Alexandria Engineering Journal 114 (2025) 681–701

Contents lists available at ScienceDirect

Alexandria Engineering Journal

journal homepage: www.elsevier.com/locate/aej

SecureIoT-FL: A Federated Learning Framework for Privacy-Preserving Real-Time Environmental Monitoring in Industrial IoT Applications

Montaser N.A. Ramadan [a,*], Mohammed A.H. Ali [a,*], Shin Yee Khoo [a,b], Mohammad Alkhedher [c]

[a] Department of Mechanical Engineering, Faculty of Engineering, Universiti Malaya, Kuala Lumpur 50603, Malaysia
[b] Centre of Research Industry 4.0 (CRI 4.0), Faculty of Engineering, Universiti Malaya, Kuala Lumpur 50603, Malaysia
[c] Mechanical and Industrial Engineering Department, Abu Dhabi University, Abu Dhabi 59911, United Arab Emirates

ARTICLE INFO

Keywords:
Federated learning
Environmental monitoring
Industrial IoT
Multi-sensor fusion
Privacy-preserving data management

ABSTRACT

Industrial environments face increasing challenges in achieving accurate environmental monitoring and maintaining data privacy. This paper presents the SecureIoT-FL framework, an innovative integration of federated learning and multi-sensor fusion for real-time, privacy-preserving prediction of environmental pollutants in industrial settings. A normalized Federated Averaging (N-FedAvg) as an enhanced version of the conventional FedAvg algorithm, is introduced for global model aggregation to balance the contributions from all clients with varying data volumes, thus stabilizing the learning process. This study contributes a scalable and secure solution for environmental monitoring and risk management in industrial IoT applications. The framework leverages IoT nodes equipped with Raspberry Pi 5 and sensors to detect pollutants (e.g., PM2.5, PM10, CO_2, VOCs, CH_2O, CO, O_3, NO_2, H_2S, O_2, CH_4, SO_2) across three distinct environments: a PCBA factory, a CNC machining factory, and a plastic injection factory. Over five months, the system collected over 516,000 data points, enabling a multi-source data fusion approach to enhance predictive accuracy. The key outcomes of the proposed work include significant improvements in the model performance over 10 biweekly training rounds, with increasing the accuracy from 70 % to 92 % and 68–89 % for LSTM and CNN models, respectively. The framework effectively predicts the pollutant concentrations over multiple timeframes (10, 30, and 60 minutes), supporting timely interventions to reduce worker exposure to harmful pollutants. The SecureIoT-FL framework employs federated learning with TLS encryption, ensuring data privacy while supporting secure local model training and weight sharing via MQTT over AWS.

1. Introduction

The importance of environmental monitoring in industrial settings cannot be overstated, particularly in the context of monitoring pollutants such as PM2.5, CO_2, NO_2, and SO_2. These pollutants pose significant risks to both environmental and human health, necessitating robust monitoring systems to ensure operational efficiency, regulatory compliance, and the safeguarding of worker health.

Workers in factories face a range of health issues due to prolonged exposure to pollution, with respiratory problems being among the most common. In steel factories, exposure to dust, fumes, and gases significantly impacts respiratory health, leading to both acute and chronic symptoms, as well as reduced lung function. This is assured by decreasing the Forced Expiratory Volume in one second (FEV1) / Forced Vital Capacity (FVC) levels, which measure lung function. FEV1 is the amount of air a person can forcefully exhale in the first second of a breath, while FVC is the total amount of air exhaled during the entire breath. A reduction ratio of FEV1/FVC indicates an obstruction of airflow, a common feature in conditions like asthma or chronic obstructive pulmonary disease (COPD) and is often observed in the workers that are exposed to industrial pollutants. Studies show an increasing of prevalence of respiratory symptoms among the exposed workers compared to the office workers [1]. The presence of metal fumes, silica, and acid gases further exacerbates these conditions, increasing the risk of silicosis, lung cancer, and cardiovascular diseases [2].

