

Air Quality Decentralized Forecasting: Integrating IoT and Federated Learning for Enhanced Urban Environmental Monitoring

Vibha Kulkarni

Department of ECE, Vasavi College of Engineering, Hyderabad, Telangana, India
d.vibhakulkarni@staff.vce.ac.in

Adepu Sree Lakshmi

Department of CSE, Geethanjali College of Engineering and Technology, Hyderabad, Telangana, India
adepu.sreelakshmi@gmail.com

Chaganti B. N. Lakshmi

Department of CSE, TKR College of Engineering and Technology, Hyderabad, Telangana, India
ch.nagalakshmi18@tkrcet.com

Sivaraj Panneerselvam

Department of EEE, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology,
Chennai, Tamilnadu, India
psivaraj@veltech.edu.in

Mohammad Kanan

Industrial Engineering Department, University of Business and Technology (UBT), College of
Engineering, Jeddah 21448, Saudi Arabia
m.kanan@ubt.edu.sa,

Aymen Flah

National Engineering School of Gabes, University of Gabes, Tunisia | MEU Research Unit, Middle East
University, Amman, 11831, Jordan | Applied Science Research Center, Applied Science Private
University, Amman, Jordan | The Private Higher School of Applied Sciences and Technologies of Gabes
(ESSAT), University of Gabes, Gabes, Tunisia
flahaymening@yahoo.fr

Mohamed F. Elnaggar

Department of Electrical Engineering, College of Engineering, Prince Sattam Bin Abdulaziz University,
Al-Kharj 11942, Saudi Arabia | Department of Electrical Power and Machines Engineering, Faculty of
Engineering, Helwan University, Helwan 11795, Egypt
mfelnaggar@yahoo.com (corresponding author)

Received: 18 May 2024 | Revised: 30 May 2024, 6 June 2024, and 9 June 2024 | Accepted: 25 June 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7869>

ABSTRACT

Air quality forecasting is a critical environmental challenge with significant implications for public health and urban planning. Conventional machine learning models, although quite effective, require data collection, which can be hampered by issues relating to privacy and data security. Federated Learning (FL) overcomes these limitations by enabling model training across decentralized data sources without compromising data privacy. This study describes a federated learning approach to predict the Air Quality

Index (AQI) based on data from several Internet of Things (IoT) sensors located in different urban locations. The proposed approach trains a model using data from different sensors while preserving the privacy of each data source. The model uses local computational resources at the sensor level during the initial data processing and training, sharing only the model updates to the central location. The results show that the performance of the proposed FL model is comparable to a centralized model and ensures better data privacy with reduced data transmission requirements. This study opens new doors to real-time, scalable, and efficient air quality monitoring systems. The proposed method is quite significant for smart city initiatives and environmental monitoring, as it provides a solid framework for using IoT technology while preserving privacy.

Keywords-IoT; air quality index; federated learning; decentralization; smart city

I. INTRODUCTION

Air Quality Index (AQI) forecasting is an indispensable tool in urban environmental management, public health, and policy-making [1]. In the traditional sense, AQI prediction is performed through centralized Machine Learning (ML) models that collect data from all kinds of sensors and sources to predict pollutant levels [2]. On the other hand, although useful for specific cases, these models operate based on centralized data collection and processing, and therefore all relevant environmental data must be collected in a single place for analysis [3]. Centralized models have some advantages, such as the ability to implement complicated algorithms that require intensive computational resources or directly control data quality and processing [4]. However, this approach suffers from significant drawbacks. Centralized data aggregation is of great concern, raising questions about privacy and security, as sensitive information must flow and be stored in a single repository location [5]. This approach also has greater vulnerabilities to data breaches and cyber-attacks. Furthermore, centralized systems often do not scale with the increasing volume and velocity of data generated by an expanding network of sensors across cities [6].

