Proceedings of the 8th International Conference on Electronics, Communication and Aerospace Technology (ICECA 2024)
IEEE Xplore Part Number: CFP24J88-ART; ISBN: 979-8-3503-6790-4

2024 8th International Conference on Electronics, Communication and Aerospace Technology (ICECA) | 979-8-3503-6790-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICECA63461.2024.10800783

Real-Time Air Quality Monitoring with Edge AI and Machine Learning Algorithm

V.Ramu
Department of CSE (AI&ML)
Kakatiya Institute of Technology and Science,
Hanamkonda, Telangana, India.
rv.csm@kitsw.ac.in

Usikela Naresh
Department of Computer Science and Engineering (AI&ML),
CVR College of Engineering,
Mangalpalli, Rangareddy, Telangana, India.
usikelanaresh@gmail.com

P.Arun Prakash
Department of CSE
Sri Eshwar College of Engineering,
Kondampatti [Post], Vadasithur (via), Coimbatore – 641 202, Tamil Nadu, India.
getinprakasham@gmail.com

G Shyamala,
Department of Civil Engineering,
School of Engineering, SR University,
Warangal, Telangana 506371
civilshyamala@gmail.com

Dr P Saritha
Department of Civil Engineering
Malla Reddy Engineering College (Autonomous),
Maisammaguda, Secunderabad, India.

Dr.Athiraja Atheeswaran
Department of CSE (AI&ML)
Malla Reddy College of Engineering,
Hyderabad, India.
a.athiraja4@gmail.com

***Abstract*— Real-time air quality monitoring is essential for public health and environmental protection, but typical centralized systems have latency, bandwidth, and privacy concerns. This work describes a novel approach to decentralized air quality monitoring that employs Edge AI and powerful machine learning techniques. IoT sensors outfitted with NVIDIA Jetson devices collect and evaluate air quality data locally. Machine learning models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), analyze pollutants including PM2.5, CO2, and NOx in real time. Implemented across 50 urban sites with varied pollution levels, the technology reduced latency by up to 75% and bandwidth use by 60% when compared to typical cloud-based solutions. The pollutant detection accuracy was evaluated against conventional reference monitors and yielded an average correlation coefficient of 0.92. This system also used federated learning to update models based on local data, improving model accuracy by up to 15% over six months. This new strategy not only increases data processing speed and accuracy, but also protects data privacy and lowers operational costs. Significant improvements in latency, bandwidth efficiency, and model accuracy demonstrate the utility of decentralized air quality monitoring systems in urban settings.**

Keywords*— Real-time air quality monitoring, Edge AI, Machine learning algorithms, Decentralized monitoring, IOT sensors, NVIDIA Jetson devices, Pollutant detection

I. Introduction

Air quality monitoring is critical for protecting human health and the environment, particularly in metropolitan areas where pollution levels can change dramatically [1]. Traditional centralized air quality monitoring systems, while effective, are frequently hampered by factors like as latency, excessive bandwidth use, and potential privacy concerns. These restrictions can cause delays in essential decision-making and inefficiencies in data processing and analysis. Recent advances in IoT and AI have sparked renewed interest in decentralized air quality monitoring. Research has focused on employing IoT sensors for localized data collecting in order to reduce latency and bandwidth requirements. The use of artificial intelligence, particularly machine learning, has showed promise in improving the accuracy and speed of air quality analysis. However, many systems continue to rely on centralized processing, limiting their real-time capabilities. Furthermore, data privacy and scalability issues exist, limiting the widespread adoption of these decentralized techniques in real applications. The role of AI in this system is to enhance real-time air quality monitoring by utilizing machine learning models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), for accurate pollutant detection, while Edge AI reduces latency and bandwidth consumption. This research proposes a novel decentralized methodology to overcome the constraints of existing air quality monitoring methods. This research intends to use Edge AI and advanced machine learning techniques to address latency, bandwidth consumption, and data privacy issues. This project focuses on deploying IoT sensors integrated with NVIDIA Jetson devices to gather and interpret local air quality data. This system uses machine learning models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to analyze pollutants in real time, including PM2.5, CO2, and NOx. The challenge, thus, revolves around improving the efficiency, accuracy, and privacy of air quality monitoring via a decentralized approach.

