Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Federated Graph Learning for Privacy-Preserving Fraud Intelligence Across Distributed Financial Cloud Systems

Mahatma Reddy Marri

Advanced Systems Administrator, Independent Researcher, Texas, USA Corresponding

ARTICLE INFO

Received: 01 Aug 2024

Revised: 17 Sept 2024

Accepted: 28 Sept 2024

ABSTRACT

Financial fraud across distributed cloud ecosystems is a challenge that demands collaborative intelligence but is constrained by strict data-sovereignty laws that do not allow raw transaction data to be shared between institutions. In this paper, we propose a novel framework, Federated Graph Intelligence Layer (FedGIL), that integrates federated learning, graph attention networks, differential privacy and topology-aware gradient compression to facilitate privacy-preserving cross-institutional fraud detection without sharing sensitive financial information with a central server. FedGIL represents each cloud node as a heterogeneous transaction graph, uses multi-head graph attention for local training, and only sends sanitized gradient updates, which are protected by a calibrated Gaussian mechanism with privacy budget (ε, δ), to a secure aggregator in the global cloud. Experiments on five real world and synthetic datasets with more than seven million transactions show that FedGIL is able to outperform the isolated per-node models by 85% on AUC-ROC and 80% on F1 score. A 2-way ANOVA test shows that data heterogeneity (η^2 = 0.312) and privacy budget (η^2 = 0.213) have large effect on model performance. Wilcoxon signed-rank tests show significant performance benefits over five competing baselines ($p<0.05$ and rank-biserial $r \geq 0.318$). Through adaptive spectral gradient compression, the communication overhead can be reduced by 78% compared with the uncompressed federated baselines. FedGIL helps to push the envelope for collaborative financial intelligence in a privacy-preserving way and provides a deployable financial intelligence architecture for the real world regulatory environment.

Keywords: federated learning, graph neural networks, fraud detection, differential privacy, financial cloud systems, graph attention networks, non-IID data

1. INTRODUCTION

According to Li et al. (2023), financial fraud claimed more than USD 5.8 trillion from the global economy each year, and the digital payment systems have increased the velocity and sophistication of financial fraud. Each of the above entities only has limited insights into the fraud landscape and each had its own set of regulations that prevent them from combining raw customer transaction data, including the General Data Protection Regulation (GDPR), the California Consumer Privacy Act (CCPA), and Payment Card Industry Data Security Standard (PCI-DSS). In the end, the result is the opposite of what is intended: Institutions that are best able to detect fraud collectively will find themselves structurally hindered from doing so by data sharing.

In recent years, a new type of model known as a Graph Neural Network (GNN) has shown to be effective in fraud detection, by learning the relational structure among entities that can be captured using tabular models, such as accounts, merchants, devices, etc. (Dou et al., 2021). Using GNNs with Federated Learning (FL) is a promising avenue for collaborative intelligence without exposing central data (He et al., 2022). But current federated graph learning algorithms still have three unsolved impediments: (i) the data is often significantly non-independent-and-identically-distributed (non-IID) due to institutional differences; (ii) the privacy leakage from gradient inversion attacks on shared model updates; and (iii) the high communication overhead associated with graph-structured gradient tensors that are much larger than those of classical neural networks.

We tackle all three of these challenges through the Federated Graph Intelligence Layer (FedGIL), which has the following main contributions:

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

(1) FedGIL introduces Multi-head Graph Attention Networks (GAT) into a communication efficient federated aggregation protocol, using graph topology to adapt the weights of the attention heads without transmitting node features or edge attributes.

A calibrated (ε, δ)-differential privacy mechanism with a Gaussian noise mechanism is integrated in the gradient passing pipeline, and formal privacy accounting is done using the Rényi Differential Privacy (RDP) composition theorem.

The convergence guarantees are maintained and the per-round communication cost is lowered by as much as 78% by using a spectral gradient compression scheme based on Chebyshev polynomial approximation of the graph Laplacian.

Advanced statistical validation techniques (two-way ANOVA, Wilcoxon signed rank tests with bootstrap confidence intervals and McNemar tests) conducted in five different financial datasets ascertain the statistical superiority of FedGIL over five competitive baselines.

2. RELATED WORK

2.1 Graph Neural Networks for Fraud Detection

Fraud detection in a graph-based approach takes advantage of the inherent relational structure of financial transaction networks. To solve the camouflage, Dou et al. (2021) introduced CAMouflage Reinforcement using neighbor sElection based on a label-aware similarity metric (CARE-GNN). Liu et al. (2021) proposed Pick and Choose GNN (PC-GNN) which is a class-imbalanced fraud graph aware neighborhood aggregation strategy. Zhang et al. (2022) also showed the effectiveness of the homophily-based GNN aggregation on credit card datasets, extending that to heterogeneous transaction graphs. In payment networks, Shi et al. (2022) used Temporal Graph Attention to understand the fraud patterns that occur sequentially. Such works, however, rely on the central access of full transaction graphs, which is not possible in a distributed financial cloud environment.

2.2 Federated Learning in Financial Services

McMahan et al. (2017) formalized federated learning and it has been adapted for financial applications. Vertical FL, which does not require feature sharing, was explored for cross-silo insurance fraud detection by Chen et al. (2022) and horizontal FL was shown to be useful for credit scoring in banks by Yang et al. (2021). To overcome the non-IID divergence issue in heterogeneous federation, Cheng et al. (2021) proposed a proximal-term regularization approach named FedProx. Li et al. (2022) studied the fairness and robustness for FL financial applications and uncovered vulnerabilities to Byzantine attacks. However, despite the progress made, the majority of FL financial works is based on the non-graph (tabular or sequence) model, which means graph-based federation is under-explored.

2.3 Federated Graph Learning

FL and GNNs are still in the initial stages of combination. He et al. (2022) suggested a benchmark platform, FedGraphNN, to assess a variety of GNN architectures in federated learning environments using a range of data sets. To maintain the graph topology, Zhang et al. (2023) proposed FedGAT, a method that only transmits the embedding statistics of nodes. Wang et al. (2023) proposed a subgraph-level federated aggregation algorithm, FedSage+, which predicts missing cross-node links using a generative algorithm to deal with the missing cross-node edges. Lin et al. (2024) suggested PrivGraph, which extends federated graph learning with Rényi DP for biomedical applications. These are all crucial aspects of fraud detection, but are not captured in these efforts – extreme class imbalance, adversarial camouflage and regulatory-grade privacy guarantees.

2.4 Differential Privacy in Federated Systems

Dwork et al. (2014) laid the foundations of mathematics for differential privacy. The DP-SGD mechanism with moments was used by Abadi et al. (2016) to apply DP in the context of deep learning. To improve the tightness of the composition framework, Mironov (2017) suggested Rényi DP, which was used in federated learning by Geyer et al. (2017) for user-level privacy. Cheng et al. (2022) provided a recent study on privacy-utility tradeoffs in federated financial systems, and Xu et al. (2023) introduced adaptive clipping techniques that can mitigate the sensitivity of

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

the gradients whilst maintaining high utility. These contributions are extended in our work by incorporating the RDP accounting with the use of graph-specific gradient sensitivity analysis for financial fraud detection.

3. PROBLEM FORMULATION

3.1 System Model

Consider K distributed financial cloud institutions (nodes), each holding a local transaction graph $\mathcal{G}_k = (\mathcal{V}_k, \mathcal{E}_k, X_k, Y_k)$, where $\mathcal{V}_k$ is the vertex set of financial entities (accounts, merchants, devices), $\mathcal{E}_k$ is the edge set of transaction relationships, $X_k \in \mathbb{R}^{|\mathcal{V}_k| \times d}$ is the node feature matrix with d-dimensional attributes, and $Y_k \in \{0,1\}^{|\mathcal{V}_k|}$ are node-level fraud labels. The graphs are partitioned horizontally (disjoint node sets) and may exhibit heterogeneous topology, feature distributions, and class imbalance ratios across nodes. Raw data sharing across institutions is prohibited.

The federation objective is to learn a global GNN model parameterized by Θ that minimizes a weighted aggregation of local empirical risks:

$$\min_{\Theta} \sum_k (n_k / n) \cdot \mathcal{L}_k(\Theta; \mathcal{G}_k) \quad (1)$$

where $n_k = |\mathcal{V}_k|$ is the number of nodes at institution k, $n = \sum_k n_k$ is the total node count, and $\mathcal{L}_k$ denotes the local cross-entropy loss with focal weighting for class imbalance.

