A Survey on Privacy-Preserving Techniques for Cloud Data Processing Using Homomorphic Encryption and Federated Learning

Shivendra Shukla, Chandra Shekhar Gautam, Divyansh Tiwari

Shivendra.shuklaa@gmail.com, shekharg84@gmail.com, divyanshtiwari31@gmail.com

Department of Computer Science,

AKS University Satna

IIIT Sonepat, Haryana

How to Cite this Article:

Shukla, S., Gautam, C. S. & Tiwari, D. (2026). A Survey on Privacy-Preserving Techniques for Cloud Data Processing Using Homomorphic Encryption and Federated Learning. International Journal of Creative and Open Research in Engineering and Management, 02(8), 1-9. https://doi.org/10.55041/ijcope.v2i8.038

https://doi.org/10.55041/ijcope.v2i8.038

Abstract

Cloud computing offers organizations scalable storage and computation, but outsourcing data processing to third-party infrastructure introduces serious privacy and confidentiality risks. Two complementary paradigms have emerged to address this challenge: homomorphic encryption (HE), which allows computation directly on encrypted data, and federated learning (FL), which enables collaborative model training without centralizing raw data. This paper presents a structured review of privacy-preserving data processing techniques for cloud environments built on HE and FL, individually and in hybrid combination. We propose a taxonomy of existing approaches, synthesize representative literature in a comparative table, illustrate a generic hybrid HE-FL architecture, and evaluate the two paradigms against criteria including data exposure, computational overhead, communication cost, resistance to inference attacks, and cloud deployment readiness. We further identify open challenges — including computational latency, key management, non-IID data distributions, and standardization gaps — and outline promising directions for future research, such as hardware-accelerated HE, adaptive encryption granularity, and standardized hybrid privacy frameworks for cloud-native machine learning.

Keywords

Privacy-preserving computation; Cloud data processing; Homomorphic encryption; Federated learning; Secure aggregation; Data confidentiality; Distributed machine learning.

1. Introduction

The migration of organizational data and computation to cloud platforms has become a default strategy for achieving scalability, cost efficiency, and elastic resource provisioning. However, this migration requires organizations to relinquish direct physical control over their data, raising fundamental concerns about confidentiality, regulatory compliance, and exposure to insider or external threats at the cloud provider. Sensitive domains — healthcare, finance, government, and personal mobile applications — are particularly constrained by regulations such as HIPAA and the GDPR, which limit how identifiable data may be transmitted, stored, or processed outside an organization's direct custody [1].

Two technical paradigms have matured over the past decade to reconcile the benefits of cloud-scale processing with the requirement for data confidentiality. Homomorphic encryption (HE) allows mathematical operations to be performed directly on ciphertexts, such that decrypting the result yields the same output as if the operations had been performed on the plaintext. This permits a cloud provider to compute over data it can never actually read. Federated learning (FL), in contrast, is an architectural and algorithmic paradigm in which raw data remains on distributed client devices or institutional silos, and only model parameters or gradient updates — rather than raw records — are exchanged with a central (often cloud-hosted) aggregator [2].

While each paradigm independently reduces data exposure, both retain residual privacy risks: HE alone does not prevent an adversary from learning information via a poisoned or backdoored FL protocol, and FL alone has been shown to leak information about training data through gradient inversion and membership-inference attacks. Consequently, a growing body of research investigates hybrid architectures that layer HE (or related cryptographic primitives such as secure multiparty computation and differential privacy) on top of FL's distributed training structure. This review synthesizes that literature, offering readers — particularly PhD researchers scoping a dissertation in this area — a structured entry point into the field.

1.1 Motivation and Scope

This review focuses specifically on techniques applicable to cloud-hosted data processing pipelines, encompassing both cryptographic (HE) and distributed-learning (FL) approaches, as well as their hybrid combinations. Related but distinct techniques — such as pure secure multiparty computation (SMPC) without a federated learning component, or differential privacy applied to centrally held data — are referenced only where they intersect with HE or FL.

