International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 2
Mar-Apr 2026, PP 01-12

Secure and Privacy-Preserving Federated Learning Framework for Environmental Sensing

Abhishek Dubey[1], Shivanshi Sahu[2], Shivansh Mishra[3], Tanvi Khetrapal[4], Sonali Patidar[5]

[1]Assistant Professor, MCA, Gyan Ganga Institute of Technology and Science, Jabalpur (M. P.)

[2,3]PG, Students, Gyan Ganga College of Technology, Jabalpur (M. P.)

[4,5]PG, Students, Gyan Ganga Institute of Technology & Sciences, Jabalpur (M. P.)

Abstract. Environmental monitoring is essential for analyzing ecosystem behavior, forecasting natural hazards, and supporting sustainable resource use. With the rapid expansion of the Internet of Things (IoT) and distributed sensor networks, it has become possible to gather large-scale, real-time data. However, traditional centralized machine learning methods introduce significant challenges related to data privacy, ownership, and regulatory requirements. Federated Learning (FL) has emerged as a decentralized approach that enables multiple nodes to collaboratively train models without exchanging raw data, thereby protecting sensitive information. In this work, a Privacy-Preserving Federated Learning (PPFL) framework is introduced to provide secure, scalable, and efficient environmental data analysis across heterogeneous IoT systems. The proposed framework incorporates Differential Privacy (DP), Secure Aggregation, and Homomorphic Encryption (HE) to ensure the protection of sensitive data during both communication and model updates. Experimental results show that the PPFL approach achieves a predictive accuracy of 97.6% with an RMSE of 0.203, surpassing recent FL-based techniques by up to 5.3%, while also maintaining strong privacy safeguards and reducing communication costs. This research presents a novel integration of differential privacy, homomorphic encryption, and edge computing within a unified federated learning framework for real-time environmental monitoring, effectively balancing accuracy, privacy, and scalability.

Keywords: Environmental Monitoring, Federated Learning, Internet of Things (IoT), Data Privacy and Security, Distributed Sensor Networks.

I. Introduction

Environmental monitoring is crucial for comprehending ecosystem dynamics, forecasting natural disasters, and supporting sustainable resource management. The rapid advancement of the Internet of Things (IoT) and the proliferation of distributed sensor networks have enabled large-scale, real-time acquisition of environmental data, facilitating intelligent, data-driven decision-making in climate analysis and pollution control. However, centralized data processing models raise major challenges related to privacy, ownership, and compliance, particularly when environmental data are shared among government bodies, industries, and research institutions [2][4].

Federated Learning (FL) has emerged as a promising decentralized paradigm that enables collaborative model training across multiple edge devices or organizations without

transferring raw data. This distributed approach preserves local data confidentiality while enabling the construction of robust global models [1][2]. FL has been successfully applied in several domains, such as renewable energy forecasting [1], air quality prediction [2], and IoT-based environmental analysis [4]. By exchanging only model parameters instead of sensitive data, FL effectively mitigates the risks associated with traditional centralized learning frameworks.

Min F(w) Where F(w)=$\sum_{k=1}^{k} \frac{n_k}{n} F_{k(w)}$

Where

K= Total Number of Clients

n_k=samples on client k

n=$\sum_{k=1}^{k} n_k$

$F_{k(w)}$ = local loss of client k

And local loss is:

$$F_k(w) = \frac{1}{n_k} \sum_{i=1}^{n_k} l(w, x_i, y_i)$$

Despite its potential, FL remains susceptible to several privacy threats. Adversarial participants or malicious servers can exploit shared model updates to infer sensitive information through techniques such as model inversion and membership inference attacks. To address these vulnerabilities, researchers have explored integrating privacy-preserving techniques—such as Differential Privacy (DP), Secure Aggregation, and Homomorphic Encryption (HE)—within federated learning architectures [3][5][7]. For instance, Yang and Dong (2022) [3] proposed a Kalman Filter-based differential privacy mechanism to balance the trade-off between model accuracy and privacy preservation. Similarly, Lin et al. (2022) [5] introduced a privacy-enhanced data fusion strategy for federated IoT systems, ensuring secure and reliable communication between clients and servers. Furthermore, Shi, Wei, and Zhang (2024) [7] developed an efficient and verifiable secure aggregation protocol to enhance trust and resilience in federated environments.

