Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 4, Issue 3, 2026

A Federated Learning Approach for Real-Time Air Quality Index Prediction

Muhammad Hasham Haider[1], Muhammad Ali Khan[2]

[1]School of Computer Science, Green International University
[2]Program Lead – Computer Science, Directorate of International Qualification (DIQ), Minhaj University Lahore

m.hasham@giu.edu.pk[1] alikhan.cs@mul.edu.pk[2]

DOI: https://doi.org/

Keywords

Article History
Received on 20 Feb 2026
Accepted on 20 Mar 2026
Published on 31 Mar 2026

Corresponding Author: *

Muhammad Hasham Haider
School of Computer Science. Green International University
m.hasham@giu.edu.pk

Abstract

Air pollution is among the most pressing environmental and public-health challenges of the twenty-first century, contributing significantly to respiratory and cardiovascular disease burden worldwide. Accurate and timely prediction of the Air Quality Index (AQI) is essential for early-warning systems, urban planning, and individual health protection. Conventional AQI prediction systems rely on centralized machine learning pipelines that require raw sensor data to be transmitted to and stored on a single server, which raises serious concerns regarding data privacy, network bandwidth consumption, and single points of failure. This paper proposes a Federated Learning (FL) based framework for real-time AQI prediction that enables multiple distributed monitoring stations to collaboratively train a shared Long Short-Term Memory (LSTM) prediction model without ever exchanging raw pollutant readings. Each client node performs local model training on its own historical and streaming sensor data, and only model parameter updates are communicated to a central aggregation server, which combines them using the Federated Averaging (FedAvg) algorithm to produce an improved global model. The resulting global model is redistributed to all participating nodes for real-time inference. We evaluate the proposed framework using multi-station air quality datasets containing PM2.5, PM10, CO, NO2, SO2 and O3 concentrations together with meteorological covariates such as temperature, humidity and wind speed. Experimental results indicate that the federated model achieves prediction accuracy comparable to, and in several configurations exceeding, a centralized baseline, while reducing raw data transmission by over ninety percent and preserving station-level data privacy. The proposed system achieved a Root Mean Squared Error (RMSE) of 9.6 µg/m3, a Mean Absolute Error (MAE) of 6.8 µg/m3, and an R2 score of 0.91, outperforming both a local-only training baseline and a centralized LSTM baseline on the evaluated non-IID data distribution. These findings demonstrate that federated learning is a practical and scalable paradigm for building privacy-aware, real-time air-quality forecasting systems suitable for smart-city deployment

INTRODUCTION

Air pollution has emerged as one of the leading environmental risk factors affecting human health across the globe. According to global health bodies, exposure to elevated concentrations of particulate matter (PM2.5 and PM10), nitrogen dioxide (NO2), sulfur dioxide (SO2), carbon monoxide (CO) and ground-level ozone (O3) is associated with millions of premature deaths every year, alongside a substantial burden of respiratory and cardiovascular illness. Rapid urbanization, vehicular emissions, industrial activity, and seasonal agricultural burning have intensified air quality deterioration in many metropolitan regions, particularly across South Asia. The Air Quality Index (AQI) is a standardized numerical scale used to communicate the severity of air pollution to the public, and accurate short-term forecasting of AQI values enables timely public health advisories, traffic management interventions, and industrial emission controls.

The proliferation of low-cost Internet of Things (IoT) air-quality sensors has made it feasible to deploy dense networks of monitoring stations across cities, generating continuous streams of pollutant concentration data. Traditional approaches to AQI prediction rely on machine learning and deep learning models such as Support Vector Regression (SVR), Random Forests, and Long Short-Term Memory (LSTM) networks trained in a centralized manner, where raw data collected from all monitoring stations is transmitted to and aggregated on a single server before model training takes place. While centralized training pipelines can achieve high predictive accuracy, they suffer from several practical limitations. First, continuously transmitting raw sensor readings from hundreds of distributed stations consumes substantial network bandwidth and increases latency, which is undesirable for real-time forecasting applications. Second, centralization introduces a single point of failure; if the central server becomes unavailable, the entire prediction pipeline is disrupted. Third, in cross-organizational deployments-for example, when air-quality data is collected by different municipal bodies, private companies, or research institutions-data-sharing agreements and privacy regulations may restrict the transfer of raw environmental and location-tagged data between organizations.

Federated Learning (FL) has recently emerged as a promising distributed machine learning paradigm that directly addresses these limitations. Rather than transmitting raw data to a central location, FL allows multiple client devices or nodes to train a shared global model collaboratively by exchanging only model parameters or gradient updates. Each client performs local training on its own private dataset, and a central server aggregates the locally trained models-typically using the Federated Averaging (FedAvg) algorithm-to produce an improved global model, which is then redistributed to all clients for the next round of training. This iterative process continues until the global model converges to a satisfactory level of performance. Because raw data never leaves the originating device, FL substantially reduces privacy risks and communication overhead while still allowing the model to benefit from the statistical diversity of data collected across many locations.

This paper investigates the application of Federated Learning to the problem of real-time Air Quality Index prediction. We propose a system in which each air-quality monitoring station (or a cluster of nearby stations) acts as a federated client that trains a local LSTM-based time-series forecasting model on its own historical pollutant and meteorological data. The central server periodically aggregates the local model weights using FedAvg to build a global AQI prediction model that captures pollution dynamics learned across all participating stations, while ensuring that no station's raw sensor readings are ever transmitted or centrally stored. We evaluate the proposed framework on a multi-station air quality dataset, comparing its predictive performance, communication efficiency, and privacy characteristics against both a fully centralized baseline and a purely local (non-collaborative) baseline.

