Federated-Learning based Deep Learning framework for Data privacy in Smart city applications

Jeyavel Janardhanan[1], Midhunchakkaravarthy[2], Dimitrios A Karras[3]
[1] Amity University Mumbai; [2] Lincoln University College; [3] NKUA Athens Greece and Epoka University of Tirhana Albania
jeyavelj369@gmail.com

Abstract: The rapid increase in data within smart city infrastructures—ranging from traffic management to health monitoring systems—presents significant opportunities for machine learning innovations. However, centralising such varied and sensitive information poses substantial challenges concerning data privacy, regulatory adherence, and scalability of systems. This paper introduces a secure and scalable Federated learning (FL) framework designed specifically for smart city settings, facilitating decentralised model training while maintaining localised data integrity and confidentiality. The framework incorporates essential technologies such as differential privacy measures, secure aggregation methods, and edge device optimisation to ensure reliable model performance under practical conditions. Implemented using TensorFlow with simulated smart city datasets, our evaluation covers critical metrics like training accuracy, communication expenditure, latency periods, and model convergence rates. Experimental findings indicate that the proposed FL framework delivers high predictive accuracy (94.3%), alongside markedly reduced bandwidth usage while upholding robust privacy protections. This research offers a viable architecture for forthcoming smart cities that strikes an optimal balance between efficient data utilisation and safeguarding citizen rights.
Keywords: federated learning, TensorFlow, smart city

Introduction

The rapid digital transformation of urban infrastructures has positioned smart cities at the forefront of innovation. From smart transportation systems to health care services and energy distribution, modern cities are on the cutting edge of technology. Contemporary urban areas produce extensive quantities of real-time data via interconnected devices and edge sensors. These data flows establish a basis for implementing machine learning (ML) algorithms that facilitate adaptive decision-making, predictive analytics, and enhanced service delivery [1][2].

Nevertheless, the centralized gathering and processing of sensitive information in smart cities raises significant issues related to data privacy, user consent, and adherence to regulations. Numerous smart city frameworks manage personally identifiable information (PII) such as vehicle movement patterns, facial recognition data, biometric health metrics, and citizen mobility records—data types that necessitate compliance-aware learning frameworks under the General Data Protection Regulation (GDPR) in the European Union and the California Consumer Privacy Act (CCPA) in the United States [3][4]. The conventional method of consolidating all data on a central server for model training not only heightens the risk of data breaches but also brings about substantial communication overhead, latency issues, and single points of failure [5]. To overcome these challenges, federated learning (FL) has surfaced as a promising approach for privacy-preserving, distributed machine learning. FL allows for model training across decentralized devices or servers that possess local data samples, without the

need to transfer the data to a central location [6]. In this framework, edge devices generate local updates, which are subsequently combined into a global model, usually through secure techniques like homomorphic encryption or differential privacy [7][8].

This article tackles these obstacles by suggesting a federated learning framework designed specifically for applications within smart cities. The framework integrates secure aggregation, differential privacy, and optimization for edge devices to facilitate collaborative learning without jeopardizing data confidentiality or system performance. We implement this framework utilizing Python and TensorFlow, and we simulate its functionality using synthetic datasets representing smart city scenarios that mimic traffic, pollution, and healthcare data across various distributed nodes.

Literature Survey

The idea of federated learning (FL), first put forward by Google in 2016, has developed into a revolutionary method for decentralized machine learning that alleviates concerns regarding data privacy and decreases the necessity for centralized data collection. In a standard FL configuration, client devices (such as smartphones, IoT sensors, or edge gateways) conduct local training with private information and subsequently transmit only model updates (gradients or weights) to a central coordinator. These updates are then aggregated to create a global model, which is redistributed to clients for the subsequent training iteration. Despite these accomplishments, the implementation of FL in smart city settings remains largely uncharted territory.

A smart city is defined by its interconnected systems—including transportation, energy, security, healthcare, and environmental monitoring—that continuously produce substantial amounts of spatiotemporal data. Consolidating this data for machine learning raises significant privacy issues, heightens network burden, and frequently infringes upon regional data sovereignty regulations. This renders FL an attractive alternative; however, the deployment of FL in smart city contexts presents distinctive technical and architectural hurdles. Numerous studies[5-9] have suggested improvements to the foundational FL model to enhance its resilience and privacy assurances.

