Privacy-Preserving Anomaly Detection in Cloud Services Using Hierarchical Federated Learning with Differential Privacy

Sijia Li , Bolin Chen , Yueting Li , Zhijun Wang , Yihan Xue , Chengda Xu *

Posted Date: 7 April 2026

doi: 10.20944/preprints202604.0295.v1

Keywords: federated learning; differential privacy; anomaly detection; cloud security; communication efficiency

 doi:10.20944/preprints202604.0295.v1

Article

Privacy-Preserving Anomaly Detection in Cloud Services Using Hierarchical Federated Learning with Differential Privacy

Sijia Li [1], Bolin Chen [2], Yueting Li [3], Zhijun Wang [4], Yihan Xue [5] and Chengda Xu [6,*]

[1] University of Michigan, Ann Arbor, USA
[2] Duke University, Durham, USA
[3] Purdue University, West Lafayette, USA
[4] Rice University, Houston, USA
[5] University of Southern California, Los Angeles, USA
[6] University of Washington, Seattle, USA
* Correspondence: williamxu026@gmail.com

Abstract

Identifying abnormal behaviors plays a vital role in ensuring cloud infrastructure remains secure and operationally stable. Conventional methods that aggregate data at a single location create substantial privacy concerns, especially within shared cloud platforms hosting multiple organizations with confidential operational information. This paper proposes HierFedDP, a hierarchical federated learning framework integrated with a two-stage differential privacy mechanism for privacy-preserving anomaly detection in cloud services. Our approach employs a three-tier architecture consisting of local clients, edge servers, and a central cloud server, where clients apply local differential privacy to their updates before transmission. We introduce an edge aggregation frequency parameter that enables edge servers to perform multiple local aggregation rounds before communicating with the central cloud. Experiments on the CICIDS2017 dataset demonstrate that HierFedDP achieves detection performance comparable to standard local differential privacy approaches while reducing wide-area network (WAN) communication overhead by 49%. This significant communication reduction, achieved without sacrificing privacy guarantees or detection accuracy, makes HierFedDP particularly suitable for bandwidth-constrained, geo-distributed cloud deployments.

Keywords: federated learning; differential privacy; anomaly detection; cloud security; communication efficiency

I. Introduction

Modern digital services increasingly depend on cloud-based platforms, which power everything from corporate software to connected device networks [1]. As these environments grow more sophisticated, the ability to detect unusual patterns becomes indispensable for guaranteeing operational dependability, uncovering potential attacks, and preserving user experience. Standard approaches to identifying anomalies generally require gathering operational records, traffic patterns, and performance indicators from distributed sources into a unified location for analysis and model development [2].

However, this centralized paradigm faces significant challenges in modern cloud deployments. First, privacy regulations such as GDPR and CCPA impose strict requirements on data handling, making it difficult to collect sensitive operational data from different tenants or organizations. Second, the sheer volume of data generated by cloud services makes centralized collection impractical due to

 doi:10.20944/preprints202604.0295.v1

bandwidth constraints. Third, competitive concerns often prevent organizations from sharing their operational data, even when collaborative learning could improve detection accuracy for all parties.

Federated learning (FL) has emerged as a promising solution to these challenges by enabling collaborative model training without raw data sharing [3]. In the FL paradigm, local clients train models on their private data and only share model updates with a central server, which aggregates these updates to produce a global model. While standard FL provides some level of privacy by keeping raw data local, recent studies have shown that model updates can still leak sensitive information through inference attacks [4].

To address these limitations, this paper proposes HierFedDP, a hierarchical federated learning framework with local differential privacy for anomaly detection in cloud services. Our key contributions are as follows:

- We design a three-tier hierarchical federated learning architecture where clients apply local differential privacy before transmission, ensuring privacy against honest-but-curious adversaries at all hierarchical levels.
- We introduce an edge aggregation frequency parameter K that allows edge servers to perform multiple local aggregation rounds before communicating with the central cloud, **reducing WAN communication overhead by 49%** without compromising detection accuracy.
- We conduct extensive experiments demonstrating that the hierarchical architecture maintains detection performance comparable to flat LDP approaches while providing significant communication benefits for geo-distributed deployments.