Federated Learning (FL) allows decentralized data processing with a paradigm shift. FL is an ML technique in which a model is trained across several decentralized devices or servers that hold local data samples, without the requirement for data exchange. This not only helps alleviate the fear of privacy issues but also reduces bandwidth consumption for the transfer of huge datasets. In the AQI prediction context, FL allows for a multi-model approach, where localized models can be developed based on data from specific sensors and then contribute to aggregation to enhance overall predictive accuracy without compromising the data at the source. Recent studies have emphasized the efficacy of FL in AQI forecasting. For example, comparative analysis of FL models could improve up to 12% in prediction accuracy against centralized models, especially in urban settings, while the data environment can be heterogeneous [7]. This study underscores the power of FL to tailor models and benefit from shared learning across the network. In addition, FL supports the use of dynamic models to adapt more rapidly to changing environmental conditions without the need for central re-training every time a change in conditions is required. Smart cities benefit from FL, allowing the integration of IoT devices and sensors without the need for constant monitoring from the center, promoting a scalable and very efficient network architecture.

II. LITERATURE SURVEY

Table I highlights the observations and the key contributions of previous studies.

TABLE I. PREVIOUS STUDIES ON FL IN AQI FORECASTING

Ref.	Observations	Key Contributions
[8]	Multi-model FL approach for AQI prediction.	Improved privacy and computational efficiency.
[9]	Nested LSTM network combined with FL for AQI forecasting.	Improved forecasting performance over traditional methods.
[10]	UAV-based data collection and CNN-LSTM model for AQI prediction.	Effective in real-time high-resolution data environments.
[11]	Detailed examination of ML algorithms for AQI prediction.	Highlights the advantage of decentralized model training.
[12]	Geographically weighted predictors for AQI using the Bangkok dataset.	Improved performance over conventional ML models.
[13]	Use of SSGPR for AQI forecasting related to industrial impacts.	Demonstrates effectiveness in handling uncertainty.
[14]	Use of decision trees, random forest, SVM, and ANN for prediction.	Effective prediction using major pollutants as predictors.
[15]	SVR model and LSTM for AQI forecasting in Chennai.	Specific interventions suggested for urban planning.
[16]	Ensemble model with wavelet packet decomposition and outlier robust ELM.	Major improvements over conventional forecasting methods.
[17]	Review of AI techniques on AQI prediction.	Improved classification accuracy in environmental monitoring.
[18]	Time-series-based modeling with auto-regressive models for AQI.	Applicable in air quality management to reduce health impacts.
[19]	Gaussian process regression and SARIMA models for AQI forecasting.	Noticeable improvement in AQI prediction.
[20]	Deep learning model in a multi-scale combination for AQI prediction.	Handles big data complexities, improves forecasting accuracy.
[21]	Machine learning method combining weather modeling and GNSS radio occultation for regional AQI forecasts.	Quickens prediction process without compromising accuracy.

III. METHOD

A. Dataset Description and Preprocessing

A dataset was obtained from [22], including particulate matter (PM_{2.5} and PM₁₀), Nitrogen Dioxide (NO₂), Sulfur Dioxide (SO₂), Carbon Monoxide (CO), and Ozone (O₃). Each entry in the dataset also features timestamp information, facilitating the temporal analysis of pollution trends and the development of time-series forecasting models. The data was subjected to several essential preprocessing steps to ensure suitability for the FL framework. Initially, data cleaning was performed to address missing values and outliers. Missing

values were imputed using forward-filling techniques to maintain data continuity, thus preventing the introduction of biases due to incomplete data points. Outliers were identified and removed based on statistical thresholds, ensuring that the dataset accurately reflected typical pollution levels without being skewed by extreme values. Subsequently, normalization was applied to all pollutant concentration values using min-max scaling. This step was crucial to standardize the data, bringing all feature values into the same range and enhancing the performance and convergence of the ML models. Feature engineering was performed to enrich the dataset with additional informative features, including rolling averages of pollutant levels to capture temporal trends and time-of-day and day-of-week indicators to account for periodic variations in pollution patterns. For the FL setup, the data was distributed among multiple clients, each representing a cluster of IoT sensors. Data were split in an 80-20 ratio for training and testing purposes at each client level. This split was designed to ensure that each client had sufficient data to train local models while enabling the evaluation of model performance on unseen data. The data were deliberately split in a non-Independent and Identically Distributed (non-IID) manner to reflect real-world scenarios where different regions exhibit varying pollution levels and trends. This non-IID distribution is typical in practical deployments, where environmental conditions differ significantly across geographic locations.