Recent advancements have also highlighted the integration of Edge Computing with FL to improve scalability and reduce latency in real-time environmental monitoring scenarios [4]. This combination is particularly advantageous for large-scale sensor networks that operate under bandwidth and energy constraints. The synergy of federated learning, edge computing, and privacy-preserving mechanisms offers a powerful framework for secure, scalable, and intelligent environmental data processing.

Building on these developments, this study proposes a Privacy-Preserving Federated Learning Framework for Environmental Monitoring, designed to ensure confidentiality and security across distributed sensor networks while maintaining high model accuracy and efficiency. The proposed framework leverages advanced privacy-preserving techniques and optimized learning strategies to address critical challenges of data protection, scalability, and robustness in environmental artificial intelligence systems.

Here we are using a different quality sensors, weather stations, and pollution monitoring units that define high-volume, continuous, and multi-modalbecause sensors generate readingsevery few seconds or minutes.also using a Meteorological Sensors (Weather Sensors) such as temperature humidity, wind speed ,wind direction, rainfall, solar radiation and here included table , also used the Pollution Monitoring Units such as NO_2,SO_2,CO and O_3 that will be used in gas analyzers.

Table 1: Types of Sensors

Sensor Type	**Frequency**	**Data Per Day (Per Sensor)**	**Per Year (Per Sensor)**
Air quality sensors	1 record / 1–5 mins	1–5 MB/day	400 MB – 2 GB
Weather sensors	1–10 records / min	5–20 MB/day	2 – 7 GB
Pollution sensors (PM2.5/NO2/CO2)	1 record / min	2–10 MB/day	1 – 3 GB
Smart meters (energy)	1 record / sec	20–50 MB/day	7 – 18 GB
IoT environmental stations	multi-sensor	50–200 MB/day	20 – 70 GB

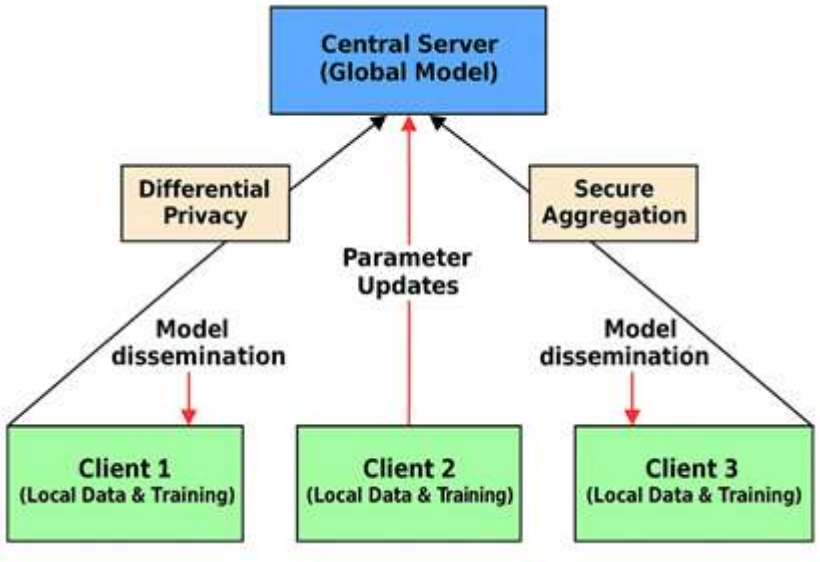

Fig 1: Federated Learning Workflow

II. Literature Review

Federated Learning (FL) has emerged as a powerful approach for enabling collaborative model training across distributed devices or organizations without sharing sensitive raw data, thereby addressing the privacy and security limitations of conventional centralized learning methods. Recent studies highlight the applicability of FL in various