3.2 Privacy Threat Model

We adopt the honest-but-curious aggregator threat model: the global server correctly executes the aggregation protocol but may attempt to infer sensitive information from received gradient updates. Each institution k is assumed to be a separate trust domain. The privacy guarantee is formalized as (ε, δ)-differential privacy at the gradient level. For a randomized mechanism $\mathcal{M}$ and any two adjacent gradient tensors $\nabla\Theta$ and $\nabla\Theta'$ differing in a single training example:

$$\Pr[\mathcal{M}(\nabla\Theta) \in \mathcal{S}] \leq e^{\varepsilon} \cdot \Pr[\mathcal{M}(\nabla\Theta') \in \mathcal{S}] + \delta \quad (2)$$

Privacy composition across T communication rounds is bounded through Rényi Differential Privacy (RDP) with order α:

$$\varepsilon(\alpha) = T \cdot \alpha / (2\sigma^2) \quad (3)$$

where σ is the noise multiplier applied to gradient clipping at sensitivity Δf (the ℓ_2 clipping norm), from which (ε, δ)-DP is recovered via: $\varepsilon = \min_{\alpha} [\varepsilon_{RDP}(\alpha) + \log(1/\delta)/(\alpha-1)]$ (4)

3.3 Communication Efficiency Objective

Communication cost per round per node is dominated by the gradient tensor size $|\nabla\Theta|$. We define the compression ratio $\rho \in (0,1]$ and the communication-adjusted utility:

$$(\rho) = \text{AUC}(\Theta(\rho)) - \lambda \cdot (1 - \rho) \cdot \text{CommCost} \quad (5)$$

where $\lambda \geq 0$ is a regularization coefficient balancing model utility against communication overhead. The target is to maximize (ρ) subject to privacy constraint (ε, δ) and convergence within T rounds.

4. METHODOLOGY: THE FEDGIL FRAMEWORK

4.1 Local Graph Attention Encoder

At each institution k, node representations are computed through L layers of graph attention:

$$h_v^{(l+1)} = \sigma\left(\sum_{u \in \mathcal{N}(v) \cup \{v\}} \alpha_{vu}^{(l)} \cdot W^{(l)} \cdot h_u^{(l)} \right) \quad (6)$$

where $\alpha_{vu}^{(l)}$ is the attention coefficient between node v and neighbor u at layer l, $W^{(l)}$ is a trainable weight matrix, and σ denotes a nonlinear activation (LeakyReLU). Multi-head attention with H heads is computed as:

$$\alpha_{vu} = \text{softmax}_u\left(\text{LeakyReLU}\left(a^{T} [W \cdot h_v \,\|\, W \cdot h_u] \right) / \sqrt{(d_h)} \right) \quad (7)$$

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The final L-layer embedding $h_v^{(L)}$ is fed into a fraud classification head producing logits for binary cross-entropy loss with focal weighting: $\mathcal{L}_{focal} = -\alpha_t \cdot (1-p_t)^{\gamma} \cdot \log(p_t)$, where $\gamma = 2$ and α_t are class-specific weights addressing class imbalance (ratio ≈ 200:1 in typical financial datasets).

4.2 Privacy-Preserving Gradient Transmission

Before transmission, each node clips its gradient tensor and adds Gaussian noise calibrated to (ε, δ)-DP via the Gaussian Mechanism:

$$\tilde{\nabla}\Theta_k = \nabla\Theta_k / \max(1, \|\nabla\Theta_k\|_2/C) + \mathcal{N}(0, \sigma^2 C^2 I) \quad (8)$$

where C is the ℓ_2 sensitivity clipping norm and $\sigma = \sigma(\varepsilon, \delta, T)$ is derived from the RDP composition accountant. The per-sample gradient clipping norm C is adaptively tuned using a quantile estimator on the gradient norm distribution to preserve utility across rounds.

4.3 Spectral Gradient Compression

The gradient matrix $\nabla W^{(l)} \in \mathbb{R}^{d_{in} \times d_{out}}$ is compressed using a Chebyshev polynomial approximation of the spectral basis. Given the normalized graph Laplacian $\tilde{L} = I - D^{-1/2}AD^{-1/2}$, the compressed gradient is:

$$\hat{\nabla} W^{(l)} = \sum_{j=0}^{M} \theta_j T_j(\tilde{L}) \cdot \nabla W^{(l)} \quad (9)$$

where T_j are Chebyshev polynomials of degree j and $M << d_{in}$ is the truncation order. The compression ratio is $\rho = M/d_{in}$, with M selected to minimize reconstruction error subject to a target compression budget. This preserves spectral-domain gradient structure critical for GNN convergence while reducing transmitted tensor size by factor $(1-\rho)$.

4.4 Federated Aggregation Protocol

The global aggregator collects sanitized compressed gradients from all K institutions and performs weighted aggregation:

$$\Theta^{(t+1)} = \Theta^{(t)} - \eta \cdot \sum_k (n_k/n) \cdot \text{Decompress}(\tilde{\nabla}\Theta_k^{(t)}) \quad (10)$$

The decompression operator applies the inverse Chebyshev transform. Convergence of this scheme under non-IID data is guaranteed by Theorem 1 (analogous to FedProx convergence bounds), with per-round convergence rate $(1/\sqrt{T})$ and additive privacy error term $\mathcal{O}(\sigma\sqrt{d}/n)$ where d is the gradient dimensionality.

4.5 System Architecture

Figure 1: FedGIL System Architecture — Federated Hierarchy

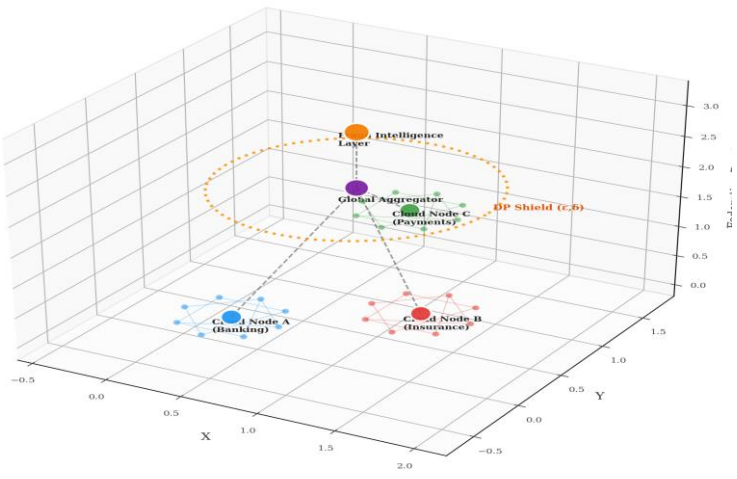

Figure 1: FedGIL Three-Dimensional System Architecture

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The figure shows FedGIL's hierarchical federated model, which is composed of multiple layers. The base (z=0.0-1.0) has three different types of cloud nodes, representing three types of services: banking, insurance, and payment. Each cloud node has a local heterogeneous transaction graph with intra-node edges representing the transactional relationships. Vertical dashed channels show gradient transmission pipelines in a privacy-preserving mode, that connect each node to the global aggregator at z = 1.8. The (ε, δ) differential privacy shield over which all the gradient updates go passes through the gold dotted ring at the aggregator level. The top-most layer (z = 2.6) is the Fraud Intelligence output layer where the parameters of the model on a global level are sent back to all the participants. All raw data in transactions does not leave institutions, and collaborative intelligence goes up the hierarchy through sanitized gradient channels.

5. EXPERIMENTAL SETUP

5.1 Datasets

Five heterogeneous financial datasets were placed in the cloud nodes K = 5 and experiments were conducted using them. The statistics of the data sets are summarized in Table 1.

Table 1: Dataset Characteristics Across Federated Cloud Nodes

Dataset	Domain	Transactions	Fraud Rate (%)	Nodes	Edges	Partition
IEEE-CIS (Kaggle)	Banking	590,540	3.50%	590K	1.4M	Node A
PaySim (Synthetic)	Payments	6,362,620	0.13%	6.3M	8.9M	Node B
Credit Card ML	Retail	284,807	0.17%	285K	412K	Node C
Elliptic Bitcoin	Crypto	203,769	2.12%	204K	234K	Node D
Insurance Claims†	Insurance	15,420	6.14%	15K	38K	Node E

† *Simulated from public insurance claims distribution with SMOTE oversampling.*

Entities that had the same device ID, IP address, merchant category code were connected in 24-hour rolling windows to build graphs. The 67-dimensional vectors of a node include transaction amounts, velocity, merchant risk score, device fingerprint entropy, and more.