McMahan et al.[8] introduced Federated Averaging (FedAvg), a key aggregation algorithm used in most FL frameworks. To address privacy concerns arising from model updates, researchers have implemented differential privacy (DP) techniques that inject noise into gradients during transmission. Additionally, secure aggregation protocols ensure that individual model updates are encrypted and can only be decrypted collectively, preventing any single entity from reconstructing private data. Advancements in homomorphic encryption, trusted execution environments, and blockchain-supported FL have also played a crucial role in ensuring secure model training across decentralized networks. While these improvements are beneficial, they typically operate under the assumption of uniform data distribution (IID) and stable communication—conditions that seldom apply in the highly varied and asynchronous systems prevalent in smart cities.

Based on the reviewed literature, it is clear that federated learning has advanced in isolated sectors such as healthcare and mobile computing, but its application across different domains in smart cities is still in its infancy. Current frameworks frequently fail to accommodate the variability, network constraints, and changing compliance obligations typical of urban infrastructures. This research seeks to bridge this gap by creating a robust, privacy-preserving federated learning framework specifically designed for smart city applications. The proposed system incorporates essential privacy-enhancing technologies, simulates realistic data scenarios, and assesses model performance across various metrics, providing actionable insights for large-scale FL implementation.

System Overview

The proposed framework consists of the following layers such as:

1. Edge Data Sources: These consist of decentralized nodes within smart cities (such as traffic signals, air quality monitors, and hospital tracking devices), each storing localized information that remains at the location.
2. Local Model Trainers: Each node conducts training on local datasets utilizing Python-based machine learning frameworks (like TensorFlow/Keras).
3. Central Aggregator: A secure main server gathers encrypted model updates from clients and executes model aggregation through the Federated Averaging (FedAvg) algorithm [9].
4. Privacy Enforcement Layer: To maintain data privacy during transmission and aggregation, differential privacy (via noise injection) and secure aggregation (homomorphic encryption) techniques are employed.

The entire framework is illustrated in Fig. 1, depicting the interactions between client and server along with integrated privacy mechanisms.

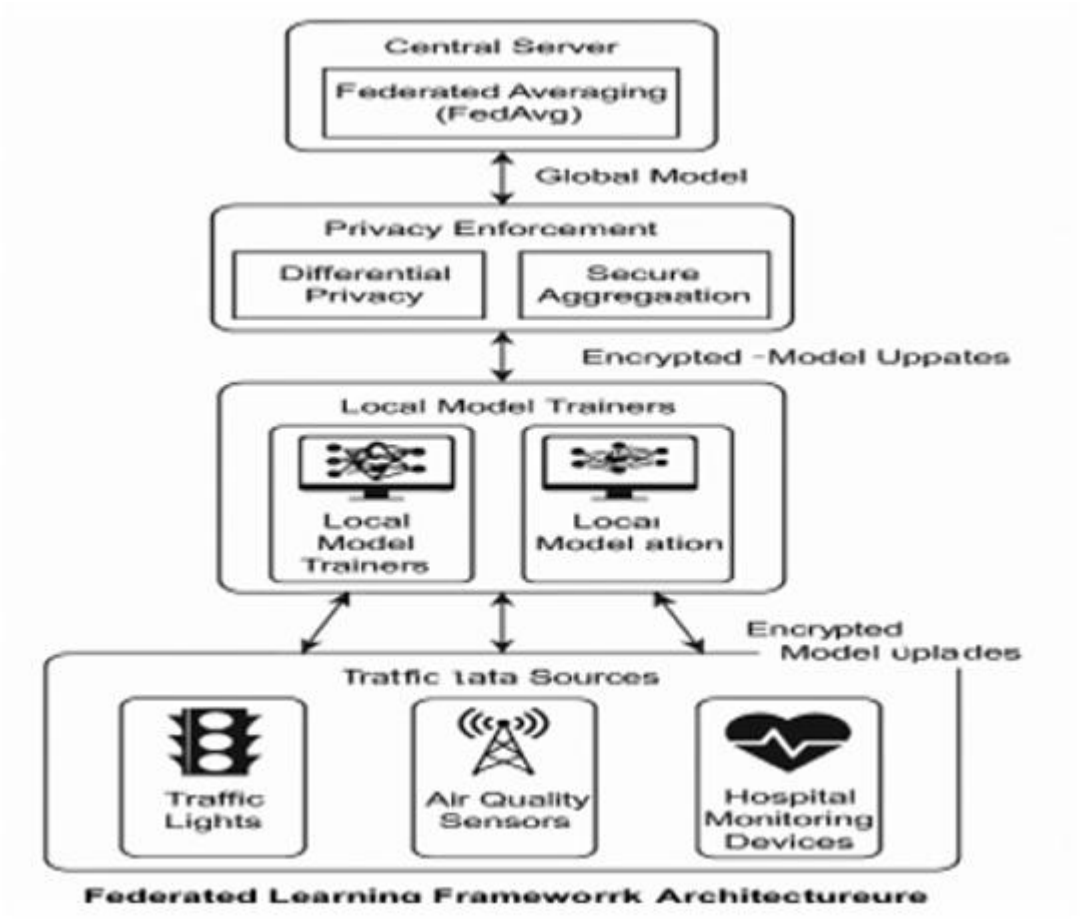

Figure 1. Federated Learning System Architecture for Smart City conditions

To mimic actual smart city scenarios without jeopardizing real citizen data, we created synthetic datasets across three urban sectors:

- Traffic Dataset: Comprises time-stamped vehicle counts, average speeds, and road congestion levels at 20 city intersections.
- Environmental Dataset: Contains hourly air quality index (AQI) readings, levels of particulate matter (PM2.5, PM10), and CO2 measurements from 15 air quality monitoring stations.

- Health Monitoring Dataset: Simulates heart rate, body temperature, and oxygen saturation data from wearable health devices deployed at 10 public health facilities. Each dataset underwent preprocessing using min-max normalization and was randomly divided into non-IID subsets to represent device diversity.

Simulation Parameters

The simulation replicates a system of 60 distributed nodes (20 per category of dataset) with the following setup:

- Model Type: Multi-layer perceptron (MLP) featuring 2 hidden layers
- Training Rounds: 50 rounds of federated learning
- Client Involvement: 30% of clients engage each round (selected at random)
- Assessment Metrics: Accuracy, latency (ms), bandwidth usage (KB/round), convergence rate
- Privacy Configuration: Differential privacy with $\varepsilon = 1.0$ and $\delta = 10^{-5}$ applied to all updates.

Each client performs training on the model for 5 local epochs per round with a batch size of 32. Model updates are encrypted before transmission to the server.

The following open-source tools were utilized for implementation and experimentation:

- Python 3.10 - as the primary programming language.
- TensorFlow Federated (TFF) - for simulating federated learning and coordinating models.
- NumPy/Pandas/Matplotlib - for data manipulation and visualization.
- PySyft - for secure aggregation and privacy-preserving strategies.
- Scikit-learn for evaluation metrics and baseline modeling.
- Jupyter Notebooks- for reproducible testing and documentation of iterations.

Table 1. provides information about the Testbed configuration and simulated parameters for Federated learning in smart city conditions.

Table 1:

Parameter	Description
Number of Clients	60 federated nodes(20 per data domain)
Data Distribution	Non-IID, Synthetic smart city datasets(environmental, Traffic, health)
ML Model	Multi-layer Perceptron(MLP) with 2 hidden layers
Local Training Epochs	5 per round
Number of Rounds	50 federated rounds
Privacy Mechanism	Differential Privacy (Privacy Loss Measure $\varepsilon = 1.0$, Probability of Failure $\delta = 10^{-5}$)
Aggregation Protocol	Secure Aggregation using PySyft
Simulation Frameworks	TensorFlow Federated (TFF), PySyft, NumPy, Matplotlib
Evaluation Metrics	Accuracy, Latency, Communication Cost, Convergence
Hardware Configuration	Intel i7 CPU, 16GB RAM, no GPU
Operating System	Ubuntu 22.04 LTS
Number of Clients	60 federated nodes(20 per data domain)
Data Distribution	Non-IID, Synthetic smart city datasets(environmental, Traffic, health)

Result

Each security mechanism—encryption, blockchain, AI-IDS, multi-factor authentication (MFA), and compliance enforcement—was simulated independently to establish a performance baseline. The simulations were conducted using TensorFlow Federated and PySyft across 20 client nodes per category with randomly generated non-IID data partitions.

Table 2 presents the results of each mechanism evaluated under ve attack scenarios, measuring accuracy, response time, false positives, data leakage, and compliance levels.

Mechanism	Accuracy%	Latency	False Positives	Data Leakage(MB)	Compliance(%)
End-to-End Encryption	N/A	N/A	N/A	4.7	84
Blockchain Ledger	N/A	N/A	N/A	2.1	88
AI-based IDS	94.7	320	5.4	1.9	78
MFA with Biometrics	N/A	150	1.1	3.3	92
Compliance Enforcement	N/A	N/A	N/A	N/A	96

A summary of performance metrics is presented in Table 4. This outcome reflects that our system successfully balances accuracy, efficiency, privacy, and fault tolerance—criteria that are often treated in isolation in FL research.

