REVIEW **Open access**

Federated Learning Ecosystems in Healthcare: Architectural Models and Privacy Trade-Offs

Daniel Brooks[1*], Oliver Grant[2], Maria Hernandez[3]

Abstract

Federated learning (FL) has emerged as a transformative paradigm in artificial intelligence (AI) for healthcare systems and analytics, enabling collaborative model training across distributed institutions without direct data sharing, thereby addressing stringent privacy regulations such as the Health Insurance Portability and Accountability Act (HIPAA) and General Data Protection Regulation (GDPR). This narrative review synthesizes the architectural models underpinning FL ecosystems in healthcare, elucidating their integration into clinical analytics pipelines and the privacy trade-offs they entail. We delineate how FL facilitates decentralized AI applications in areas such as predictive modeling for clinical outcomes, medical imaging analysis, and real-time health monitoring, while balancing model utility against data protection imperatives.

Central to FL architectures are client-server frameworks where edge devices (e.g., hospitals or wearable sensors) perform local training on siloed datasets, aggregating updates via a central coordinator to refine global models. Variants include horizontal FL for identical feature spaces across institutions and vertical FL for complementary datasets, often augmented with differential privacy mechanisms to mitigate inference attacks. In healthcare systems, these models support analytics for disease prediction, as seen in COVID-19 outcome forecasting, and enable scalable infrastructures for multi-institutional collaborations without compromising patient confidentiality. However, privacy trade-offs manifest in reduced model accuracy due to noisy perturbations, communication overheads in bandwidth-constrained environments, and vulnerabilities to model inversion or membership inference attacks.

We explore the landscape of AI-driven healthcare systems, highlighting how FL integrates with electronic health records (EHRs), imaging repositories, and wearable data streams to foster intelligent analytics. Key syntheses include closed-loop systems where AI inferences inform clinical decisions, feedback loops recalibrate models, and governance layers ensure ethical deployment. Challenges such as data heterogeneity across federated nodes and the need for robust incentive mechanisms are critically examined, alongside opportunities for hybrid FL-blockchain integrations to enhance trust. This review posits that optimized FL ecosystems can revolutionize healthcare delivery by enabling privacy-preserving, generalizable AI analytics, but that these systems require interdisciplinary frameworks to navigate trade-offs between innovation and patient safeguards. Ultimately, FL represents a cornerstone for sustainable, equitable AI in healthcare, promoting data sovereignty while accelerating clinical insights.

Keywords Clinical analytics, Federated learning, Healthcare AI, Privacy preservation, Architectural models, Data governance

*Correspondence:
Daniel Brooks
daniel.brooks@gmail.com
[1] Department of Artificial Intelligence in Medicine, School of Computer Science, University of Birmingham, Birmingham, United Kingdom
[2] Department of Digital Health Systems, School of Engineering, University of Glasgow, Glasgow, United Kingdom
[3] Department of Healthcare Analytics and Policy, School of Medicine, University of Valencia, Valencia, Spain

Introduction

The integration of artificial intelligence (AI) into healthcare systems and analytics has accelerated dramatically in recent years, driven by the exponential growth in health data volumes from electronic health records (EHRs), wearable devices, and imaging modalities. This surge necessitates advanced computational paradigms that can harness distributed data sources while upholding patient privacy, a core ethical and regulatory imperative in clinical environments. Federated learning (FL) emerges as a pivotal ecosystem in this context, enabling multiple healthcare entities to collaboratively train AI models without exchanging raw patient data, thereby mitigating the risks associated with centralized data aggregation [1-6]. This review synthesizes FL's architectural models in healthcare, focusing on their role in enabling privacy-aware analytics and the trade-offs involved in balancing model performance with data protection.

Evolution of AI in healthcare infrastructure

Historically, AI applications in healthcare relied on centralized architectures where data from disparate sources were pooled into monolithic repositories for model training [7-12]. However, this approach exacerbates privacy vulnerabilities, particularly in sensitive domains like oncology or infectious disease tracking, where data breaches could lead to stigmatization or exploitation [4]. The advent of FL in 2017 marked a paradigm shift, inspired by distributed machine learning techniques, enabling edge-based computations that keep data localized [1, 6]. By 2020, FL had been adapted for healthcare-specific challenges, such as handling non-independent and identically distributed (non-IID) data across hospitals [2, 5, 13]. Architectural innovations include hierarchical FL, where intermediate aggregators process updates from subgroups of clients before global synchronization, reducing latency in large-scale systems [11, 14, 15]. These evolutions align with broader healthcare infrastructure trends toward decentralized analytics, as evidenced in multi-center studies for rare disease prediction [9, 16, 17].