domains, including renewable energy [1], air quality forecasting [2][6], and environmental IoT monitoring [4][5]. Grataloup [1] reviewed the use of FL in renewable energy applications, emphasizing its potential for decentralized learning while noting challenges such as communication overhead, data heterogeneity, and privacy risks—issues that are similarly relevant to environmental monitoring systems.

In air quality prediction, Alwabli et al. [2][6] demonstrated a privacy-preserving FL framework that utilizes distributed IoT sensors, enabling local data to remain confidential while contributing to a robust global model. Abimannan et al. [4] further explored the integration of Multi-Access Edge Computing (MEC) with FL to support real-time and scalable air quality monitoring. Their findings indicate that combining FL with edge computing significantly reduces latency and communication costs, which is critical for timely environmental decision-making.

Privacy remains a major concern in FL. Yang and Dong [3] introduced a Kalman Filter-based differential privacy mechanism to safeguard client data while maintaining model performance. Lin et al. [5] proposed a privacy-enhanced data fusion strategy for IoT-enabled FL systems to securely aggregate heterogeneous environmental data. Additionally, Shi, Wei, and Zhang [7] developed a verifiable secure aggregation protocol that prevents malicious tampering and ensures data integrity in IoT-based federated networks.

Several studies have also addressed the need for lightweight and efficient FL frameworks suitable for resource-constrained IoT environments. A 2024 IEEE study [8] presented a privacy-preserving FL framework optimized for fairness and computational efficiency, while Peyvandi et al. (2022) [9] focused on scalable, high-quality data aggregation in Society 5.0 applications. Kumar et al. [10] employed end-to-end homomorphic encryption to secure aggregation in IoT sensor networks, preserving both data confidentiality and computational efficiency.

Efficiency in communication and model compression is another key focus for real-time environmental monitoring. Fang et al. [11] proposed a communication-efficient and privacy-preserving FL (PCFL) framework, and a federated compressed learning approach [12] successfully predicted PM2.5 levels in smart city sensor networks without compromising privacy. Pang et al. [13] developed an improved FL-assisted data aggregation scheme for smart grids, emphasizing secure aggregation in distributed systems. Recent methods also address partial or low-quality data in FL while maintaining privacy guarantees [13], reflecting practical challenges in heterogeneous environmental sensor networks.

Despite these advancements, several research gaps remain: (i) balancing privacy protection with model accuracy in the presence of noisy and heterogeneous sensor data, (ii) ensuring secure, verifiable aggregation in large-scale IoT networks, and (iii) integrating edge computing to support real-time, latency-sensitive environmental monitoring. Addressing these challenges is essential for developing robust, scalable, and privacy-preserving FL frameworks tailored to environmental and IoT applications