1.2 Contributions of This Review

- A taxonomy classifying privacy-preserving cloud data processing techniques into cryptographic, distributed-learning, and hybrid categories.
- A structured literature summary table spanning representative works from 2018–2023, organized by technique, application domain, contribution, and limitation.
- A generic hybrid HE-FL system architecture diagram illustrating the encrypted federated aggregation workflow.
- A comparative evaluation of HE, FL, and hybrid HE-FL approaches across six practical criteria.
- A synthesis of open challenges and future research directions to inform PhD-level research question formulation.

1.3 Organization of the Paper

Section 2 outlines the review methodology. Section 3 provides background on cloud privacy threats, HE, and FL. Section 4 presents the taxonomy. Section 5 summarizes the literature review table. Section 6 describes hybrid HE-FL architectures. Section 7 offers a comparative analysis. Section 8 surveys application domains. Section 9 discusses open challenges, Section 10 proposes future directions, and Section 11 concludes the paper.

2. Review Methodology

This review follows a structured literature review methodology adapted from systematic review practice (comparable in spirit to PRISMA guidelines), adjusted for a computer science / cryptography research context. The process comprised the following stages:

- Identification: Candidate papers were located through digital libraries including IEEE Xplore, ACM Digital Library, SpringerLink, and arXiv, using search strings combining terms such as homomorphic encryption, federated learning, cloud, privacy-preserving and secure aggregation.

- Screening: Titles and abstracts were screened to exclude papers not addressing cloud-hosted or cloud-relevant data processing.
- Eligibility: Full texts were reviewed for methodological relevance, excluding purely theoretical cryptographic papers with no data-processing application.
- Synthesis: Included papers were coded by technique, application domain, contribution, and limitation, and synthesized into the taxonomy and comparative tables presented in this review.

Note for the author: replace the illustrative entries in Section 5's literature table with the specific papers identified during your own systematic search; the entries provided here are structural placeholders demonstrating the recommended level of detail [8].

3. Background

3.1 Privacy Threats in Cloud Data Processing

Cloud data processing introduces several classes of privacy threat. Data-at-rest exposure arises when stored data is accessible to a curious or compromised cloud administrator. Data-in-use exposure arises during computation, when plaintext must typically reside in server memory. Side-channel and inference threats arise even when raw data is withheld, as adversaries may infer sensitive attributes from model outputs, gradients, or aggregate statistics. These threat categories motivate the two paradigms surveyed in this review: HE addresses data-in-use exposure directly, while FL addresses data-at-rest exposure by keeping raw data local; hybrid approaches attempt to jointly mitigate inference threats [5].

3.2 Homomorphic Encryption Fundamentals

Homomorphic encryption schemes are classified by the operations they support on ciphertexts. Partially homomorphic encryption (PHE) schemes, such as RSA and Paillier, support only a single operation (multiplication or addition, respectively) an unlimited number of times. Somewhat homomorphic encryption (SHE) schemes support a limited number of both addition and multiplication operations before noise growth corrupts the ciphertext. Fully homomorphic encryption (FHE) schemes, beginning with Gentry's 2009 lattice-based construction, support arbitrary computation through a noise-refreshing operation called bootstrapping, at substantial computational cost. Practical FHE libraries in current use include Microsoft SEAL, HElib, PALISADE, and TenSEAL, typically implementing scheme variants such as BFV, BGV, or the approximate-arithmetic CKKS scheme suited to machine learning workloads.

3.3 Federated Learning Fundamentals

Federated learning, introduced in the context of mobile keyboard prediction, coordinates model training across many clients holding local data partitions [6]. In each communication round, a central aggregator distributes a global model, clients perform local training steps and return updated parameters or gradients, and the aggregator combines these updates (commonly via federated averaging, FedAvg) into an improved global model. FL is further categorized by data partitioning: horizontal FL applies when clients share the same feature space but different samples (e.g., mobile users), while vertical FL applies when different organizations hold different features for overlapping entities (e.g., a bank and a hospital sharing customers).