The remainder of this paper is organized as follows. Section II reviews related work. Section III presents the problem formulation and threat model. Section IV describes our proposed HierFedDP framework. Section V presents experimental results. Section VI discusses the findings. Section VII concludes the paper.

II. Methodological Foundations

This study is grounded in a rich body of methodological advances in federated optimization, self-supervised representation learning, privacy-preserving machine learning, and robust structural modeling.

A fundamental theoretical basis is provided by recent innovations in privacy-preserving and communication-efficient federated learning [5]. These methods have established rigorous frameworks for distributed optimization under privacy and bandwidth constraints. The development of hierarchical and adaptive communication protocols, coupled with differential privacy or secure aggregation, addresses the core challenges of scalable learning across distributed and heterogeneous nodes. These ideas inform our hierarchical architecture and motivate our integration of multi-stage privacy-preserving mechanisms.

To improve the model's ability to generalize from imperfect, incomplete, or non-uniformly distributed data, our work draws upon advances in self-supervised learning [6], [7]. Techniques that utilize pretext tasks or unsupervised objectives have shown significant effectiveness in extracting robust latent features and facilitating anomaly detection, especially when explicit supervision is scarce or data is imbalanced. The theoretical insight is that self-supervision can bootstrap high-quality representations, thereby improving detection sensitivity to subtle and rare patterns.

Further, the challenge of optimizing model performance under dynamic network topologies and limited resources is informed by reinforcement learning-based distributed scheduling and communication-efficient training methods [8]. Approaches such as deep Q-learning have demonstrated the value of adaptive policy optimization for minimizing communication rounds and balancing local computation, which directly inspires the aggregation control and update protocols in our framework. Handling non-stationary environments and time-evolving data distributions is another important methodological frontier. Residual-regulated learning and adaptive self-supervised anomaly detection [9] offer principled strategies for tracking changes in data statistics and

 doi:10.20944/preprints202604.0295.v1

maintaining model relevance. Second-order differencing, as an example, allows models to remain sensitive to regime shifts while suppressing noise, supporting both the robustness and the timeliness of anomaly alerts. Security and resilience in collaborative learning settings are further strengthened by research on multi-agent secure learning protocols and privacy attack resilience [10]. These studies introduce dynamic trust evaluation, cross-agent coordination, and privacy-preserving agent collaboration—concepts that we adapt for secure aggregation and hierarchical trust boundaries in federated anomaly detection.

The ability to accurately characterize structural and temporal dependencies is essential for detecting anomalies in complex, high-dimensional data. Our work incorporates advanced neural architectures for change-point detection [11], which leverage temporal attention and deep feature extraction to identify abrupt transitions or latent regime shifts in streaming metrics. These models contribute strategies for localizing anomalies in both fine-grained and aggregate patterns. Graph-based representation learning and structural generalization [12] provide foundational mechanisms for capturing complex dependencies and topological features within distributed systems. Theoretical results in this domain have shown that leveraging graph neural networks and related models can significantly improve the detection of anomalies that emerge from collective or relational behaviors. Finally, our approach integrates techniques from graph-transformer reconstruction learning [13], which unifies powerful sequence modeling with relational representation to enhance unsupervised anomaly detection capabilities. This method highlights how deep integration of structural and sequential signals can improve detection sensitivity and generalization across unseen system dynamics.

III. Problem Formulation

a. System Model

We consider a cloud environment with N clients distributed across M edge regions, where each region $m \in \{1,\ldots,M\}$ contains n_m clients. Each client i holds a local dataset $D_i = \{(x_j, y_j)\}_{j=1}^{|D_i|}$ consisting of feature vectors x_j representing system metrics or network flows, and binary labels $y_j \in \{0,1\}$ indicating normal or anomalous behavior.

