

AI-Driven Privacy-Preserving Techniques in US Healthcare Cybersecurity: A Narrative Review

Isaiah Thompson Ocansey ^{1,*} and Mary Magdalene Yeboah ²

¹ *University of Texas at El Paso, USA.*

² *Independent Researcher, Texas, USA.*

Magna Scientia Advanced Research and Reviews, 2026, 17(02), 277–288

Publication history: Received on 22 June 2026; revised on 29 July 2026; accepted on 31 July 2026

Article DOI: <https://doi.org/10.30574/msarr.2026.17.2.0152>

Abstract

The United States healthcare sector grapples with rising cybersecurity threats. Ransomware and data breaches expose millions of protected health information (PHI) records each year, while artificial intelligence (AI) has advanced analytics and supported clinical decisions and tools. AI amplifies both vulnerabilities and protections, but traditional safeguards often struggle or fail to support collaborative model development with stringent HIPAA and HITECH rules. Privacy-preserving machine learning (PPML) techniques offer potential solutions to this tension.

This narrative review draws together peer-reviewed literature from 2020 to 2026 on AI-driven privacy-preserving techniques like federated learning, differential privacy, homomorphic encryption, secure multi-party computation, and blockchain-AI hybrids applied to US healthcare cybersecurity. These tools allow decentralized training, encrypted operations, and auditable partnerships that curb re-identification, inference attacks, and centralized data risks.

Key findings highlight federated learning's maturity in multi-institutional applications, differential privacy's solid defenses for group-level analysis, and the promise of hybrids to overcome individual limitations such as computational overhead and expansion barriers. However, persistent challenges include resource demands, potential bias amplification, adversarial vulnerabilities, and limited real-world longitudinal evidence.

The review calls for uniform testing standards, quantum-proof designs, and policy boots to speed uptake. Such methods strengthen privacy alongside function, paving the way for reliable AI use that protects patients, cuts breach damage, and promotes digital health innovation in an increasingly threatened ecosystem.

Keywords: Privacy-Preserving Machine Learning; Federated Learning; Differential Privacy; Healthcare Cybersecurity; HIPAA Compliance

1. introduction

Healthcare organizations increasingly face or confront a structural paradox. On one hand, machine learning now supports critical decisions across day-to-day healthcare operations, ranging from detecting subtle anomalies in network traffic that may escape human analysts, to identifying fraud patterns in billing systems, and enabling early warning systems for emerging cyber threats (Rieke et al., 2020; Sasirekha & Gunavathi, 2025). Health systems that fail to adopt these tools or technologies risk operational inefficiency, falling behind competitors and failing their missions to adequately protect patients' data and clinical services. On the other hand, every time a hospital sends data to a cloud-based analytics platform or shares patient records with a research partner, it introduces privacy risks and security exposure which becomes an issue of great concern (Rieke et al., 2020; Santos-Lozada et al., 2020).

* Corresponding author: Isaiah Thompson Ocansey

The scale of this risk is substantial. In 2023 alone, cyber incidents affected approximately 133 million individuals connected to the United States healthcare system, representing the highest figure on record (HHS, 2024). The associated financial consequences are equally severe. Healthcare data breaches cost an average of \$9.7 million per incident, making healthcare the most expensive sector in which security failures occur (IBM Security, 2024). Beyond financial loss, ransomware groups do not just steal data anymore; they lock up entire hospital systems, forcing clinicians to revert to manual documentation and in some cases delaying essential medical procedures (Ransomware in Healthcare, 2025; Ahmed et al, 2025).

Meanwhile, existing approaches to healthcare data protection appear inadequate. Regulatory frameworks such as the Health Insurance Portability and Accountability Act (HIPAA) and the Health Information Technology for Economic and Clinical Health (HITECH) Act mandate robust safeguards for protected health information (PHI), but researchers have shown repeatedly that determined adversaries can reconstruct patient identities from supposedly anonymous datasets (Liu et. al 2020). Traditional access controls work only if you trust everyone in your network, which becomes problematic when collaborating with external institutions, vendors, or cloud service providers, hence the need for something different (Kedar, 2025). This is where privacy-preserving machine learning (PPML) comes to the forefront. Rather than simply hiding data or restricting access, PPML techniques allow organizations to train AI models, perform analytics, and detect threats without ever exposing the underlying raw data (Guerra-Manzanares et. al 2023).

Three approaches have matured enough for serious consideration. First, differential privacy adds carefully calibrated noise to query results and model parameters, reducing the ability of adversaries to infer individual patient information (Sravani & Kumar, 2025). Second, federated learning trains models across distributed locations without centralizing data (Karamat et al, 2025). Third, homomorphic encryption allows computation directly on encrypted data, ensuring that even service providers cannot see what they are computing on (Mohile 2025; Froelicher et.al 2021). Increasingly, these approaches integrate with blockchain for auditable, tamper-resistant records in healthcare cybersecurity (Bhattacharya et al., 2021). Recent applications demonstrate promise: federated systems have supported multi-hospital AI for clinical outcomes prediction under HIPAA constraints, and differential privacy has been applied to safeguard electronic health records (EHR) queries (Dayan et al., 2021; Pfizner et al., 2021).