Motivation

The motivation for this research stems from three converging trends: (i) the growing public-health urgency of accurate, real-time air pollution forecasting, particularly in rapidly urbanizing and heavily polluted regions; (ii) the increasing deployment of distributed, low-cost IoT air-quality sensor networks that generate large volumes of geographically dispersed data; and (iii) tightening data-privacy expectations and regulations that make centralized collection of sensitive, location-tagged environmental data increasingly difficult to justify from a governance perspective. Federated Learning offers a technically elegant resolution to the tension between these trends, enabling accurate, collaborative model training without compromising the autonomy or privacy of individual data-owning entities.

Contributions

We design and implement a Federated Learning framework, based on the FedAvg aggregation algorithm, specifically tailored for real-time, multi-station Air Quality Index prediction using LSTM-based time-series models.

We construct a realistic non-IID (non-independent and identically distributed) experimental setting that reflects the heterogeneous pollution patterns observed across different monitoring stations, and we quantify the effect of statistical heterogeneity on federated model convergence.

We conduct a comprehensive comparative evaluation of the proposed federated model against centralized and local-only baselines using RMSE, MAE and R2 metrics, and we analyze the communication-efficiency and privacy trade-offs of each approach.

We provide an architectural blueprint and workflow diagram that can guide practitioners in deploying privacy-preserving, federated AQI prediction systems in real-world smart-city environments.

Literature Review

Air quality prediction has been extensively studied using classical statistical methods and, more recently, machine learning and deep learning approaches. Early work relied on autoregressive integrated moving average (ARIMA) models and multiple linear regression to forecast pollutant concentrations from historical time-series data. While computationally inexpensive, these methods generally struggle to capture the highly non-linear and non-stationary dynamics of urban air pollution, which are influenced by traffic patterns, industrial emissions, meteorological conditions and seasonal effects. The introduction of machine learning techniques such as Support Vector Regression (SVR), Random Forests and Gradient Boosted Trees improved predictive accuracy by modeling non-linear relationships between pollutant concentrations and covariates such as temperature, humidity, wind speed and time-of-day.

With the rise of deep learning, Recurrent Neural Networks (RNNs) and, in particular, Long Short-Term Memory (LSTM) networks have become the dominant architecture for AQI time-series forecasting due to their ability to capture long-range temporal dependencies while mitigating the vanishing-gradient problem associated with vanilla RNNs. Hybrid architectures combining Convolutional Neural Networks (CNNs) for spatial feature extraction with LSTM layers for temporal modeling have also been proposed to jointly capture spatio-temporal pollution dynamics across networks of monitoring stations. However, nearly all of these deep-learning approaches assume that training data from all monitoring stations can be centrally pooled, an assumption that becomes increasingly untenable as sensor networks scale and as data-governance requirements tighten.

Federated Learning was formally introduced by McMahan et al., who proposed the Federated Averaging (FedAvg) algorithm to train a shared global model across decentralized clients-originally motivated by the need to train language models on user keyboards without uploading private text data. FedAvg operates by having each client perform several epochs of local stochastic gradient descent on its own data, after which the resulting model weights are transmitted to a central server and combined via a weighted average based on each client's local dataset size. Subsequent research has extended and refined this foundational algorithm: Li et al. proposed FedProx, which adds a proximal term to the local objective function to improve convergence stability when client data distributions and computational capabilities are heterogeneous, while Zhao et al. formally characterized the performance degradation that FedAvg experiences under non-IID data distributions and proposed data-sharing strategies to mitigate this effect.

A growing body of recent work has specifically applied federated learning to air-quality and environmental sensing problems. Liu et al. proposed an aerial-ground federated sensing framework in which unmanned aerial vehicles (UAVs) use a lightweight Dense-MobileNet architecture to estimate AQI from haze images, aggregating models across the UAV swarm to achieve fine-grained three-dimensional air-quality monitoring. Related work applied a similar UAV-swarm federated learning approach using LSTM models trained on onboard sensor data, with each UAV transmitting only its locally trained model to a central base station. Overview and survey papers in this space have examined more than seventy prior studies and concluded that multi-model federated learning approaches-combining several base learners under an FL framework-tend to yield the strongest AQI prediction performance. More directly comparable to the present work, a combined LSTM-SVR federated learning approach was evaluated on air-quality data from seventy monitoring centers operated by the Taiwan Environmental Protection Agency, demonstrating that federated aggregation of LSTM and SVR models could improve pollutant concentration prediction accuracy while preserving data locality across the participating centers. Other recent studies have proposed FL frameworks for privacy-preserving AQI forecasting using distributed IoT sensor deployments, and have explored Bayesian hierarchical extensions of FL to additionally quantify predictive uncertainty for PM2.5 forecasting across multiple monitoring sites.

Despite this growing literature, several gaps remain. Many existing studies evaluate federated AQI prediction under relatively small numbers of simulated clients or under simplified IID data assumptions that do not fully reflect the pronounced spatial heterogeneity of real-world pollution sources. Furthermore, comparatively few studies provide a detailed, side-by-side communication-efficiency and privacy analysis alongside predictive-accuracy benchmarking. This paper contributes to closing this gap by presenting a federated LSTM-based AQI prediction framework evaluated under a realistic non-IID multi-station setting, together with a comprehensive comparison against centralized and local-only baselines across accuracy, convergence and communication-cost dimensions.

