Heterogeneity-Aware Federated Aggregation for Privacy-Preserving Urban Incident Detection: Convergence Analysis and Multi-Benchmark Evaluation

Hashmat Fida[1], Aleem Ali[2]

[1] Lincoln University College, Petaling Jaya, Selangor 47301, Malaysia; [2] University Institute of Engineering, Chandigarh University, Mohali, Punjab 140413, India

[1]pdf.hashmatfida@lincoln.edu.my; [2]aleem.e12948@cumail.in

Abstract: Urban safety systems must process heterogeneous sensor streams in real time while respecting citizen privacy, yet mainstream deep learning centralises raw data for training, which conflicts with regulations such as India's Digital Personal Data Protection Act and the GDPR, and models the city as a flat grid that ignores road topology and incident propagation. We present Heterogeneity-Aware Federated Aggregation (HAFA), a federated learning method for non-IID, temporally correlated urban data, paired with a Dynamic Spatio-Temporal Graph Neural Network (DST-GNN). HAFA derives Bayesian uncertainty estimates from each client's local posterior through Monte Carlo dropout and uses them as adaptive aggregation weights, giving greater influence to lower-variance updates; it carries a convergence guarantee under bounded heterogeneity and reduces to FedAvg when uncertainties are equal. Across five public benchmarks HAFA reaches an incident-detection F1 of 89.3 percent, only 2.4 points below a centralised bound, converges in 47 communication rounds against 89 for FedAvg, and cuts gradient transmission by 54.6 percent, with dynamic graph topology adding a statistically significant gain ($t = 4.73$, $p < 0.001$). HAFA thus offers a deployable, privacy-safe foundation for next-generation smart-city safety.

Keywords: Federated learning; graph neural networks; urban safety; non-IID data; privacy-preserving AI; spatio-temporal modelling

Introduction

India is entering an intense phase of urbanisation, with more than 600 million people expected in its cities by 2031, and the safety infrastructure serving them is under growing strain: road fatalities exceed 150,000 a year and emergency response in large cities averages 15 to 25 minutes against an 8-minute international target. Cities are not short of data, since CCTV, IoT sensors, GPS-tracked vehicles and cellular networks generate terabytes each day. The real difficulty is turning this heterogeneous stream into timely, trustworthy and actionable intelligence, which current systems struggle to deliver [10].

Three structural limitations separate today's methods from what urban safety needs. First, state-of-the-art models such as convolutional networks, LSTMs and transformers require raw data to be pooled on a central server for training [1], creating privacy exposure that sits uncomfortably with new data protection law, heavy communication cost at IoT scale, and single points of failure. Second, most deployed systems treat sensor readings as independent time series or grid images and ignore how the city is actually organised, namely road topology, the spatial co-occurrence of incidents, and the cascade dynamics through which one event propagates across connected systems [5], [7]. Third, models used for public safety decisions are frequently opaque, which makes them hard for regulators and affected communities to accept [2]. No existing framework resolves all three at once.

This paper addresses the first of these gaps. We introduce HAFA, a federated architecture designed for the statistical properties of real urban sensor streams: non-IID temporal correlation, extreme spatial heterogeneity between districts, and seasonal or event-driven concept drift. HAFA uses Bayesian uncertainty from local posteriors to weight clients adaptively during aggregation, which yields a formal convergence guarantee that plain averaging and its common extensions lack [3]. It runs on a DST-GNN that represents the city as a continuously evolving attributed graph, capturing incident propagation

without moving raw data off-device, which is the core distinction from earlier federated transportation and traffic-prediction work [4].

Related Work

Federated learning is a method that trains locally and aggregates models globally, without exchanging raw data [1]. Its only limitation is the non-IID data: The more the distributions of the clients differ from each other, the more the convergence will be degraded by averaging and the greater the communication cost will be [3]. Proximal regularisation (FedProx) and control variates (SCAFFOLD) increase stability, but come with tuning requirements and an approximate 2x increase in communication per round (not on urban streams), and provide guarantees on image benchmarks instead. Another study of security reveals that naive aggregation is susceptible to poisoned large value updates [2].