Industries such as CNC manufacturing, Printed Circuit Board Assembly (PCBA), and Plastic Injection are particularly susceptible to these pollutants, making continuous environmental monitoring essential. For instance, a study on controlling industrial air-pollutant emissions

* Corresponding authors.
E-mail addresses: 23063016@siswa.um.edu.my (M.N.A. Ramadan), hashem@um.edu.my (M.A.H. Ali).

https://doi.org/10.1016/j.aej.2024.11.069
Received 6 September 2024; Received in revised form 10 November 2024; Accepted 19 November 2024
Available online 12 December 2024

highlights the critical need for improved energy efficiency and cleaner production methods to mitigate emissions from key sectors like electricity supply and chemical products [3]. Additionally, the use of nanotechnology and wireless sensor networks has been proposed to enhance the monitoring of toxic gases, offering high sensitivity and spatio-temporal resolution, which are crucial for effective environmental management in industrial settings [4]. te

However, while the importance of environmental monitoring is clear, current solutions predominantly focus on real-time data collection and reporting, with limited emphasis on forecasting potential risks. This reactive approach can hinder timely interventions and exacerbate environmental and health impacts [5,6]. Moreover, the reliance on centralized data processing raises significant data privacy concerns, as sensitive industrial information is vulnerable to breaches during transmission and storage. These limitations highlight the need for a more advanced system that not only ensures data privacy but also enhances predictive capabilities through real-time forecasting.

The integration of Internet of Things (IoT) technologies presents a promising solution to these challenges [7]. IoT enables decentralized data collection and edge processing, which enhances data privacy by allowing local data storage. This approach not only mitigates privacy concerns but also improves the robustness of data management systems. Furthermore, IoT frameworks can incorporate predictive analytics, enabling real-time forecasting of environmental conditions. For example, an IoT-cloud-based air pollution forecasting model has demonstrated the ability to predict air quality indices effectively, thereby informing policy decisions and enabling proactive environmental management [8]. Similarly, the use of deep learning in predictive maintenance within the Industrial Internet of Things (IIoT) has shown potential in predicting impending failures and mitigating unplanned downtime, which can be adapted for environmental monitoring to anticipate and address pollution events before they occur [9].

Studies have demonstrated the effectiveness of IoT in industrial environmental monitoring. For instance, the application of a hybrid-factorial environmental input-output model in Fujian province has provided insights into optimal strategies for controlling emissions, emphasizing the role of technology upgrades and clean energy adoption [3]. Additionally, the development of intelligent monitoring systems that incorporate decision-making algorithms can enhance the reliability and accuracy of environmental assessments, balancing well-being and energy efficiency objectives [10]. In response to these challenges, the integration of federated learning (FL) within IoT frameworks offers a novel solution. FL allows predictive models to be trained locally at each industrial site, with only the updated model parameters being shared with a central server. This method ensures that raw data remains within the local environment, significantly reducing the risk of data breaches while still enabling the global model to improve based on the collective data insights from multiple sources. The proposed approach aims to address the gaps in existing systems, providing a more secure, efficient, and proactive solution for industrial environmental monitoring.

The proposed SecureIoT-FL framework is designed to adapt the varying conditions across different industrial environments. The method's flexibility allows to accommodate diverse diagnosis objects, acquisition equipment, and data distribution characteristics specific to each client. This adaptability ensures that the framework remains effective in the aspect of varying the operational challenges encountered in the real-world scenarios, enhancing its practical applicability.

The primary objective of this study is to develop and validate a novel IoT-based sensing system that combines real-time environmental monitoring with predictive forecasting capabilities, all while maintaining strict data privacy via FL technology. This system is designed to provide industrial environments with the ability to monitor pollutants continuously and predict potential hazards before they occur, ensuring a safer and more compliant operational framework.

Secondary objectives include improving the accuracy of predictions over time through the application of FL techniques, which allow models to learn from data across multiple industrial sites without compromising data privacy. Another key objective is to demonstrate the scalability of the system across different industrial environments, proving its effectiveness in a variety of settings and pollutant profiles.

This research has significant implications for industrial safety, environmental protection, and worker health. By enabling real-time monitoring and forecasting, the proposed system adopts a proactive approach to managing industrial pollutants, potentially reducing the incidence of pollution-related health issues and environmental damage.