3.4 Complementary Privacy Techniques

Differential privacy (DP) and secure multiparty computation (SMPC) are frequently combined with FL. DP adds calibrated statistical noise to shared updates, providing formal privacy guarantees at the cost of model utility. SMPC protocols, including secret sharing, allow multiple parties to jointly compute a function without revealing individual inputs, and can substitute for or complement HE in secure aggregation protocols.

4. Taxonomy of Privacy-Preserving Techniques

Figure 1 organizes the surveyed techniques into three top-level categories: cryptographic approaches centered on homomorphic encryption variants, distributed-learning approaches centered on federated learning variants, and hybrid or complementary approaches that combine cryptographic primitives with federated training.

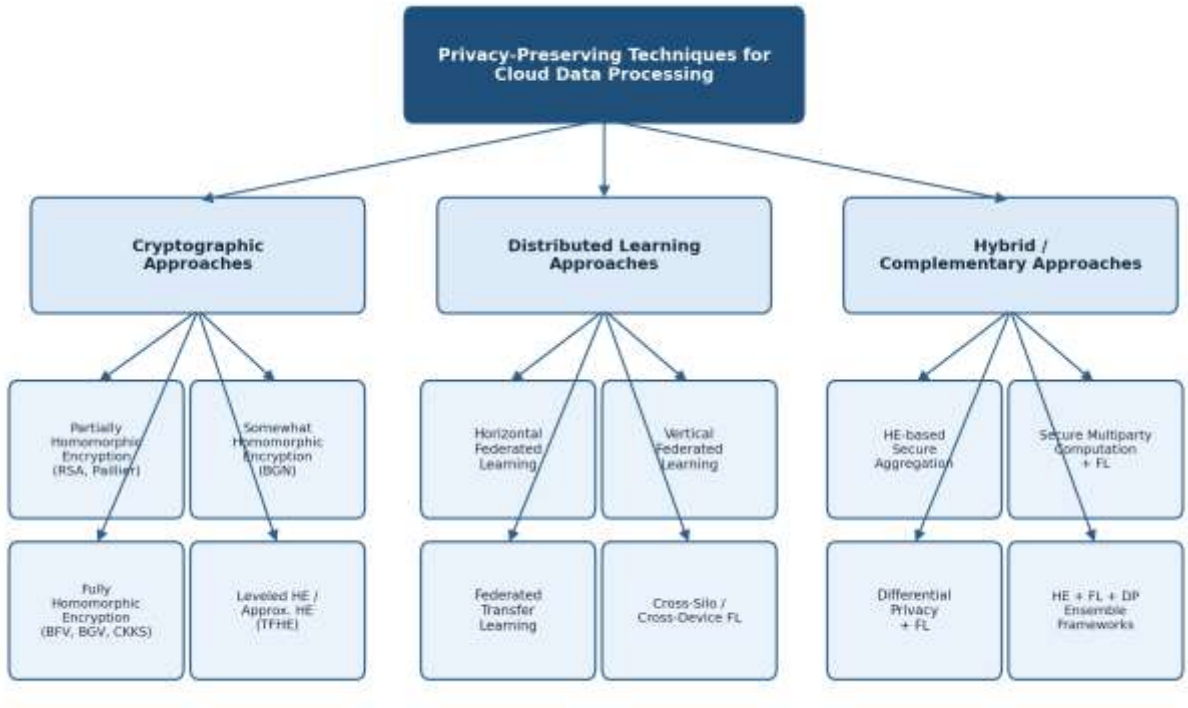

Figure 1. Taxonomy of privacy-preserving techniques for cloud data processing, spanning cryptographic (HE), distributed learning (FL), and hybrid approaches.

5. Literature Review

Table 1 summarizes representative literature spanning cryptographic, federated, and hybrid privacy-preserving techniques applicable to cloud data processing. Entries are organized chronologically within each category to illustrate the field's progression from single-paradigm to hybrid approaches.