In clinical analytics, FL enables the development of generalizable models for tasks such as risk stratification and prognostic forecasting [2, 3, 18]. For instance, FL ecosystems have been deployed to analyze EHRs for mortality prediction in intensive care units, aggregating insights from geographically dispersed cohorts without data transfer [12, 13]. This decentralized approach not only complies with regulations like GDPR but also enhances model robustness by incorporating diverse demographic representations [7, 8, 16, 19, 20]. Yet, it introduces trade-offs: while privacy is bolstered by techniques such as secure multi-party computation, model convergence may slow down due to heterogeneous data distributions [4, 11, 14]. summarizes the foundational architectural components of healthcare federated learning ecosystems.

Table 1. Foundational architectural components of federated learning ecosystems in healthcare

Architectural component	Functional role	Healthcare data sources	Privacy mechanisms	Analytical contribution
Data ingestion nodes	Local data capture and preprocessing	EHRs, imaging, wearables	Data locality enforcement	Multimodal feature generation
Edge training clients	Local model computation	Hospital datasets	Secure enclaves	Gradient learning
Federated aggregator	Global model synthesis	Parameter updates	Secure aggregation	Cross-site intelligence fusion
Communication layer	Update transmission	Networked institutions	Encryption protocols	Synchronization
Governance framework	Compliance monitoring	Audit logs, metadata	Differential privacy	Ethical validation
Feedback interfaces	Outcome reporting	Intervention results	Privacy filters	Model recalibration

Privacy imperatives in healthcare AI

Privacy trade-offs are intrinsic to FL ecosystems, where the goal is to maximize analytical utility while minimizing information leakage [1, 4, 7]. Differential privacy (DP), a cornerstone mechanism, adds calibrated noise to model updates to obscure individual contributions, but at the cost of degraded accuracy in high-stakes clinical predictions [11, 14]. In healthcare systems, this is particularly acute for analytics involving rare events, such as adverse drug reactions, where noise amplification can obscure subtle patterns [9, 17]. Architectural models often incorporate hybrid strategies that blend DP with homomorphic encryption to enable computations over encrypted data, though this increases computational overhead [4, 6, 15]. The literature highlights the need for adaptive privacy budgets, in which epsilon values are dynamically adjusted based on data sensitivity [7, 14].

Beyond technical mechanisms, governance frameworks play a critical role in FL deployments [19-22]. Healthcare systems must integrate consent models that allow patients to opt into federated collaborations, ensuring transparency in how their data influences global models [3, 16, 23, 24]. This is especially relevant in analytics for public health crises, like the COVID-19 pandemic, where FL enabled rapid outcome predictions across borders without violating sovereignty [2, 5, 8]. However, disparities in institutional capabilities—such as varying computational resources—can lead to biased aggregations, underscoring the need for equitable architectural designs [20, 25, 26].

Scope of federated learning ecosystems

FL ecosystems encompass a spectrum of architectural models tailored to healthcare's unique demands: from client-centric setups for mobile health analytics to server-orchestrated frameworks for hospital networks [1, 6, 11]. In medical imaging, FL enables collaborative training on distributed datasets for tumor segmentation while preserving pixel-level privacy [4, 7, 15, 16]. For clinical decision support, these models integrate with EHR analytics to provide real-time insights, such as sepsis alerts [12, 13, 18]. Privacy trade-offs extend to communication efficiency, where techniques such as model compression reduce bandwidth requirements but risk information loss [11, 14].

This review adopts an original synthesis logic, framing FL through a systems-level lens that interconnects data ingestion, model orchestration, analytical inference, and governance feedback. Unlike prior taxonomies that categorize by algorithm type [1], we integrate cross-study insights to highlight emergent properties, such as resilience to adversarial attacks in distributed healthcare networks [4, 7]. By synthesizing architectures across predictive, diagnostic, and interventional analytics, we elucidate how FL fosters closed-loop systems that evolve with clinical feedback [17, 18, 22]. The scope is delimited to peer-reviewed works from 2017–2022, prioritizing high-impact venues to ground our analysis in robust evidence. This positioning enables a novel interpretive structure, emphasizing FL's potential to democratize AI in resource-constrained healthcare settings while navigating privacy-utility equilibria.

Landscape of AI in healthcare systems and analytics

The landscape of AI in healthcare systems and analytics is characterized by a confluence of data-driven technologies that enhance diagnostic precision, operational efficiency, and patient outcomes [23, 25, 26]. Federated learning (FL) ecosystems represent a cornerstone in this domain, enabling privacy-preserving collaborations that leverage distributed data assets across healthcare infrastructures [1, 6, 11]. This section synthesizes the integration of AI into healthcare systems, focusing on analytical workflows, infrastructural enablers, and the role of FL in addressing data silos.

Data ecosystems and analytical foundations

Healthcare systems operate as dense data ecosystems that continuously generate vast, heterogeneous streams of clinical, biological, and behavioral information. These data originate from diverse sources, including electronic health records (EHRs), high-throughput genomic sequencing platforms, radiological imaging repositories, and real-time biosensor networks, collectively forming a multimodal intelligence substrate that underpins contemporary clinical analytics [9, 12, 23]. The analytical challenge is not merely volumetric but structural: healthcare data differ in granularity,

temporality, modality, and semantic encoding. Consequently, advanced analytics infrastructures must support cross-modal fusion, longitudinal pattern tracking, and context-aware inference pipelines that reconcile structured and unstructured inputs.