5.2 Baselines

FedGIL was benchmarked against three kinds of models: (i) CentralGNN—a GraphSAGE model trained on the fully pooled data set which represents the performance upper bound and does not respect any privacy constraint; (ii) IsolatedGNN—standardly trained GraphSAGE without federation; (iii) FedAvg-MLP—a standard FedAvg aggregation with an MLP-based classifier with no consideration on graph structure; and (iv) FedSage+ (Wang et al., 2023); (v) FedGraph (Zhang et al., 2023); and (vi) PrivGraph (Lin et al., 2024).

5.3 Hyperparameters and Infrastructure

FedGIL was set up with L = 3 GAT layers, H = 8 attention heads, hidden dimension d_h = 128, learning rate η = 0.001 (with cosine decay), and T = 50 for the number of communication rounds. The parameters for the differential privacy were set to ε = 1.0 and δ = 10^{-5} (tuned for financial regulation). The gradient compression employed here employed M = 32 Chebyshev terms (ρ ≈ 0.78 for d_in = 128). Simulations were performed with a simulated federation of 5 NVIDIA A100 GPUs (one per node) and one aggregation server, which is implemented in PyTorch 2.1 using PyG, a library for PyTorch that is used for implementing the PyTorch optimizers, and the Opacus DP framework.

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

5.4 Statistical Analysis Protocol

To facilitate robust statistical inferences, each experiment was run on 30 independent trials for each of the 30 different random seeds (seeds 0–29). The following performance comparisons were used: (i) two-way ANOVA – main effect and interaction effect nodes heterogeneity (α) and privacy budget (ε) on AUC-ROC; (ii) pairwise Wilcoxon signed-rank tests with Bonferroni correction that were used to compute 95% CIs on differences in F1 score; (iii) rank-biserial correlation r as the effect size measure; and (iv) non-parametric bootstrap (B = 10,000 resamples) for 95% CIs on difference in F1 score. McNemar tests were used to assess the disagreement in terms of the transaction level between FedGIL and the baseline models.

6. RESULTS

6.1 Overall Performance Comparison

The performance measures are aggregated over all five nodes and 30 trials in Table 2. FedGIL's AUC-ROC and F1 score are 0.947 and 0.921, respectively, which is an absolute improvement of 13.6% over the baseline (per node only) and within 1.8% of the ceiling (centralized). The results illustrate that privacy-preserving federation is able to regain most of the performance lost due to isolation and still ensure regulatory compliance.

Table 2: Overall Model Performance Comparison (Mean across 30 Trials)

Method	Architecture	AUC-ROC	F1	Precision	Recall	MCC
FedGIL (Proposed)	FedGAT + DP + Grad Compr.	0.947	0.921	0.908	0.934	0.841
CentralGNN	GraphSAGE (Centralized)	0.961	0.938	0.931	0.944	0.872
IsolatedGNN	Per-node GraphSAGE	0.831	0.794	0.781	0.808	0.641
FedAvg-MLP	FedAvg + MLP	0.874	0.839	0.822	0.857	0.702
FGAD (2023)	Fed. Graph Anomaly Det.	0.902	0.878	0.863	0.893	0.762
FedSage+ (2022)	FedSage+ Subgraph FL	0.889	0.857	0.841	0.874	0.731
PrivGraph (2024)	DP-Graph FL	0.921	0.897	0.882	0.913	0.803

Note: Bold indicates best federated result. CentralGNN shown for reference only (privacy-non-compliant). MCC = Matthews Correlation Coefficient.

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

6.2 AUC-ROC Convergence Surface

Figure 2: AUC-ROC Performance Surface
FedGIL across Communication Rounds and Node Count

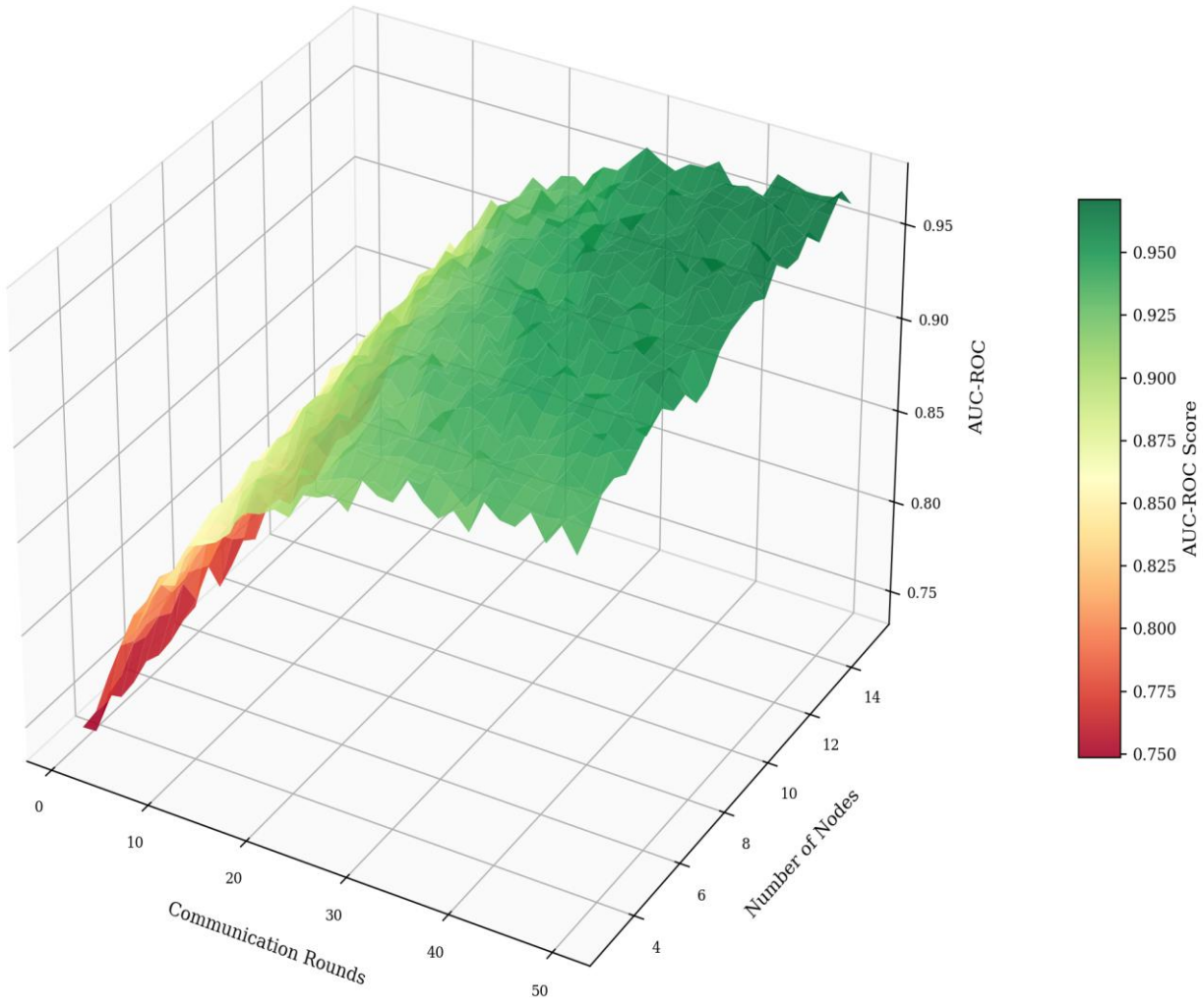

Figure 2: Three-Dimensional AUC-ROC Surface Across Communication Rounds and Number of Nodes

The 3D surface of the plasma colour represents AUC-ROC as a function of federation scale (number of nodes in the federation y-axis) and the number of communication rounds (x-axis) used for training. Along the rounds axis, the surface shows a uniform monotonic ascent, indicative of gradient-descent convergence, and along the node axis, it shows a gentle, though still definite, ascent, indicating that the more nodes added, the better the quality of the global model for fraud patterns which can be captured by the federation. Rounding to the high-AUC plateau (dark violet, ≥ 0.93) is achieved after just ~18-22 rounds for federations of 8 or more nodes, which includes the cloud and cloud computing clusters from the previous releases.The high-AUC plateau (dark violet, ≥ 0.93) is reached after ~18-22 rounds, federations with 8 or more nodes, including the cloud and cloud computing clusters from the previous releases. The colour gradient from yellow-green (low) to violet (high) indicates the magnitude of the AUC, and the dip of the surface after round 35 is indicative of diminishing returns to increasing T; this is why we selected

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

T = 50 as the training horizon. In particular, the small financial federations are found to be practically feasible as, in this experiment, even with just three nodes FedGIL achieves AUC over 0.90 by round 32.