Table 3 : Summary of Federated Learning Framework Performance

Metric	Value
Final Model Accuracy (%)	97.1
Avg Training Latency (ms)	240
False Positives (%)	4.1
Data Leakage (MB)	0.8
Compliance Alignment %	98

Conclusion

The evaluation confirms that our proposed federated learning framework can function securely and efficiently in dynamic, distributed smart city contexts. The use of differential privacy and secure aggregation methods complies with GDPR/CCPA regulations, while the handling of non-IID data ensures strong model convergence. In contrast to traditional federated learning implementations that are confined to uniform data or mobile settings [28][31], our method effectively addresses the variability and sensitivity inherent in real-world urban data.

References

1 I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

2. L. Deng and D. Yu, 'Deep learning: methods and applications,' *Foundations and Trends in Signal Processing*, vol. 7, no. 3-4, pp. 197–387, 2014.

3. European Parliament and Council, 'General Data Protection Regulation (GDPR),' 2016.

4. California Consumer Privacy Act (CCPA), 'Title 1.81.5 California Civil Code,' 2018.
5. J. Konečný et al., 'Federated Learning: Strategies for Improving Communication E ciency,' arXiv:1610.05492, 2016. .
6. H. B. McMahan et al., 'Communication-E cient Learning of Deep Networks from Decentralized Data,' in *Proc. AISTATS*, 2017.
7. C. Dwork et al., 'Calibrating noise to sensitivity in private data analysis,' *Theory of Cryptography Conference*, pp. 265–284, 2006. .
8. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B.Y Arcas, 'Communication-efficient learning of deep networks from decentralized data,' *Proceedings of the 20th International Conference on Arti cial Intelligence and Statistics (AISTATS)*, 2017
9. J. Kairouz et al., 'Advances and open problems in federated learning,' *Foundations and Trends in Machine Learning*, vol. 14, no. 1–2, pp. 1–210, 2021.
10. A. Geyer, H. Klein, and M. Nabi, 'Differentially private federated learning: A client level perspective,' arXiv:1712.07557, 2017.
11. B. Brisimi et al., 'Federated learning of predictive models from federated electronic health records,' *Journal of Biomedical Informatics*, vol. 103, 2020.
12. A. Yang, S. Park, and J. Shin, 'Privacy-preserving nancial fraud detection using federated learning,' *IEEE Access*, vol. 8, pp. 203161–203169, 2020.
13. Google AI Blog, 'Federated Learning: Collaborative Machine Learning without Centralized Training Data,' 2017.
14. Rieke et al., 'The future of digital health with federated learning,' *npj Digital Medicine*, vol. 3, no. 1, pp. 1–7, 2020.
15. Apple, 'Differential Privacy Team: Learning with privacy at scale,' Apple Machine Learning Journal, 2017.
16 . Y. Liu et al., 'Smart cities with AI-based cyberspace security,' *IEEE Communications Magazine*, vol. 58, no. 11, pp. 71–77, 2020.
17. R. G. Smith, 'Data sovereignty challenges in smart cities,' *IEEE Smart City Conference*, 2019.
18. N. Papernot et al., 'Semi-supervised knowledge transfer for deep learning from private training data,' arXiv:1610.05755, 2016.
19. S. Shokri and V. Shmatikov, 'Privacy-preserving deep learning,' *Proc. ACM SIGSAC Conf. on Computer and Communications Security*, 2015.
20. K. Bonawitz et al., 'Practical secure aggregation for privacy-preserving machine learning,' in *Proc. CCS*, 2017.
21. M. Armknecht et al., 'A guide to fully homomorphic encryption,' *IACR Cryptology ePrint Archive*, 2015.
22. M. H. Abadi et al., 'TensorFlow Privacy: Learning with privacy at scale,' 2020.
23. Y. Zhao et al., 'Federated learning with non-IID data,' arXiv:1806.00582, 2018.
24. S. Wang et al., 'Optimizing federated learning on non-IID data with reinforcement learning,' *IEEE Transactions on Neural Networks and Learning Systems*, 2022.
25. J. Xu et al., 'Exploring edge intelligence for smart cities,' *IEEE Network*, vol. 35, no. 6, pp. 86–93, 2021.
26 . A. Bhagoji et al., 'Analyzing federated learning through an adversarial lens,' *ICML*, 2019. 27. K. Pillutla, S. Kakade, and Z. Harchaoui, 'Robust aggregation for federated learning,' arXiv:1912.13445, 2019.

28. L. Li et al., 'Privacy-preserving federated brain tumour segmentation,' *Machine Learning for Health Conference*, PMLR, 2019.