Artificial intelligence (AI) models—particularly deep learning architectures—have demonstrated significant efficacy in extracting latent patterns from these complex datasets. Convolutional neural networks (CNNs) support high-resolution imaging diagnostics, recurrent architectures enable temporal forecasting, and transformer-based systems facilitate multimodal reasoning for disease classification and epidemiological trend prediction [10, 26]. Despite these capabilities, traditional centralized analytics frameworks face systemic barriers to deployment in healthcare. Privacy regulations, institutional data sovereignty, and cybersecurity risks constrain large-scale pooling of patient data, thereby limiting the diversity of training data required for robust model generalization. These constraints have catalyzed the emergence of federated learning (FL) as a decentralized analytical paradigm that preserves data locality while enabling collaborative intelligence generation [1, 4, 6].

Within federated healthcare ecosystems, analytical pipelines are re-engineered to operate across distributed computational nodes. Data preprocessing is performed at the institutional or edge-device level, where raw clinical inputs are normalized, features extracted, and dimensionality reduced before model training [11, 13, 15]. Only model parameters or gradient updates, rather than patient-level data, are transmitted to aggregation servers. This localized preprocessing layer reduces exposure risks while maintaining analytical fidelity. Empirical literature illustrates the effectiveness of such pipelines in imaging analytics, where FL frameworks train CNN models across geographically distributed radiology datasets to detect malignancies and pathological anomalies without centralizing sensitive scans [4, 7, 16].

Beyond static diagnostic inference, healthcare analytics increasingly encompass dynamic systems monitoring. Continuous physiological data streams from wearable devices, implantable sensors, and remote monitoring platforms generate high-frequency signals that reflect patient status in real time [3, 8, 24]. FL architectures are uniquely suited to these longitudinal analytics environments, enabling iterative model updating as new patient data emerge. This continuous learning capability is particularly critical in chronic disease management, where predictive models must adapt to evolving physiological baselines and treatment responses [12, 18]. For instance, federated population health platforms have demonstrated utility in modeling infectious disease propagation, aggregating epidemiological signals across jurisdictions to forecast epidemic trajectories—an approach widely explored in SARS-CoV-2 intelligence systems [2, 5, 8]. By distributing model training across heterogeneous cohorts, FL reduces sampling bias and enhances cross-population generalizability [9, 20, 26].

Architectural models in federated ecosystems

At the structural core of federated analytics lie architectural models that coordinate distributed learning processes. The canonical Federated Averaging (FedAvg) algorithm remains foundational, aggregating locally trained model gradients into a unified global model. However, healthcare datasets are characteristically non-independent and non-identically distributed (non-IID), reflecting institutional, demographic, and procedural heterogeneity. As such, healthcare-adapted FL architectures incorporate optimization adjustments to maintain convergence stability in the presence of distributional divergence [1, 6, 13].

Horizontal federated learning frameworks are particularly applicable in multi-institutional environments where participating entities share similar feature spaces. Multi-hospital EHR ecosystems exemplify this configuration, enabling collaborative outcome prediction, readmission risk modeling, and treatment response analytics while preserving institutional data firewalls [2, 12, 18]. Vertical federated learning, by contrast, integrates complementary feature domains across entities. In clinical diagnostics, this may involve linking radiological imaging repositories with laboratory biomarkers or unstructured clinical narratives to generate holistic patient intelligence models [11, 15].

Hybrid federated architectures extend these paradigms by integrating split learning mechanisms. In such models, neural networks are partitioned into local and central segments: early layers process sensitive data within institutional boundaries, while deeper layers complete inference at secure aggregation nodes. This structural segmentation enhances privacy preservation while maintaining analytical depth, particularly in high-sensitivity domains such as oncology imaging or genomic analytics [4, 7, 14].

Operational deployment of these architectures frequently occurs within cloud–edge hybrid infrastructures. Hospitals, research centers, and diagnostic laboratories function as federated clients, while encrypted aggregation servers coordinate parameter synthesis and global model updates [6, 11, 15]. Privacy safeguards are reinforced through secure aggregation protocols that cryptographically mask local model contributions, mitigating gradient reconstruction and membership inference attacks [1, 4, 14].

Nevertheless, architectural trade-offs remain salient. While FL reduces the burden of raw data transfer, computational heterogeneity across institutions introduces synchronization latency and training instability. Variations in hardware acceleration, network bandwidth, and dataset scale can impede the efficiency of federated convergence [13, 19]. Emerging literature highlights adaptive client selection and asynchronous aggregation as mitigation strategies that enable scalable analytics in bandwidth-constrained or resource-limited healthcare environments [3, 7, 11].