B. Federated Learning Setup

TensorFlow Federated (TFF), an open-source framework developed by Google, was used in this study. TFF allows collaborative training of a global model without necessitating the exchange of raw data between clients, thereby preserving data privacy. In the federated setup, each client was assigned a subset of the AQI data, ensuring a non-IID distribution. The core model was a neural network with two hidden layers, utilizing Rectified Linear Unit (ReLU) activation functions. ReLU functions were chosen for their efficacy in training deep neural networks to mitigate issues such as vanishing gradients. Model training was carried out locally on each client using historical data windows, and different configurations were tested to identify the optimal setup. These configurations included varying the historical data window (7, 30, and 180 days), the number of clients (2 and 10), and the batch sizes (8 and 32). Each of these configurations aimed to capture different temporal trends and patterns in the pollution data while balancing computational load and learning stability.

After local training, model updates were aggregated using a weighted averaging method to form a new global model. This aggregation step was designed to harmonize the contributions of each local model, irrespective of the variations in their local data distributions. The use of weighted averaging ensured that the global model was representative of the broader decentralized data landscape, improving its generalizability and robustness. This approach not only underscores the feasibility and effectiveness of FL for AQI forecasting but also highlights its potential for scalable and privacy-preserving environmental monitoring in smart city initiatives.

C. Hyperparameter Variations and Model Evaluation

Various configurations were tested to explore the optimal settings for the FL model.

- Number of days for past data reference: The models were trained with different historical windows: 7, 30, and 180 days.
- Number of clients: Experiments were carried out with different numbers of clients (2 and 10) to understand how the distribution of data among more clients affects the learning process and the final model performance.
- Batch size: The models were trained using different batch sizes (8 and 32). Smaller batch sizes can provide a more frequent update but might result in a less stable learning process, while larger batch sizes offer more stable updates but at the cost of slower convergence.

This is typically performed using gradient descent. The update rule can be expressed as:

$$w_i^{(t+1)} = w_i^t - \eta \nabla L_i(w_i^{(t)}) \quad (1)$$

where w_i^t is the local model parameters at client i at iteration t , η is the learning rate, and $\nabla L_i(w_i^{(t)})$ is the gradient of the loss function with respect to the model parameters. After each round of local updates, the central server aggregates these updates to form a new global model. A common method for aggregation is weighted averaging:

$$w^{(t+1)} = \sum_{i=1}^N \frac{n_i}{n} w_i^{(t+1)} \quad (2)$$

where N is the number of clients, n_i is the number of data points at client i , n is the total number of data points across all clients, and $w_i^{(t+1)}$ are the updated model parameters from each client. The loss function L used at each client to evaluate the model's performance is typically a function of the predictions and the true values. For regression tasks, such as AQI prediction, a common choice is the Mean Squared Error (MSE).

$$L_i(w) = \frac{1}{n_i} \sum_{j=1}^{n_i} (y_j - f(x_j; w))^2 \quad (3)$$

where y_j are the true values, x_j are the input features, and $f(x_j; w)$ is the model's prediction. The convergence of the FL process can be evaluated by monitoring the change in global loss or the change in model parameters over iterations:

$$\text{Converge if } \|w^{(t+1)} - w^t\| < \varepsilon \quad (4)$$

where ε is the threshold. By this approach, FL enables multiple clients to collaboratively train a model while keeping their data localized.

IV. RESULTS AND DISCUSSION

Figures 1 and 2 show the distribution of the weights of the linear layers of two FL models. The range of the weight values extends from approximately -0.2 to 0.6. Figure 2 shows the second model in a series, hinting at an FL process involving multiple models, possibly representing different clients or data segments in the federated network. The use of linear weights

specifies that they belong to a linear component of the model, which is typically associated with the last layer in regression tasks, where the final output is a weighted sum of the input features. The method obtains the results of training individual models in a federated setting, where each model was exposed to a distinct subset of the data.