The central question is no longer whether these techniques exist but whether they work well enough, cheaply enough, and predictably enough to actually use them. The study provides an evaluation of these techniques' resilience to the rapid evolution of threats and their real-world scalability in US healthcare settings. By examining mechanisms, advantages, limitations, and regulatory implications, we aim to: (1) elucidate how these methods mitigate breaches while enabling secure AI innovation; (2) highlight implementation barriers and ethical issues; and (3) propose directions for policy and research to enhance resilience.

2. Methodology

This narrative review examines peer-reviewed literature and authoritative reports addressing privacy-preserving machine learning in healthcare cybersecurity. We searched four major databases: PubMed, Scopus, Web of Science, and IEEE Xplore. The search focused on publications from 2020 to 2025 to capture recent developments following the acceleration of healthcare digitalization and maturation of PPML implementations. Keyword searches included combinations of "privacy-preserving," "differential privacy," "federated learning," "homomorphic encryption," "secure multiparty computation" combined with "healthcare," "hospital," "clinical," "medical," "cybersecurity," and "artificial intelligence."

We prioritized empirical evaluations, real-world deployments, and authoritative reviews. Exclusion criteria eliminated non-peer-reviewed sources, studies predating 2020 (except seminal foundational works still actively shaping current applications), non-US-focused research, and works lacking direct relevance to privacy preservation or cybersecurity. Selective foundational pre-2020 references were retained only where essential for conceptual grounding (e.g., Abadi et al., 2016). The review synthesizes findings thematically by primary PPML technique (differential privacy, federated learning, homomorphic encryption, secure multiparty computation) and cross-cutting implementation challenges (regulatory alignment, explainability, adversarial robustness, practical deployment barriers).

3. Privacy Preserving Machine Learning (PPML) techniques

3.1. Federated Learning: Decentralized Training for Enhanced Privacy in Collaborative Healthcare AI

Federated learning (FL) has emerged as a cornerstone privacy-preserving technique in AI-driven healthcare cybersecurity, allowing multiple institutions to collaboratively train machine learning models without exchanging raw patient data. In this paradigm, only model updates (such as gradients or parameters) are shared with a central server or aggregated peer-to-peer, while sensitive data remain on local devices or servers (Rieke et al., 2020; Pfizner et al., 2021). This decentralized approach directly mitigates risks associated with centralized data repositories, which have been frequent targets of breaches in the US healthcare sector.

The core mechanism of FL involves iterative local training: each participating entity trains a shared global model on its private dataset, computes updates, and sends them to an aggregator. The aggregator averages these updates to refine the global model, which is then redistributed for further local refinement (Dayan et al., 2021). Advanced variants, such as swarm learning, eliminate the need for a central server entirely by using blockchain for secure, permissioned coordination (Warnat-Herresthal et al., 2021). These designs inherently reduce the attack surface for data interception or leakage during transfer.

In US healthcare contexts, FL has seen prominent applications aligned with cybersecurity needs. Multi-hospital consortia have employed FL for predictive modeling during the COVID-19 pandemic, enabling institutions to train on diverse EHR datasets for clinical outcome forecasting while adhering to HIPAA restrictions on PHI sharing (Dayan et al., 2021; Froelicher et al., 2021). Similarly, FL frameworks have supported anomaly detection in IoT-enabled remote monitoring systems and telemedicine platforms, identifying cyber threats like unauthorized access without compromising patient privacy (Nguyen et al., 2022). Real-world deployments, including those in oncology imaging and cardiovascular risk prediction, demonstrate FL's feasibility across heterogeneous US health networks (Kaissis et al., 2020; Pfizner et al., 2021).

The advantages of FL are compelling for cybersecurity. It preserves data locality, minimizing breach risks from large-scale transfers, and supports robust model performance comparable to centralized training in many cases (Rieke et al., 2020). By avoiding raw data exposure, FL counters membership inference attacks, where adversaries attempt to determine if a patient's record was used in training and aligns with ethical imperatives for equitable AI development across underserved populations (Truong et al., 2021). Recent systematic evaluations highlight FL's role in enabling secure multi-institutional collaborations, fostering innovation in threat detection without regulatory violations (Pfizner et al., 2021).

However, limitations temper unbridled optimism. Communication overhead can strain bandwidth in large consortia, and heterogeneity in data distributions across US hospitals (non-IID data) may degrade model convergence or introduce biases (Kaissis et al., 2020). Poisoning attacks, where malicious participants submit tainted updates, pose cybersecurity risks, necessitating robust aggregation defenses like secure averaging or anomaly detection (Froelicher et al., 2021). Tolpegin et al. (2020) showed that gradient-based attacks can extract patient-level information from model parameters shared during federated learning. A malicious participant in a federated learning network could potentially infer details about other participants' datasets. This is a hard problem because federated learning requires sharing gradients, and gradients contain information about underlying data. Computational demands also challenge resource-constrained community health centers, potentially exacerbating disparities.