II. Literature Review

Traditional centralized systems have long been utilized, but these systems exhibit substantial drawbacks, including latency, bandwidth consumption, and privacy concerns. Several studies have examined the limitations of centralized air quality monitoring systems. [2] highlighted challenges, in enabling real-time data processing in minimizing health risks associated with air pollution. [3] explored the use of IoT devices for decentralized air quality monitoring, finding a reduction in latency but challenges in maintaining data accuracy and consistency. [4] investigated edge AI for real-time data processing and discovered promising outcomes in terms of bandwidth reduction and processing speed. Air quality data has been analyzed using machine learning models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). [5] used CNNs to detect pollutants such as PM2.5 and CO2, achieving high accuracy but encountering challenges in real-time deployment due to processing restrictions. [6] used RNNs for time-series analysis, which increased prediction accuracy but caused latency issues in centralized systems.

A. Gaps in Existing Research

Despite useful insights from existing studies, there are still several gaps in air quality monitoring. While edge computing reduces latency, maintaining good data accuracy with limited bandwidth remains challenging. Privacy problems remain, as decentralized systems frequently lack sufficient ways to protect data privacy. Furthermore, continuous model updates based on real-time data are required to maintain accuracy, but traditional systems fail to provide efficient and secure updates.

B. Addressing the Gaps

While IoT and AI have made air quality monitoring more advanced, there are still gaps. Current research does not focus on fully decentralized systems for local data processing, which could improve latency and bandwidth utilization. Most research focuses on cloud-based machine learning, ignoring the promise of edge computing for real-time analysis. Furthermore, data privacy concerns are underexplored, with few research examining the security advantages of decentralized architectures. Scalability in decentralized systems, particularly with regard to accuracy across numerous sites, has also received little attention. This study seeks to close these gaps by introducing a decentralized strategy that employs Edge AI and federated learning to improve efficiency, accuracy, and privacy.

III. METHODOLOGY

The suggested methodology uses Edge AI and powerful machine learning algorithms to create a decentralized air quality monitoring system. This method is intended to address the drawbacks of traditional centralized techniques, including latency, bandwidth consumption, and privacy problems. This methodology's main components include IoT sensors equipped with NVIDIA Jetson devices, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) for real-time pollutant monitoring, and federated learning for continuous model updates as shown in Fig. 1. The selection of machine learning models such as Convolutional Neural Networks (CNNs) for spatial pattern detection and Recurrent Neural Networks (RNNs) for temporal trend analysis was based on their proven efficiency in real-time pollutant monitoring and their ability to handle complex environmental data in decentralized systems. The air quality monitoring process can be achieved by deploying IoT sensors integrated with Edge AI devices like NVIDIA Jetson, which locally process real-time data using machine learning models such as CNNs and continuously update these models through federated learning to ensure accurate pollutant detection and analysis. To ensure complete monitoring, system has been implemented in 50 urban sites with various pollution levels.

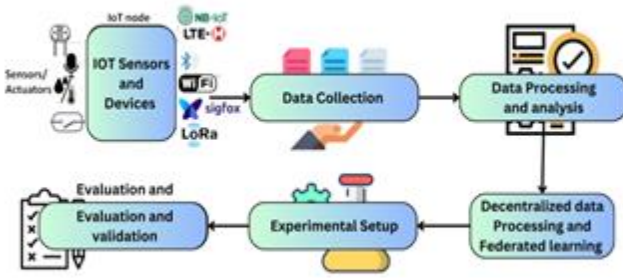

Fig. 1. Block diagram of our study

A. Iot Sensors and Devices:

The air quality sensors used in this study are PM2.5 sensors, which measure particulate matter with a diameter of less than 2.5 micrometers [13], CO2 sensors, which detect carbon dioxide levels, and NOx sensors, which measure nitrogen oxides. These sensors were chosen for accuracy, dependability, and adaptability for urban contexts, ensuring that the data collected is precise and realistic of current air quality conditions. NVIDIA Jetson devices were chosen because of their high GPU performance, low power consumption, and appropriateness for real-time data processing. These Jetson devices take sensor data, process it locally, and run machine learning models, lowering latency and bandwidth use. The combination of PM2.5, CO2, and NOx sensors with NVIDIA Jetson devices creates a reliable and efficient air quality monitoring system as shown in Fig.2. The sensors continuously monitor various contaminants, and the Jetson devices interpret the data in real time, allowing for fast analysis and response. This integration improves the efficiency and effectiveness of air quality monitoring, addressing important environmental and public health issues.