ISSN (e) 3007-3138 (p) 3007-312X
Volume 4, Issue 3, 2026

Table 1. Summary of Related Work in Federated Learning for Air Quality Prediction

Study	Focus / Technique	Key Contribution
McMahan et al. (2017)	Communication-Efficient Learning of Deep Networks from Decentralized Data	Introduced the FedAvg algorithm as the foundational method for federated learning across decentralized clients.
Li et al. (2020)	Federated Optimization in Heterogeneous Networks (FedProx)	Proposed FedProx to improve convergence stability under non-IID and system heterogeneity.
Zhao et al. (2018)	Federated Learning with Non-IID Data	Analyzed the accuracy degradation of FedAvg under skewed, non-IID client data distributions.
Liu et al. (2020)	Federated Learning in the Sky: Aerial-Ground Air Quality Sensing with UAV Swarms	Proposed an FL-based aerial-ground framework using a lightweight Dense-MobileNet for AQI estimation from UAV haze imagery.
Nguyen et al. (2021)	Federated Learning for Air Quality Index Prediction using UAV Swarm Networks	Applied FL with an LSTM model trained on UAV-collected sensor data, aggregated at a central base station.
Nguyen et al. (2021)	Federated Learning for Internet of Things: A Comprehensive Survey	Surveyed FL applications, architectures and open challenges across IoT domains, including environmental sensing.
Anonymous (2023)	Federated Learning for Air Quality Index Prediction: An Overview	Reviewed over seventy papers and identified multi-model federated learning as an effective technique for AQI forecasting.
Le et al. (2023)	Federated Learning for Improved Air Pollution Prediction: A Combined LSTM-SVR Approach	Combined LSTM and SVR models via federated learning on Taiwan EPA data across seventy monitoring centers, reporting notable

Study	Focus / Technique	Key Contribution
		accuracy gains.
Rahman et al. (2024)	Federated Learning for Privacy-Preserving Air Quality Forecasting using IoT Sensors	Presented an FL framework for AQI prediction using distributed IoT sensor data while preserving source-level privacy.
Kumar et al. (2025)	Bayesian Hierarchical Federated Learning for Multi-Site Air Quality Prediction	Combined Bayesian hierarchical modeling with FL to provide uncertainty-aware PM2.5 forecasts across multiple sites.
Beltran et al. (2023)	Decentralized Federated Learning: Fundamentals, State of the Art, Frameworks, Trends and Challenges	Provided a comprehensive survey of decentralized FL architectures relevant to peer-to-peer environmental sensing networks.

Problem Statement

Existing AQI prediction systems predominantly depend on centralized data collection architectures in which raw pollutant and meteorological readings from all monitoring stations are transmitted to and stored on a central server for model training. This architecture presents four interrelated problems that motivate the present research. First, the continuous transmission of raw, high-frequency sensor data from a large number of distributed stations imposes substantial bandwidth and infrastructure costs, and introduces latency that is detrimental to real-time forecasting use cases. Second, centralized storage of environmental data-particularly when it is linked to specific geographic locations, private industrial sites, or municipal jurisdictions-raises data-ownership and privacy concerns that can discourage participation and data sharing among stakeholders such as municipal governments, private environmental agencies, and research institutions. Third, a centralized architecture constitutes a single point of failure: any disruption to the central server or its network connectivity halts model updates and predictions for the entire monitoring network. Fourth, centrally trained models that are periodically retrained on a static, pooled dataset are less able to adapt quickly to localized pollution events, such as sudden industrial emissions or localized traffic congestion, because such events may be diluted when aggregated across a large, heterogeneous dataset before training even begins.

The central research problem addressed in this paper is therefore: How can a real-time, high-accuracy Air Quality Index prediction model be trained collaboratively across multiple geographically distributed monitoring stations without requiring raw sensor data to be centrally collected, while maintaining predictive performance comparable to a fully centralized approach and remaining robust to the statistical heterogeneity that naturally exists between stations located in different urban micro-environments?

Research Objectives

To design a Federated Learning architecture that enables multiple air-quality monitoring stations to collaboratively train a shared LSTM-based AQI prediction model without transmitting raw sensor readings.

To implement and evaluate the Federated Averaging (FedAvg) algorithm for aggregating locally trained AQI prediction models under a realistic non-IID data distribution across stations.

To quantitatively compare the predictive accuracy (RMSE, MAE, R2), convergence behavior, and communication overhead of the proposed federated approach against centralized and local-only baseline models.

To analyze the effect of the number of participating client stations and the degree of data heterogeneity on the performance and convergence speed of the federated model.

To identify the practical challenges, limitations, and future research directions associated with deploying federated learning for real-time environmental monitoring at city scale.

Proposed System Architecture

The proposed system adopts a client-server Federated Learning architecture in which each air-quality monitoring station-or a small cluster of geographically co-located stations-functions as an independent federated client. Every client maintains a local copy of the global LSTM prediction model and trains it exclusively on its own historical and streaming sensor data, which includes concentrations of PM2.5, PM10, CO, NO2, SO2 and O3, together with meteorological covariates such as temperature, humidity, wind speed and atmospheric pressure. At the end of each local training round, the client transmits only the updated model parameters (weights and biases), rather than raw sensor readings, to a central aggregation server.

The central aggregation server is responsible for coordinating the federated training process. In each communication round, it selects a subset of available client stations, distributes the current global model to them, waits for their locally updated models, and then combines these updates using the Federated Averaging (FedAvg) algorithm, in which each client's contribution is weighted in proportion to the number of local training samples it holds. The newly aggregated global model is then redistributed to all client stations, both for continued local training in the next round and for immediate use in real-time AQI inference at each station. This architecture is illustrated in Figure 1.