Furthermore, this study is pioneering in its application of FL to real-world IoT data in industrial settings. Unlike existing studies that often rely on centralized data processing or synthetic datasets, this research applies advanced machine learning techniques to actual data collected in operational environments. This innovative approach not only sets this work apart but also provides a scalable and efficient solution for enhancing environmental monitoring practices in industries worldwide.

2. Related work

The integration of Internet of Things (IoT) technologies in environmental monitoring has significantly advanced the detection and management of pollutants across various ecosystems[11]. IoT systems leverage a network of sensors and devices to collect, transmit, and analyze environmental data, providing real-time insights into pollution levels and enabling proactive management strategies. These technologies have been extensively employed in monitoring air, water, and soil pollutants, demonstrating their versatility and effectiveness. For air quality monitoring, systems often utilize a combination of sensors to measure parameters such as temperature, humidity, carbon dioxide (CO_2), particulate matter (PM2.5 and PM10), volatile organic compounds (VOCs), and gases like nitrogen dioxide (NO_2), ozone (O_3), and sulfur dioxide (SO_2) [12]. These sensors, integrated with microcontrollers like the ESP32, enable seamless data transmission to cloud servers for real-time analysis and remote access [12].

The recent advancements in IoT-based environmental monitoring have significantly enhanced the capability to analyze complex pollutant profiles in real-time through the integration of advanced sensor networks and AI technologies. These systems are increasingly employed in various environments, including industrial and urban areas, to monitor the air quality by leveraging sophisticated sensors and machine learning algorithms, enabling more accurate detection and prediction of pollutants and thereby improving environmental safety and health outcomes. The IoT systems nowadays incorporate various sensors to monitor hazardous compounds such as VOCs, particulate matter, and other pollutants such as in the chrome plating industry, IoT sensors detect pollutants like hexavalent chromium and VOCs, which are critical for workplace safety [13]. Multi-sensor networks is able to capture a broad spectrum of air pollutants such as NO_2, CO, and SO_2, in real-time, offering comprehensive data for analysis [14]. AI techniques, including machine learning algorithms [15], [16] and time series models, are essential for processing and analyzing data from IoT sensors which enhance the prediction and proactive response to pollution events [17]. The advanced AI models like Graph Neural Networks (GNNs) have shown a superior performance in capturing the spatial and temporal variations in the environmental data, leading to more accurate predictions and effective monitoring systems [18]. Additionally, the environmental monitoring by IoT and AI is extended to cover a wide range of applications in agriculture and urban planning, highlighting their versatility and potential for sustainable development [19]. While such advancements present significant benefits, the challenges are remained which include data privacy, sensor calibration, and the need for robust infrastructure to support large-scale deployment. The addressing of the above-mentioned challenges is essential to maximize the potential of IoT and AI in environmental monitoring.

Several IoT frameworks have been developed for environmental monitoring, each with its strengths and weaknesses. For instance, a standard-based IoT platform using the IEEE 1451 standard facilitates stable data flow and communication between network nodes, ensuring reliable long-term monitoring [20]. This framework is particularly effective in urban environments where continuous data collection is crucial. In industrial settings, IoT infrastructures are tailored to integrate with existing systems, providing precise monitoring of air quality and enabling predictive analytics through machine learning [21]. However, these frameworks face challenges such as the complexity of sensor networks, the need for regular maintenance and calibration, and issues related to data sharing and privacy [12], [22]. Furthermore, the lack of standardized protocols for certain sensor types, such as electronic noses, complicates their widespread adoption [23]. Despite these challenges, IoT technologies have revolutionized environmental monitoring by providing comprehensive, real-time data on pollutants across various ecosystems. The integration of advanced sensors and data processing techniques enables effective pollution management and fosters environmental sustainability. The continued development of IoT frameworks promises to enhance the accuracy and reliability of environmental monitoring systems, supporting informed decision-making and proactive environmental protection strategies.