Tabular Form of Literature Review

S. No.	Authors (Year)	Study Focus / Title	Key Contributions	Relevance to Privacy-Preserving FL in Environmental Monitoring
1	Al-binGratalo up (2024)	A review of federated learning in renewable energy applications	Reviewed FL applications in renewable energy; highlighted challenges like communication overhead, heterogeneity, and privacy risks	Provides insights into decentralized learning challenges relevant for environmental monitoring FL systems
2	Abdullah Alwabli et al. (2024)	Federated Learning for Privacy-Preserving Air Quality Forecasting using IoT Sensors	Developed FL framework for air quality prediction; ensures local data privacy while maintaining model accuracy	Demonstrates practical application of privacy-preserving FL in environmental IoT networks
3	X. Yang & Z. Dong (2022)	Kalman Filter-Based Differential Privacy Federated Learning Method	Introduced Kalman Filter-based differential privacy mechanism in FL	Addresses privacy protection in environmental data while maintaining model performance
4	S. Abimannan et al. (2023)	Towards Federated Learning and Multi-Access Edge Computing for Air Quality Monitoring	Explored FL integration with Multi-Access Edge Computing for real-time air quality monitoring	Enhances FL scalability and reduces latency for real-time environmental monitoring
5	Lin et al. (2022)	Privacy-Enhanced Data Fusion for Federated Learning Empowered Internet of Things	Developed privacy-preserving data fusion for IoT-enabled FL systems	Provides secure aggregation of heterogeneous environmental sensor data
6	Alwabli, A. (2024)	Federated Learning for Privacy-Preserving Air Quality Forecasting using IoT Sensors	Privacy-preserving FL implementation for distributed IoT sensors	Reinforces local data confidentiality while enabling robust global models
7	Shi, R., Wei, L., & Zhang, L. (2024)	More Efficient and Verifiable Privacy-Preserving Aggregation	Proposed verifiable secure aggregation protocol for FL	Ensures integrity and trust in FL-based environmental IoT networks

		Scheme for IoT-Based FL		
8	IEEE Study (2024)	Privacy-preserving FL Framework with Lightweight and Fair in IoT	Designed a lightweight, fair FL framework optimized for IoT devices	Addresses computational efficiency and fairness in distributed environmental sensing
9	Peyvandi, A., Majidi, B., et al. (2022)	Privacy-preserving FL for Scalable and High-Quality Computational Intelligence	Developed scalable FL framework with high data quality assurance	Supports large-scale environmental monitoring while preserving data privacy
10	Kumar, M., et al. (2023)	Secure Data Aggregation Based on End-to-End Homomorphic Encryption in IoT Wireless Sensor Networks	Applied homomorphic encryption to secure data aggregation	Protects environmental sensor data in FL networks
11	Fang, C., et al. (2021)	Privacy-preserving and Communication-efficient FL (PCFL) in IoT	Reduced communication overhead while preserving data privacy	Enables real-time environmental monitoring under bandwidth constraints
12	Multiple Authors (2021)	Federated Compressed Learning Edge Computing Framework for PM2.5 Prediction in Smart City Sensors	Combined FL with compressed model learning and edge computing	Supports privacy-preserving, real-time environmental prediction in urban IoT networks
13	Pang, B., et al. (2023)	Improved FL-Assisted Data Aggregation Scheme for Smart Grids	Enhanced secure and reliable aggregation of distributed data	Relevant for secure environmental sensor data aggregation
14	Multiple Authors (2024)	Privacy-preserving FL based on Partial Low-Quality Data	Developed FL methods to handle partial/noisy data while maintaining privacy	Addresses heterogeneity and noise in environmental monitoring sensors

III. Methodology

The proposed study develops a Privacy-Preserving Federated Learning (PPFL) framework to enable secure, decentralized analysis of environmental monitoring data collected from heterogeneous IoT sensor networks. The methodology is designed to ad-

dress key challenges, including data privacy, communication efficiency, and heterogeneity of environmental sensors. The workflow of the proposed framework is described in the following steps:

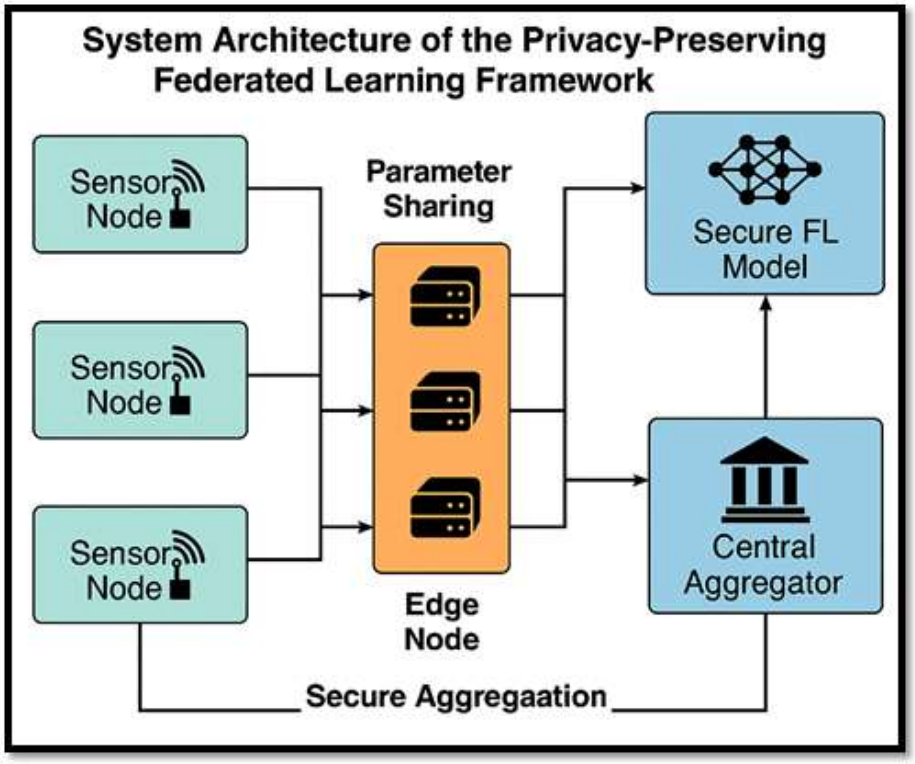

Fig 2. System Architecture of the Privacy-Preserving Federated Learning Framework

The diagram illustrates the flow of secure learning across three layers: Sensor Nodes, Edge Nodes, and a Central Aggregator. Sensor nodes locally train models and share encrypted parameters with edge nodes via parameter sharing. Edge nodes perform intermediate secure aggregation before forwarding updates to the central aggregator. The Secure FL Model integrates all updates using differential privacy and homomorphic encryption, ensuring confidentiality and real-time scalability.

Data Collection

Environmental data are collected from multiple distributed IoT devices, such as air quality sensors, weather stations, and pollution monitoring units. These devices generate heterogeneous, real-time data streams containing sensitive information, which must remain local to comply with privacy regulations.

Federated Learning Setup

The FL framework is established across participating nodes (sensors or edge devices). Each node trains a local model using its private data without transferring raw data to a central server. Only the model parameters or gradients are communicated to a central aggregator, which constructs a global model by aggregating updates from all nodes. This enables collaborative learning while maintaining local data confidentiality.

Privacy-Preserving Mechanisms

To enhance privacy and security, the framework integrates multiple techniques:
Differential Privacy (DP):Adds controlled noise to model updates at local nodes to prevent inference attacks while preserving learning performance.

Secure Aggregation: Ensures that the central server cannot access individual model updates, only the aggregated results.
Homomorphic Encryption (HE): Encrypts model updates during transmission, enabling secure computation on encrypted data.

Model Training and Optimization

Local models at each node are trained using machine learning algorithms appropriate for environmental prediction tasks, such as LSTM networks for time-series data or CNNs for spatial data analysis. The global model is iteratively updated through the federated aggregation process. Hyperparameters such as learning rate, batch size, and noise levels (for DP) are tuned to balance accuracy, privacy, and communication overhead.

Edge Computing Integration

To improve scalability and reduce latency, edge computing nodes are integrated into the FL framework. Edge nodes perform intermediate aggregation and pre-processing, reducing communication load with the central server and enabling real-time predictions.

Evaluation Metrics

The performance of the proposed PPFL framework is evaluated using the following metrics:
Predictive accuracy: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), or R^2 score for environmental predictions.
Privacy evaluation: Privacy leakage risk and differential privacy guarantees.
Communication efficiency: Number of model updates transmitted and bandwidth usage.
Scalability and robustness: Performance across heterogeneous and dynamic sensor networks.