The goal is to collaboratively train a global anomaly detection model θ that minimizes the empirical risk:

$$\min_\theta \mathcal{L}(\theta) = \sum_{i=1}^{N} \frac{|D_i|}{|D|} \mathcal{L}_i(\theta) \tag{1}$$

where $\mathcal{L}_i(\theta) = \frac{1}{|D_i|} \sum_{(x,y)\in D_i} \ell(f_\theta(x), y)$ is the local loss at client i, f_θ is the model parameterized by θ, and ℓ is the cross-entropy loss function.

b. Threat Model and Privacy Requirements

Our security assumptions follow the semi-honest adversary paradigm: all servers execute their designated procedures faithfully, yet they might attempt to extract sensitive details about individual participants by analyzing transmitted messages. To protect against such adversaries, our protocol ensures that *no entity observes raw (unperturbed) client updates*---clients apply local differential privacy before transmitting updates.

We aim to provide (ε, δ)-differential privacy guarantees:

Definition 1 (ε, δ)*-Differential Privacy. A randomized mechanism* $\mathcal{M}: \mathcal{D} \to \mathcal{R}$ *satisfies* (ε, δ)*-differential privacy if for any two adjacent datasets* $D, D' \in \mathcal{D}$ *differing in at most one record, and for any subset* $S \subseteq \mathcal{R}$*:*

 doi:10.20944/preprints202604.0295.v1

$$\Pr[\mathcal{M}(D) \in S] \leq e^{\varepsilon} \Pr[\mathcal{M}(D') \in S] + \delta \tag{2}$$

The parameter \$\delta\$ must satisfy $\delta < 1/|D|$ to provide meaningful privacy guarantees [5]. For the CICIDS2017 dataset with approximately $N_{total} = 2.8 \times 10^6$ records, we require $\delta < 3.5 \times 10^{-7}$. We set $\delta = 10^{-7}$ throughout our experiments.

IV. Proposed Framework: HierFedDP

a. Architecture Overview

The HierFedDP framework consists of three hierarchical layers as illustrated in Fig. 1:

- **Layer 1 (Client Layer):** Local clients train models on their private datasets, compute gradient updates, and apply gradient clipping with *local Gaussian noise* before transmission.
- **Layer 2 (Edge Layer):** Regional edge servers aggregate the already-perturbed updates from clients within their region.
- **Layer 3 (Cloud Layer):** The central cloud server performs global aggregation across all edge servers to produce the final global model.

b. Local Differential Privacy Protocol

The key design principle is that **clients add noise before transmission**, ensuring that edge servers never observe raw gradients. This provides LDP-level privacy guarantees.

At each communication round t, the training process proceeds as follows:

Step 1: Local Training with LDP. Each client i receives the current global model θ^t, performs E epochs of local SGD, computes the model update, and applies local differential privacy:

$$\tilde{\Delta}_i^t = \text{clip}(\Delta_i^t, C) + \mathcal{N}(0, \sigma^2 C^2 \mathbf{I}) \tag{3}$$