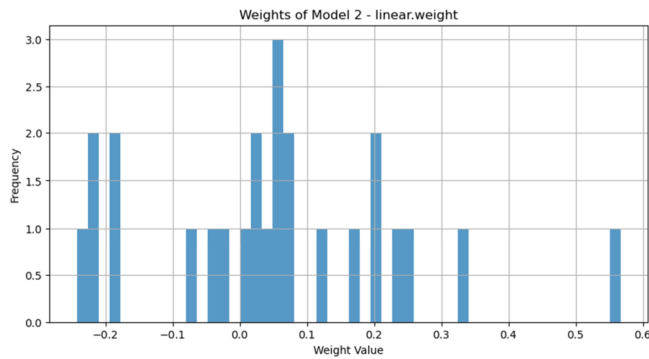


Fig. 1. Weights observed in client 1 for AQI forecasting in the FL training model.

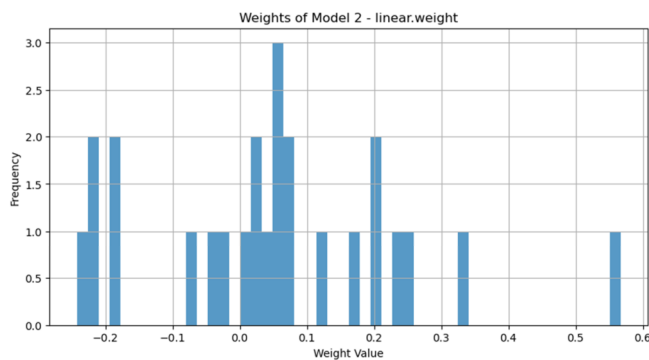


Fig. 2. Weights observed in client 2 for the AQI forecasting in the FL training model.

The federated averaging process shows that the weights of the models are averaged to form the global model's parameters. The averaging process is designed to balance the contributions of each local model, irrespective of the variations in their local data distributions. This step is crucial in FL as it harmonizes local updates, allowing for a model that is representative of the broader, decentralized data landscape. The training procedure adopted a reduced number of epochs and clients for simplicity. This approach was taken to demonstrate the feasibility and effectiveness of FL in a controlled environment before scaling to more complex and realistic scenarios. It should be noted that even with these simplifications, the resultant model exhibits a sophisticated understanding of the input features, as evidenced by the weight distributions.

By carefully scrutinizing the weight distribution graph, valuable insights into the model's interpretability and the significance of different input features in AQI prediction are obtained. The results of this model not only advance the field of FL in environmental science but also underscore the importance of feature weighting in understanding complex,

real-world phenomena such as air quality. The implications of these results are profound, as they provide a window into how models can be collectively trained across diverse data sources in a privacy-preserving manner, while still yielding a rich and nuanced understanding of the underlying predictive factors for crucial environmental metrics such as AQI. This sets the stage for deploying FL models in live and heterogeneous environments, harnessing the power of IoT networks while upholding the stringent privacy requirements demanded in today's data-sensitive world. The performance metrics of the model are critical to validate the efficacy of the different configurations tested.

TABLE II. RMSE COMPARISON

Days of data taken for input processing	Number of clients used in FL	Batches of data taken for processing	RMSE
7	10	8	32.265
180	10	32	29.27
30	2	8	15.93

TABLE III. MSE COMPARISON

Days of data taken for input processing	Number of clients used in FL	Batches of data taken for processing	MSE
7	10	8	1041.06
180	10	32	857.23
30	2	8	21.70

TABLE IV. MAE COMPARISON

Days of data taken for input processing	Number of clients used in FL	Batches of data taken for processing	MAE
7	10	8	25.21
180	10	32	21.70
30	2	8	12.42

The performance of the model was evaluated using MSE, Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). These metrics provide a comprehensive understanding of the model's accuracy and error magnitude, with RMSE being particularly sensitive to large errors due to its squaring of residuals. The first configuration used a historical window of 7 days, with 10 clients, each processing data in batches of 8. The RMSE for this model was recorded at 32.265, with an MSE of 1041.057 and an MAE of 25.217. Although these results represent a reasonable starting point, the relatively high error metrics indicate that a shorter historical data window may not capture the necessary trends and patterns needed for more precise AQI forecasting.