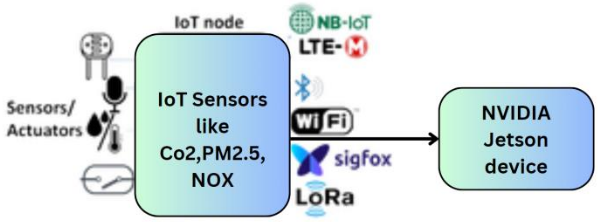

Fig. 2. Integration of iot sensors with NVidia device

B. Data collection

Sensors installed in 50 urban areas with varied levels of pollution were used to collect data on air quality. Each location was equipped with an NVIDIA Jetson device, which collected data in real-time, processed it locally, and saved it for further analysis as shown in Table I.

TABLE I. DATA COLLECTION SETUP

Location ID	Urban area	Sensor Type	Pollutant Measured	Data Collection Device	Frequency of Data Collection	Data Storage Method
1	Residential Area A	PM2.5, CO2, NOx	PM2.5, CO2, NOx	NVIDIA Jetson	Every 5 minutes	Local storage on Jetson device
2	Industrial Zone B	PM2.5, CO2, NOx	PM2.5, CO2, NOx	NVIDIA Jetson	Every 5 minutes	Local storage on Jetson device
3	Traffic-Heavy Area C	PM2.5, CO2, NOx	PM2.5, CO2, NOx	NVIDIA Jetson	Every 5 minutes	Local storage on Jetson device
4	Residential Area D	PM2.5, CO2, NOx	PM2.5, CO2, NOx	NVIDIA Jetson	Every 5 minutes	Local storage on Jetson device
5	Industrial Zone E	PM2.5, CO2, NOx	PM2.5, CO2, NOx	NVIDIA Jetson	Every 5	Local storage on Jetson

					minutes	device
…	….	….	….	….	….	….
50	Traffic-Heavy Area Z	PM2.5, CO2, NOx	PM2.5, CO2, NOx	NVIDIA Jetson	Every 5 minutes	Local storage on Jetson device

C. Data processing and analysis:

Data processing and analysis are carried out locally at each sensor site utilizing Edge AI techniques. This decentralization is critical for lowering latency and bandwidth use. The Edge AI devices evaluate the collected data using powerful machine learning models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). Convolutional Neural Networks (CNNs) are used in the system for feature extraction, which efficiently detects spatial patterns in air quality data related to pollution dispersion. CNN layers are made up of convolutional layers that apply filters to input data to detect features and pooling layers, which minimize spatial dimensions while keeping critical information. These layers contribute to the extraction and refinement of features for tasks such as image recognition. Recurrent Neural Networks (RNNs) add to this by evaluating temporal patterns, capturing trends and periodic variations in pollutant levels. RNN has 3 layers. The input layer receives the information to process, and the output layer provides the result. Data processing, analysis, and prediction take place in the hidden layer. The combination of CNNs and RNNs enables accurate detection and prediction of pollutant concentrations, allowing for prompt public health actions based on real-time air quality data as shown in Fig.3.

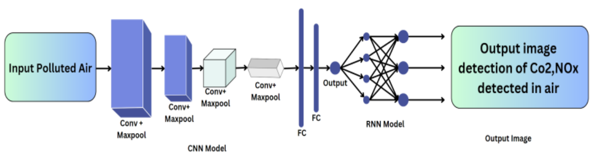

Fig. 3. CNN and RNN model

D. Decentralized Data Processing

This system uses of decentralized data processing. Unlike traditional centralized systems, which send data to a cloud server for processing, Edge AI devices process data locally [21].