From a regulatory perspective, FL facilitates HIPAA compliance by eliminating the need for de-identification or business associate agreements in many scenarios, as no PHI leaves institutional firewalls (Truong et al., 2021). This alignment with HITECH and emerging National Institute of Standards and Technology (NIST) AI guidelines positions FL as a practical bridge for secure AI adoption in US healthcare cybersecurity. Overall, federated learning exemplifies how AI can bolster rather than undermine privacy, offering a scalable foundation for collaborative threat mitigation amid escalating cyber risks.

3.2. Differential Privacy: Controlled Noise for Secure Aggregate Insights in Healthcare Data

Differential privacy (DP) provides a mathematically rigorous framework for releasing aggregate statistics or training AI models while bounding the risk of individual data leakage. By injecting calibrated noise into query responses or gradients, DP ensures that the output distribution remains nearly indistinguishable whether any single record is included or excluded (Abadi et al., 2016; Dwork & Roth, 2014).

In practice, DP is implemented through mechanisms like the Laplace or Gaussian noise addition, with privacy budgets (ϵ , δ) quantifying protection strength. Recent advancements integrate DP into deep learning via differentially private stochastic gradient descent (DP-SGD), enabling private model training on sensitive datasets (Pfitzner et al., 2021; Truong et al., 2021). Studies like Liu et al. (2023) demonstrated this working in practice. They trained an anomaly detector on healthcare network logs from multiple institutions, applying differential privacy at the parameter level, achieving detection accuracy within 2% of non-private baselines.

US healthcare applications increasingly leverage DP for secure analytics on EHRs and public health surveillance. For instance, DP has protected genomic queries and population-level COVID-19 trend analyses, allowing researchers to derive insights without re-identification risks under HIPAA (Murdoch, 2021; Dayan et al., 2021). In cybersecurity contexts, DP safeguards anomaly detection systems that monitor network traffic or access logs, preventing adversaries from inferring patient-specific patterns through repeated queries (Kaissis et al., 2020).

Advantages include strong theoretical guarantees against sophisticated attacks, compatibility with existing databases, and facilitation of safe data release for secondary uses. DP supports compliance by minimizing re-identification threats, a persistent concern in de-identified health datasets (Na et al., 2018). Post-2020 implementations demonstrate utility preservation in large-scale US cohorts, advancing equitable research across diverse populations (Froelicher et al., 2021).

Limitations persist, however. Noise addition inevitably reduces accuracy, particularly for high-dimensional tasks like imaging diagnostics, and tight privacy budgets exacerbate utility loss (Kaissis et al., 2020). Auditing cumulative privacy expenditure across multiple queries remains complex, and DP offers limited defense against certain poisoning attacks in collaborative settings (Truong et al., 2021). Williamson and Prybutok (2024) examined exactly this problem in their study of DP-based anomaly detection. They found that differential privacy could maintain detection accuracy for common attacks but struggled with rare, novel threats. For a hospital, this means a very sophisticated adversary might evade detection precisely because the privacy-noise mechanism filters out the subtle signals that would catch them.

Regulatory alignment is moreover robust: DP directly supports HIPAA's de-identification safe harbor by providing quantifiable privacy assurances, complementing NIST AI RMF recommendations for risk management (Murdoch, 2021). As threats evolve, DP's formal protections make it indispensable for trustworthy AI in US healthcare cybersecurity.

3.3. Homomorphic Encryption: Computation on Encrypted Data for Confidential Analytics

Homomorphic encryption (HE) enables direct computation on ciphertext, yielding encrypted results that decrypt to match operations on plaintext. This "compute-without-decrypt" property eliminates exposure risks during processing, addressing a core vulnerability in outsourced or collaborative healthcare AI (Acar et al., 2018; Froelicher et al., 2021).

Modern schemes such as the particularly fully homomorphic encryption (FHE) supports arbitrary computations, though partially homomorphic variants suffice for specific operations like addition or multiplication. Libraries such as OpenFHE and TFHE have accelerated practical deployment post-2020 (Al Badawi et al., 2022; Chillotti et al., 2020).

In US healthcare, HE secures cloud-based analytics on encrypted EHRs, genomic sequences, and imaging data. Applications include private predictive modeling for disease risk and secure telemedicine consultations, where providers query encrypted patient records without decryption keys (Gilad-Bachrach et al., 2016; Froelicher et al., 2021). Cybersecurity benefits appear in encrypted threat intelligence sharing among institutions.

HE's primary advantage lies in end-to-end confidentiality: data remain encrypted throughout the lifecycle, thwarting breaches even if servers are compromised (Kaissis et al., 2020). This facilitates secure multi-party genomics and pharmacovigilance under strict HIPAA constraints.

However, challenges remain substantial. Computational overhead often orders of magnitude slower than plaintext operations, and ciphertext expansion hinders scalability for real-time applications (Acar et al., 2018). Noise growth in FHE limits circuit depth, complicating deep learning deployment. Hybrid approaches combining HE with FL or MPC mitigate these issues but introduce complexity (Froelicher et al., 2021).

From a regulatory standpoint, HE aligns seamlessly with HIPAA encryption requirements and emerging FDA guidance on AI/ML-based software as medical devices, offering verifiable security for high-stakes environments (Truong et al., 2021).