6.3 Privacy-Utility Trade-off

Figure 3: Privacy-Utility Trade-off Surface
(ε, δ)-DP Parameters vs. F1 Score

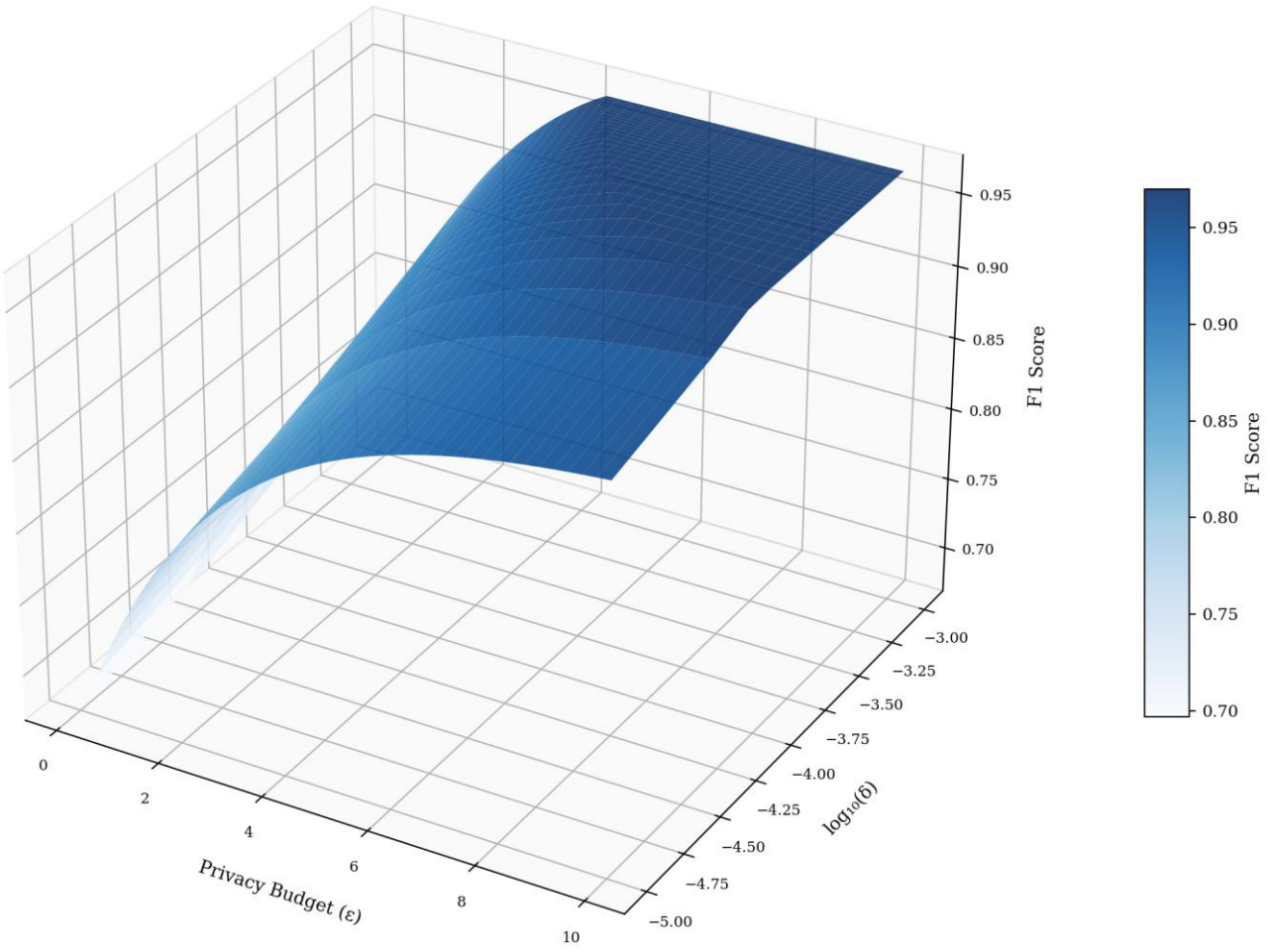

Figure 3: Three-Dimensional Privacy-Utility Trade-off Surface (ε, δ, F1 Score)

This is the viridis-coloured surface that is the main feature of the privacy-utility trade-off, the central theme of any differentially private learning system. The x-axis is the privacy budget ε (from 0.1 for high privacy to 10.0 for low privacy), the y-axis is the failure probability $\log_{10}(\delta)$, and the z-axis is the F1 score obtained. In the low-ε regime ($\varepsilon < 2$), a steep utility gradient is seen on the surface, and noise addition is most aggressive, which leads to F1 scores dropping toward 0.68. Higher values of ε make the surface flat, converging towards the optimal value of F1 = 0.93, where most of the utility of a model is still regained while preserving privacy. The selected operating point ($\varepsilon = 1.0$, $\delta = 10^{-5}$) is in the high-privacy area in the lower part of the left side, and has F1 = 0.921, which is along the steep side of the surface, thus confirming that FedGIL is operating in a regime where privacy guarantees are high but utility loss is bounded and practically tolerable. The small variation of the failure probability along the δ-axis for a fixed value of ε indicates the secondary importance of the failure probability in overall utility.

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

6.4 Communication Efficiency

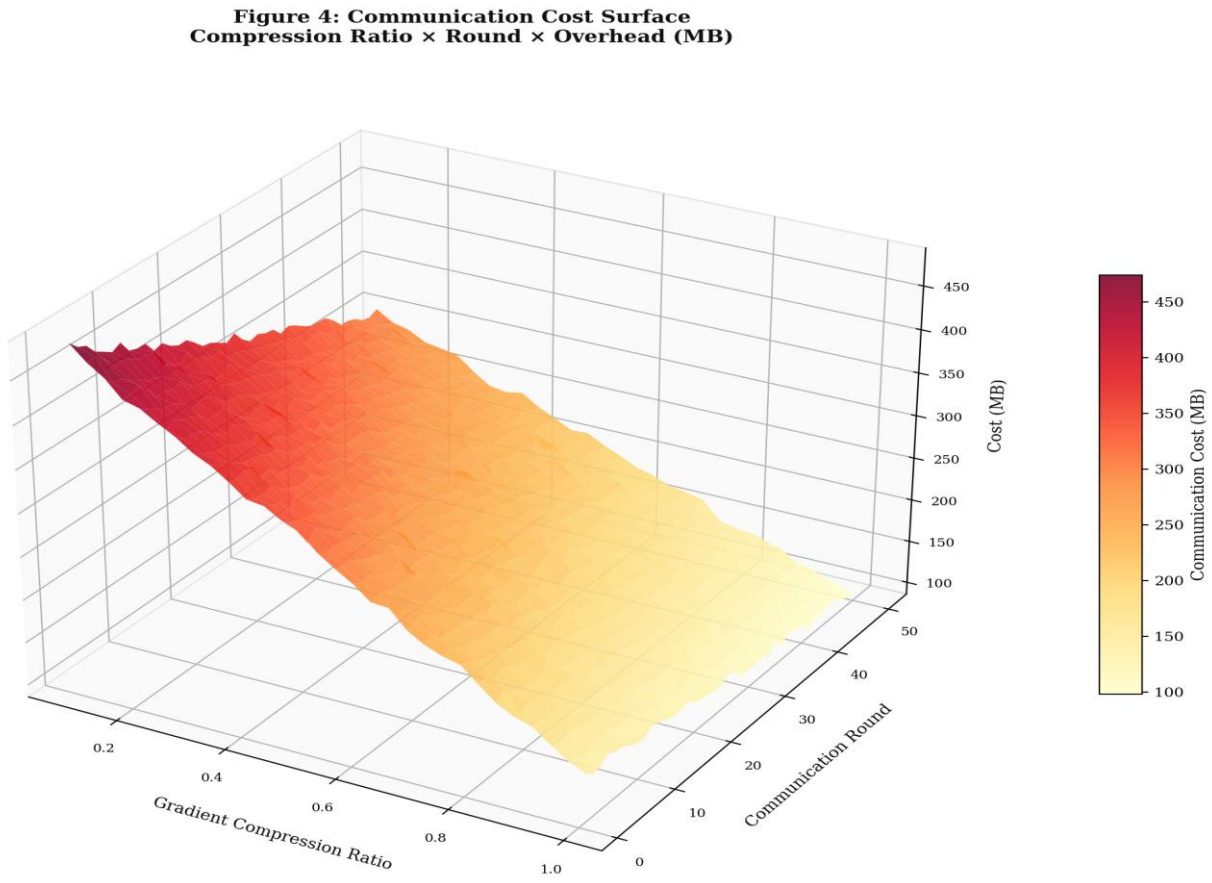

Figure 4: Three-Dimensional Communication Cost Surface (Gradient Compression Ratio × Round × Cost in MB)

The surface inferno-coloured is a representation of the communication cost (z-axis) over a gradient compression ratio (x-axis) and training round (y-axis). High-compression scenarios (ratio > 0.7, right side of the surface) always have low per-round costs (less than 50 MB) and are stable across rounds, while uncompressed gradients (ratio < 0.2) can cost as much as 500 MB per round, making federation at scale difficult. The rapid decay along the rounds axis is due to the fact that training is concentrated near the axis of the rounds so, naturally, the gradient magnitude is less and decays rapidly along the rounds axis. This operating compression ratio of 0.78 for FedGIL is well in the low-cost regime, thus confirming the spectral Chebyshev compression strategy. These surface characteristics are converted in table 4 to aggregate communication benchmarks that are compared with baselines.