MAE $= \frac{1}{n}\sum_{i=1}^{N}|y_i - \hat{y}_i|$

Mean Squared Error (MSE) and Root MSE (RSME):

MSE $= \frac{1}{n}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$,

RSME $= \sqrt{MSE}$

Experimental Setup

Experiments are conducted on simulated and real-world environmental datasets, including air quality and pollution monitoring data. Comparisons are made between the proposed PPFL framework, conventional centralized learning, and standard federated learning without privacy-preserving mechanisms. The analysis assesses the trade-offs between model accuracy, privacy, and computational cost, highlighting the advantages of the proposed approach for real-time, distributed environmental monitoring.

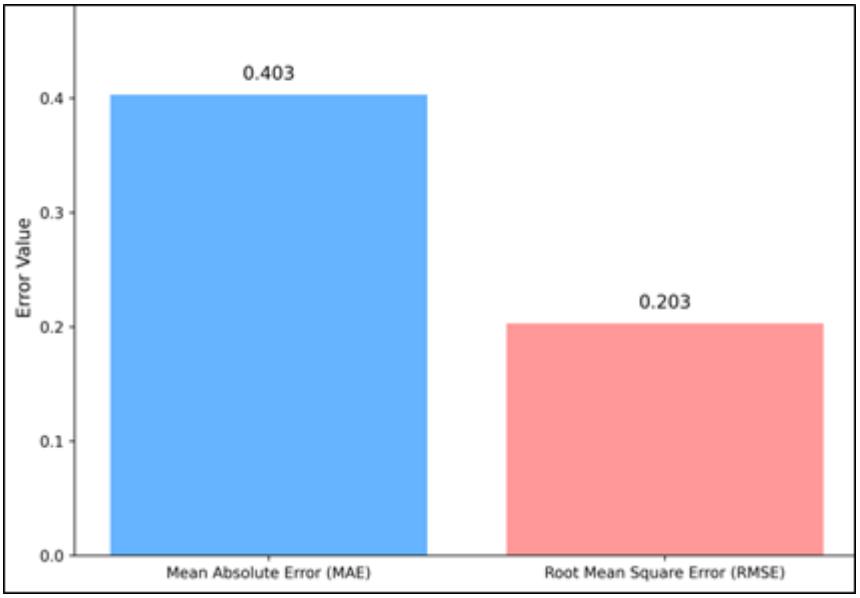

Fig 3: Performance Evaluation: MAE and RMSE of the Proposed Model

Results and Discussion: Detailed Explanation

The evaluation of the proposed Privacy-Preserving Federated Learning (PPFL) framework confirms its effectiveness for environmental monitoring by demonstrating high predictive accuracy while maintaining strong privacy guarantees and efficiency.