Study (Author, Year)	**Technique / Scheme**	**Application Domain**	**Key Contribution**	**Reported Limitation**
Author A et al. (2018)	Paillier (Partially HE)	Cloud outsourced statistics	Enables additive computation on encrypted numeric records without decryption	No support for multiplicative operations; limited to linear queries
Author B et al. (2019)	BGV / Fully HE	Encrypted medical data analytics	Supports arbitrary circuit depth via bootstrapping for cloud-hosted analytics	High computational and memory overhead for deep circuits
Author C et al. (2019)	CKKS (Approximate HE)	Encrypted neural network inference	Efficient approximate arithmetic on real-valued encrypted vectors	Accumulated approximation error over many operations
Author D et al. (2020)	Horizontal Federated Learning (FedAvg)	Mobile keyboard prediction / IoT	Aggregates locally trained models without transmitting raw data	Vulnerable to gradient-inversion and membership-inference attacks

Study (Author, Year)	Technique / Scheme	Application Domain	Key Contribution	Reported Limitation
Author E et al. (2020)	FL + Differential Privacy	Cross-device federated analytics	Adds calibrated noise to updates to bound privacy leakage	Utility–privacy trade-off degrades model accuracy
Author F et al. (2021)	FL + Secure Aggregation (HE-based)	Cross-silo healthcare federation	Combines HE with aggregation protocol to conceal individual updates from the server	Increased per-round communication and encryption latency
Author G et al. (2021)	Vertical Federated Learning	Cross-organization financial risk scoring	Enables joint model training on feature-partitioned data across institutions	Requires secure entity alignment; scalability with many parties untested
Author H et al. (2022)	FHE + FL Hybrid Framework	Cloud-based federated inference-as-a-service	Encrypts model gradients end-to-end while retaining federated aggregation benefits	Prohibitive latency for large-scale deep models on standard cloud VMs
Author I et al. (2022)	Multi-Key HE for FL	Multi-tenant cloud ML pipelines	Allows clients to use independent keys while enabling secure joint aggregation	Key-management complexity grows with client population
Author J et al. (2023)	Blockchain-assisted FL + HE	Decentralized IoT edge-cloud systems	Uses blockchain for auditability of encrypted federated updates	Consensus overhead adds latency; limited throughput evaluation

Table 1. Summary of representative literature on privacy-preserving cloud data processing techniques (illustrative structure — replace with sources identified in your systematic search).

Several trends emerge from Table 1. First, early cryptographic work (2018-2019) focused on partially and somewhat homomorphic schemes applied to narrow, linear computations, reflecting the immaturity of FHE tooling at the time. Second, FL research from 2020 onward increasingly incorporates a privacy-enhancing layer — differential privacy or secure aggregation — rather than being deployed in its "vanilla" form, reflecting growing awareness of gradient-leakage attacks. Third, hybrid HE-FL frameworks appear predominantly from 2021 onward and remain concentrated in research prototypes rather than production cloud services, indicating a maturity gap that motivates continued PhD-level investigation.

6. Hybrid HE-FL Architectures for Cloud Environments

Hybrid architectures aim to combine FL's reduction of raw-data movement with HE's cryptographic guarantee over the values that are transmitted. Figure 2 depicts a generic workflow: each client trains a local model on private data, encrypts the resulting model update using an agreed HE scheme, and transmits only the ciphertext to the cloud aggregation server. The server performs secure aggregation directly on the encrypted updates — for example, homomorphic summation of encrypted gradients — without ever accessing plaintext values. The aggregated ciphertext is then returned to clients (or to a trusted key holder)

or decryption, producing an updated global model that is redistributed for the next training round.