Predictive modeling in environmental systems, particularly for air pollutants, has seen significant advancements with the application of deep learning techniques such as Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN). These models have been employed for both short-term and long-term forecasting, each with its own set of challenges and limitations, especially in industrial settings. LSTM models are particularly effective for time series prediction due to their ability to capture temporal dependencies in data. For instance, a study by Wu et al. demonstrated the use of an ISSA-LSTM model to predict the Air Quality Index (AQI) with high accuracy, outperforming traditional models in terms of RMSE and R^2 metrics [24]. Similarly, LSTM models have shown superior performance in long-term predictions, effectively incorporating auxiliary data such as forest fire disturbances and climate data to enhance prediction accuracy [25]. These models are adept at handling complex, nonlinear patterns in air quality data, which is crucial for accurate forecasting [26]. CNN models, often integrated with LSTM, are used to extract spatial features from environmental data. Liu et al. developed a CNN-LSTM model that effectively predicted PM2.5 concentrations and its chemical components, achieving a mean R^2 above 0.9 for several pollutants [27]. This integration leverages CNN's feature extraction capabilities and LSTM's temporal modeling strengths, providing robust predictions.

Despite these advancements, there are notable limitations. One major challenge is the handling of spatio-temporal data, which is crucial for accurate predictions in industrial settings. Zhang et al. addressed this by proposing a GATBL-Learning model that combines Graph Attention Networks (GAT) with Bi-LSTM to effectively capture spatio-temporal features, demonstrating improved adaptability and prediction accuracy [28]. However, the complexity of these models can lead to increased computational costs and the need for extensive data preprocessing. Another limitation is the uncertainty in model predictions. Cabaneros and Hughes reviewed methods for handling uncertainty in ANN models, noting that while input uncertainty is often addressed, structural and parameter uncertainties receive less attention. Techniques like bootstrapping and Bayesian methods are underutilized, suggesting a need for more comprehensive approaches to quantify and manage uncertainty in predictive models [29].

In industrial settings, the application of these models is further complicated by the variability in pollutant sources and the need for real-time data processing. Imam et al. emphasized the importance of hyperparameter tuning in machine learning models to achieve high prediction accuracy, which is critical in dynamic environments like industrial areas [39]. In conclusion, while LSTM and CNN models offer promising solutions for air pollutant forecasting, their application in industrial settings is limited by challenges in handling spatio-temporal data, managing model uncertainty, and ensuring computational efficiency. Future research should focus on developing models that can effectively integrate these aspects, potentially through hybrid approaches that combine the strengths of different modeling techniques.

Data privacy in IIoT systems, particularly for pollutants monitoring, presents significant challenges due to the sensitive nature of the data collected and the complex, interconnected environments in which these systems operate. The primary challenges include the risk of data breaches and unauthorized access, which can lead to the exposure of sensitive information and compromise the integrity of the data collected for environmental monitoring [40], [41]. The heterogeneity and scale of IIoT systems further complicate privacy protection, as they involve numerous devices across different domains that must collaborate, raising trust and privacy concerns [41,42]. To address these challenges, several methods have been developed. Encryption techniques are fundamental, providing a basic layer of security by ensuring that data is only accessible to authorized users [43]. Differential privacy is another approach that adds noise to the data, thus protecting individual data points while allowing for aggregate data analysis [42], [43]. Blockchain technology has also been employed to enhance data integrity and privacy by creating immutable records and enabling secure, decentralized data management [41–43].

Among these approaches, FL emerges as a novel solution to these privacy challenges. It allows multiple devices to collaboratively train machine learning models without sharing raw data, thus maintaining data privacy while still benefiting from collective insights [42,44]. This approach is particularly advantageous in IIoT environments, where data from various sources can be used to improve system performance without compromising individual privacy. FL, combined with secure communication protocols like the CKKS cryptosystem, ensures that model parameters are encrypted during the training process, further safeguarding data privacy [44].

FL is a decentralized machine learning approach that enables multiple devices to collaboratively train a model without sharing their raw data, thus preserving privacy and reducing data transfer costs. Originating from the need to protect data privacy while leveraging distributed data sources, FL has become a prominent paradigm in various domains, particularly in IoT systems. In IoT, FL allows devices to train models locally and only share model updates with a central server, which aggregates these updates to improve the global model [45]. In the context of IIoT, FL is utilized to enhance system intelligence while addressing challenges such as data isolation and cybersecurity. An example is the Trusted Device - Model Decision Boundary (TD-MDB) strategy, which is employed in IIoT systems. TD-MDB uses FL to ensure model accuracy by selecting trusted industrial edge devices (IEDs) for training, thereby resisting model poison attacks and improving system revenue [46]. Similarly, federated temporal learning (FTL) is applied for cyber-attack detection in IIoT, effectively extracting temporal features from distributed data to enhance security [47]. FL's application extends to the medical IoT domain, where it is used to optimize privacy and accuracy in chronic disease monitoring. The Adaptive FL for Chronic Disease Prediction (AFL-CDP) framework exemplifies this by integrating secure aggregation techniques to protect sensitive medical data while achieving high predictive accuracy [48]. Additionally, FL has been applied in healthcare to build predictive models using Findable, Accessible, Interoperable, and Reusable (FAIR) health data, demonstrating its capability to maintain privacy across multiple organizations without data sharing [49].