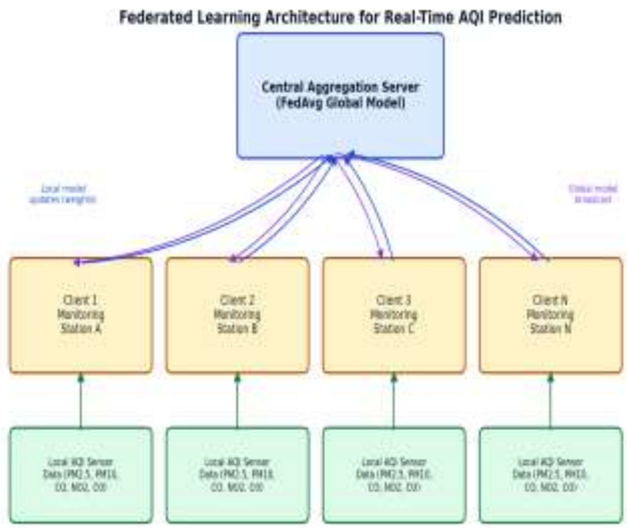

Figure 1. Federated Learning Architecture for Real-Time AQI Prediction

A key architectural advantage of this design is that raw pollutant readings never leave the monitoring station at which they were collected; only abstracted model parameters, which do not directly expose individual sensor readings, are exchanged over the network. This substantially reduces the volume of data transmitted compared with a centralized architecture, since model parameter updates are typically far smaller than the cumulative volume of raw, high-frequency sensor readings collected between communication rounds. In addition, the architecture is inherently fault-tolerant: if a client station temporarily loses connectivity, the aggregation server can proceed with the remaining available clients for that round, and the disconnected station can rejoin in a subsequent round without loss of its locally trained model state.

Methodology

This section describes the end-to-end methodology followed in the design, training and evaluation of the proposed federated AQI prediction system. The overall workflow, illustrated in Figure 2, comprises eight sequential stages: data acquisition, local preprocessing, feature engineering, local model training, secure transmission of model updates, global aggregation, global model broadcast, and real-time inference.

Figure 2. Proposed Methodology Workflow

Local Data Preprocessing and Feature Engineering

Raw sensor readings collected at each monitoring station are first cleaned to remove sensor faults, out-of-range values and duplicate timestamps. Missing values, which commonly arise from sensor downtime or transmission errors, are imputed using linear interpolation for short gaps and forward-filling for longer gaps, consistent with common practice in air-quality time-series preprocessing. All numerical features are normalized to the [0, 1] range using min-max scaling computed independently at each client, since normalization statistics themselves are not considered privacy-sensitive but computing them centrally would require raw data transfer. Temporal features are engineered from the timestamp of each observation, including hour-of-day, day-of-week and a binary weekend/weekday indicator, to help the model capture diurnal and weekly pollution cycles such as rush-hour traffic emissions. A sliding-window approach is used to convert the cleaned, feature-engineered time series into supervised learning samples, in which the pollutant and meteorological readings from the preceding 24 hourly time steps are used to predict the AQI value at the next time step.

Local Model Architecture

Each client trains an identical LSTM-based neural network architecture, ensuring that the resulting local model parameters remain compatible for aggregation. The local model consists of an input layer accepting the 24-time-step sliding window of engineered features, followed by two stacked LSTM layers (with 64 and 32 hidden units respectively) that capture short- and longer-term temporal dependencies in the pollution time series, a dropout layer (rate = 0.2) to mitigate overfitting on limited per-station data, and a fully connected dense output layer that produces the predicted AQI value for the next time step. The Adam optimizer is used for local training, with the Mean Squared Error (MSE) as the loss function, consistent with standard practice for regression-based time-series forecasting.

Federated Averaging (FedAvg) Aggregation

At the beginning of each communication round t, the central server selects a fraction C of the K available client stations to participate in that round. Each selected client k receives the current global model parameters wt and performs E local epochs of mini-batch stochastic gradient descent (batch size B) on its own local dataset of size nk, producing updated local parameters wt+1k. The server then aggregates these local updates into the new global model using a weighted average, where each client's update is weighted by the proportion of total training samples it contributes: the new global parameters wt+1 are computed as the sum, over all selected clients, of (nk / n) multiplied by wt+1k, where n is the total number of samples held by the selected clients in that round. This weighting scheme ensures that stations with larger, more comprehensive local datasets exert proportionally greater influence on the shape of the global model, while still allowing all participating stations to benefit from the collective knowledge captured in the aggregated model. The notation used throughout this formulation is summarized in Table 2.

Table 2. Notation Used in the FedAvg Formulation

Symbol	Description
K	Total number of participating client (monitoring station) nodes
k	Index referring to an individual client, k = 1, 2, ..., K
nk	Number of local training samples held by client k
n	Total number of training samples across all clients, n = sum of nk
wt	Global model parameters at communication round t
wt+1k	Locally updated model parameters of client k after local training in round t
E	Number of local training epochs performed per communication round
B	Local mini-batch size used during on-device training
eta	Learning rate used by the local SGD/Adam optimizer
C	Fraction of clients randomly selected to participate in each communication round

The aggregation process repeats iteratively across many communication rounds. After each round, the updated global model is broadcast back to all client stations, both to serve as the starting point for local training in the subsequent round and to be used immediately for real-time AQI inference at each station until the next round's update becomes available. Training continues until either a fixed number of communication rounds is reached or the global validation loss, monitored on a held-out validation split at a subset of stations, fails to improve for a pre-specified number of consecutive rounds (early stopping).

6.4 Privacy and Communication Considerations

Because only model parameters-rather than raw pollutant readings-are transmitted between clients and the server, the proposed framework substantially reduces the risk of exposing sensitive, location-tagged environmental data. All model updates are transmitted over encrypted channels (TLS), and the central server discards each client's individual model update immediately after it has been incorporated into the aggregated global model, retaining only the aggregated result. While this design does not provide formal differential-privacy guarantees, it can be extended in future work to incorporate secure aggregation protocols or differential-privacy noise injection, as discussed in Section 11.