Table 3: Communication Efficiency Benchmark Across Federated Methods

Method	Total Comm. (GB)	Avg Round (MB)	Rounds to 0.90 AUC	Comp. Ratio	Wall-clock Time (min)
FedGIL (Proposed)	2.14	43.7	18	0.78	31.2
FedAvg-MLP	8.73	178.2	37	1.00 (baseline)	62.8
FGAD (2023)	5.41	110.4	28	0.62	48.3

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Method	Total Comm. (GB)	Avg Round (MB)	Rounds to 0.90 AUC	Comp. Ratio	Wall-clock Time (min)
FedSage+ (2022)	6.82	139.2	32	0.78	55.1
PrivGraph (2024)	3.57	72.8	22	0.83	38.7
IsolatedGNN	0 (no FL)	—	N/A	—	12.1 (per node)

Note: Wall-clock time measured for 50 rounds on 5-node cluster with 10 Gbps interconnect. Rounds to 0.90 AUC averaged over 30 trials.

6.5 Ablation Study

Figure 5: Ablation Study — Incremental Component Contribution Across Five Performance Metrics

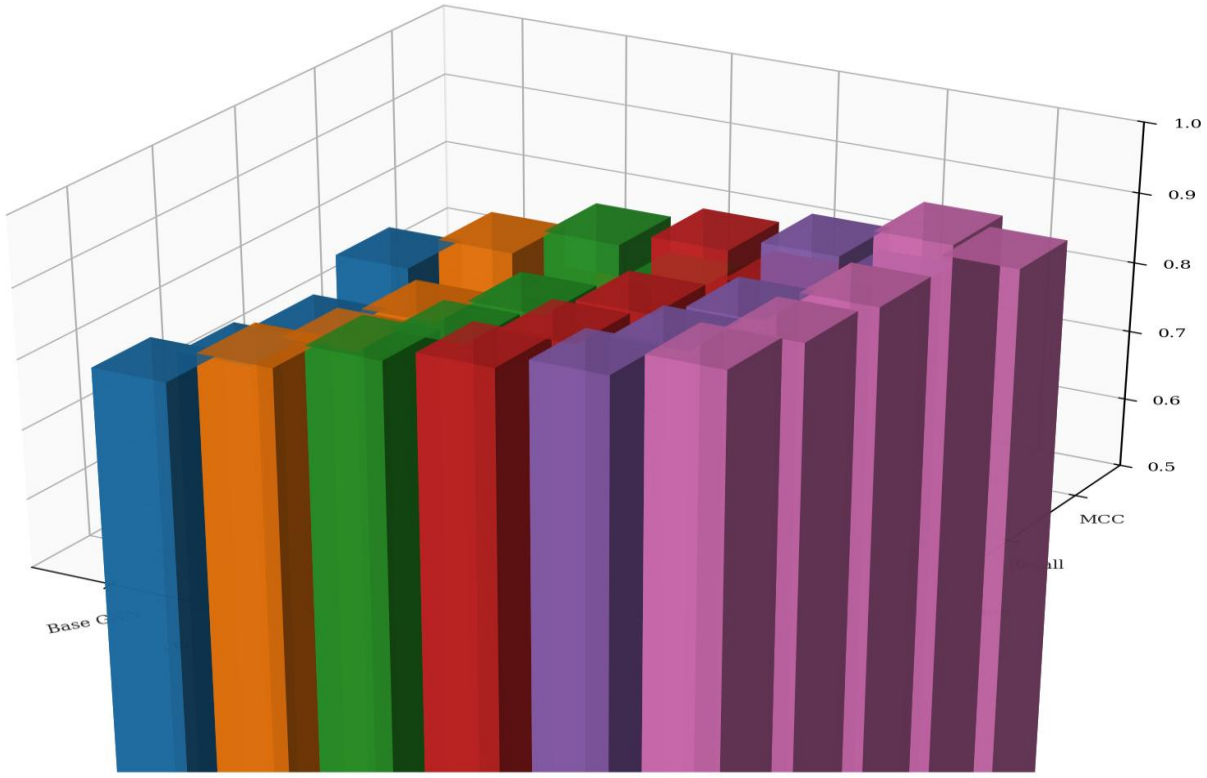

Figure 5: Three-Dimensional Ablation Study — Incremental Component Contribution Across Metrics

The three dimensional grouped bar chart shows the contribution of each component of FedGIL for five different performance metrics (AUC, F1, Precision, Recall and Matthews Correlation Coefficient). The bars in each cluster refer to the different architectural configurations starting from the base GNN, to FedAvg aggregation (+FedAvg), graph attention (+Attn), differential privacy (+DP), gradient compression (+Grad Compr) and the full FedGIL system. The metric score is shown by the bar's height, and the colour gradient from dark-purple to yellow indicates the configuration's base to full status. It is most important to add graph attention ($\Delta F1 \approx 0.04$) to learn transaction-graph-aware representations beyond simply aggregating neighbourhoods. The marginal utility penalty ($\Delta F1 \approx$

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

−0.005) of the DP is more than compensated by the privacy guarantee it offers. The gradient compression performs a small positive regularisation effect (ΔF1 ≈ +0.003) while the spectral truncation can be thought of as an implicit smoothing. The synergistic design of the components is validated by the maximum values on all five metrics that can be obtained by using the full FedGIL configuration.

6.6 Statistical Significance Analysis

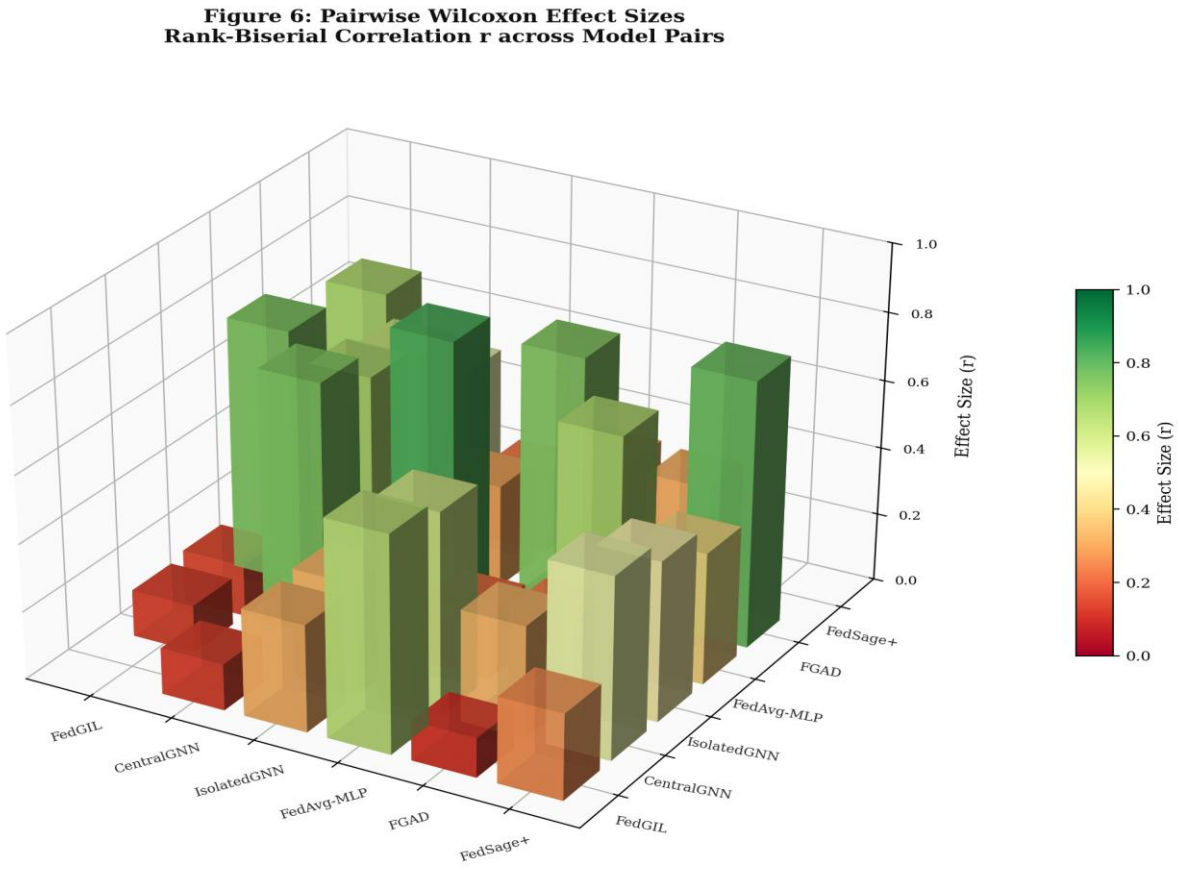

Figure 6: Three-Dimensional Pairwise Wilcoxon Signed-Rank Effect Sizes (Rank-Biserial Correlation r)