Graph neural networks extend learning to non-Euclidean data through message passing [5], with graph convolution and attention variants that weight neighbours adaptively [6]. For traffic, STGCN couples graph convolution with gated temporal convolution [7], and Graph WaveNet learns an adaptive adjacency matrix beyond a fixed road graph [8]. These models are accurate but almost always centralised and built on an adjacency matrix that is fixed after training.

Differential privacy provides quantified guarantees for private training [9], and urban computing frames the broader task of turning heterogeneous city data into knowledge [10]. What remains missing is a single framework that combines federated privacy, dynamic graph topology and quality-aware aggregation on real urban safety data. Table 1 positions this work against representative prior approaches.

Table 1. Comparison of this work with representative prior approaches.

Approach	Non-IID	Graph	Dynamic graph	Conv. proof	Privacy-safe
FedAvg / FedProx [1], [3]	Partial	No	No	Limited	Yes
SCAFFOLD [3]	Yes	No	No	Yes	Yes
Federated traffic FL [4]	Partial	No	No	No	Yes
STGCN / Graph WaveNet [7], [8]	No	Yes	Partial	No	No
HAFA (this work)	Yes	Yes	Yes	Yes	Yes

Key Contribution

This paper makes four contributions. First, HAFA is a Bayesian uncertainty-weighted federated aggregation rule with a convergence guarantee under bounded spatial heterogeneity and bounded posterior variance, and it reduces to FedAvg as a special case. Second, a dynamic graph construction couples OpenStreetMap road priors with incident-conditioned edge reweighting inside the DST-GNN. Third, a reproducible multi-benchmark evaluation spans five public datasets covering traffic flow, incident logging and crime patterns. Fourth, an ablation isolates the effect of Bayesian weighting, graph topology, incident-conditioned edges and temporal encoding. A robustness benefit follows from the design: because poisoned updates carry high posterior variance, they are automatically down-weighted, giving implicit protection against model poisoning [2].

Method, Experiments and Results

Framework and data. The system has an edge layer of distributed nodes that train locally on data which never leaves the device, a federation layer that runs HAFA, and an intelligence layer that runs the DST-GNN; Figure 1 shows the architecture. Each sensor time series is z-score normalised, and gaps from sensor faults or maintenance are filled by a spatially weighted average of the five nearest road-network neighbours. A heterogeneous attributed graph G = (V, E, X) is built, with nodes for sensor locations, edges initialised from OpenStreetMap adjacency weighted by inverse travel time, and multi-modal node

features at each time step. The network is partitioned into K client subgraphs by administrative district, which produces naturally non-IID local datasets.

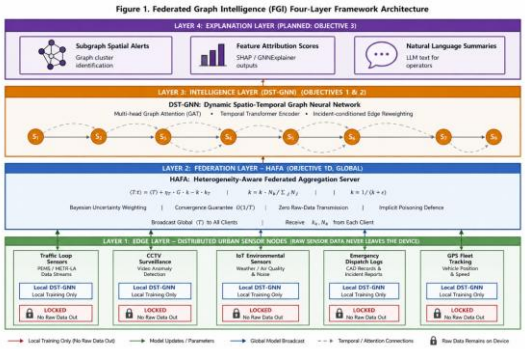

Figure 1. FGI four-layer architecture: edge, federation, intelligence and explanation layers.