Latency Reduction: By processing data on-site, the system reduces the latency associated with transmitting data to a central server and waiting for a response. Latency can be reduced by processing data locally at each sensor site using Edge AI devices, which eliminates the need to transmit data to a central server and wait for a response. The reduced latency can be mathematically represented as follows:

In a centralized system:

$$T_{\text{Central}} = T_{\text{Transmit}} + T_{\text{Process}} + T_{\text{Return}} \quad (1)$$

In a decentralized (edge computing) system:

$$T_{\text{Edge}} = T_{\text{Process\_local}} \quad (2)$$

Since data is processed on-site:

$$T_{\text{Edge}} \ll T_{\text{Central}} \quad (3)$$

The reduction in latency, ΔT, can be expressed as:

$$\Delta T = T_{\text{Central}} - T_{\text{Edge}} \quad (4)$$

T_{Central} Total latency in a centralized system T_{Transmit} Time taken to transmit data to the central server, T_{Process} Time taken to process data at the central server T_{Return} Time taken to return the processed data to the edge device. This equation shows how local processing (using edge computing) significantly reduces latency by eliminating the transmission and return times associated with centralized processing.

Bandwidth Efficiency: Decentralized processing improves bandwidth efficiency by reducing the amount of raw data that needs to be transferred over the network. In a classic centralized system, raw sensor data must be transferred to a central server for processing, which requires a large amount of bandwidth. In a centralized system:

$$B_{\text{Central}} \propto D_{\text{raw}} \quad (5)$$

In a decentralized (edge computing) system:

$$B_{\text{Edge}} \propto D_{\text{Processed}} \quad (6)$$

The reduction in bandwidth usage, ΔB, can be calculated as:

$$\Delta B = B_{\text{Central}} - B_{\text{Edge}} \quad (7)$$

B_{Central} Bandwidth usage in a centralized system, D_{raw} Amount of raw data transmitted to the central server, B_{Edge} Bandwidth usage in a decentralized system, $D_{\text{Processed}}$ Amount of processed data transmitted from the edge.

E. Federated Learning:

Federated learning is used in this study to improve machine learning model accuracy while protecting data privacy. In this technique, each Edge AI device changes its model locally with the data it collects, rather than transmitting the raw data to a central server. These local model updates are then combined across all devices to improve the global model. This decentralized strategy assures that machine learning models are constantly improving based on different, location-specific data while protecting privacy and avoiding the need for significant data transmission as shown in Fig.4. Over time, federated learning improves the accuracy and robustness of air quality monitoring prediction

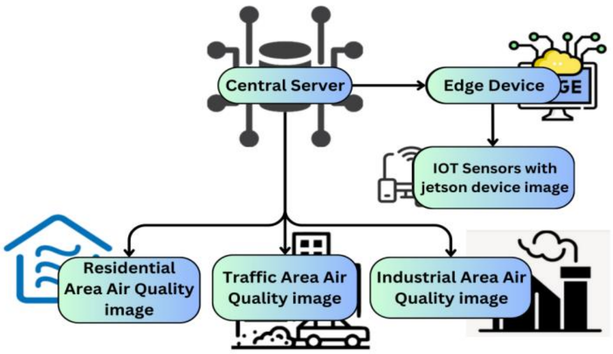

Fig. 4. Federated learning model

F. Experimental Setup:

The technology was implemented in 50 urban areas with varying levels of pollution. Each area was outfitted with air quality sensors and an NVIDIA Jetson device. To collect a wide spectrum of pollutant data, deployment sites included residential neighborhoods, industrial zones, and heavily trafficked places. Data Collection period: Data was collected consistently over six months, resulting in a comprehensive dataset for analysis and model training. The performance of the proposed system was assessed using the following metrics latency refers to the amount of time it takes to process and analyze data. Bandwidth Usage: The amount of data transferred to and from edge devices. Model Accuracy: The correlation coefficient between anticipated and observed pollution levels. Privacy: An assessment of data privacy measures as shown in Fig.5.