In stark contrast, the model trained on a substantially longer window of 180 days, also with 10 clients and a larger batch size of 32, demonstrated improved accuracy. The RMSE decreased to 29.279, along with a lower MSE of 857.234 and an MAE of 21.703. This improvement suggests that incorporating a more extended range of historical data can enhance the model's predictive capability, possibly by capturing more long-term trends and seasonal effects. However, the most significant improvement in model performance was observed in the configuration with 30 days of historical data, 2 clients, and a batch size of 8. This model achieved an RMSE of 15.933, an MSE of 253.851, and an MAE of 12.423, exceeding the other configurations by a

considerable margin. This configuration offers a balance between capturing relevant temporal patterns and maintaining a manageable computational workload.

These results elucidate some key points. First, there appears to be an optimal historical window (30 days, in this case) that provides enough data for the model to learn effectively without being overwhelmed by the volume or noise of the information. Second, the number of clients and batch size also play a crucial role, as too few clients may not provide enough diversity in the data, while too many may introduce complexity without corresponding benefits in accuracy. Lastly, larger batch sizes seem to offer diminishing returns in terms of model accuracy. Transitioning seamlessly from the examination of model parameters to performance metrics, it becomes clear that the configurations of FL networks must be carefully tailored to the specific forecasting task at hand. The chosen configuration must strike a balance between data comprehensiveness, model complexity, and computational efficiency to achieve optimal forecasting accuracy. These insights have profound implications for the deployment of FL models in real-world settings, where balancing these factors is crucial for both performance and practicality. In the realm of predictive analytics for environmental data, the reliability and accuracy of forecasts are paramount.

Figure 3 shows a comparative plot with the actual AQI values traced in blue and the predicted values in an orange dashed line, utilizing a 30-day window with two clients and a batch size of 8. The model's ability to trace the volatility and trends in AQI levels is evident, although with some divergence from the actual values, particularly in capturing the peaks and troughs. Despite this, the overall trend closely mirrors the actual data, suggesting a degree of synchronicity between the model outputs and the true AQI values. Considering performance metrics, the model with a 30-day data window, 2 clients, and a batch size of 8 demonstrates a strikingly lower RMSE, MSE, and MAE compared to the other configurations. These metrics are indicative of its superior performance and suggest that this model configuration is adept at discerning the underlying patterns within the data.

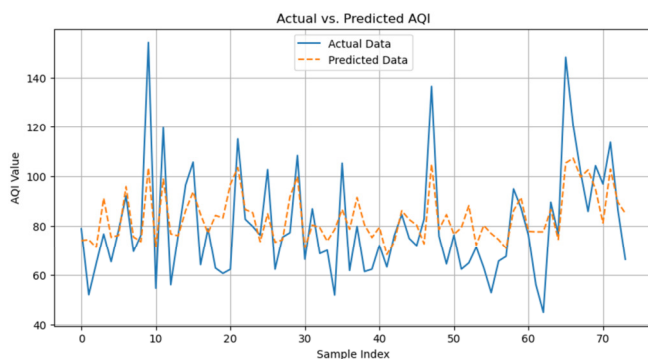


Fig. 3. AQI forecasting using the FL model with the minimal RMSE using a 30-day data window, 2 clients, and a batch size of 8.

In conclusion, the results derived from the FL models strongly advocate the adoption of the 30-day, 2-client, 8-batch

size model as the preferred for AQI forecasting in an FL framework. This model not only excels in quantitative performance metrics but also qualitatively in its fidelity to actual environmental patterns.

V. CONCLUSION

The imperative for real-time, privacy-preserving air quality monitoring has never been more acute, as urban environments continue to face the growing complexities of environmental pollution. This study illustrates a significant step forward in this domain, positing FL as a superior method for AQI forecasting using IoT technology. The resultant FL framework allows for the seamless integration of diverse, distributed sensor data, facilitating smart city initiatives and comprehensive environmental monitoring without compromising individual data integrity. The validation of the FL model approach indicates that the optimized model, taking into account 30 days of data with 2 clients and processing in batches of 8, demonstrates superior performance with lower error metrics. This finding advocates for the custom application of FL in environmental sensing, ensuring the scalability, efficiency, and robustness of predictive models.