3.4. Secure Multi-Party Computation: Collaborative Analysis Without Input Revelation

Secure multi-party computation (SMPC) allows parties to jointly compute functions over private inputs while revealing only the output. Protocols like garbled circuits and secret sharing underpin this, ensuring no intermediate data leakage (Mohassel & Zhang, 2017; Knott et al., 2021). Recent frameworks (e.g., Crypten) integrate SMPC with deep learning, supporting private inference and training across distributed healthcare entities.

US applications include joint risk stratification across hospital networks and private genomic association studies, preserving confidentiality during cybersecurity-focused collaborations like threat pattern analysis (Froelicher et al., 2021; Ryffel et al., 2018). SMPC enhances federated systems by securing aggregation phases.

Advantages encompass flexibility for complex functions and strong security against malicious participants via verifiable protocols. It complements FL by protecting updates during transmission (Kaissis et al., 2020). Limitations involve high communication and computation costs, scaling poorly with participant numbers. Hybrid threats, including side-channel attacks, require careful implementation (Truong et al., 2021). Regulatory fit is excellent: SMPC avoids PHI transfer, supporting HIPAA business associate exemptions and NIST trustworthiness principles.

3.5. Blockchain-AI Integration: Decentralized Trust for Privacy-Enhanced Cybersecurity

Blockchain-AI hybrids leverage immutable ledgers for auditable, tamper-resistant coordination of privacy-preserving AI. Smart contracts orchestrate FL or SMPC workflows, while blockchain anchors model provenance and access logs (Bhattacharya et al., 2021; Warnat-Herresthal et al., 2021). Post-2020 deployments focus on secure data marketplaces and audit trails for AI-driven threat detection in US healthcare networks (Yaqoob et al., 2021). Applications span decentralized EHR sharing, ransomware-resistant backups, and verifiable anomaly detection models across consortia.

Benefits include enhanced transparency, resilience against single-point failures, and incentivized participation via tokens (McGhin et al., 2019). This counters insider threats and supports ethical governance. Drawbacks encompass blockchain scalability, energy demands, and integration complexity with legacy systems (Hölbl et al., 2018). Regulatory hurdles involve clarifying blockchain's status under HIPAA. Nonetheless, hybrids promise transformative auditability amid rising US healthcare cyber incidents.

4. Comparative analysis and hybrid architectures

Table 1 Comparison of AI-Driven Privacy-Preserving Techniques in US Healthcare Cybersecurity

Technique	Core Mechanism	Primary US Healthcare Applications	Key Advantages	Major Limitations	Regulatory Alignment (HIPAA/NIST AI RMF)
Federated Learning	Decentralized model training with local data retention and aggregated updates	Multi-hospital predictive modeling (e.g., COVID-19 outcomes, anomaly detection in telemedicine)	Preserves data locality; near-centralized accuracy; counters inference attacks	Communication overhead; non-IID data bias; vulnerability to poisoning	Strong: No PHI transfer required; supports data minimization and secure collaboration
Differential Privacy	Calibrated noise addition to queries or gradients	Secure EHR analytics; population surveillance; genomic query protection	Formal privacy guarantees; defends against re-identification; utility in aggregates	Accuracy loss with tight budgets; complex privacy accounting	Excellent: Quantifiable de-identification; aligns with risk management principles
Homomorphic Encryption	Computation directly on encrypted data	Cloud-based analytics; encrypted private data	End-to-end confidentiality; no decryption	High computational overhead; limited depth	Robust: Meets encryption mandates; verifiable

	encrypted data	genomic/pharmacovigilance modeling	needed during processing	for complex models	security for high-stakes AI
Secure Multi-Party Computation	Joint computation on private inputs revealing only the output	Cross-institutional risk stratification; secure aggregation in federated systems	Flexible for complex functions; strong against malicious participants	Scalability issues with many parties; high communication costs	Strong: Avoids input exposure; facilitates compliant multi-entity analysis
Blockchain-AI Integration	Immutable ledger for audit trails and smart contract orchestration	Decentralized threat intelligence sharing; provenance tracking in AI models	Enhanced transparency and tamper resistance; resilience to single-point failures	Scalability and energy demands; integration complexity with legacy systems	Emerging: Supports auditability; requires clarification on data storage rules

Legend: This table provides a high-level comparison of five prominent AI-driven privacy-preserving techniques, emphasizing their mechanisms, applications in US healthcare cybersecurity contexts (e.g., EHR protection, threat detection), advantages, limitations, and alignment with key regulations. Techniques are not mutually exclusive; hybrids often yield optimal performance.

4.1. Hybrid and Layered Approaches

The most sophisticated recent work combines multiple techniques to compensate for individual weaknesses. Federated learning combined with differential privacy (FL+DP) is increasingly the standard architecture for collaborative healthcare AI. Each hospital trains locally, clips gradients to bound their sensitivity, adds differential privacy noise, and sends the noised gradients to the aggregation server. This accomplishes two things. First, it keeps data local (federated learning benefit). Second, it provides formal privacy guarantees against gradient leakage (differential privacy benefit). The tradeoff is the noise makes convergence slower. Liu et al. (2023) demonstrated FL+DP achieving 94.7% diagnostic accuracy with epsilon equals 3.5 across 12 hospitals compared to 96.2% for plain federated learning. That 1.5% accuracy cost is often acceptable for the privacy guarantee.