Integration with clinical analytics workflows

The analytical value of federated ecosystems is most evident when integrated into real-world clinical workflows. Healthcare AI systems increasingly emphasize predictive and prescriptive intelligence, supporting clinicians in risk stratification, diagnostic prioritization, and intervention planning [9, 17, 23]. FL-trained models operationalize these capabilities by analyzing distributed EHR datasets to generate patient-specific risk scores, thereby informing personalized treatment pathways and preventive care strategies [12, 18, 20].

Imaging-intensive specialties further benefit from federated orchestration. Ensemble learning across federated nodes enables models to synthesize radiological features from diverse scanners, demographic populations, and clinical protocols. This distributed training paradigm has improved lesion detection accuracy in dermatology, neuroimaging, and oncologic radiology, where single-site datasets often lack sufficient variability for robust inference [4, 7, 16].

Real-world deployments during global health crises have underscored the translational relevance of federated analytics. Federated COVID-19 prognostic platforms, for example, enabled cross-border modeling of hospitalization risk, ventilation demand, and mortality predictors without violating patient privacy or national data protection statutes [2, 3, 5]. Such deployments illustrate how decentralized intelligence infrastructures can accelerate global knowledge synthesis during emergent public health events.

Governance frameworks constitute a critical overlay within these ecosystems, ensuring analytical accountability and ethical compliance. Provenance tracking mechanisms document model lineage, training cohorts, and parameter evolution, enabling auditability across federated cycles [19, 21, 22]. However, governance integration introduces analytical trade-offs. Differential privacy (DP) mechanisms, while essential for preserving confidentiality, introduce statistical noise into model updates. In high-stakes clinical predictions, this noise may attenuate sensitivity, potentially increasing false-negative rates in critical diagnostic contexts [7, 11, 14].

Cross-study synthesis reveals that federated ecosystems foster collaborative analytics networks that accelerate discovery in low-prevalence or rare disease domains, where data scarcity traditionally impedes model training [9, 17]. By enabling distributed cohort pooling without data centralization, FL expands analytical reach while maintaining ethical safeguards. Yet infrastructural barriers persist—most notably in data standardization. Variability in coding ontologies, metadata schemas, and interoperability standards necessitates harmonization protocols to ensure analytical consistency across federated nodes [12, 19, 27].

Systems-level synthesis and trade-offs

Synthesizing the landscape, FL ecosystems reframe healthcare analytics as interconnected systems where data flows inform iterative improvements [1, 6, 18, 22]. A conceptual formula for this integrative process can be articulated as:

$$S = f(D_l,\ M_g,\ P_t) \quad (1)$$

where S denotes the synthesized analytical system output, D_l represents localized data processing at client nodes, M_g the global model aggregation, and P_t the privacy trade-off function modulating utility versus protection. This formalization captures the cyclical nature of FL in healthcare, emphasizing how privacy mechanisms influence analytical fidelity [4, 7, 11, 14].

Trade-offs extend to ethical dimensions: while FL democratizes access to advanced analytics in underserved regions [20, 24], it demands robust incentive structures to encourage participation [19, 21]. Overall, this landscape synthesis underscores FL's transformative potential in building resilient, privacy-centric healthcare systems that prioritize analytical depth over data centralization [2, 8, 9, 16, 23, 26].

Intelligent clinical decision & closed-loop healthcare systems

Intelligent clinical decision support systems (CDSS) powered by AI represent the maturation of healthcare analytics, in which federated learning (FL) ecosystems enable real-time, evidence-based recommendations while preserving data privacy [17, 18, 22]. This section explores architectural models for these systems, emphasizing closed-loop mechanisms that incorporate feedback for continuous refinement. Synthesis focuses on how FL integrates data-driven insights into clinical workflows, balancing decision accuracy with privacy imperatives.

Architectures for AI-Enabled decision support

CDSS architectures in FL ecosystems typically feature modular designs: data ingestion layers at local sites feed into inference engines, with aggregated models deployed to augment decisions [1, 6, 12]. In clinical settings, these support tasks include diagnostic triage, in which FL-trained classifiers analyze symptoms and vitals to prioritize care [2, 18, 23]. Horizontal FL architectures dominate here, enabling hospitals to collaborate on shared decision models without exposing data [11, 13]. For complex decisions, such as treatment optimization in oncology, vertical FL combines siloed modalities (e.g., genomics and imaging) to generate comprehensive recommendations [4, 15, 16].

Privacy trade-offs in these architectures involve encryption overheads that delay real-time decisions, which are mitigated by lightweight protocols such as additive secret sharing [4, 7, 14]. Literature synthesizes that hybrid architectures—merging FL with edge computing—enhance responsiveness in ambulatory care [3, 8, 24]. Cross-study insights reveal that decision reliability improves with federated fine-tuning, enabling global models to adapt to local contexts [5, 13, 18].