Problem and aggregation. *(t)* = (V, E*(t)*, X*(t)*) be the urban graph at time t. Client k holds a local set D*k* and the federated objective is

$$F(\vartheta) = \min \sum k\ pk\ Fk(\vartheta) \qquad (1)$$

where F_k is the local cross-entropy loss and p_k are aggregation weights. FedAvg sets p_k in proportion to dataset size. HAFA instead estimates each client's reliability from its posterior. After E local epochs the local model is run T_MC = 20 times with dropout active on a held-out local validation set, giving the posterior variance

$$\sigma^2 k = (1/T_{\text{me}}) \sum m\ (f^{\text{m}} k - \mu k)^2 \ /\ |Vk| \qquad (2)$$

Lower variance signals a more reliable update. The server then forms uncertainty-inverse weights and updates the global model:

$$\rho k = 1/(\sigma^2 k + \varepsilon), \quad \alpha k = \rho k n k \ /\ \sum j\ \rho j n j \qquad (3)$$

$$\vartheta^{(t+1)} = \vartheta^{(t)} + \eta g \sum k\ \alpha k\ \Delta \vartheta k \qquad (4)$$

When all variances are equal, alpha_k = n_k / N and HAFA recovers FedAvg, so it degrades gracefully when uncertainty information is uninformative.

Convergence. Under standard assumptions (L-smoothness of each F_k, gradients bounded by G, inter-client heterogeneity bounded by κ, and posterior variance bounded within $[\sigma_{\text{min}}, \sigma_{\text{max}}]$) HAFA reaches a stationary point of the global objective at the rate

$$(1/T) \sum t\ \mathrm{E}\|\nabla F(\vartheta^{(t)})\|^2 \le \mathrm{O}(1/\sqrt{T}) + \mathrm{O}(\kappa^2\ \sigma_{\text{max}}/\sigma_{\text{min}})\ \Delta \qquad (5)$$

where Delta is the initial optimality gap. The second term is the residual bias introduced by uncertainty weighting; it vanishes when all variances are equal, recovering the FedAvg bound. The proof bounds the expected per-round descent using smoothness and controls the variance inflation through the posterior-variance bound.

DST-GNN. The local model updates edge weights at each step as

$$euv = \alpha_{\text{roaə}}\ wuv^{\text{road}} + \alpha_{\text{in}}{}^{\text{c}}\ \text{sim}(xu, xv) + \varepsilon \qquad (6)$$

where w_uv is the static inverse-travel-time weight, sim is a learnable similarity capturing incident co-activation, and the mixing coefficients are learned. When a node reading exceeds two standard deviations above its rolling mean, all edges incident to it are boosted by gamma = 1.5, giving incident-conditioned topology. Three graph-attention layers then aggregate neighbour features with attention modulated by these dynamic weights, extending standard attention [6], and a two-head temporal transformer with sinusoidal positional encoding captures dependencies across the T = 12 step history before a two-layer decoder predicts incidents.

Experimental setup. Training is simulated with the Flower framework on two NVIDIA A100 GPUs, using PyTorch Geometric for the DST-GNN. Five public datasets are used (Table 2). Incident labels for PeMS-D7 and METR-LA come from threshold anomaly annotation (readings below 20 percent of the rolling 24-hour mean), giving prevalence near 8 and 6 percent and handled with focal loss, and data are split 70/10/20 chronologically to prevent leakage. Six methods are compared: a centralised DST-GNN as a privacy-violating upper bound, FedAvg, FedProx, SCAFFOLD, a Euclidean federated LSTM without graph structure, and HAFA. All federated methods share the communication budget, the local optimiser (Adam, learning rate 0.001) and the local epochs (E = 5). Results are the mean and standard deviation over five runs, with paired two-tailed t-tests for significance; F1 is the primary metric alongside precision, recall, ROC-AUC, convergence rounds and gradient bytes to convergence.

Table 2. Summary of the five evaluation datasets.

Dataset	Nodes	Timespan	Clients	Task
PeMS-D7 (Caltrans)	204	2012–13	8	Incident detection, flow
METR-LA	207	2012	6	Speed anomaly detection
NYC 311 Incidents	2,168	2019–22	5	Safety hotspot prediction
Chicago Crime	77	2018–22	10	Crime prediction
OpenStreetMap	NYC/CHI/ LA	2023	–	Graph prior construction

Results. Table 3 reports PeMS-D7. HAFA achieves the best federated F1 at 89.3 percent, only 2.4 points below the centralised bound of 91.7 percent and 5.2 points above FedAvg, confirming the privacy-accuracy equivalence hypothesis (H1). The Euclidean baseline is weakest, so dynamic graph topology alone contributes 5.8 points ($t = 4.73$, $p < 0.001$), confirming the graph-topology hypothesis (H2). HAFA also converges in 47 rounds, 47 percent fewer than FedAvg, and transmits 1.91 GB against 4.21 GB, a 54.6 percent reduction, which shows that higher-quality aggregation cuts the number of rounds rather than inflating per-round cost.