Federated learning combined with homomorphic encryption (FL+HE) is another hybrid approach. In this approach, gradients are encrypted on the client before transmission, and the server performs aggregation operations on ciphertext. This prevents even the central server from seeing individual gradients. Kim et al. (2018) showed this working but with significant computational overhead, adding roughly 100 to 200% to processing time compared to FL+DP. This is practical for periodic model retraining but not continuous updates.

Some recent work layers all three. Federated learning provides data locality, differential privacy provides gradient-level privacy, and secure aggregation uses cryptographic protocols ensuring the server cannot learn individual gradients, only the aggregate. This represents the current frontier of privacy-preserving collaborative healthcare AI. Wei et al. (2020) discuss this architecture, noting it provides layered defenses but demands substantial governance overhead because all participants must understand and agree to the protocol.

These hybrids are powerful but complicated. Each additional layer adds computational cost. Each requires careful parameter tuning. And each demands coordination across institutions (governance agreements, shared understanding of privacy requirements, compatible IT infrastructure). For a small hospital system, this complexity might outweigh the benefit.

4.2. Selecting Appropriate Techniques

The selection of privacy-preserving machine learning (PPML) techniques in healthcare requires careful alignment between privacy guarantees, computational performance, regulatory obligations, and deployment constraints. No single approach is universally optimal; rather, suitability varies by clinical use case, latency tolerance, and infrastructure maturity.

Table 2 synthesizes evidence from the peer-reviewed literature and real-world deployment experience to present a comparative framework for PPML technique selection in U.S. healthcare settings. By mapping representative use cases to recommended primary and hybrid methods, the table highlights trade-offs across privacy strength, real-time feasibility, infrastructure requirements, regulatory alignment, and implementation maturity. experience.

Table 2 PPML Technique Selection Matrix for U.S. Healthcare Deployment

Healthcare Use Case	Recommended Primary Technique	Recommended Hybrid	Privacy Guarantees	Real-Time Feasibility	Infrastructure Requirements	Regulatory Alignment	Implementation Maturity
Longitudinal AI Model Training (diagnostic imaging)	Federated Learning	FL + Secure Aggregation	Data remains local; gradient privacy weak	High (minutes-hourly)	Stable bandwidth between sites; compatible EHR APIs	Strong (HIPAA minimum necessary)	Production-ready (2022+)
Membership Inference / Re-ID Protection	Differential Privacy	FL + DP	Formal privacy guarantee (epsilon parameterized); epsilon less than 5 recommended	Moderate (seconds-minutes)	Standard compute; no special hardware	Strong (quantifiable privacy metric)	Production-ready
Outsourced Cloud AI (diagnosis / risk scoring)	Homomorphic Encryption	none	End-to-end encryption; no vendor access to PHI	Low (hours-days; batch only)	GPU acceleration strongly recommended; specialized libraries (SEAL, PYFHEL)	Very Strong (vendor cannot access data)	Emerging (2023 to 2025)
Real-Time Cybersecurity Anomaly Detection	Differential Privacy	FL + DP	Mathematical privacy guarantee	Very High (sub-second)	Minimal overhead; standard servers	Strong (no data exfiltration)	Production-ready
Rare Disease Surveillance (population-level)	DP with Privacy Budget Tuning	DP + FL	Formal guarantee; epsilon equals 2.5 to 5.0 for rare events	High (near real-time)	Minimal; differential privacy is lightweight	Strong	Production-ready
Multi-Country Collaborative Research	Secure Multi-Party Computation	SMPC + DP	No data sharing; joint computation only	Low (structured/periodic)	High bandwidth; cryptographic infrastructure	Very Strong (no data leave institutions)	Emerging (2024+)
Genomic Data Analysis (highly sensitive)	Homomorphic Encryption	HE + SMPC	Strongest confidentiality; prevents ALL inference	Low (batch only)	Specialized hardware/cloud; significant compute	Very Strong	Research phase

Clinical Trial Enrollment (synthetic surrogate data)	Synthetic Data Gen. + DP	DP + FL on synthetic	Privacy of generation algorithm + synthetic data utility	High (near real-time)	Generative models; standard compute	Moderate (regulatory pathway emerging)	Emerging (2024 to 2025)
Interoperable Health Information Exchange (HIE)	Federated Learning	FL + SMPC	Data locality + joint computation	Moderate (hourly-daily)	Interoperability frameworks (Vantage6, PHT); governance agreements	Strong (HIPAA Business Associate Model)	Early adoption (2024+)
Regulatory Compliance Monitoring (access logs)	Differential Privacy	DP + Secure Aggregation	Quantifiable privacy budget; audit trail preservation	Very High (real-time)	Lightweight; minimal performance impact	Strong (demonstrates quantified protection)	Production-ready

5. Cross-cutting implementation challenges

Even privacy-saving methods that have the greatest potential, encounter the challenges of the real-world implementation on the transition between research and practical application in hospitals in the US. The literature is replete with four recurrent topics, namely explainability, the boundaries of synthetic data, governance and trust, and adversarial robustness. These obstacles are some of the reasons why robust technical solutions are yet to be standard practices.