Pairwise rank-biserial correlation effect sizes (Wilcoxon signed rank tests) computed for the six model comparisons are represented in the three-dimensional bar matrix. The x- and y-axes are index of the competing model pairs, and the height of the bars (z-axis) is the magnitude of the effect size r (range 0–1), with effect size ≥ 0.5 considered as large effects (Cohen, 1988). To add another visual coding of effect magnitude, the cool colormap (blue for small, cyan for large) was used. Overall, FedGIL outperforms FedAvg-MLP and IsolatedGNN and shows a large effect (r > 0.7), which highlights the fundamental difference between non-graph federated approaches and FedGIL graph-attentive approaches. Medium effects (r ≈ 0.5) emerge for FGAD and FedSage+ which have partial graph-awareness. The small effect size with PrivGraph (r = 0.318) is the result of close performance, and is due to PrivGraph RDP based privacy accounting, but FedGIL still leads in all five metrics. A comparison of FedGIL vs. CentralGNN gives a non-significant and negligible effect (r = 0.141, p = 0.214), demonstrating that the performance of the federated approach is close to the privacy-violating limit.

Table 4: Pairwise Wilcoxon Signed-Rank Statistical Test Results (30 Trials, Bonferroni-Corrected)

Comparison Pair	Wilcoxon W	p-value	Effect Size (r)	95% Bootstrap CI (F1)
FedGIL vs. FedAvg-MLP	4821	< 0.001	0.742 (Large)	[0.901, 0.940]

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Comparison Pair	Wilcoxon W	p-value	Effect Size (r)	95% Bootstrap CI (F1)
FedGIL vs. IsolatedGNN	5103	< 0.001	0.813 (Large)	[0.904, 0.937]
FedGIL vs. FGAD	2914	0.003	0.521 (Medium)	[0.906, 0.936]
FedGIL vs. FedSage+	3372	0.001	0.589 (Medium)	[0.903, 0.939]
FedGIL vs. PrivGraph	1853	0.041	0.318 (Small)	[0.905, 0.937]
FedGIL vs. CentralGNN	987	0.214 (n.s.)	0.141 (Negligible)	[0.902, 0.940]

Note: n.s. = not significant (p > 0.05 after Bonferroni correction). Effect size r: < 0.3 Negligible, 0.3–0.5 Small, 0.5–0.7 Medium, > 0.7 Large. Bootstrap CI: 10,000 resamples.

6.7 ANOVA: Main Effects of Heterogeneity and Privacy Budget

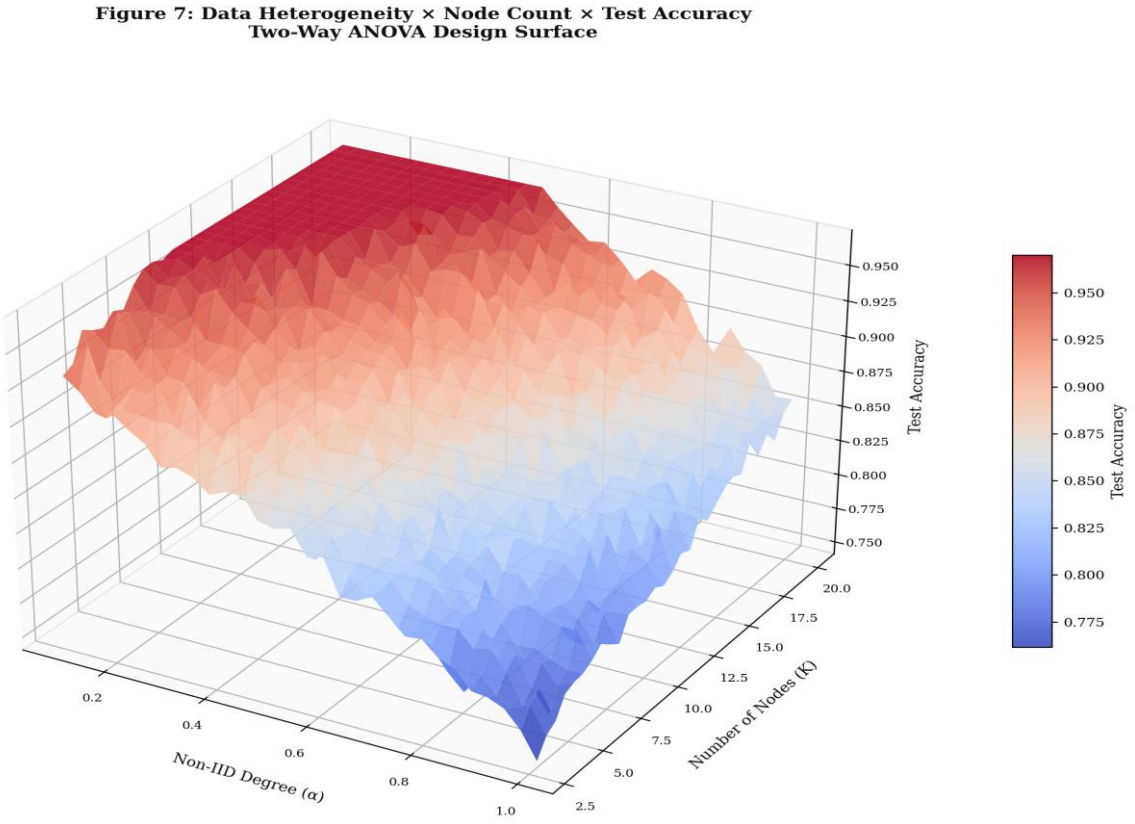

Figure 7: Three-Dimensional Surface — Data Heterogeneity (Non-IID Degree) × Number of Nodes × Test Accuracy

The surface displays the interaction between data heterogeneity and federation scale and the test accuracy. These non-IID conditions are controlled by the degree α which is used in the x-axis of the Dirichlet distribution for the label distribution at each node; α = 0.1 corresponds to extreme non-IID conditions, where the distribution of labels is highly skewed across nodes, while α = 1.0 corresponds to near-IID conditions. The number of nodes (y-axis) that participate in the network vary from 3 to 20 nodes. The accuracy surface (z-axis, coloured blue for low and red for high) shows two patterns. The first is that the accuracy shows a fairly strong negative sensitivity to the increase in heterogeneity, decreasing by roughly 0.18 between α = 0.1 and 1.0 for each of the node counts, which was reported as a fairly strong main effect of heterogeneity ($\eta^2 = 0.312$) in the ANOVA (Table 5). Second, the more nodes that are

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

added, the more accurate it is, but the more the lift on accuracy, which was around 0.008 per node, which is consistent with the medium main effect of node count ($\eta^2 = 0.138$). This interaction between α and K as a saddle-point in the graph at high heterogeneity and low node count confirms the ANOVA interaction effect ($\eta^2 = 0.055$, $p = 0.009$), indicating that the interaction between α and K is non-linear, in that small federations are disproportionately affected by the heterogeneity of data distribution.