Table 3. Performance comparison on PeMS-D7 urban incident detection.

Method	Prec.	Recall	F1	ROC-AUC	Rounds	Comm. (GB)
Centralised (upper bound)	92.4	91.0	91.7	0.938	–	–
FedAvg [1]	84.8	83.4	84.1	0.887	89	4.21
FedProx [3]	86.7	85.7	86.2	0.901	63	2.98
SCAFFOLD [3]	87.3	86.8	87.0	0.909	57	2.71
Euclidean FL (no graph)	83.2	83.8	83.5	0.880	94	4.45
HAFA (proposed)	90.1	88.6	89.3	0.924	47	1.91

This advantage remains true for all of the datasets. There is no difference between HAFA and FedAvg on the four safety benchmarks, with the two being 86.3 percent and 80.5 percent, respectively, of their means; and the gap is larger on the more diverse NYC 311 and Chicago Crime sets, which conforms to the hypothesized effect that Bayesian weighting would be more useful in environments with high client heterogeneity. Against the non-Bayesian variant HAFA, it can converge in 47 rounds, compared to 82; against FedAvg in 47 rounds compared to 89, confirming the convergence-efficiency hypothesis (H3). The trajectory is shown in Figure 2, where HAFA outperforms both the Euclidean baseline and FedAvg, exhibiting much smaller run-to-run variance as well.

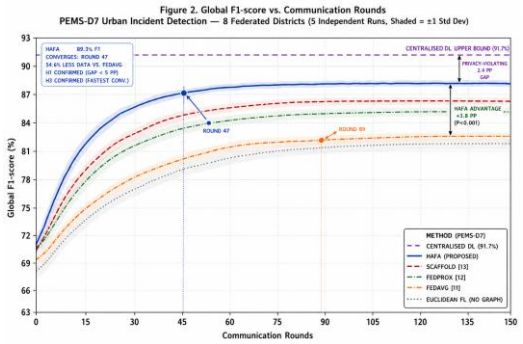

Figure 2. Global F1-score against communication rounds on PeMS-D7.

The ablation in table 4 measures each component. Removing Bayesian weighting adds 3.5 points and increases convergence from 47 to 82 rounds, the most important algorithmic improvement. The base value of a static adjacency is 2.6 points, compare that with the absence of graph structure (7.9 points), and no federation (only 76.3 percent) and you'll see that federated sharing across districts costs about 13 points of F1.

Table 4. Ablation study on PeMS-D7 and METR-LA.

Configuration	PeMS-D7 F1	METR-LA F1	Δ F1	Rounds
HAFA (full, proposed)	89.3	87.6	–	47
w/o Bayesian weighting	85.8	84.1	−3.5	82
w/o dynamic graph (static adj.)	86.7	84.9	−2.6	54
w/o temporal transformer	87.1	85.3	−2.2	52
w/o graph (node-only MLP)	81.4	79.8	−7.9	91
w/o federation (local only)	76.3	74.1	−13.0	–