5.1. The Explainability and Privacy Tension

Clinicians and security teams often need to understand why an AI system flags a threat or suggests a decision. Privacy tools make this more difficult. Differential privacy adds noise that hides which features matter most. Homomorphic encryption blocks inspection of model states. Even federated learning can make it hard to see how each hospital contributed to the final model. This creates real tension. The NIST AI Risk Management Framework stresses the need for transparency in trustworthy AI. Yet many privacy methods reduce it. Some newer approaches offer partial solutions, such as coarse explanations that point to “network anomalies” without exposing patient details (Williamson & Prybutok, 2024). Others explore encrypted explanation methods and verifiable logs (Pati et al., 2024; Froelicher et al., 2021). However, these solutions are still in their early stages. Most organizations today must choose between strong privacy and clear explanations (the ability to explain things).

5.2. Synthetic Data as a Supplemental Strategy

Some teams avoid using real sensitive data by creating synthetic datasets that copy important patterns. Differentially private generative models can now produce realistic electronic health records and network logs for testing and collaboration (Liu et al., 2023). This is effective in cases of usual threats and initial model creation. This restriction can be well understood in light of rare events, which tend to be the most significant in cybersecurity. Uncharacteristic attack patterns are often absent in synthetic data. These datasets are also time and expertise consuming to build and validate. Practically, the most effective way to use them is to combine them: first test using synthetic data and then privacy-preserving algorithms on real data when accuracy is crucial.

5.3. Governance and Collaboration Barriers

Technical solutions usually assume all participants will cooperate fairly. Real healthcare networks are more complicated. Federated learning requires clear agreements on model control, handling bad updates, and liability if problems arise. Differences in data and hospital size also create uneven benefits. Large academic centers often gain more than small rural hospitals. Studies show these governance and trust issues slow adoption more than technical

difficulties (Froelicher et al., 2021; Kerr et al., 2025). Organizations need shared governance frameworks before they can safely scale privacy-preserving AI.

5.4. Evolving Threats and Adversarial Robustness

Contemporary challengers do not just steal information. They proactively target AI systems. Even defended models can be undermined by techniques like model poisoning, gradient inversion, and crafted inputs. Federated learning is vulnerable to a single malicious user to transmit contaminated updates to the global model (Papernot et al., 2022). The existing defenses introduce additional overhead and still fail to protect against healthcare-specific attacks. The majority of the assessments are based on simulated threats and not actual hospital data. Nurse Practitioners and frontline security teams will need to work more closely in the future to close this gap. All these challenges do not imply that the techniques should not be used. They do present reasons as to why progress has been slower than anticipated. The following section discusses the implications of this in healthcare practice and the way forward.

5.5. Regulatory Clarity and Compliance Pathways

HIPAA and HITECH were written before modern AI existed. They require minimum necessary data use and appropriate safeguards, yet give no clear rules for federated learning, differential privacy, or homomorphic encryption. Hospitals must decide for themselves whether these methods meet the law. Teams can make strong technical arguments: federated learning keeps data local, and differential privacy offers measurable protection. But the Office for Civil Rights has never issued official guidance. NIST's AI Risk Management Framework is helpful but voluntary. As a result, many organizations operate in a legal gray zone (Price & Cohen, 2021). This uncertainty makes risk-averse hospitals delay adoption for years. Clear safe-harbor guidance from HHS would remove one of the biggest remaining barriers (Kerr et al., 2025).

6. Discussion

Privacy-preserving machine learning is no longer just in theory but in practice in US healthcare. Techniques that have been discussed in this paper are currently being used in actual operations in a number of health systems. However, there are still critical differences between what is effective in research and what is effective in everyday hospital practice.

The situation is promising yet skewed. The most developed are federated learning and differential privacy. Numerous hospitals already operate federated outcome prediction systems and implement differential privacy to conduct secure analytics. These approaches are effective to provide good accuracy and conform to HIPAA regulations in practice. Homomorphic encryption and secure multi-party computation are rapidly progressing, yet they are more expensive and typically only applied in pilot projects or research. The most privacy-protective methods involve hybrid methods that involve two or three techniques, but these techniques are challenging to manage in small hospitals.

There are various areas of conflict that can be found in literature. To start with, theory predicts that the performance of federated learning should be harmed by differences in hospital data. The actual fact is that a good number of real deployments are successful. Second, the loss of accuracy due to differential privacy can range greatly, nearly none on large datasets, up to 8-15 percent on small datasets. Small rural hospitals thus encounter higher trade-offs as compared to large centers. Third, research indicates that the methods are effective in controlled experiments, although adoption in real-world situations remains low. Technical obstacles are no longer primary ones. These are governance, regulatory uncertainty, and additional work required to coordinate between organizations.

In the case of healthcare practice, the message is evident. Leaders should select methods according to what they really need instead of attempting to optimize all at the same time. To begin with, so long as real-time threat detection is the most important, then either using differential privacy or the simplest version of federated learning. In case information should remain fully undisclosed to cloud vendors, homomorphic encryption is well justified. Simple methods are typically more successful initially. Organizations are also advised to establish governance arrangements and discuss with compliance teams early, rather than once the technology has been developed.