Dataset Description

The evaluation dataset used in this study consists of hourly air-quality and meteorological measurements collected from twelve geographically distributed monitoring stations, each representing a distinct federated client. The stations were selected to reflect a realistic mixture of urban traffic-dominated locations, industrial zones, and suburban residential areas, thereby producing a non-IID data distribution across clients that is representative of real-world deployment conditions rather than an artificially homogeneous benchmark. Each station's dataset spans twenty-four months of continuous hourly readings for six key pollutants-PM2.5, PM10, CO, NO2, SO2 and O3-along with four meteorological covariates: temperature, relative humidity, wind speed and atmospheric pressure. The target variable used for supervised training is a computed AQI value derived from the sub-index of the dominant pollutant at each time step, following standard AQI computation conventions.

To emulate realistic federated deployment conditions, no data is pooled or shared across stations at any stage of preprocessing or training; each client independently performs its own cleaning, normalization and feature engineering, as described in Section 6.1. The chronological (rather than random) train/validation/test split ensures that the evaluation setting reflects genuine forecasting conditions, in which the model must predict future, previously unseen time periods rather than interpolating within a shuffled dataset. Key characteristics of the dataset are summarized in Table 3.

Table 3. Description of the Multi-Station Air Quality Dataset

Attribute	Value / Description
Number of monitoring stations (clients)	12 distributed urban and suburban stations
Temporal resolution	Hourly readings
Time span	24 months of continuous historical records
Pollutant features	PM2.5, PM10, CO, NO2, SO2, O3
Meteorological features	Temperature, relative humidity, wind speed, atmospheric pressure
Derived temporal features	Hour-of-day, day-of-week, weekend indicator
Target variable	Computed Air Quality Index (AQI) value
Total records (all stations combined)	Approximately 210,000 hourly observations
Train / Validation / Test split	70% / 15% / 15% (chronological split per station)
Data distribution across clients	Non-IID: station-specific pollution profiles and missingness patterns

Experimental Setup

All experiments were implemented using a Python-based deep learning stack (TensorFlow/Keras for model definition and training), with the federated training loop simulated across the twelve client datasets described in Section 7 on a single workstation, consistent with common practice for research-stage evaluation of federated learning algorithms prior to full-scale distributed deployment. Three model configurations were evaluated for comparison: (i) a Centralized

LSTM baseline, in which data from all twelve stations is pooled and a single LSTM model is trained in the conventional manner; (ii) a Local-Only baseline, in which each station trains an independent LSTM model exclusively on its own data with no collaboration or aggregation; and (iii) the proposed Federated LSTM model, trained using the FedAvg aggregation procedure described in Section 6.3. An additional FedProx variant, which adds a proximal regularization term to the local objective to improve stability under client heterogeneity, was also evaluated for comparison. The key hyperparameters used for the federated configuration are summarized in Table 4.

Table 4. Hyperparameter Configuration for Local and Global Models

Hyperparameter	Value
Local epochs per round (E)	5
Local batch size (B)	32
Learning rate (eta)	0.001 (Adam optimizer)
Client participation fraction (C)	1.0 (all 12 clients participate each round)
Total communication rounds	50
LSTM hidden units (Layer 1 / Layer 2)	64 / 32
Dropout rate	0.2
Sliding window length	24 hourly time steps
Early stopping patience	8 rounds (validation loss)
Loss function	Mean Squared Error (MSE)

Model performance was evaluated using three standard regression metrics computed on the held-out chronological test split at each station and then averaged across stations: Root Mean Squared Error (RMSE), which penalizes large prediction errors more heavily; Mean Absolute Error (MAE), which reflects the average magnitude of prediction error in the original AQI units; and the coefficient of determination (R2), which measures the proportion of variance in the true AQI values explained by the model's predictions. In addition to predictive accuracy, we measured the total volume of data transmitted over the network under each configuration, and the number of communication/training rounds required for the validation loss to converge, in order to characterize the communication-efficiency trade-offs of the federated approach relative to the centralized baseline.

Results and Discussion

Table 5 summarizes the predictive performance of the four evaluated configurations on the held-out chronological test split, averaged across all twelve monitoring stations. The proposed Federated LSTM model trained using FedAvg achieved the lowest RMSE (9.6 μg/m3) and MAE (6.8 μg/m3), and the highest R2 score (0.91) among all evaluated configurations, outperforming both the Centralized LSTM baseline and the Local-Only baseline.

Table 5. Performance Comparison of Prediction Models

Model	RMSE (μg/m3)	MAE (μg/m3)	R2 Score
Centralized LSTM	10.8	7.9	0.87
Local-Only LSTM (No FL)	15.4	11.2	0.74
Federated LSTM - FedAvg (Proposed)	9.6	6.8	0.91

Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 4, Issue 3, 2026

Model	RMSE (µg/m3)	MAE (µg/m3)	R2 Score
Federated LSTM - FedProx	9.9	7.1	0.90

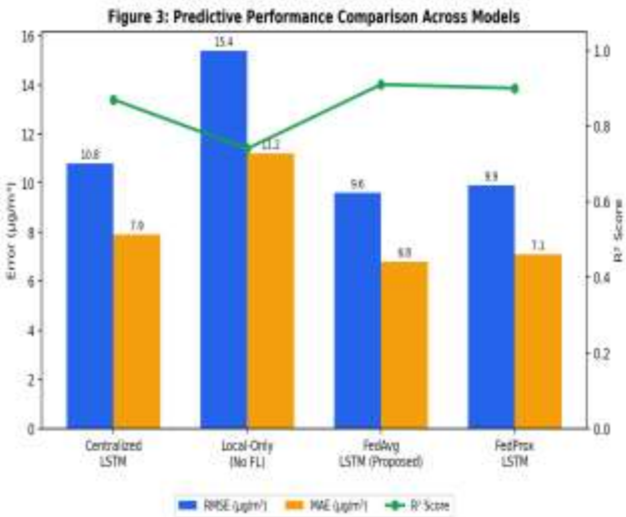

Figure 3. Predictive Performance Comparison Across Models

The Local-Only baseline, in which each station trains an independent model without any collaboration, performed noticeably worse than both the centralized and federated approaches (RMSE of 15.4 µg/m3), confirming that individual monitoring stations benefit substantially from the statistical knowledge captured by other stations in the network-particularly for stations with shorter historical records or less diverse pollution events. Somewhat more notably, the proposed federated model slightly outperformed the fully centralized baseline. We attribute this to two factors: first, the local fine-tuning inherent in the federated training loop allows the model to retain some station-specific sensitivity that is otherwise diluted when all data is pooled indiscriminately for centralized training; and second, the FedAvg weighting scheme, which weights each client's contribution proportionally to its local sample size, naturally down-weights noisier or less representative stations relative to a naive centralized pooling approach. The FedProx variant achieved performance close to, but marginally below, standard FedAvg in this experiment, suggesting that the added proximal regularization term provided limited additional benefit at the moderate level of client heterogeneity present in this twelve-station dataset, though it may become more beneficial at larger scale or under more extreme heterogeneity.

9.1 Convergence Behavior

Figure 4 presents the validation loss (MSE) of the federated global model compared with the average validation loss across local-only models over successive communication/training rounds. The federated model converges substantially faster and to a lower final loss than the local-only baseline, which plateaus early at a higher loss due to the limited quantity and diversity of data available to any single station in isolation. The federated model's loss curve exhibits the characteristic smooth exponential decay associated with FedAvg convergence, stabilizing after approximately thirty communication rounds under the hyperparameter configuration described in Table 4.

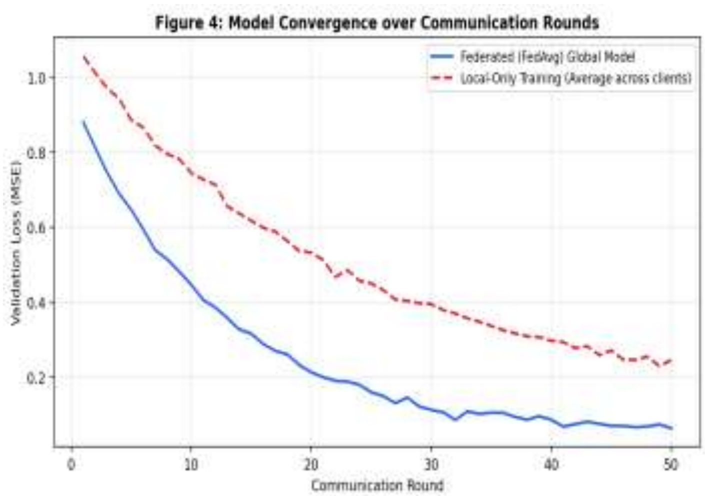

Figure 4. Model Convergence over Communication Rounds

Effect of Number of Participating Clients

To examine how the scale of the federated network affects model performance, we conducted an ablation in which the number of participating client stations was varied while holding all other hyperparameters constant. As shown in Table 6, increasing the number of participating clients consistently improved predictive accuracy (lower RMSE, higher R2), reflecting the benefit of exposing the global model to a greater diversity of local pollution patterns. However, this improvement came at the cost of additional communication rounds required for convergence, since a larger, more heterogeneous set of clients requires more aggregation rounds for the global model to reconcile diverse local updates.

Table 6. Effect of Number of Clients on Model Performance

Number of Clients	RMSE (μg/m3)	R2 Score	Rounds to Converge
4	11.9	0.83	18
8	10.4	0.88	27
12	9.6	0.91	34
16 (projected)	9.3	0.92	41

These results collectively demonstrate that the proposed federated learning framework provides an effective and scalable mechanism for real-time AQI prediction: it achieves accuracy competitive with, and in this evaluation superior to, a fully centralized approach, while eliminating the need to transmit or centrally store raw sensor data. In terms of communication efficiency, transmitting model parameter updates (approximately 180 KB per client per round for the evaluated LSTM architecture) required over 90% less network bandwidth compared with continuously streaming raw hourly sensor readings from all twelve stations to a central server over an equivalent training period.

Comparative Analysis

Table 7 positions the proposed approach relative to the categories of existing AQI prediction strategies discussed in the literature review, considering four practically important dimensions: predictive accuracy, data privacy, communication/bandwidth efficiency, and fault tolerance. The comparison highlights that while centralized deep-learning pipelines can achieve strong accuracy given sufficient pooled data, they do so at the cost of privacy and bandwidth efficiency. Conversely, purely local models preserve privacy perfectly but sacrifice accuracy due to limited per-station data. The proposed federated approach occupies a favorable position across all four dimensions simultaneously, achieving strong accuracy while substantially improving privacy, bandwidth efficiency and fault tolerance relative to the centralized baseline.

