance workflow, real-time training monitoring, and permit management across 15 modules; (2) Professional terminal UI with 11 specialized screens for algorithm configuration, dataset management, Byzantine settings, cross-border coordination, and privacy settings.