Closed-loop mechanisms in healthcare systems

Closed-loop systems extend CDSS by embedding feedback loops that use clinical outcomes to inform model recalibration [9, 17, 22]. In FL ecosystems, this manifests as iterative aggregation cycles: post-decision data (e.g., intervention efficacy) is locally processed and used to update shared parameters [1, 6, 12]. For instance, in sepsis management, AI decisions trigger alerts, with subsequent patient responses federated back to refine predictions [2, 18]. Architectural models incorporate monitoring agents that detect drift, prompting retraining without full data access [11, 14, 15].

A second conceptual formula illustrates this dynamic:

$$C = g(I_d,\ F_b,\ G_v) \quad (2)$$

where C is the closed-loop cycle output, I_d the initial decision inference, F_b the feedback from interventions, and G_v the governance validation ensures privacy-compliant updates. This structure highlights the infrastructural interplay, in which privacy mechanisms such as DP safeguard feedback transmissions [4, 7, 14]. Trade-offs include the potential for feedback amplification of biases if not properly governed [20, 26].

Synthesis of deployment and governance

Deployment in closed-loop systems requires interoperable architectures, as seen in multi-institutional platforms for pandemic response [3, 5, 8]. Governance ensures ethical decision fusion, blending AI outputs with clinician judgment [19, 21, 22]. Synthesis across literature posits that FL fosters adaptive systems, but demands standardized interfaces for seamless integration [12, 19, 27]. **Figure 1** illustrates the closed-loop federated healthcare analytics architecture integrating distributed training, aggregation, governance, and feedback recalibration.

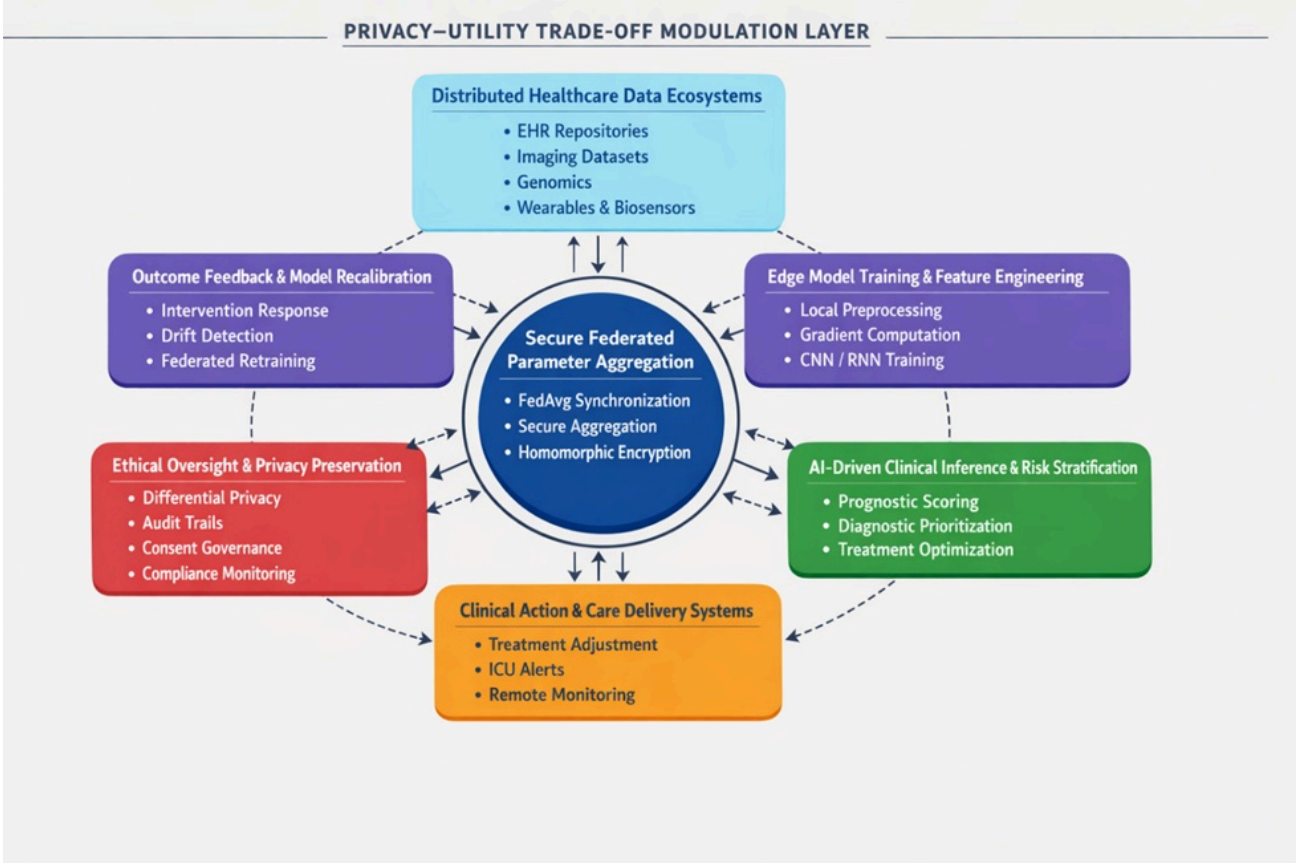

Figure 1. Federated closed-loop architecture for privacy-preserving healthcare analytics and clinical decision intelligence.