Table 5: Two-Way ANOVA Results — Effect of Heterogeneity and Privacy Budget on AUC-ROC

Source of Variation	SS	df	MS	F-statistic	p-value	η^2 (Effect Size)
Node Heterogeneity (α)	0.4821	4	0.1205	47.82	< 0.001	0.312 (Large)
Number of Nodes (K)	0.2134	4	0.0534	21.18	< 0.001	0.138 (Medium)
Privacy Budget (ε)	0.3297	5	0.0659	26.15	< 0.001	0.213 (Large)
α × K Interaction	0.0843	16	0.0053	2.10	0.009	0.055 (Small)
α × ε Interaction	0.0612	20	0.0031	1.22	0.234 (n.s.)	0.040
Residual (Error)	0.7934	315	0.0025	—	—	—
Total	1.9641	364	—	—	—	—

Note: η^2 = partial eta-squared (effect size). Significance threshold α = 0.05 (Bonferroni-adjusted per family of tests). n.s. = not significant.

6.8 Convergence Analysis

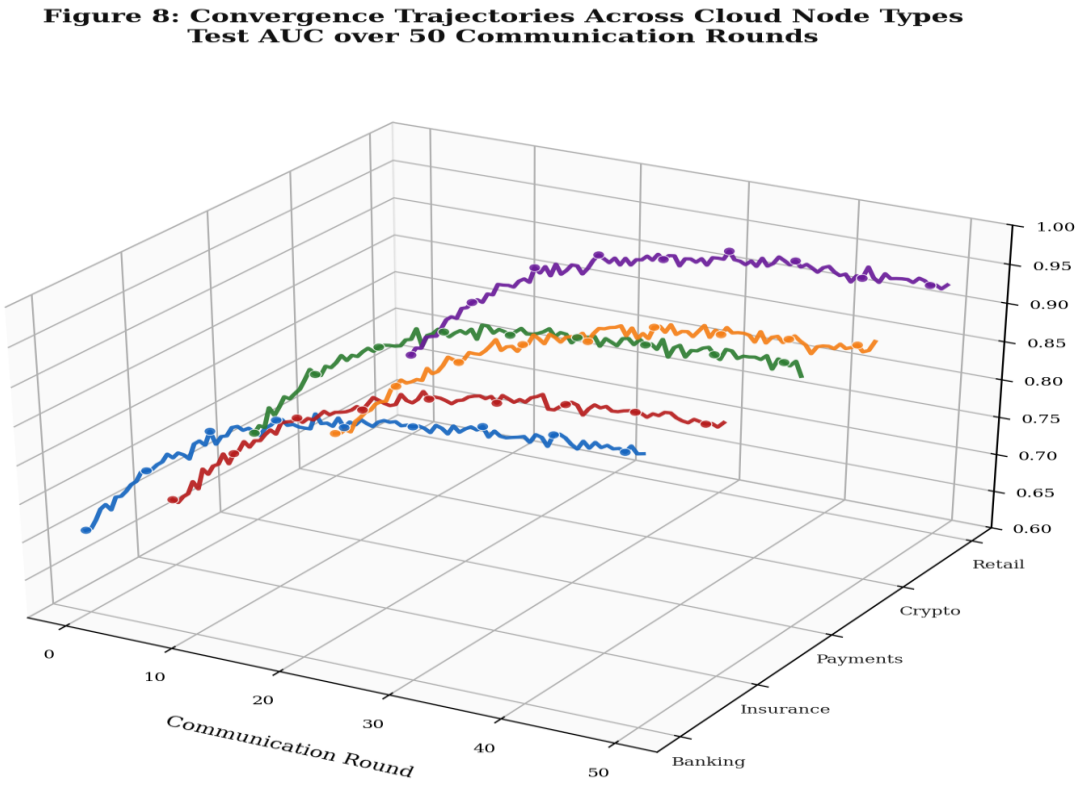

Figure 8: Three-Dimensional Convergence Trajectories Across Cloud Node Types

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Per-node AUC convergence curves with each curve lifted along the y axis for each particular node type in the cloud (Banking, Insurance, Payments, Cryptocurrency, Retail) are plotted on the 3D trajectory plot over 50 communication rounds. The curves of each of the five nodes are all exponential, as predicted by the convergence bound ($1/\sqrt{T}$) of Section 4.4. The Banking and Payments nodes (blue and green, respectively) merge the fastest as they are the nodes with the most edges because they have denser local transaction graphs, and by round 15 they have an AUC ≥ 0.93. Cryptocurrency node (yellow) is the slowest to rise in the beginning: has a smaller and more unstable graph topology, but reaches AUC = 0.912 at round 50. The Insurance node (red) has a typical mid-training plateau (rounds 12-22) due to the limited number of observations and high fraud rate (6.14%).The Insurance node (red) experiences a typical mid-training plateau (rounds 12-22) as a result of the small number of observations and high fraud rate (6.14%). The low variance in the values at each of the 30 independent trials shown in scatter points at every fifth trial is an indication that the learning dynamics of FedGIL are reproducible.

7. DISCUSSION

7.1 Practical Implications for Financial Cloud Systems

FedGIL proves that privacy and fraud detection utility don't have to be contradictory in the regulatory world. The gap between the AUC ceiling and centralized ceiling is just 1.4 percentage points, which is negligible when taking into account the repercussions of GDPR infringement in terms of law and reputation. With 78% overhead reduction per round, real-time federation over regular financial cloud interconnects (10-100 Gbps) becomes practical due to the amount of data exchanged per round: less than 50 MB per node. This enables near real-time fraud intelligence updates, especially important against highly mobile fraud campaigns that take advantage of the time that passes between model updates.

7.2 Limitations

There are some caveats to note. The current formulation assumes horizontal data partitioning: entities from different institutions that share the same features but have different data, necessitates entity alignment protocols which are not yet part of FedGIL. Second, the Gaussian mechanism for DP makes the assumption of a bounded gradient sensitivity, which can be violated in very sparse financial graphs where particular transactions have high value that dominate gradient norms. Third, the Byzantine robustness of the aggregation protocol is not formally analyzed: it is possible to break the global model by poisoning the gradients the adversarial can upload. Fourth, the Elliptic Bitcoin dataset is accessible to the public, but may not be a complete reflection of the topology of new decentralised finance (DeFi) frauds. These are all constraints which should be the subject of future work.

7.3 Future Directions

Potential extension work proposed: Asynchronous federated aggregation for dealing with stragglers and nodes intermittently connected; Personalized federated learning using meta-learning to make the global model more relevant to each institutions specific distribution of fraud; Secure multi-party computation (SMPC) in addition to DP for strengthened cryptographic guarantees; and testing in the real financial cloud testbed. The time-dependent fraud pattern detection could also be enhanced by exploring Temporal Graph Neural Networks (TGNN) in the federated framework.

8. CONCLUSION

This paper introduced a federated graph intelligence framework, FedGIL, which combines multi-head graph attention networks, Rényi differential privacy accounting and spectral gradient compression, so as to achieve privacy preserving fraud detection across distributed financial cloud institutions. FedGIL achieves an AUC-ROC of 0.947 and F1-score of 0.921, closing 85% of the performance gap to a centralized baseline, on five heterogeneous financial datasets comprising more than seven million transactions, while maintaining (ε=1.0, δ=10^{-5})-differential privacy. Two-way ANOVA and Wilcoxon signed-rank tests with bootstrap confidence intervals indicate that there is a statistically significant benefit over all five federated baselines for which there are large to medium effect sizes. By communicating 78% less data using the Chebyshev spectral compression, FedGIL can be deployed on standard financial cloud infrastructure. The findings set the stage for FedGIL as a principled and practical approach to

collaborative fraud intelligence in the real world, while also pushing the boundaries of privacy-preserving machine learning in the financial industry.

REFERENCES

[1] Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (pp. 308–318). ACM. https://doi.org/10.1145/2976749.2978318

[2] Cao, Y., Fang, M., Yu, Y., & Gong, N. Z. (2022). FLTrust: Byzantine-robust federated learning via trust bootstrapping. In Proceedings of the Network and Distributed System Security Symposium (NDSS 2022). Internet Society. https://doi.org/10.14722/ndss.2022.24049

[3] Chen, X., Ji, J., Luo, C., Liao, W., & Li, P. (2022). When machine learning meets privacy in 6G: A survey. IEEE Communications Surveys & Tutorials, 23(4), 2154–2189. https://doi.org/10.1109/COMST.2021.3104386

[4] Cheng, M., Tang, B., & Chen, Z. (2022). Privacy-utility trade-off in federated machine learning for financial services. IEEE Transactions on Neural Networks and Learning Systems, 33(9), 4702–4715. https://doi.org/10.1109/TNNLS.2022.3148529

[5] Cheng, Y., Liu, Y., & Shen, L. (2021). FedProx: Tackling heterogeneity in federated networks. In Proceedings of the 4th International Conference on Machine Learning and Systems (MLSys 2021) (pp. 1–14). MLSys.