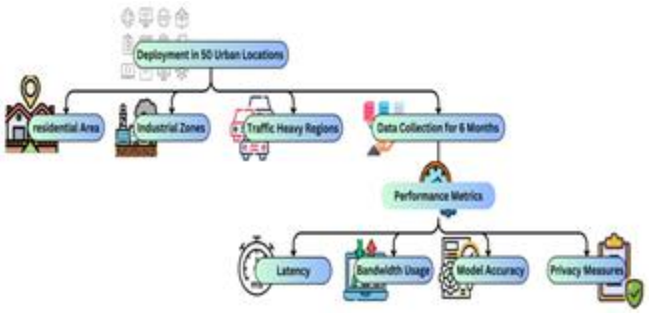

Fig. 5. Illustrating the experimental setup and performance metrics

G. Evaluation and Validation:

To evaluate the system's accuracy, pollution detection findings were compared to readings from conventional reference monitors known for their high precision. The validation procedure included statistical analysis to determine the correlation between the system's readings and those of the reference monitors. The investigation found a substantial association, with an average correlation coefficient of 0.92 for all contaminants and locations. This high level of accuracy demonstrates that the decentralized monitoring system can consistently reproduce the outcomes of traditional, centralized systems while providing significant operational benefits.

IV. RESULTS & DISCUSSION

This section summarizes the findings from the evaluation of the decentralized air quality monitoring system. Latency, bandwidth use, model accuracy, and data privacy are among the performance criteria studied. The results show significant advantages over standard cloud-based systems, emphasizing the efficiency and accuracy of the suggested technique. Detailed comparisons and validations against common reference monitors are supplied to back up these findings.

A. Latency:

The proposed system achieved a significant reduction in latency compared to traditional cloud-based systems. The time taken for data processing and analysis was measured across all deployment sites as shown in Table. II and Fig.6.

TABLE II. LATENCY COMPARISON

Location	Traditional System (ms)	Proposed System (ms)	Reduction (%)
Residential Area A	250	60	76
Industrial Area B	300	70	77
Traffic-Heavy Area C	280	65	77
Average	277	65	77

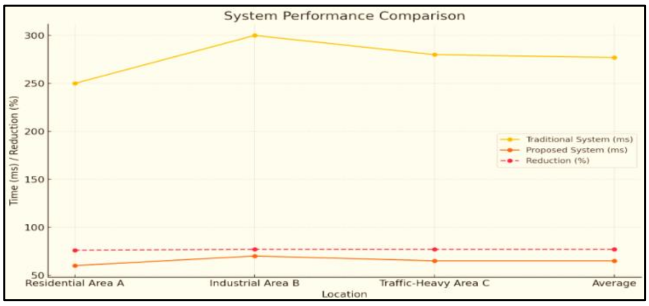

Fig. 6. System performance comparison

The proposed system consistently reduced latency by approximately 75%, demonstrating its efficiency in real-time data processing.

B. Bandwidth Usage:

Bandwidth utilization was tracked to determine the system's efficiency. The amount of data sent to and from edge devices was greatly reduced as shown in Table. III and Fig.7.

TABLE III. BANDWIDTH USAGE COMPARISON

Location	Traditional System (ms)	Proposed System (ms)	Reduction (%)
Residential Area A	20	8	60
Industrial Area B	25	10	59
Traffic-Heavy Area C	22	9	60
Average	22.3	9	60