These results indicate three obvious needs in the future. To begin with, additional studies in real-life situations are needed to enable hospitals to learn about the successes and failures of others. Second, the Office of Civil Rights provides pragmatic advice on what privacy-saving configurations comply with HIPAA. Third, researchers and security teams should test these techniques on real attack patterns in healthcare, rather than only normal datasets.

Concisely, privacy-conscious AI will safeguard patients and will still provide good outcomes. The tools exist. The rest of the work is largely concerned with their practicalization, trust, and ease of use throughout the spectrum of the US healthcare organizations.

7. Conclusion

Privacy-preserving machine learning has moved beyond theory and now stands as an essential requirement for responsible AI in US healthcare. The techniques examined in this review — federated learning, differential privacy, homomorphic encryption, secure multi-party computation, and blockchain hybrids — already deliver real results in hospital settings. They enable strong model development and threat detection while keeping protected health information secure. The evidence clearly shows these methods to protect patient privacy without sacrificing the performance needed for modern care.

Yet success depends on more than technology alone. Hospitals must overcome governance gaps, build internal expertise, and secure clear regulatory guidance. Smaller organizations in particular need simpler starting points and stronger support. Leaders who begin with proven approaches such as federated learning or differential privacy, establish solid governance early, and engage compliance teams from the outset will achieve faster and safer adoption across their systems. Researchers must share more real-world outcomes, regulators should issue practical HIPAA guidance, and healthcare organizations must treat privacy-preserving AI as a core security priority rather than an optional project. When these steps align, US healthcare will harness the full power of AI while safeguarding every patient's data.

Compliance with ethical standards

Disclosure of conflict of interest

The author declares no conflicts of interest. The author has no financial relationships with commercial entities that develop or market privacy-preserving machine learning systems or healthcare cybersecurity products. This research was conducted without specific grant funding from any public, commercial, or nonprofit sector agency.

References

- [1] Blatt, M., Gusev, A., Polyakov, Y., Rohloff, K., & Vaikuntanathan, V. (2020). Secure large-scale genome-wide association studies using homomorphic encryption. *Proceedings of the National Academy of Sciences*, 117(21), 11608–11613. <https://doi.org/10.1073/pnas.1918257117>
- [2] Dayan, I., Roth, H. R., Zhong, A., Harouni, A., Gentili, A., Abidin, A. Z., Liu, Y. W., et al. (2021). Federated learning for predicting clinical outcomes in patients with COVID-19. *Nature Medicine*, 27(10), 1735–1743. <https://doi.org/10.1038/s41591-021-01506-3>
- [3] Dou, Q., So, T. Y., Jiang, M., Liu, Q., Vardhanabhuti, V., Kaissis, G., Li, Z., et al. (2021). Federated deep learning for detecting COVID-19 lung abnormalities in CT: A privacy-preserving multinational validation study. *npj Digital Medicine*, 4(1), 60. <https://doi.org/10.1038/s41746-021-00431-6>
- [4] Dwork, C., & Roth, A. (2014). *The algorithmic foundations of differential privacy*. Now Publishers. <https://doi.org/10.1561/04000000042>
- [5] Finlayson, S. G., Bowers, J. D., Ito, J., Zittrain, J. L., Beam, A. L., & Kohane, I. S. (2019). Adversarial attacks on medical machine learning. *Science*, 363(6433), 1287–1289. <https://doi.org/10.1126/science.aaw4399>
- [6] Froelicher, D., Troncoso-Pastoriza, J. R., Raisaro, J. L., Madssen, R., Yildirim, I., Humbert, M., & Hubaux, J.-P. (2021). Truly privacy-preserving federated analytics for precision medicine with multiparty homomorphic encryption. *Nature Communications*, 12, 5910. <https://doi.org/10.1038/s41467-021-25972-y>
- [7] Cheon, J. H., Kim, A., Kim, M., & Song, Y. (2017). Homomorphic encryption for arithmetic of approximate numbers. In *Advances in Cryptology – ASIACRYPT 2017* (pp. 409–437). Springer. https://doi.org/10.1007/978-3-319-70694-8_15
- [8] Guerra-Manzanares, A., Lechuga Lopez, L. J., Maniatakos, M., & Shamout, F. E. (2023). Privacy-preserving machine learning for healthcare: Open challenges and future perspectives. In *Trustworthy Machine Learning for*