Although the model achieved the best results with only two clients, future work should focus on expanding its effectiveness to accommodate more clients. Strategies to improve scalability and handle the increased complexity of larger federated networks will be crucial. Techniques such as hierarchical FL, where multiple levels of aggregation are employed, or the incorporation of more sophisticated model architectures, may enhance the model's ability to generalize across a broader range of clients. Furthermore, exploring adaptive client participation, where clients with the most significant data updates are prioritized, could further optimize the training process. Looking ahead, the promising results laid out here pave the way for expansive future work. Further research should investigate the integration of more diverse datasets, the refinement of client selection strategies, and the application of the framework to other domains that require decentralized data processing. The ultimate goal is to develop a robust, scalable, and efficient FL system capable of providing accurate and timely air quality predictions in increasingly complex and diverse urban environments.

ACKNOWLEDGMENT

This study was funded by Prince Sattam bin Abdulaziz University project number PSAU/2024/R/1445.

REFERENCES

- [1] A. Pant, R. C. Joshi, S. Sharma, and K. Pant, "Predictive Modeling for Forecasting Air Quality Index (AQI) Using Time Series Analysis," *Avicenna Journal of Environmental Health Engineering*, vol. 10, no. 1, pp. 38–43, Jun. 2023, <https://doi.org/10.34172/ajehe.2023.5376>.
- [2] P. Mullangi *et al.*, "Assessing Real-Time Health Impacts of outdoor Air Pollution through IoT Integration," *Engineering, Technology & Applied Science Research*, vol. 14, no. 2, pp. 13796–13803, Apr. 2024, <https://doi.org/10.48084/etasr.6981>.
- [3] S. Abimannan *et al.*, "Towards Federated Learning and Multi-Access Edge Computing for Air Quality Monitoring: Literature Review and Assessment," *Sustainability*, vol. 15, no. 18, Jan. 2023, Art. no. 13951, <https://doi.org/10.3390/su151813951>.