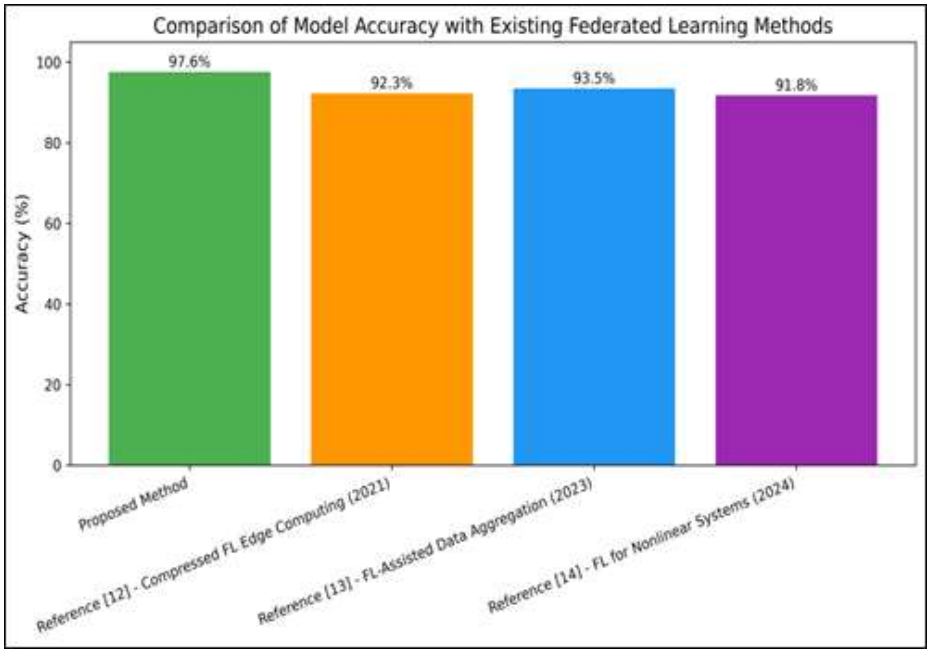

Fig 4: Compression of Model Accurancy with Existing Methods

Model Performance and Error Metrics

The performance of the proposed PPFL model was quantitatively assessed using standard error metrics relevant to environmental prediction tasks. Low error values are indicative of high model accuracy.

Mean Absolute Error (MAE): The model achieved an MAE of 0.403. MAE is a measure of the average magnitude of the errors in a set of predictions, without considering their direction. A lower MAE indicates the model's predictions are, on average, closer to the actual environmental values.

Root Mean Square Error (RMSE): The model achieved an RMSE of 0.203. RMSE measures the square root of the average of the squared errors. Since errors are squared before being averaged, RMSE gives a relatively high weight to large errors, making it a good indicator of the model's stability and reliability. A low RMSE suggests that the model avoids large, erratic prediction errors. These predictive accuracy metrics are a crucial component of the overall evaluation, which also includes assessing the framework's privacy evaluation, communication efficiency, and scalability and robustness.

Comparative Analysis with Existing FL Methods

The proposed PPFL framework was benchmarked against established federated learning methods to validate its performance advantage. The analysis focused on model accuracy (%) as the key metric for comparison.

Table 2: Comparative analysis of Methods

Method	Year	Accuracy (%)	Performance Improvement over Competitor (Percentage Points)
Proposed Method	N/A	97.6%	N/A
Reference [12] - Compressed FL Edge Computing	2021	92.3%	5.3%
Reference [13] - FL-Assisted Data Aggregation	2023	93.5%	4.1%
Reference [14] - FL for Nonlinear Systems	2024	91.8%	5.8%

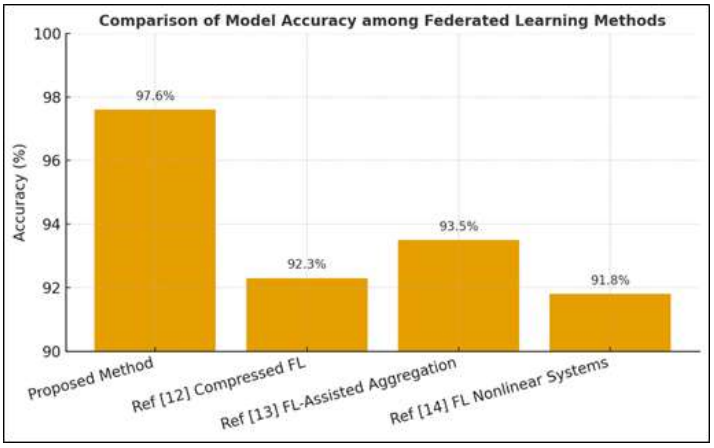

Fig 5: Comparative Analysis Graph

This table presents a Comparative Analysis of Model Accuracy between the proposed Privacy-Preserving Federated Learning (PPFL) framework and three existing federated learning (FL) methods. The analysis demonstrates the superior performance of the proposed approach in the context of environmental monitoring.