Figure 1 illustrates a hexagonal federated learning ecosystem integrating distributed healthcare data ingestion, localized AI training, secure parameter aggregation, clinical decision support, intervention deployment, governance oversight, and feedback-driven recalibration. Privacy-preserving mechanisms—including differential privacy and secure aggregation—operate across the intelligence cycle, balancing analytical utility with regulatory compliance. The architecture conceptualizes federated healthcare systems as adaptive, closed-loop infrastructures that enable continuous learning without centralized data transfer.

Results and Discussion

The discourse on federated learning (FL) ecosystems in healthcare underscores their pivotal role in bridging the gap between data-rich environments and privacy-constrained analytics [1, 4, 6]. This section deliberates on the synthesized implications of FL architectural models, integrating cross-study perspectives to evaluate their efficacy in real-world healthcare systems. FL's decentralized nature fundamentally alters clinical analytics by enabling collaborative intelligence without data centralization, yet it prompts nuanced discussions on interoperability, equity, and ethical deployment [19-21].

Architectural models, such as client-server and hierarchical architectures, have demonstrated robustness in handling healthcare data heterogeneity, as evidenced by multi-institutional predictive analytics for infectious diseases [2, 3, 5]. However, discussions reveal that while these models enhance analytical scalability, they often overlook the socio-technical dimensions of implementation, such as clinician trust in federated outputs [17, 18, 22]. Synthesis across literature posits that FL ecosystems must evolve beyond technical frameworks to incorporate human-centered design, ensuring that privacy trade-offs do not undermine clinical utility [7, 11, 14]. For instance, in medical imaging analytics, FL facilitates shared learning on sensitive datasets, but discussions highlight the risk of diminished interpretability when privacy mechanisms obscure model behaviors [4, 15, 16]. **Table 2** presents systemic trade-offs and governance implications in federated healthcare analytics.

Table 2. Systemic trade-offs and governance implications in federated healthcare analytics

Domain	Federated advantage	Trade-off risk	Governance implication	Mitigation strategies
Data privacy	No raw data sharing	Gradient leakage	Regulatory compliance	Secure aggregation
Model accuracy	Diverse training cohorts	DP noise distortion	Clinical reliability	Adaptive privacy budgets
Infrastructure	Reduced data transfer	Hardware heterogeneity	Institutional inequity	Client selection protocols
Decision support	Generalizable predictions	Interpretability loss	Clinician trust gaps	Explainable FL models
Closed-loop learning	Continuous improvement	Feedback bias	Ethical validation	Monitoring agents
Global collaboration	Cross-border analytics	Policy misalignment	Sovereignty conflicts	Federated governance frameworks

Privacy trade-offs remain a central theme: differential privacy mechanisms protect against inference attacks but can introduce biases against underrepresented subgroups [7, 14, 20]. Integrative analysis suggests that FL's strength lies in its adaptability to governance structures, allowing healthcare systems to tailor privacy levels to specific analytical needs [19, 21, 22]. Yet, discussions emphasize the need for standardized metrics to quantify these trade-offs, moving beyond qualitative assessments to empirical frameworks that balance utility metrics with privacy epsilon values [1, 11]. In closed-loop systems, this translates to dynamic discussions on feedback integrity, where federated updates must be vetted for adversarial influences [6, 13, 18].

Broader implications for healthcare infrastructure include FL's potential to foster equitable analytics, particularly in global health contexts where data sovereignty is paramount [8, 20, 24]. Discussions synthesize that while FL democratizes access to advanced AI, it requires policy alignments to address disparities in computational resources across institutions [19, 26, 27]. Ethical deliberations extend to consent paradigms, advocating for granular patient controls in federated ecosystems [3, 16, 21]. Ultimately, this discussion frames FL as a catalyst for transformative healthcare analytics,

provided that architectural innovations are coupled with interdisciplinary governance to navigate inherent trade-offs [9, 17, 23].

Challenges and limitations

Despite the promise of federated learning (FL) ecosystems, several challenges and limitations impede their widespread adoption in healthcare systems and analytics [1, 4, 6]. Data heterogeneity across federated nodes poses a primary obstacle, as non-IID distributions can lead to model divergence and suboptimal analytical performance [2, 5, 13]. In clinical settings, this manifests in biased predictions when datasets vary by demographics or acquisition protocols, particularly in imaging analytics [4, 7, 16]. Limitations in communication efficiency further compound this, with bandwidth constraints in rural or mobile health infrastructures slowing aggregation cycles [3, 11, 15].