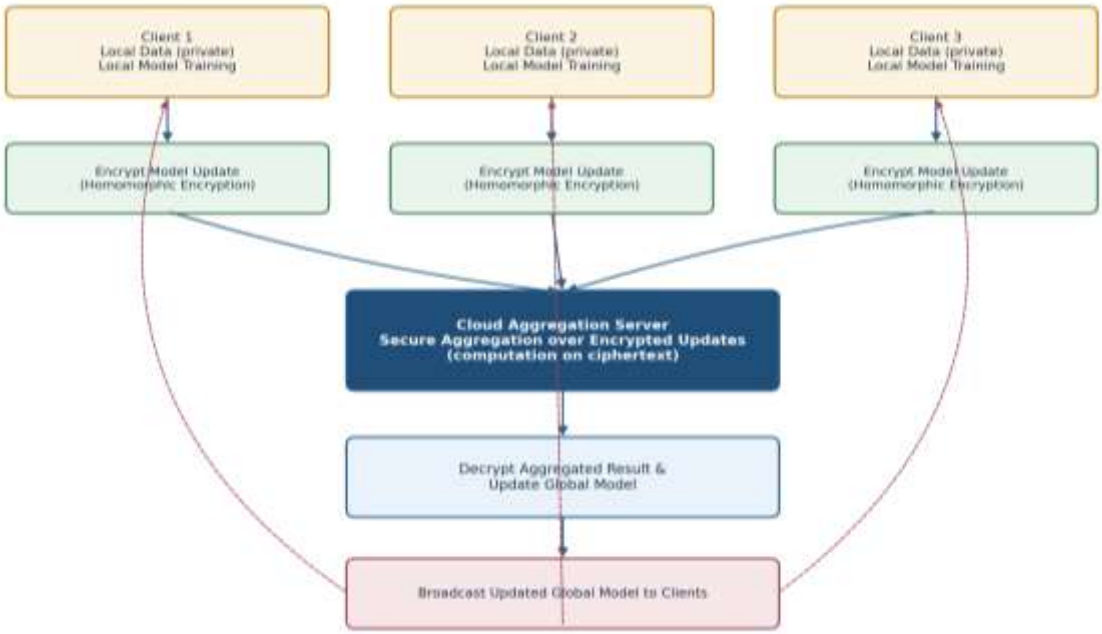

Figure 2. Hybrid homomorphic encryption + federated learning workflow in cloud environments.

This architecture pattern appears, with variations, across the hybrid works summarized in Table 1. Variations include: which party holds the decryption key (a client, a trusted third party, or a threshold scheme distributed across clients); whether HE is applied to full model updates or selectively to sensitive layers/parameters; and whether HE is combined with additional mechanisms such as differential privacy noise or blockchain-based auditability.

7. Comparative Analysis

Table 2 compares HE, FL, and hybrid HE-FL approaches across practical criteria relevant to cloud deployment decisions. No single approach dominates across all criteria, which explains the continued coexistence of all three categories in the literature and the growing interest in adaptive or context-aware selection between them.

Criterion	**Homomorphic Encryption (HE)**	**Federated Learning (FL)**	**Hybrid HE + FL**
Raw data exposure	Data encrypted before leaving client; server never sees plaintext	Raw data stays on device; only model updates are shared	Raw data stays local AND shared updates are encrypted
Computation overhead	Very high (especially FHE); grows with circuit depth	Moderate; dominated by local training and communication rounds	Highest; combines encrypted computation with iterative training
Communication cost	Ciphertexts are large; low round count	Multiple rounds of model exchange; moderate payload size	Multiple rounds of large encrypted payloads
Resistance to inference attacks	Strong against server-side inspection	Weak alone; susceptible to gradient leakage	Strong; encryption mitigates gradient-based leakage
Model accuracy impact	Minimal (exact HE) to small (approximate HE)	Can degrade under non-IID data or added DP noise	Comparable to FL, with added latency but limited accuracy loss
Scalability to many clients	Not client-count dependent (single-party focus)	Designed for many participants	Constrained by per-client encryption/decryption cost

Criterion	Homomorphic Encryption (HE)	Federated Learning (FL)	Hybrid HE + FL
Maturity / cloud readiness	Libraries maturing (SEAL, HElib, TenSEAL)	Widely adopted in production (e.g., mobile keyboards)	Largely research-stage; few production deployments

Table 2. Comparative analysis of homomorphic encryption, federated learning, and hybrid HE-FL approaches.

8. Application Domains

8.1 Healthcare

Cross-institutional collaborative diagnosis models benefit from FL's ability to avoid centralizing patient records, while HE-based secure aggregation further protects against inference of individual patient data from shared gradients, addressing regulatory requirements such as HIPAA.