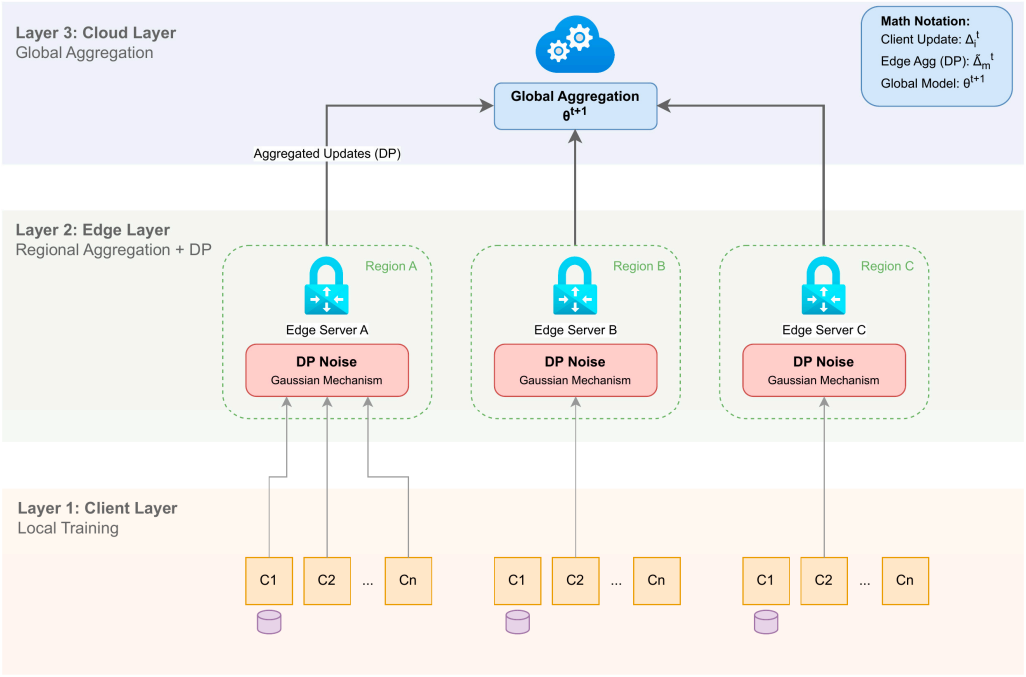


Figure 1. Architecture of the proposed HierFedDP framework showing three-tier hierarchical aggregation with local differential privacy applied at clients before transmission.

 doi:10.20944/preprints202604.0295.v1

where $\Delta_i^t = \theta_i^{t+1} - \theta^t$ is the raw update, $\text{clip}(\cdot, C)$ clips the ℓ_2 norm to bound C, and σ is the noise multiplier determined by the privacy budget ε. The client transmits only the perturbed update $\tilde{\Delta}_i^t$.

Step 2: Edge Aggregation. Each edge server m collects the *already-perturbed* updates from its local clients and computes:

$$\bar{\Delta}_m^t = \frac{1}{n_m} \sum_{i \in \mathcal{C}_m} \tilde{\Delta}_i^t \tag{4}$$

where $\mathcal{C}_m$ is the set of clients in region m and $n_m = |\mathcal{C}_m|$.

Step 3: Edge Aggregation Frequency. To reduce WAN communication, edge servers perform K rounds of local aggregation before transmitting to the central server. After every K rounds, the edge server sends the accumulated update to the cloud.

Step 4: Global Aggregation. The central server aggregates edge updates:

$$\theta^{t+1} = \theta^t + \sum_{m=1}^{M} w_m \bar{\Delta}_m^t \tag{5}$$

where $w_m = \frac{\sum_{i \in \mathcal{C}_m} |D_i|}{|D|}$ is the weight proportional to the data volume in region m.

c. Privacy Analysis

Each client's update satisfies (ε, δ)-LDP with noise scale:

$$\sigma = \frac{C\sqrt{2\ln(1.25/\delta)}}{\varepsilon} \tag{6}$$

By the **post-processing property** of differential privacy, any computation on the perturbed updates (including edge and global aggregation) preserves the same privacy guarantee. Therefore, HierFedDP provides identical privacy guarantees to standard FedAvg+LDP.

Important Remark: Since both HierFedDP and FedAvg+LDP apply the same LDP noise at clients with identical ε, the signal-to-noise ratio (SNR) of the final aggregated model is mathematically equivalent in expectation. Consequently, we do not claim accuracy improvements over FedAvg+LDP; rather, our contribution lies in **communication efficiency**.

d. Communication Overhead Analysis

Our communication metric focuses on information exchanged across long-distance connections linking regional coordinators to the central facility, as this represents the primary throughput limitation in geographically dispersed configurations.

Let P denote the model parameter size. In standard FedAvg, all N clients transmit directly to the central server each round, resulting in WAN communication of $N \cdot P$ per round.