- [4] A. A. Siyal, S. R. Samo, Z. A. Siyal, K. C. Mukwana, S. A. Jiskani, and A. Mengal, "Assessment of Air Pollution by PM10 and PM2.5 in Nawabshah City, Sindh, Pakistan," *Engineering, Technology & Applied Science Research*, vol. 9, no. 1, pp. 3757–3761, Feb. 2019, <https://doi.org/10.48084/etasr.2440>.
- [5] S. Senthivel and M. Chidambaranathan, "Machine Learning Approaches Used for Air Quality Forecast: A Review - ProQuest," *Revue d'Intelligence Artificielle*, vol. 36, no. 1, pp. 73–78, <https://doi.org/10.18280/ria.360108>.
- [6] M. Alquraish and K. Abuhasel, "Sustainable Hybrid Design to Ensure Efficiency and Air Quality of Solar Air Conditioning," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 11036–11041, Jun. 2023, <https://doi.org/10.48084/etasr.5907>.
- [7] 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*, vol. 8, no. 12, pp. 9827–9837, Jun. 2021, <https://doi.org/10.1109/JIOT.2020.3021006>.
- [8] L. D. Labzovskii *et al.*, "Who should measure air quality in modern cities? The example of decentralization of urban air quality monitoring in Krasnoyarsk (Siberia, Russia)," *Environmental Science & Policy*, vol. 140, pp. 93–103, Feb. 2023, <https://doi.org/10.1016/j.envsci.2022.11.016>.
- [9] N. Lazrak, J. Ouarzazi, J. Zahir, and H. Mousannif, "Enabling distributed intelligence in Internet of Things: an air quality monitoring use case," *Personal and Ubiquitous Computing*, vol. 27, no. 6, pp. 2043–2053, Dec. 2023, <https://doi.org/10.1007/s00779-020-01483-3>.
- [10] J. C. Jiang, B. Kantarci, S. Oktug, and T. Soyata, "Federated Learning in Smart City Sensing: Challenges and Opportunities," *Sensors*, vol. 20, no. 21, Jan. 2020, Art. no. 6230, <https://doi.org/10.3390/s20216230>.
- [11] A. Khan, S. Aslam, K. Aurangzeb, M. Alhussein, and N. Javaid, "Multiscale modeling in smart cities: A survey on applications, current trends, and challenges," *Sustainable Cities and Society*, vol. 78, Mar. 2022, Art. no. 103517, <https://doi.org/10.1016/j.scs.2021.103517>.
- [12] N. Ge, G. Li, L. Zhang, and Y. Liu, "Failure prediction in production line based on federated learning: an empirical study," *Journal of Intelligent Manufacturing*, vol. 33, no. 8, pp. 2277–2294, Dec. 2022, <https://doi.org/10.1007/s10845-021-01775-2>.
- [13] D. D. Le, A. K. Tran, M. S. Dao, M. S. H. Nazmudeen, V. T. Mai, and N.-H. Su, "Federated Learning for Air Quality Index Prediction: An Overview," in *2022 14th International Conference on Knowledge and Systems Engineering (KSE)*, Nha Trang, Vietnam, Oct. 2022, pp. 1–8, <https://doi.org/10.1109/KSE56063.2022.9953790>.
- [14] D. D. Le *et al.*, "Insights into Multi-Model Federated Learning: An Advanced Approach for Air Quality Index Forecasting," *Algorithms*, vol. 15, no. 11, Nov. 2022, Art. no. 434, <https://doi.org/10.3390/a15110434>.
- [15] N. Jin, Y. Zeng, K. Yan, and Z. Ji, "Multivariate Air Quality Forecasting With Nested Long Short Term Memory Neural Network," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 12, pp. 8514–8522, Sep. 2021, <https://doi.org/10.1109/TII.2021.3065425>.
- [16] P. Chhikara, R. Tekchandani, N. Kumar, M. Guizani, and M. M. Hassan, "Federated Learning and Autonomous UAVs for Hazardous Zone Detection and AQI Prediction in IoT Environment," *IEEE Internet of Things Journal*, vol. 8, no. 20, pp. 15456–15467, Jul. 2021, <https://doi.org/10.1109/JIOT.2021.3074523>.
- [17] R. R. Relkar, "Prediction of Air Quality Index Using Supervised Machine Learning," *International Journal for Research in Applied Science and Engineering Technology*, vol. 10, no. 6, pp. 1371–1382, Jun. 2022, <https://doi.org/10.22214/ijraset.2022.43993>.
- [18] Q. C. Thi, N. T. T. Hoa, and N. T. C. Ngoan, "Prediction of Air Quality Index using genetic programming," *Journal of Military Science and Technology*, vol. 91, pp. 85–95, Nov. 2023, <https://doi.org/10.54939/1859-1043.j.mst.91.2023.85-95>.
- [19] N. Phruksahiran, "Improvement of air quality index prediction using geographically weighted predictor methodology," *Urban Climate*, vol. 38, Jul. 2021, Art. no. 100890, <https://doi.org/10.1016/j.uclim.2021.100890>.
- [20] H. Alkabbani, A. Ramadan, Q. Zhu, and A. Elkamel, "An Improved Air Quality Index Machine Learning-Based Forecasting with Multivariate Data Imputation Approach," *Atmosphere*, vol. 13, no. 7, Jul. 2022, Art. no. 1144, <https://doi.org/10.3390/atmos13071144>.
- [21] M. Hardini, R. A. Sunarjo, M. Asfi, M. H. R. Chakim, and Y. P. A. Sanjaya, "Predicting Air Quality Index using Ensemble Machine Learning," *ADI Journal on Recent Innovation*, vol. 5, no. 1S, pp. 78–86, Aug. 2023, <https://doi.org/10.34306/ajri.v5i1Sp.981>.
- [22] "Air quality monitoring, emission inventory and source apportionment study for Indian cities." Central Pollution Control Board, Ministry of Environment, Forest and Climate Change, Government of India.