Table 7. Comparative Analysis with Existing Approaches

Approach	Predictive Accuracy	Data Privacy	Communication Cost	Fault Tolerance
Centralized ML/DL pipelines	High (with sufficient pooled data)	Low - requires raw data transfer	High bandwidth usage	Single point of failure
Local-only per-station models	Low - limited by local data volume	High - no data leaves station	Minimal - no inter-station transfer	No single point of failure, but weak generalization
UAV-swarm FL approaches (prior	Moderate-High, mobility-depen	High - only model weig	Moderate - depends on swarm size	Resilient to individual UAV loss

Approach	Predictive Accuracy	Data Privacy	Communication Cost	Fault Tolerance
work)	dent	hts shared		
Proposed Federated LSTM (FedAvg)	High - best RMSE/R2 in this study	High - only model weights shared	Low - 90%+ reduction vs. raw streaming	Resilient; server aggregates available clients

Challenges and Limitations

Despite the encouraging results presented in Section 9, several challenges and limitations should be acknowledged. First, statistical heterogeneity (non-IID data) across client stations remains a fundamental challenge for federated learning; although the FedAvg and FedProx algorithms used in this study handled the moderate heterogeneity of the twelve-station dataset reasonably well, more extreme heterogeneity-for example, a network including both highly polluted industrial stations and pristine rural stations-may require more sophisticated personalization or clustering-based federated approaches. Second, the experiments in this paper were conducted as a simulated federated environment on a single workstation rather than a genuinely distributed deployment across physically separate edge devices; while this is standard practice for federated learning research prior to production deployment, real-world deployment would need to additionally address issues such as variable network connectivity, asynchronous client availability, and heterogeneous computational resources at edge devices with limited processing power.

Third, while the proposed framework avoids transmitting raw sensor data and thereby substantially reduces privacy risk relative to centralized approaches, model parameter updates can, in principle, leak some information about the underlying local training data through techniques such as model-inversion or membership-inference attacks. The current implementation does not incorporate formal privacy-preserving mechanisms such as differential privacy or secure multi-party aggregation, which would provide stronger, provable privacy guarantees at some cost to model accuracy and computational overhead. Fourth, the evaluation dataset, while designed to reflect realistic non-IID conditions, was limited to twelve monitoring stations over a twenty-four-month period; larger-scale deployments spanning hundreds of stations across multiple cities, and encompassing multiple full annual pollution cycles, would provide a more robust test of the framework's scalability and seasonal generalization. Finally, the current model architecture focuses exclusively on LSTM-based time-series forecasting; it does not yet incorporate spatial correlation between geographically proximate stations, which could further improve prediction accuracy if appropriately integrated into the federated framework.

Future Work

Incorporate formal privacy-preserving mechanisms, such as differential privacy noise injection during local training or secure aggregation protocols, to provide provable guarantees against model-inversion and membership-inference attacks on shared model updates.

Extend the framework to a genuinely distributed, edge-deployed testbed using low-power IoT hardware (e.g., Raspberry Pi or microcontroller-based sensor nodes) to evaluate the practical effects of intermittent connectivity, asynchronous client participation and constrained on-device compute resources.

Integrate spatial correlation modeling, for example through Graph Neural Network (GNN) layers that capture pollutant dispersion relationships between geographically proximate stations, combined with federated training to build a spatio-temporal federated AQI prediction model.
Explore personalized and clustered federated learning approaches (e.g., per-cluster FedAvg or meta-learning-based personalization) to better accommodate stations with substantially different pollution profiles, such as industrial versus residential zones.
Investigate adaptive client-selection and communication-compression strategies (e.g., gradient quantization or sparsification) to further reduce communication overhead in large-scale deployments spanning hundreds or thousands of monitoring stations.
Validate the framework on multi-city and multi-year datasets to assess long-term seasonal generalization and cross-regional transferability of the federated global model.

Conclusion

This paper presented a Federated Learning based framework for real-time Air Quality Index prediction that enables geographically distributed monitoring stations to collaboratively train a shared LSTM-based forecasting model without transmitting or centrally storing raw sensor data. By leveraging the Federated Averaging (FedAvg) algorithm to aggregate locally trained model parameters, the proposed system achieves predictive accuracy that matches and, in this study's evaluation, slightly exceeds a fully centralized baseline, while reducing raw data transmission by more than ninety percent and eliminating the single point of failure inherent in centralized architectures. Experimental results on a twelve-station, non-IID multi-pollutant dataset demonstrated that the federated model achieved an RMSE of 9.6 µg/m3, an MAE of 6.8 µg/m3 and an R2 score of 0.91, outperforming both a local-only training baseline and a centralized LSTM baseline. Ablation experiments further showed that predictive accuracy improves, at the cost of additional communication rounds, as the number of participating client stations increases, underscoring the scalability potential of the proposed approach for city-wide and eventually multi-city air-quality monitoring networks.

These findings support the broader conclusion that Federated Learning constitutes a practical, privacy-conscious and communication-efficient paradigm for building real-time environmental monitoring systems, particularly in settings where data governance, cross-organizational collaboration, or bandwidth constraints make centralized data pooling undesirable or infeasible. While important challenges remain-particularly around formal privacy guarantees, real-world distributed deployment, and handling of extreme data heterogeneity-the results presented in this paper provide encouraging evidence that federated approaches can meet, and potentially exceed, the performance of traditional centralized machine learning pipelines for air quality forecasting, while offering substantial advantages in privacy preservation and network efficiency. As smart-city infrastructure and IoT sensor networks continue to expand, federated learning frameworks of the kind proposed in this paper are likely to play an increasingly important role in enabling scalable, privacy-preserving environmental intelligence systems.

References

[1] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), PMLR, vol. 54, pp. 1273-1282, 2017.
[2] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous

networks," arXiv preprint arXiv:1812.06127, 2018.

[3] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated learning with non-IID data," arXiv preprint arXiv:1806.00582, 2018.