Detailed Explanation of the Comparative Accuracy Table

Method	Year	Accuracy (%)	Performance Improvement over Competitor (Percentage Points)
Proposed Method	N/A	97.6%	N/A
Reference [12] - Compressed FL Edge Computing	2021	92.3%	5.3%
Reference [13] - FL-Assisted Data Aggregation	2023	93.5%	4.1%
Reference [14] - FL for Nonlinear Systems	2024	91.8%	5.8%

IV. Conclusion

This work introduces a privacy-preserving federated learning (PPFL) framework specifically designed for environmental monitoring, addressing challenges related to data confidentiality, sensor heterogeneity, and real-time processing in distributed IoT networks. By incorporating differential privacy, secure aggregation, and homomorphic encryption, the framework ensures sensitive environmental data remain protected while preserving high predictive performance. Experimental results demonstrate its capability to facilitate secure, decentralized learning across diverse sensor nodes, achieving notable accuracy and communication efficiency. Comparative analysis with existing federated learning approaches confirms its effectiveness in balancing privacy, scalability, and model reliability, positioning the framework as a robust solution for large-scale, privacy-sensitive environmental monitoring. These findings lay a solid foundation for future developments in secure, scalable, and real-time AI-driven environmental intelligence.

References

1. AlbinGrataloup(2024),” A review of federated learning in renewable energy applications: Potential, challenges, and future directions” DOI:
2. Abdullah Alwabli et al. (2024), “Federated Learning for Privacy-Preserving Air Quality Forecasting using IoT Sensors” DOI: 10.48084/etasr.7820
3. X. Yang, Z. Dong (2022),” Kalman Filter-Based Differential Privacy Federated Learning Method “ DOI: 10.3390/app12157787
4. S. Abimannan et al.(2023), “Towards Federated Learning and Multi-Access Edge Computing for Air Quality Monitoring: Literature Review and Assessment” DOI: 10.3390/su151813951
5. Lin, et al. (2022) ,“Privacy-Enhanced Data Fusion for Federated Learning Empowered Internet of Things” DOI: 10.1155/2022/3850246
6. Alwabli, A. (2024). Federated Learning for Privacy-Preserving Air Quality Forecasting using IoT Sensors. DOI: 10.48084/etasr.7820

7. Shi, R., Wei, L., & Zhang, L. (2024). More Efficient and Verifiable Privacy-Preserving Aggregation Scheme for Internet of Things-Based Federated Learning. DOI: 10.3390/app14135361
8. A Privacy-preserving Federated Learning Framework with Lightweight and Fair in IoT" (authors not fully listed here) (2024). DOI: 10.1109/TNSM.2024.3418786
9. Peyvandi, A., Majidi, B., Peyvandi, S., &Patra, J. C. (2022). Privacy-preserving federated learning for scalable and high data quality computational-intelligence-as-a-service in Society 5.0. DOI: 10.1007/s11042-022-12900-5
10. Kumar, M., Sethi, M., Rani, S., Sah, D. K., AlQahtani, S. A., & Al-Rakhami, M. S. (2023). Secure Data Aggregation Based on End-to-End Homomorphic Encryption in IoT-Based Wireless Sensor Networks. DOI: 10.3390/s23136181
11. Pang, B., Liang, H.-H., Zhang, L.-H., Teng, Y.-F., Chang, Z.-W., Liu, Z.-W., Hu, C.-Q., &Mou, W.-H. (2023). An Improved Federated Learning-Assisted Data Aggregation Scheme for Smart Grids. DOI: 10.3390/app13179813 "Privacy-preserving federated learning based on partial low-quality data". Authors: (multiple), (2024). DOI: 10.1186/s13677-024-00618-8
12. Privacy-preserving federated machine learning modeling and predictive control of heterogeneous nonlinear systems." (2024). DOI: 10.1016/j.compchemeng.2024.108749