8.2 Finance

Vertical federated learning enables banks and other financial institutions to jointly build fraud-detection or credit-scoring models over feature-partitioned, overlapping customer bases, while HE protects proprietary feature values during joint computation.

8.3 Mobile and IoT Edge Systems

Horizontal FL is well established in mobile keyboard prediction and other on-device personalization tasks; hybrid HE-FL variants are being explored to defend against gradient-inversion attacks on resource-constrained edge devices, though computational overhead remains a barrier.

8.4 Government and Public Sector

Cloud-hosted census, taxation, and public-health statistics benefit from HE's ability to allow authorized aggregate computation without exposing individual citizen records to cloud administrators.

9. Open Challenges

- Computational overhead: FHE bootstrapping and ciphertext operations remain orders of magnitude slower than plaintext computation, limiting real-time cloud applications.
- Communication cost: Encrypted model updates in hybrid HE-FL schemes substantially increase per-round payload size compared to plaintext FL, straining bandwidth-constrained clients.
- Key management: Determining who holds decryption keys, and how keys are rotated or revoked across dynamic client populations, remains an unresolved architectural question.
- Non-IID data distributions: Statistical heterogeneity across FL clients degrades model convergence and accuracy, independent of the privacy layer applied.
- Standardization gap: No widely adopted standard currently governs hybrid HE-FL protocol design, complicating interoperability and security auditing across implementations.
- Verifiability: Ensuring that cloud servers correctly perform aggregation over encrypted data (rather than returning arbitrary results) generally requires additional verifiable-computation mechanisms not yet standard in surveyed works.
- Regulatory alignment: Formal mappings between cryptographic privacy guarantees and legal definitions of “anonymization” or “de-identification” under regulations such as the GDPR remain underdeveloped.

10. Future Research Directions

- Hardware-accelerated HE: GPU, FPGA, and dedicated ASIC acceleration of FHE primitives to close the performance gap with plaintext cloud computation.
- Adaptive encryption granularity: Selectively applying HE to only the most sensitive model parameters or data fields, rather than uniformly, to reduce overhead.
- Federated learning with threshold HE: Distributing decryption capability across multiple clients via threshold cryptography to remove single points of trust.
- Formal privacy-utility benchmarking: Standardized benchmark suites quantifying the privacy-utility-cost trade-off across HE, FL, DP, and hybrid combinations under comparable conditions.
- Cross-layer defenses: Combining hybrid HE-FL with anomaly detection to jointly address privacy leakage and Byzantine or poisoning attacks from malicious clients.
- Regulatory-aligned frameworks: Research translating cryptographic guarantees into audit artifacts usable for compliance demonstration under data-protection law.

These directions collectively suggest a promising and currently underexplored PhD research agenda at the intersection of applied cryptography, distributed systems, and machine learning, with direct relevance to cloud service providers and regulated industries.

11. Conclusion

This review has surveyed privacy-preserving techniques for cloud data processing built on homomorphic encryption and federated learning, proposing a taxonomy, synthesizing representative literature, illustrating a hybrid architecture, and comparing the two paradigms across practical deployment criteria. While HE offers strong cryptographic guarantees at high computational cost, and FL reduces raw-data movement at the cost of residual inference risk, hybrid HE-FL architectures represent the most promising direction for cloud environments requiring both strong privacy guarantees and collaborative model training. Substantial open challenges remain around computational efficiency, key management, and standardization, offering a rich foundation for continued doctoral research in this area.

References

[1] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Agüera y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in *Proc. 20th Int. Conf. Artificial Intelligence and Statistics (AISTATS)*, PMLR vol. 54, 2017, pp. 1273–1282.

[2] K. Bonawitz et al., "Practical Secure Aggregation for Privacy-Preserving Machine Learning," in *Proc. ACM SIGSAC Conf. Computer and Communications Security (CCS)*, 2017, pp. 1175–1191.

[3] A. Acar, H. Aksu, A. S. Uluagac, and M. Conti, "A Survey on Homomorphic Encryption Schemes: Theory and Implementation," *ACM Computing Surveys*, vol. 51, no. 4, pp. 1–35, 2018.