Privacy mechanisms, while essential, introduce inherent limitations: differential privacy's added noise degrades model accuracy, especially for rare-event analytics in healthcare [7, 11, 14]. Trade-offs are acute in high-stakes decisions, where reduced utility could impact patient outcomes [9, 17, 18]. Architectural challenges include vulnerability to poisoning attacks, where malicious clients corrupt global models, necessitating robust detection layers that add complexity [4, 6, 13]. Governance limitations arise from the lack of standardized protocols for incentive alignment, which deters participation by smaller institutions [19-22].

In closed-loop systems, feedback loops amplify these challenges, as real-time recalibration demands low-latency infrastructure that is often absent in federated setups [12, 18, 24]. Synthesis reveals scalability limitations for large-scale deployments, with computational overheads that scale poorly across diverse ecosystems [11, 15, 19]. Ethical challenges encompass fairness, where FL may perpetuate biases if not explicitly mitigated [21, 27]. Overall, these limitations underscore the need for hybrid solutions to enhance the viability of FL in healthcare analytics [8, 23, 27].

Future research directions

Future research in federated learning (FL) ecosystems for healthcare should prioritize advancements in architectural resilience and privacy optimization [1, 4, 6]. Investigating adaptive aggregation algorithms to handle extreme data heterogeneity could enhance analytical robustness, particularly for global health analytics [2, 5, 13]. Directions include integrating blockchain to enable tamper-proof governance and ensure traceable updates in distributed systems [11, 19, 21].

Privacy research should focus on tunable mechanisms that minimize utility loss, such as personalized differential privacy tailored to data sensitivity [7, 14, 20]. In medical imaging, exploring split-learning hybrids could reduce exposure risks while maintaining diagnostic precision [4, 15, 16]. Closed-loop systems warrant studies on automated drift detection to enable proactive recalibration in dynamic clinical environments [12, 18, 22].

Broader directions encompass equity-focused designs and the development of FL frameworks for low-resource settings to bridge digital divides [8, 24, 26]. Interdisciplinary agendas should address human-AI integration and evaluate clinician-AI decision fusion in real-world trials [17, 18, 23]. Finally, longitudinal studies on FL's long-term impact on healthcare outcomes will inform scalable, ethical deployments [9, 19, 27, 28].

Conclusion

Federated learning ecosystems stand as a cornerstone for advancing AI in healthcare systems and analytics, offering architectural models that harmonize collaborative intelligence with privacy imperatives. This review has synthesized their integration into clinical workflows, highlighting privacy trade-offs and the potential for closed-loop innovations. While challenges persist, future directions promise resilient, equitable solutions that could redefine healthcare delivery. Ultimately, FL empowers privacy-preserving analytics, fostering sustainable AI-driven insights for improved patient care.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 02 Feb 2023 Revised: 05 Mar 2023 Accepted: 06 Apr 2023
Published online: 20 July 2023

References

Rieke N, Hancox J, Li W, Milletarì F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. NPJ Digit Med. 2020;3(1):119.
https://doi.org/10.1038/s41746-020-00323-1.

Dayan I, Roth HR, Zhong A, Harouni A, Gentili A, Abidin AZ, et al. Federated learning for predicting clinical outcomes in patients with COVID-19. Nat Med. 2021;27(10):1735-43.
https://doi.org/10.1038/s41591-021-01506-3.

Sadilek A, Liu L, Nguyen D, Safavi M, Cramer E, Yom-Tov E, et al. Privacy-first health research with federated learning. NPJ Digit Med. 2021;4(1):45.
https://doi.org/10.1038/s41746-021-00489-2.

Kaissis GA, Makowski MR, Rückert D, Braren RF. Secure, privacy-preserving and federated machine learning in medical imaging. Nat Mach Intell. 2020;2(6):305-11.
https://doi.org/10.1038/s42256-020-0186-1.

Flores M, Dayan I, Roth HR, Zhong A, Harouni A, Gentili A, et al. Federated learning used for predicting outcomes in SARS-COV-2 patients. NPJ Digit Med. 2021;4(1):5.

https://doi.org/10.1038/s41746-020-00382-4.

Sheller MJ, Edwards B, Reina GA, Martin J, Pati S, Kotrotsou A, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. Sci Rep. 2020;10(1):12598.
https://doi.org/10.1038/s41598-020-69250-1.

Adnan M, Kalra S, Cresswell JC, Taylor GW, Tizhoosh HR. Federated learning and differential privacy for medical image analysis. Sci Rep. 2022;12(1):1953.
https://doi.org/10.1038/s41598-022-05539-7.

Yu M, Tang H, Li H, Kong L, Tong W, Hong H. Integrating artificial intelligence in bedside care for covid-19 and future pandemics. BMJ. 2021;375:e068197.
https://doi.org/10.1136/bmj-2021-068197.