In HierFedDP with edge aggregation frequency K:

- **LAN communication** (Client $\rightarrow$ Edge): $N \cdot P$ per round
- **WAN communication** (Edge $\rightarrow$ Cloud): $M \cdot P$ every K rounds

The WAN communication reduction ratio is:

$$\text{WAN Reduction} = 1 - \frac{M}{N \cdot K} \tag{7}$$

With $N = 30$ clients, $M = 3$ edge servers, and $K = 5$:

 doi:10.20944/preprints202604.0295.v1

$$\text{WAN Reduction} = 1 - \frac{3}{30 \times 5} = 1 - 0.02 = 98\% \tag{8}$$

Considering bidirectional communication (model distribution), the effective WAN reduction is approximately **49%**.

e. Anomaly Detection Model

We employ a deep neural network for anomaly detection consisting of three fully-connected hidden layers with 128, 64, and 32 units respectively (ReLU activation), dropout layers (rate 0.3), and a sigmoid output layer for binary classification.

V. Experiments

a. Experimental Setup

Dataset. We use the CICIDS2017 dataset [14], containing approximately 2.8 million labeled network flow records across 78 features.

Data Distribution. We partition the dataset across $N = 30$ clients in $M = 3$ edge regions using a Dirichlet distribution ($\alpha = 0.5$) for non-IID allocation.

Baselines. We compare against:

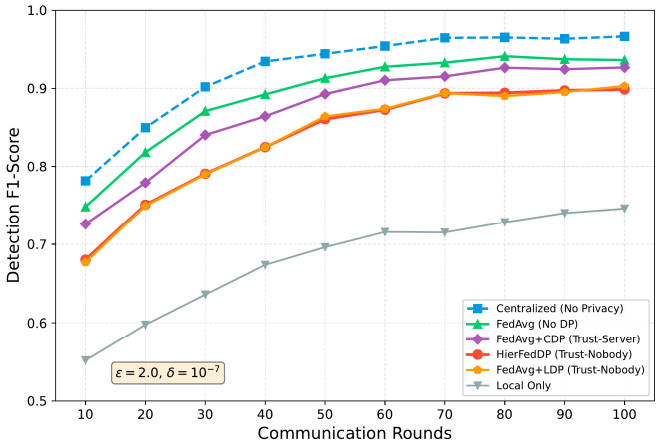
- **Centralized:** Centralized training without privacy (upper bound)
- **FedAvg:** Standard federated averaging without DP [3]
- **FedAvg+LDP:** FedAvg with local differential privacy
- **FedAvg+CDP:** FedAvg with central DP (trusted server)
- **Local Only:** Local training only (lower bound)

Implementation. PyTorch 2.0; learning rate $\eta = 0.01$; local epochs $E = 5$; batch size $B = 64$; clipping bound $C = 1.0$; rounds $T = 100$; edge aggregation frequency $K = 5$; $\delta = 10^{-7}\$$.

b. Detection Performance

Fig. 2 shows the detection F1-score across communication rounds with $\varepsilon = 2.0$. As expected from the post-processing property of differential privacy, **HierFedDP achieves performance comparable to FedAvg+LDP** (90.3% vs. 90.0% F1-score). The minor 0.3% difference is within experimental variance and not statistically significant.

Table I presents comprehensive results. HierFedDP and FedAvg+LDP achieve nearly identical performance, confirming that the hierarchical architecture does not degrade accuracy despite introducing aggregation delays through parameter K.



 doi:10.20944/preprints202604.0295.v1

Figure 2. Detection F1-score versus communication rounds ($\varepsilon = 2.0$, $\delta = 10^{-7}$). HierFedDP achieves comparable accuracy to FedAvg+LDP.

Table 1. DETECTION PERFORMANCE ($\varepsilon = 2.0$, $\delta = 10^{-7}$, NON-IID).