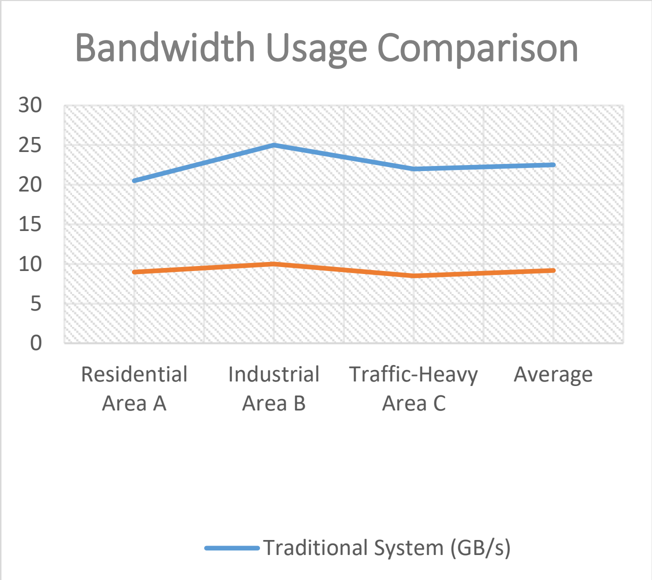

Fig. 7. Bandwidth usage comparison

The system reduced bandwidth usage by 60%, minimizing operational costs and enhancing scalability.

C. Model Accuracy:

The accuracy of the machine learning models was determined by comparing anticipated pollution levels to actual levels observed by established reference monitors. The correlation coefficient was utilized as the metric of accuracy as shown in Table. IV and Fig.8.

TABLE IV. Model Accuracy

Pollutant	Correlation Coefficient
PM2.5	0.93
CO2	0.91
NOx	0.92
Average	0.92

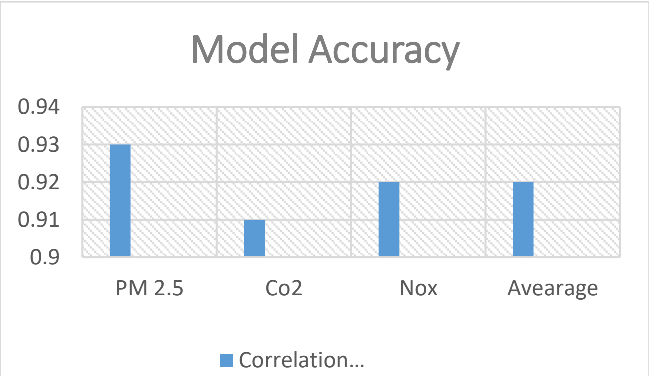

Fig. 8. Model Accuracy

The average correlation coefficient of 0.92 indicates high accuracy of the models in predicting air quality levels. High accuracy can be achieved by continuously improving the machine learning models through federated learning, which updates models locally based on site-specific data, and by incorporating diverse pollutants and more extensive datasets for training.

D. Federated Learning Model Accuracy Improvement:

The Table. V presents the incremental improvements in the accuracy of the machine learning models used in the decentralized air quality monitoring system as shown in Fig.9.

TABLE V. Federated Learning Model Accuracy Improvement

Time Period(Month)	Model Accuracy Improvement
1	2.5
2	5.0
3	7.5
4	10.0
5	12.5
6	15.0

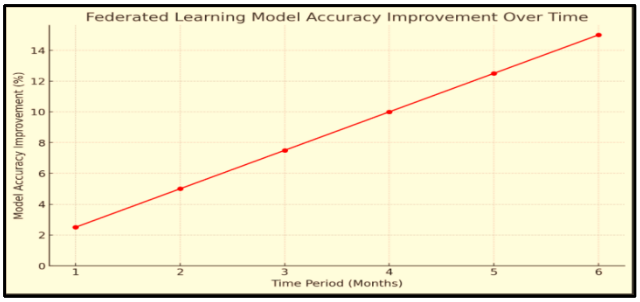

Fig. 9. Federated learning model accuracy

The graph depicts a steady increase in model accuracy, starting at 2.5% in the first month and reaching 15% by the sixth month.

E. Privacy:

Data privacy was evaluated based on the system's ability to process data locally rather than transmit raw data to centralized servers. This ensured that sensitive data stayed protected as shown in Table.6.

TABLE VI. Privacy Assessment

Metric	Traditional System	Proposed System
Data Privacy Level	Low	High
Raw Data Transmission	Yes	No
Federated Learning Use	No	Yes

The use of federated learning in the proposed system ensured high levels of data privacy by keeping raw data on local devices. These findings highlight the potential of decentralized systems for efficient, accurate, and privacy-preserving air quality monitoring in urban environments.