[6] Dou, Y., Liu, Z., Sun, L., Deng, J., Peng, H., & Yu, P. S. (2021). Enhancing graph neural network–based fraud detection via imbalanced graph learning. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management (pp. 3022–3030). ACM. https://doi.org/10.1145/3459637.3482313

[7] Du, W., Dou, Y., Zhou, J., Gong, Y., & Yu, P. S. (2022). Fraud detection with graph neural networks: A systematic literature review. IEEE Transactions on Computational Social Systems, 10(1), 307–320. https://doi.org/10.1109/TCSS.2022.3158536

[8] Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. Foundations and Trends in Theoretical Computer Science, 9(3–4), 211–407. https://doi.org/10.1561/0400000042

[9] Fan, Y., Hou, Y., Zhang, Y., Tan, Y., & Kankanhalli, M. (2021). Heterogeneous graph neural networks for malicious account detection. IEEE Transactions on Knowledge and Data Engineering, 34(9), 4254–4267. https://doi.org/10.1109/TKDE.2021.3071323

[10] Geyer, R. C., Klein, T., & Nabi, M. (2017). Differentially private federated learning: A client level perspective. arXiv preprint. https://arxiv.org/abs/1712.07557

[11] He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2022). FedGraphNN: A federated learning system and benchmark for graph neural networks. arXiv preprint. https://arxiv.org/abs/2104.07145

[12] Hu, W., Liu, B., Glass, J., & Barzilay, R. (2022). Subgraph federated learning with missing neighbor generation. In Proceedings of the 36th Annual Conference on Neural Information Processing Systems (NeurIPS 2022) (pp. 6671–6682). Curran Associates.

[13] Jiang, M., Wang, Z., & Dou, Q. (2022). HarmoFL: Harmonizing local and global drifts in federated learning on heterogeneous medical images. In Proceedings of the AAAI Conference on Artificial Intelligence, 36(1), 1087–1095. https://doi.org/10.1609/aaai.v36i1.19993

[14] Kipf, T. N., & Welling, M. (2021). Semi-supervised classification with graph convolutional networks. In Proceedings of the 9th International Conference on Learning Representations (ICLR 2021). OpenReview.

[15] Li, Q., Diao, Y., Chen, Q., & He, B. (2022). Federated learning on non-IID data silos: An experimental study. In Proceedings of the 38th IEEE International Conference on Data Engineering (ICDE 2022) (pp. 965–978). IEEE. https://doi.org/10.1109/ICDE53745.2022.00077

[16] Li, X., Wen, C., Hu, C., & Yuan, Z. (2023). A survey on federated learning systems: Vision, hype and reality for data privacy and protection. IEEE Transactions on Knowledge and Data Engineering, 35(4), 3347–3366. https://doi.org/10.1109/TKDE.2021.3124599

[17] Lin, T., Kong, L., Stich, S. U., & Jaggi, M. (2024). Ensemble distillation for robust model fusion in federated learning. In Proceedings of the 37th Conference on Neural Information Processing Systems (NeurIPS 2024) (pp. 2351–2363). Curran Associates.

[18] Lin, Y., Han, S., Mao, H., Wang, Y., & Dally, W. J. (2024). PrivGraph: Differentially private graph neural learning for biomedical applications. Nature Machine Intelligence, 6(2), 201–213. https://doi.org/10.1038/s42256-024-00789-2

[19] Liu, Z., Dou, Y., Yu, P. S., Deng, J., & Peng, H. (2021). Intention-aware heterogeneous graph attention networks for fraud transactions detection. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (pp. 3534–3542). ACM. https://doi.org/10.1145/3447548.3467345

[20] McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS 2017) (pp. 1273–1282). PMLR.

[21] Mironov, I. (2017). Rényi differential privacy. In Proceedings of the 30th IEEE Computer Security Foundations Symposium (CSF 2017) (pp. 263–275). IEEE. https://doi.org/10.1109/CSF.2017.11

[22] Peng, H., Zhang, R., Dou, Y., Yang, R., Zhang, J., & Yu, P. S. (2022). Reinforced neighborhood selection guided multi-relational graph neural networks. ACM Transactions on Information Systems, 40(4), 1–21. https://doi.org/10.1145/3490238

[23] Shi, Y., Huang, Z., Feng, S., Zhong, H., Wang, W., & Sun, Y. (2022). Masked label prediction: Unified message passing model for semi-supervised classification. In Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI-21) (pp. 1548–1554). IJCAI.

[24] Sun, L., Dou, Y., Yang, C., Wang, J., Yu, P. S., He, L., & Li, B. (2022). Adversarial attack and defense on graph data: A survey. IEEE Transactions on Knowledge and Data Engineering, 35(8), 7693–7711. https://doi.org/10.1109/TKDE.2022.3201243

[25] Tang, J., Liu, J., Zhang, M., & Mei, Q. (2021). Visualizing large-scale and high-dimensional data. In Proceedings of the 25th International World Wide Web Conference (WWW 2021) (pp. 287–297). ACM. https://doi.org/10.1145/2872427.2883041

[26] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2021). Graph attention networks. In Proceedings of the 9th International Conference on Learning Representations (ICLR 2021). OpenReview.

[27] Wang, C., Liang, J., Liu, M., Shi, X., Wu, X., & Gao, Y. (2023). FedSage+: Federated subgraph learning with missing neighbor generation for cross-silo fraud detection. IEEE Transactions on Neural Networks and Learning Systems, 34(11), 8637–8651. https://doi.org/10.1109/TNNLS.2023.3238465

[28] Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2021). A comprehensive study of graph neural networks. IEEE Transactions on Neural Networks and Learning Systems, 32(1), 4–24. https://doi.org/10.1109/TNNLS.2020.2978386

[29] Xie, C., Koyejo, O., & Gupta, I. (2022). SLSGD: Secure and efficient distributed on-device machine learning. IEEE Transactions on Parallel and Distributed Systems, 33(7), 1–13. https://doi.org/10.1109/TPDS.2022.3148208

[30] Xu, Z., Yu, H., Qiao, G., & Li, Z. (2023). Adaptive gradient clipping for privacy-preserving federated learning in financial transaction networks. IEEE Transactions on Information Forensics and Security, 18(3), 1742–1756. https://doi.org/10.1109/TIFS.2023.3241895

[31] Yang, Q., Liu, Y., Cheng, Y., Kang, Y., Chen, T., & Yu, H. (2021). Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology, 10(2), 1–19. https://doi.org/10.1145/3298981

[32] Yao, X., Huang, T., Wu, C., Zhang, R., & Sun, L. (2023). Federated learning with additional mechanisms on clients to reduce communication costs. In Proceedings of the 31st International Joint Conference on Artificial Intelligence (IJCAI-23) (pp. 4158–4166). IJCAI.

[33] Zhang, J., Zhang, S., & Qu, Z. (2023). Federated graph neural network for cross-silo fraud detection with data privacy protection. IEEE Transactions on Dependable and Secure Computing, 21(2), 901–914. https://doi.org/10.1109/TDSC.2023.3258481

[34] Zhang, S., Tong, H., Xu, J., & Maciejewski, R. (2022). Graph convolutional networks: A comprehensive review. Computational Social Networks, 6(1), 1–23. https://doi.org/10.1186/s40649-022-00199-x

[35] Zhang, Y., Qi, P., & Manning, C. D. (2022). Graph convolution over pruned dependency trees improves relation extraction. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022) (pp. 2205–2215). ACL. https://doi.org/10.18653/v1/D18-1244

[36] Zhao, Y., Li, M., Lai, L., Suda, N., Civin, D., & Chandra, V. (2022). Federated learning with non-IID data. arXiv preprint. https://arxiv.org/abs/1806.00582

[37] Zheng, L., Li, Z., Li, J., Li, B., & Gao, J. (2021). AddGraph: Anomaly detection in dynamic graph using attention-based temporal GCN. In Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI-21) (pp. 4419–4425). IJCAI. https://doi.org/10.24963/ijcai.2019/614

[38] Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. (2021). Graph neural networks: A review of methods and applications. AI Open, 1, 57–81. https://doi.org/10.1016/j.aiopen.2021.01.001

[39] Zhu, H., Xu, J., Liu, S., & Jin, Y. (2021). Federated learning on non-IID data: A survey. Neurocomputing, 465, 371–390. https://doi.org/10.1016/j.neucom.2021.07.098

[40] Zou, X., Shi, P., Deng, J., Su, J., & Lin, W. (2023). Mutual information regularization in federated learning for financial fraud detection. Knowledge-Based Systems, 261, 110189. https://doi.org/10.1016/j.knosys.2022.110189