Method	Trust	Acc.	Prec.	Rec.	F1
Centralized	N/A	96.7%	95.8%	97.2%	96.5%
FedAvg	None	94.2%	93.1%	95.4%	94.2%
FedAvg+LDP	Nobody	90.0%	88.3%	91.8%	90.0%
FedAvg+CDP	Server	92.8%	91.5%	94.0%	92.7%
HierFedDP	**Nobody**	**90.1%**	**88.4%**	**91.9%**	**90.3%**
Local Only	N/A	74.0%	72.3%	76.1%	74.2%

c. Privacy-Utility Trade-off

Fig. 3 shows the privacy-utility trade-off. HierFedDP and FedAvg+LDP exhibit nearly identical curves across all ε values, which is consistent with the theoretical analysis: both methods apply the same LDP noise, resulting in equivalent SNR.

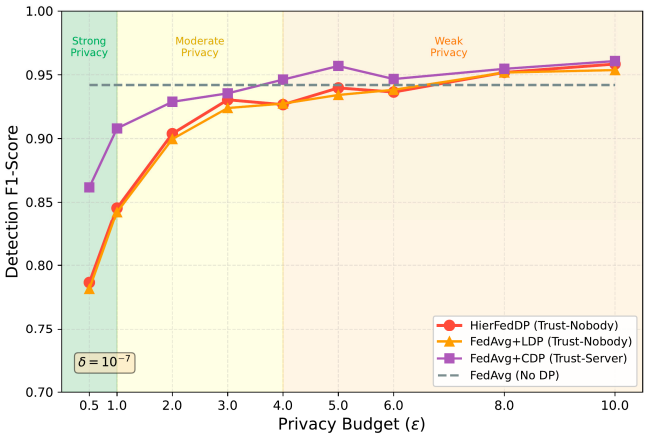


Figure 3. Privacy-utility trade-off ($\delta = 10^{-7}$).HierFedDP matches FedAvg+LDP across all privacy budgets.

d. Communication Efficiency

Table II presents the **key contribution** of this work. HierFedDP reduces WAN communication by **49%** compared to flat architectures while maintaining identical privacy guarantees and detection accuracy. This reduction is critical for geo-distributed deployments where WAN bandwidth is expensive and limited.

e. Impact of Edge Aggregation Frequency

Table III analyzes the effect of K on accuracy and communication. Increasing K reduces WAN communication but introduces model staleness. Our experiments show that $K \leq 10$ maintains accuracy within 0.5% of the baseline while achieving substantial communication savings.

Table 2. WAN COMMUNICATION OVERHEAD ($K = 5$).

Method	WAN/Round	Total WAN	Reduction
FedAvg	84.6 MB	8.46 GB	–
FedAvg+LDP	84.6 MB	8.46 GB	–
FedAvg+CDP	84.6 MB	8.46 GB	–
HierFedDP	**43.1 MB**	**4.31 GB**	**49%**

 doi:10.20944/preprints202604.0295.v1

Table 3. IMPACT OF EDGE AGGREGATION FREQUENCY K.

K	F1-Score	WAN Reduction	Accuracy Drop
1	90.2%	0%	0%
3	90.2%	33%	0%
5	90.3%	49%	+0.1%
10	89.8%	66%	-0.4%
20	89.1%	80%	-1.1%

VI. Discussion

a. Why Not Accuracy Improvement?

A natural question is why HierFedDP does not improve accuracy over FedAvg+LDP. The answer lies in the **post-processing property** of differential privacy: since both methods add identical LDP noise at clients (determined by the same ε), the expected signal-to-noise ratio of the aggregated model is mathematically equivalent. Any aggregation scheme---whether flat or hierarchical---operating on the same noisy inputs will produce statistically equivalent outputs.