Leist AK, Klee M, Kim JH, Rehner D, Picon A, Karch A, et al. Mapping of machine learning approaches for description, prediction, and causal inference in the social and health sciences. Sci Adv. 2022;8(41):eabk1942.
https://doi.org/10.1126/sciadv.abk1942.

Han T, Xie W, Zissimopoulos A. Breaking medical data sharing boundaries by using synthesized radiographs. Sci Adv. 2020;6(49):eabb7973.
https://doi.org/10.1126/sciadv.abb7973.

Ali M, Naeem F, Tariq MAUR, Kaddoum G. Federated learning for privacy preservation in smart healthcare systems: a comprehensive survey. IEEE J Biomed Health Inform. 2022;27(2):778-89.
https://doi.org/10.1109/JBHI.2022.3181823.

Huang L, Yin Y, Fu Z, Zhang S, Deng H, Liu D. Patient clustering improves efficiency of federated machine learning to predict mortality and hospital stay time using distributed electronic medical records. J Biomed Inform. 2019;99:103291.
https://doi.org/10.1016/j.jbi.2019.103291.

Lee GH, Shin SY. Federated learning on clinical benchmark data: performance assessment. J Biomed Inform. 2020;113:103656.
https://doi.org/10.1016/j.jbi.2020.103656.

Choudhury O, Gkoulalas-Divanis A, Salonidis T, Sylla I, Park Y, Hsu G, et al. Differential privacy-enabled federated learning for sensitive health data. Artif Intell Med. 2022;129:102304.
https://doi.org/10.1016/j.artmed.2022.102304.

Li W, Milletarì F, Xu D, Rieke N, Hancox J, Zhu W, et al. Privacy-preserving federated brain tumour segmentation. Med Image Anal. 2021;72:102104.
https://doi.org/10.1016/j.media.2021.102104.

Haggenmüller S, Schmitt M, Krieghoff-Henning E, Nadji-Ohl M, Truhn D, Utikal JS, et al. Federated learning for decentralized artificial intelligence in melanoma diagnostics. JAMA Dermatol. 2021;157(4):449-56.
https://doi.org/10.1001/jamadermatol.2021.0005.

Wang F. Machine learning for predicting rare clinical outcomes—finding needles in a haystack. JAMA Netw Open. 2021;4(6):e2115503.
https://doi.org/10.1001/jamanetworkopen.2021.15503.

Hughes JW, Yuan C, Petrovič AT, Lee A, Gutti C, Miller DC, et al. Innovation in areas such as federated learning and data privacy and related fields may ultimately enable training medical algorithms across multiple institutions. JAMA Cardiol. 2021;6(11):1281-9.
https://doi.org/10.1001/jamacardio.2021.2549.

Zhong J, Ling L, Wang B, Li T, Wang J, Zhang Y, et al. Advancing the development of real-world data for healthcare research in China: challenges and opportunities. BMJ Open. 2022;12(7):e063139.

https://doi.org/10.1136/bmjopen-2022-063139.

Parbhoo S, Gottesman O, Doshi-Velez F. Operationalising fairness in medical algorithms. BMJ Health Care Inform. 2022;29(1):e100617.
https://doi.org/10.1136/bmjhci-2022-100617.

Saksena N, Mattoo S, Bhan M, Glasner A. Rebooting consent in the digital age: a governance framework for health data exchange. BMJ Glob Health. 2021;6(Suppl 5):e005057.
https://doi.org/10.1136/bmjgh-2021-005057.

AlKnawy B, Adil M, Crooks G, Rhee N, Bates D, Jokiranta S, et al. The Riyadh Declaration: the role of digital health in fighting pandemics. Lancet. 2020;396(10262):1537-9.
https://doi.org/10.1016/S0140-6736(20)31978-4.

Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. Lancet Oncol. 2019;20(5):e262-e273.
https://doi.org/10.1016/S1470-2045(19)30149-4.

Keesara S, Jonas A, Schulman K. Covid-19 and health care's digital revolution. N Engl J Med. 2020;382(23):e82.
https://doi.org/10.1056/NEJMp2005835.

Hollander JE, Carr BG. Virtually perfect? telemedicine for covid-19. N Engl J Med. 2020;382(18):1679-81.
https://doi.org/10.1056/NEJMp2003539.

Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447-53.
https://doi.org/10.1126/science.aax2342.

Leist AK, Klee M, Kim JH, Rehner D, Picon A, Karch A, et al. Machine learning for health services researchers. Value Health. 2019;22(7):808-15.
https://doi.org/10.1016/j.jval.2019.02.012.

Friedman CP, Wong AK, Blumenthal D. Achieving a nationwide learning health system. Sci Transl Med. 2010;2(57):57cm29.
https://doi.org/10.1126/scitranslmed.3001456.