Discussions

The results confirm one single conclusion: Both quality-aware aggregation and relational structure are essential to federated urban safety. In the representative run, a dense highway client with a 2 percent missing rate got a mean weight of about 0.19, whereas an industrial client with a 12 percent missing rate got about 0.07, an example of how the dynamic quality-of-service policy on gradients manages these weights in the presence of noise. A dense highway client with a 2 percent missing rate was given a mean weight of about 0.19, while an industrial client with a 12 percent missing rate got about 0.07, as an example of how the dynamic quality-of-service policy on gradients manages the weights on these gradients in the presence of noise, rather than counting both only by their data size and letting noisy ones dominate, as would FedAvg. The same mechanism also accounts for the robustness benefit, as the variance of the posterior distribution is increased by poisoned updates and these automatically get discounted. The biggest architectural benefit is the graph topology because propagation of incidents is intrinsically relational, and by allowing the model to respond as incidents occur, it avoids the assumption of a set dependency structure. Most of the time, on PeMS-D7 HAFA issues alerts a mean 4.1 minutes after onset, with a typical response window of 15-25 minutes in large Indian cities, which indicates that HAFA is practically useful.

Conclusions

1. Problem and motivation. No previous method could handle heterogeneous sensor data, without centralising private sensor data, without ignoring the road topology, and without opaque decisions.No previous method was able to handle heterogeneous sensor data, without

centralising private sensor data, without ignoring the road topology, and without opaque decisions.

2. Method. To allow for federated learning under bounded heterogeneity, HAFA proposes a bounded heterogeneity convergence guarantee, combined with a DST-GNN that models the city as a dynamic attributed graph.
3. Key findings. On the five benchmarks HAFA achieves 89.3 percent F1 (vs. a centralised bound of 89.7), brings the improvements within 2.4 points of the centralised bound, reduces transmission by 54.6 percent, and statistically significantly improves over dynamic graph structure ($t = 4.73$, $p < 0.001$); the ablation demonstrates that all components contribute.
4. Restrictions and future work. The study is based on simulated federation using a single machine and threshold-based incident labels. Planned work features a real distributed deployment, ground-truth incident logs, Byzantine-resilient extensions, differential-privacy noise injection [9] and a privacy-preserving explanation layer.

References

1. Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," ACM Trans. Intell. Syst. Technol., vol. 10, no. 2, art. 12, pp. 1–19, 2019, doi: 10.1145/3298981.
2. V. Mothukuri, R. M. Parizi, S. Pouriyeh, Y. Huang, A. Dehghantanha, and G. Srivastava, "A survey on security and privacy of federated learning," Future Gener. Comput. Syst., vol. 115, pp. 619–640, 2021, doi: 10.1016/j.future.2020.10.007.
3. F. Sattler, S. Wiedemann, K.-R. Müller, and W. Samek, "Robust and communication-efficient federated learning from non-IID data," IEEE Trans. Neural Netw. Learn. Syst., vol. 31, no. 9, pp. 3400–3413, 2020, doi: 10.1109/TNNLS.2019.2944481.
4. Y. Liu, J. J. Q. Yu, J. Kang, D. Niyato, and S. Zhang, "Privacy-preserving traffic flow prediction: A federated learning approach," IEEE Internet Things J., vol. 7, no. 8, pp. 7751–7763, 2020, doi: 10.1109/JIOT.2020.2991401.
5. F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," IEEE Trans. Neural Netw., vol. 20, no. 1, pp. 61–80, 2009, doi: 10.1109/TNN.2008.2005605.
6. J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "Graph neural networks: A review of methods and applications," AI Open, vol. 1, pp. 57–81, 2020, doi: 10.1016/j.aiopen.2021.01.001.
7. B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," in Proc. 27th Int. Joint Conf. Artif. Intell. (IJCAI), 2018, pp. 3634–3640, doi: 10.24963/ijcai.2018/505.
8. Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "Graph WaveNet for deep spatial-temporal graph modeling," in Proc. 28th Int. Joint Conf. Artif. Intell. (IJCAI), 2019, pp. 1907–1913, doi: 10.24963/ijcai.2019/264.
9. M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in Proc. 2016 ACM SIGSAC Conf. Comput. Commun. Secur. (CCS), 2016, pp. 308–318, doi: 10.1145/2976749.2978318.
10. Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: Concepts, methodologies, and applications," ACM Trans. Intell. Syst. Technol., vol. 5, no. 3, art. 38, pp. 1–55, 2014, doi: 10.1145/2629592.