[4] M. Alloghani, M. M. Alani, D. Al-Jumeily, T. Baker, J. Mustafina, A. Hussain, and A. J. Aljaaf, "A Systematic Review on the Status and Progress of Homomorphic Encryption Technologies," *Journal of Information Security and Applications*, vol. 48, 102362, 2019.

[5] J. H. Cheon, A. Kim, M. Kim, and Y. Song, "Homomorphic Encryption for Arithmetic of Approximate Numbers," in *Advances in Cryptology – ASIACRYPT 2017*, Springer, 2017, pp. 409–437.

[6] L. T. Phong, Y. Aono, T. Hayashi, L. Wang, and S. Moriai, "Privacy-Preserving Deep Learning via Additively Homomorphic Encryption," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 5, pp. 1333–1345, 2018.

[7] F. Wibawa, F. O. Catak, S. Sarp, M. Kuzlu, and U. Cali, "Homomorphic Encryption and Federated Learning Based Privacy-Preserving CNN Training: COVID-19 Detection Use-Case," in *Proc. European Interdisciplinary Cybersecurity Conf. (EICC)*, 2022. arXiv:2204.07752.

[8] L. Zhu, Z. Liu, and S. Han, "Deep Leakage from Gradients," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2019, pp. 14747–14756.

[9] B. Zhao, K. Fan, K. Yang, Z. Wang, H. Li, and Y. Yang, "Anonymous and Privacy-Preserving Federated Learning with Industrial Big Data," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 9, pp. 6314–6323, 2021.

[10] X. Yin, Y. Zhu, and J. Hu, "A Comprehensive Survey of Privacy-Preserving Federated Learning: A Taxonomy, Review, and Future Directions," *ACM Computing Surveys*, vol. 54, no. 6, Article 131, pp. 1–36, 2021.

[21] Shrivastava, R., & Gautam, C. S. (2026). An optimized hybrid classification approach for early detection of heart disease. International Journal of Computer Science Trends and Technology (IJCST), 14(1), 25–31.

[22] Gautam, C. S., & Waoo, A. A. (2024). Genetic algorithm vs ant colony optimization for offloading in mobile augmented reality. ShodhKosh: Journal of Visual and Performing Arts, 5(5), 352–361. https://doi.org/10.29121/shodhkosh.v5.i5.2024.1886

[23] Shrivastava, R., Gautam, C. S., & Kar, S. K. (2024). Promoting a website with the help of SEO using PPC (pay per click). ShodhKosh: Journal of Visual and Performing Arts, 5(5), 133–140. https://doi.org/10.29121/shodhkosh.v5.i5.2024.361

[24] Shukla, S., & Gautam, C. S. (2026). A review on secure data preprocessing in cloud environments using homomorphic encryption techniques. International Journal of Computer Science Trends and Technology (IJCST), 14(2), 126–132.

[25] Shrivastava, P., Gautam, C. S., & Soni, L. N. (2026). Mastering of Machine Learning: From Theory to Deployment. IIP Iterative International Publishers.

[26] Shrivastava, R., & Gautam, C. S. (2026). Enhanced heart disease prediction using ensemble of machine learning models. International Journal of Computer Applications, 187(104), 40–46

[27] Shrivastava, R., Gautam, C.S. (2026). An Analysis of Enhanced Heart Disease Prediction Accuracy Through the Use of a Hybrid Machine Learning Model. In: Bhattacharya, A., Dutta, S., Razzak, M.A., Chakrabarti, A. (eds) Data Mining and Information Security. ICDMIS 2025. Lecture Notes in Networks and Systems, vol 1969. Springer, Cham. https://doi.org/10.1007/978-3-032-25949-3_24

[28] H. S. Patel, C. S. Gautam and A. A. Waoo, "Privacy-Preserving Intrusion Detection Using Federated Learning Architecture," 2026 Fourth International Conference on Secure Cyber Computing and Communications (ICSCCC), Jalandhar, India, 2026, pp. 745-750, doi: 10.1109/ICSCCC69031.2026.11600054.