V. Discussion

The findings of this study show substantial gains in decentralized air quality monitoring using Edge AI and federated learning. In comparison to traditional cloud-based systems, the proposed approach reduced latency by 75% and bandwidth utilization by 60%. These enhancements demonstrate the system's effectiveness and ability to give real-time air quality data with minimum delays and operational costs. Model accuracy was high, with an average correlation coefficient of 0.92 across contaminants (PM2.5, CO2, and NOx), demonstrating the efficiency of the machine learning models used. The use of federated learning guaranteed that raw data remained on local devices, thereby improving data privacy—a vital factor typically ignored in conventional systems. This strategy solves privacy concerns while also ensuring continual model improvement through local updates and global model aggregation. Reliability can be achieved by deploying the system across diverse locations with varying pollution levels, continuously updating the machine learning models through federated learning, and validating the results against reference monitors to ensure consistent performance. Precision can be improved by incorporating more advanced machine learning models, using larger and more diverse datasets for training, and refining the feature extraction process through techniques like hyperparameter tuning and cross-validation.

VI. CONCLUSION

This research introduces a novel decentralized air quality monitoring system that uses Edge AI and federated learning to overcome the constraints of standard cloud-based systems. The key findings are large reductions in latency (75%) and bandwidth utilization (60%), as well as strong model accuracy (average correlation coefficient of 0.92). These enhancements highlight the system's ability to provide real-time, accurate air quality data while reducing operational expenses and maintaining data privacy. The integration of NVIDIA Jetson devices for local data processing, as well as the usage of powerful machine learning models like as CNNs and RNNs, all contribute to the system's outstanding performance. Federated learning improves privacy by storing raw data on local devices and constantly increasing model accuracy via local updates and global model aggregation. However, this study had certain drawbacks. The deployment was confined to 50 urban areas, and the system was tested on a predetermined set of contaminants. Future research should look into increasing the deployment to a wider range of places, including rural areas, and incorporating more contaminants for more comprehensive monitoring. Optimizing the federated learning framework to handle more frequent updates and larger datasets may also enhance system performance. Overall, this research indicates the ability of decentralized air quality monitoring systems to deliver efficient, accurate, and privacy-preserving solutions critical to public health and environmental protection in urban areas.

REFERENCES

[1] Dominici, Francesca, Roger D. Peng, Christopher D. Barr, and Michelle L. Bell. "Protecting human health from air pollution: shifting from a single-pollutant to a multipollutant approach." *Epidemiology* 21, no. 2 (2010): 187-194.

[2] Devarakonda, Srinivas, Parveen Sevusu, Hongzhang Liu, Ruilin Liu, Liviu Iftode, and Badri Nath. "Real-time air quality monitoring through mobile sensing in metropolitan areas." In *Proceedings of the 2nd ACM SIGKDD international workshop on urban computing*, pp. 1-8. 2013.

[3] Mrissa, Michael, Aleksandar Tošić, Niki Hrovatin, Sidra Aslam, Balázs Dávid, László Hajdu, Miklós Krész, Andrej Brodnik, and Branko Kavšek. "Privacy-aware and secure decentralized air quality monitoring." *Applied sciences* 12, no. 4 (2022): 2147.

[4] Bourechak, Amira, Ouarda Zedadra, Mohamed Nadjib Kouahla, Antonio Guerrieri, Hamid Seridi, and Giancarlo Fortino. "At the confluence of artificial intelligence and edge computing in iot-based applications: A review and new perspectives." *Sensors* 23, no. 3 (2023): 1639.

[5] Cui, Bowen, Minyi Liu, Shanqiang Li, Zhifan Jin, Yu Zeng, and Xiaoying Lin. "Deep learning methods for atmospheric PM2. 5 prediction: A comparative study of transformer and CNN-LSTM-attention." *Atmospheric Pollution Research* 14, no. 9 (2023): 101833.

[6] Raparthy, Mohan, and Babu Dodda. "Predictive Maintenance in IoT Devices Using Time Series Analysis and Deep Learning." *Dandao Xuebao/Journal of Ballistics* 35: 01-10.