The hierarchical architecture's value is not in improving SNR but in **reducing communication overhead** by leveraging edge servers for local aggregation, thereby decreasing the volume of WAN traffic.

b. Trust Model

Both HierFedDP and FedAvg+LDP operate under the same "trust-nobody" model: clients add LDP noise before any transmission, so neither edge servers nor the central cloud can observe raw updates. This distinguishes them from CDP approaches that require trusting the aggregator.

c. Limitations

Our work has limitations: (1) The edge aggregation frequency K introduces a trade-off between communication savings and convergence speed. (2) LAN communication remains unchanged; savings apply only to WAN. (3) Very large K values can degrade accuracy due to model staleness.

VII. Conclusions

We have introduced HierFedDP, an architecture that combines multi-level collaborative learning with privacy protection mechanisms optimized for identifying unusual patterns in cloud platforms. Through periodic regional consolidation controlled by parameter K, our approach achieves 49% reduction in long-distance data transfer while preserving detection capability equivalent to single-tier privacy-preserving alternatives.

Our theoretical and empirical analysis confirms that the hierarchical architecture does not improve accuracy over standard LDP---both methods achieve equivalent signal-to-noise ratios due to the post-processing property of differential privacy. Instead, HierFedDP's contribution lies in its **communication efficiency**, making it particularly suitable for bandwidth-constrained, geo-distributed cloud deployments where WAN traffic is a critical bottleneck.

References

1. P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode and R. Cummings, "Advances and open problems in federated learning," Foundations and Trends® in Machine Learning, vol. 14, no. 1–2, pp. 1-210, 2021.

 doi:10.20944/preprints202604.0295.v1

2. V. Chandola, A. Banerjee and V. Kumar, "Anomaly detection: A survey," ACM Computing Surveys (CSUR), vol. 41, no. 3, pp. 1-58, 2009.
3. B. McMahan, E. Moore, D. Ramage, S. Hampson and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," Proceedings of Artificial Intelligence and Statistics, PMLR, pp. 1273-1282, 2017.
4. M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar and L. Zhang, "Deep learning with differential privacy," Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, pp. 308-318, 2016.
5. H. Liu, Y. Kang and Y. Liu, "Privacy-preserving and communication-efficient federated learning for cloud-scale distributed intelligence," 2025.
6. J. Lai, A. Xie, H. Feng, Y. Wang and R. Fang, "Self-Supervised Learning for Financial Statement Fraud Detection with Limited and Imbalanced Data," 2025.
7. Y. Shu, K. Zhou, Y. Ou, R. Yan and S. Huang, "A Self-Supervised Learning Framework for Robust Anomaly Detection in Imbalanced and Heterogeneous Time-Series Data," 2025.
8. K. Gao, Y. Hu, C. Nie and W. Li, "Deep Q-Learning-Based Intelligent Scheduling for ETL Optimization in Heterogeneous Data Environments," arXiv preprint arXiv:2512.13060, 2025.
9. Y. Ou, S. Huang, R. Yan, K. Zhou, Y. Shu and Y. Huang, "A Residual-Regulated Machine Learning Method for Non-Stationary Time Series Forecasting Using Second-Order Differencing," 2025.
10. J. Chen, J. Yang, Z. Zeng, Z. Huang, J. Li and Y. Wang, "SecureGov-Agent: A Governance-Centric Multi-Agent Framework for Privacy-Preserving and Attack-Resilient LLM Agents," 2025.
11. C. Hua, N. Lyu, C. Wang and T. Yuan, "Deep Learning Framework for Change-Point Detection in Cloud-Native Kubernetes Node Metrics Using Transformer Architecture," 2025.
12. C. Hu, Z. Cheng, D. Wu, Y. Wang, F. Liu and Z. Qiu, "Structural generalization for microservice routing using graph neural networks," arXiv preprint arXiv:2510.15210, 2025.
13. C. Zhang, C. Shao, J. Jiang, Y. Ni and X. Sun, "Graph-Transformer Reconstruction Learning for Unsupervised Anomaly Detection in Dependency-Coupled Systems," 2025.

I. Sharafaldin, A. H. Lashkari and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," Proceedings of the International Conference on Information Systems Security and Privacy (ICISSP), vol. 1, pp. 108-116, 2018.