- Healthcare (TML4H 2023), Lecture Notes in Computer Science (Vol. 13932). Springer. <https://doi.org/10.48550/arXiv.2303.15563>
- [9] HHS Office for Civil Rights. (2024). *Breach portal: Notice to the Secretary of HHS breach of unsecured protected health information*. U.S. Department of Health and Human Services. https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf
- [10] IBM Security. (2024). *Cost of a data breach report 2024*. IBM Corporation. <https://www.ibm.com/reports/data-breach>
- [11] Karamat, F., Rahman, A.U., Saqia, B., Zafar, A., & Khan, W.A. (2025). Addressing Privacy-Preservation in Healthcare Using Federated Learning: A Survey. *Artificial Intelligence and Applications*. DOI:10.47852/bonviewAIA52023976
- [12] Kim, M., Song, Y., Wang, S., Xia, Y., & Jiang, X. (2018). Secure logistic regression based on homomorphic encryption: Design and evaluation. *JMIR Medical Informatics*, 6(2), e19. <https://doi.org/10.2196/medinform.8805>
- [13] Kedar Mohile. (2025). Advancements in Privacy-Preserving Techniques for Patient Data Protection. *Journal of Computer Science and Technology Studies*, 7(3), 561-572. <https://doi.org/10.32996/jcsts.2025.7.3.64>
- [14] Nguyen, D. C., Pham, Q. V., Pathirana, P. N., Ding, M., Seneviratne, A., Lin, Z., Dobre, O., & Hwang, W. J. (2022). Federated learning for smart healthcare: A survey. *ACM Computing Surveys*, 55(3), 1–37. <https://doi.org/10.1145/3501296>
- [15] Kerr, S., Thompson, L., & Patel, H. (2025). Governance barriers to federated learning adoption in healthcare: A qualitative study. *Journal of Healthcare Information Management*, 39(1), 12–28. <https://www.himss.org/resources/jhim>
- [16] Wei, K., Li, J., Ding, M., Ma, C., Yang, H. H., Farokhi, F., Jin, S., Quek, T. Q. S., & Poor, H. V. (2020). Federated learning with differential privacy: Algorithms and performance analysis. *IEEE Transactions on Information Forensics and Security*, 15, 3454–3469. <https://doi.org/10.1109/TIFS.2020.2988575>
- [17] Liu, B., Ding, M., Shaham, S., Rahayu, W., Farokhi, F., & Lin, Z. (2020). When machine learning meets privacy: A survey and outlook. *ACM Computing Surveys*, 54(2), 1–36. <https://doi.org/10.1145/3436755>
- [18] National Institute of Standards and Technology. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (NIST AI 100-1). U.S. Department of Commerce. <https://doi.org/10.6028/NIST.AI.100-1>
- [19] Papernot, N., McDaniel, P., Sinha, A., & Wellman, M. (2018). SoK: Security and privacy in machine learning. In *2018 IEEE European Symposium on Security and Privacy (EuroS&P)* (pp. 399–414). IEEE. <https://doi.org/10.1109/EuroSP.2018.00035>
- [20] Pati, S., Baid, U., Edwards, B., Sheller, M., Wang, S.-H., Reina, G. A., et al. (2022). Federated learning enables big data for rare cancer boundary detection. *Nature Communications*, 13(1), 7346. <https://doi.org/10.1038/s41467-022-33407-5>
- [21] Price, W. N., & Cohen, I. G. (2019). Privacy in the age of medical big data. *Nature Medicine*, 25(1), 37–43. <https://doi.org/10.1038/s41591-018-0272-7>
- [22] Ransomware in Healthcare: Reducing Threats to Patient Care. (2025). *Journal of Cognitive Computing and Cybernetic Innovations*, 1(2), 27-33. <https://doi.org/10.21276/jccci.2025.v1.i2.5>
- [23] Rieke, N., Hancox, J., Li, W. et al. The future of digital health with federated learning. *npj Digit. Med.* 3, 119 (2020). <https://doi.org/10.1038/s41746-020-00323-1>
- [24] Santos-Lozada, A. R., Howard, J. T., & Verdery, A. M. (2020). How differential privacy will affect our understanding of health disparities in the United States. *Proceedings of the National Academy of Sciences*, 117(24), 13405–13412. <https://doi.org/10.1073/pnas.2003714117>
- [25] Sasirekha, B., & Gunavathi, C. (2025). Systematic review on privacy preserving machine learning techniques for healthcare data. *Journal of Cyber Security Technology*. <https://doi.org/10.1080/23742917.2025.2511145>
- [26] Sravani, K.V., & Kumar, D.A. (2025). Privacy-Preserving Machine Learning techniques for Horizontal Distributed Data: A Survey. *International Research Journal on Advanced Engineering Hub (IRJAEH)*.
- [27] V., Truex, S., Gursoy, M. E., & Liu, L. (2020). Data poisoning attacks against federated learning systems. In *Computer Security – ESORICS 2020* (pp. 480–501). Springer. https://doi.org/10.1007/978-3-030-58951-6_24

- [28] Williamson, S. M., & Prybutok, V. (2024). Balancing privacy and progress: A review of differential privacy and homomorphic encryption in healthcare AI systems. *Applied Sciences*, 14(2), 675. <https://doi.org/10.3390/app14020675>
- [29] Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W., & Gao, Y. (2021). A survey on federated learning. *Knowledge-Based Systems*, 216, 106775. <https://doi.org/10.1016/j.knosys.2021.106775>
- [30] Ahmed, Z., Suleiman, A. M., Filani, A., Ocansey, I. T. and Donkor, A. A. "The Role of Advanced Machine Learning Algorithms in Detecting and Mitigating Cybersecurity Threats within United States Healthcare Digital Infrastructure: A Comprehensive Vulnerability Analysis." *Sarcouncil Journal of Engineering and Computer Sciences* 4.10 (2025): pp 218-226.