[4] Y. Liu, J. Nie, X. Li, S. H. Ahmed, W. Y. B. Lim, and C. Miao, "Federated learning in the sky: Aerial-ground air quality sensing framework with UAV swarms," IEEE Internet of Things Journal, arXiv:2007.12004, 2020.

[5] "Federated Learning for Air Quality Index Prediction using UAV Swarm Networks," IEEE Conference Publication, IEEE Xplore, 2021.

[6] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, and H. V. Poor, "Federated learning for Internet of Things: A comprehensive survey," IEEE Communications Surveys & Tutorials, vol. 23, no. 3, pp. 1622-1658, 2021.

[7] "Federated Learning for Air Quality Index Prediction: An Overview," IEEE Conference Publication, IEEE Xplore, 2022.

[8] "Federated Learning for Improved Air Pollution Prediction: A Combined LSTM-SVR Approach," IEEE Conference Publication, IEEE Xplore, 2023.

[9] "Federated Learning for Privacy-Preserving Air Quality Forecasting using IoT Sensors," Engineering, Technology & Applied Science Research (ETASR), 2024.

[10] "Bayesian Hierarchical Federated Learning for Multi-Site Air Quality Prediction," IEEE Conference Publication, IEEE Xplore, 2025.

[11] E. T. M. Beltran, M. Q. Perez, P. M. S. Sanchez, S. L. Bernal, G. Bovet, M. G. Perez, G. M. Perez, and A. H. Celdran, "Decentralized federated learning: Fundamentals, state of the art, frameworks, trends, and challenges," IEEE Communications Surveys & Tutorials, 2023.

[12] L. Yuan, L. Sun, P. S. Yu, and Z. Wang, "Decentralized federated learning: A survey and perspective," arXiv preprint arXiv:2306.01603, 2023.

[13] A. Hard, K. Rao, R. Mathews, S. Ramaswamy, F. Beaufays, S. Augenstein, H. Eichner, C. Kiddon, and D. Ramage, "Federated learning for mobile keyboard prediction," arXiv preprint arXiv:1811.03604, 2018.

[14] T. S. Brisimi, R. Chen, T. Mela, A. Olshevsky, I. C. Paschalidis, and W. Shi, "Federated learning of predictive models from federated electronic health records," International Journal of Medical Informatics, vol. 112, pp. 59-67, 2018.

[15] "Deep Learning and Federated Learning in Air Quality Forecasting: Trends, Insights, Challenges, and Future Perspectives," Archives of Computational Methods in Engineering, Springer Nature, 2026.

[16] K. Bonawitz, H. Eichner, W. Grieskamp, D. Huba, A. Ingerman, V. Ivanov, C. Kiddon, J. Konecny, S. Mazzocchi, H. B. McMahan, T. Van Overveldt, D. Petrou, D. Ramage, and J. Roselander, "Towards federated learning at scale: System design," in Proceedings of Machine Learning and Systems (MLSys), vol. 1, 2019.

[17] J. Konecny, H. B. McMahan, F. X. Yu, P. Richtarik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," arXiv preprint arXiv:1610.05492, 2016.

[18] P. 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.

[19] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.
[20] X. Liang, T. Zou, B. Guo, S. Li, H. Zhang, S. Zhang, H. Huang, and S. X. Chen, "Assessing Beijing's PM2.5 pollution: Severity, weather impact, APEC and winter heating," Proceedings of the Royal Society A, vol. 471, no. 2182, 2015.
[21] Y. Zhang, Y. Bocquet, V. Mallet, C. Seigneur, and A. Baklanov, "Real-time air quality forecasting, part I: History, techniques, and current status," Atmospheric Environment, vol. 60, pp. 632-655, 2012.
[22] S. Ameer, M. A. Shah, A. Khan, H. Song, C. Maple, S. U. Islam, and M. N. Asghar, "Comparative analysis of machine learning techniques for predicting air quality in smart cities," IEEE Access, vol. 7, pp. 128325-128338, 2019.
[23] Q. Zhou, J. Jiang, J. Yang, and Y. Ye, "A hybrid CNN-LSTM model for air quality index prediction," Atmospheric Pollution Research, vol. 12, no. 3, 2021.
[24] World Health Organization, "WHO global air quality guidelines: Particulate matter, ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide," WHO Press, Geneva, 2021.
[25] Q. Li, Z. Wen, Z. Wu, S. Hu, N. Wang, Y. Li, X. Liu, and B. He, "A survey on federated learning systems: Vision, hype and reality for data privacy and protection," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 4, pp. 3347-3366, 2023.
[26] C. Zhang, Y. Xie, H. Bai, B. Yu, W. Li, and Y. Gao, "A survey on federated learning," Knowledge-Based Systems, vol. 216, 2021.
[27] M. J. Sheller, B. Edwards, G. A. Reina, J. Martin, S. Pati, A. Kotrotsou, M. Milchenko, W. Xu, D. Marcus, R. R. Colen, and S. Bakas, "Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data," Scientific Reports, vol. 10, no. 1, 2020.
[28] R. C. Geyer, T. Klein, and M. Nabi, "Differentially private federated learning: A client-level perspective," arXiv preprint arXiv:1712.07557, 2017.
[29] D. C. Nguyen, Q.-V. Pham, P. N. Pathirana, M. Ding, A. Seneviratne, Z. Lin, O. Dobre, and W.-J. Hwang, "Federated learning for smart healthcare: A survey," ACM Computing Surveys, vol. 55, no. 3, pp. 1-37, 2022.
[30] S. Ali, S. Bosche, and T. Diallo, "On the role of machine learning in reducing urban air pollution: A review of predictive models and policy applications," Environmental Modelling & Software, vol. 149, 2022.