Despite significant advancements in IoT-based environmental monitoring and predictive modeling, a critical gap remains in integrating these technologies with robust data privacy measures, such as FL. Current studies often overlook the importance of data privacy in industrial settings, where sensitive information is frequently generated and transmitted. While FL has shown promise in protecting data privacy by enabling local model training without sharing raw data, its application in industrial IoT systems for environmental monitoring is still largely unexplored. Existing research tends to focus on individual benefits of IoT, machine learning, and FL, rather than investigating their combined potential in real-world, scalable industrial applications.

To address this gap, this study introduces SecureIoT-FL, a novel IoT-based sensing system that integrates real-time environmental monitoring with predictive forecasting capabilities, enhanced by FL to ensure strict data privacy. By deploying SecureIoT-FL across multiple industrial environments, its scalability and effectiveness in monitoring key pollutants such as PM2.5, CO_2, NO_2, and SO_2 are demonstrated. This approach not only enhances predictive accuracy over time but also safeguards sensitive data, setting a new standard for secure and efficient environmental monitoring in industrial settings.

2.1. Study contribution and novelty

This paper makes several significant contributions to the field of industrial environmental monitoring:

- **Innovative SecureIoT-FL Framework:** The study presents SecureIoT-FL, a system that integrates real-time pollutant monitoring with advanced sensors, ensuring continuous and accurate data collection in varied industrial settings.
- **Mathematical Modeling for SecureIoT-FL Framework:** Developed a comprehensive mathematical model for the SecureIoT-FL framework, incorporating factory online status checks, data accumulation dynamics, local model training, secure communication via Message Queuing Telemetry Transport (MQTT) over Amazon Web Services (AWS), and global model aggregation using N-FedAvg (Normalized Federated Averaging)—an enhanced version of the conventional FedAvg algorithm. This optimized version incorporates normalization, balancing data contributions and stabilizing the learning process for secure decentralized learning in industrial IoT environments.
- **Comprehensive Data Collection:** IoT sensing nodes were deployed across three distinct industrial environments—PCBA, CNC machining, and plastic injection—over five months, amassing over 516,000 data points on critical pollutants like PM2.5, CO_2, and SO_2. This diverse dataset offers valuable insights for environmental monitoring.
- **Advanced Predictive Forecasting:** The system uses LSTM and CNN models to predict pollutant levels (e.g., PM2.5, CO_2, O_3) for the next 10, 30, and 60 minutes, enabling proactive environmental management and timely hazard prevention.
- **Enhanced Privacy and Predictive Accuracy via FL:** The SecureIoT-FL framework employs FL to enable local model training, enhancing data privacy and reducing the risk of breaches. This approach also improves AI model accuracy over time, adapting to new data from different environments.
- **Scalability and Security:** SecureIoT-FL demonstrates scalability across various industrial environments, maintaining high accuracy and operational efficiency. By integrating federated learning, it sets a new standard for secure, efficient, and proactive environmental management.

3. Material and methods

3.1. System description and overview

The system developed in this study is designed to enable real-time environmental monitoring and forecasting across multiple industrial environments. It leverages a combination of Internet of Things (IoT) technologies and the novel SecureIoT-FL framework, which integrates FL to enhance both data privacy and predictive accuracy, making it a robust solution for industrial applications.

At the core of the system are three IoT sensing nodes, each strategically deployed in a different industrial setting: a PCBA factory, a CNC machining factory, and a plastic injection factory, as illustrated in Fig. 1. Each node is equipped with a comprehensive array of air quality sensors

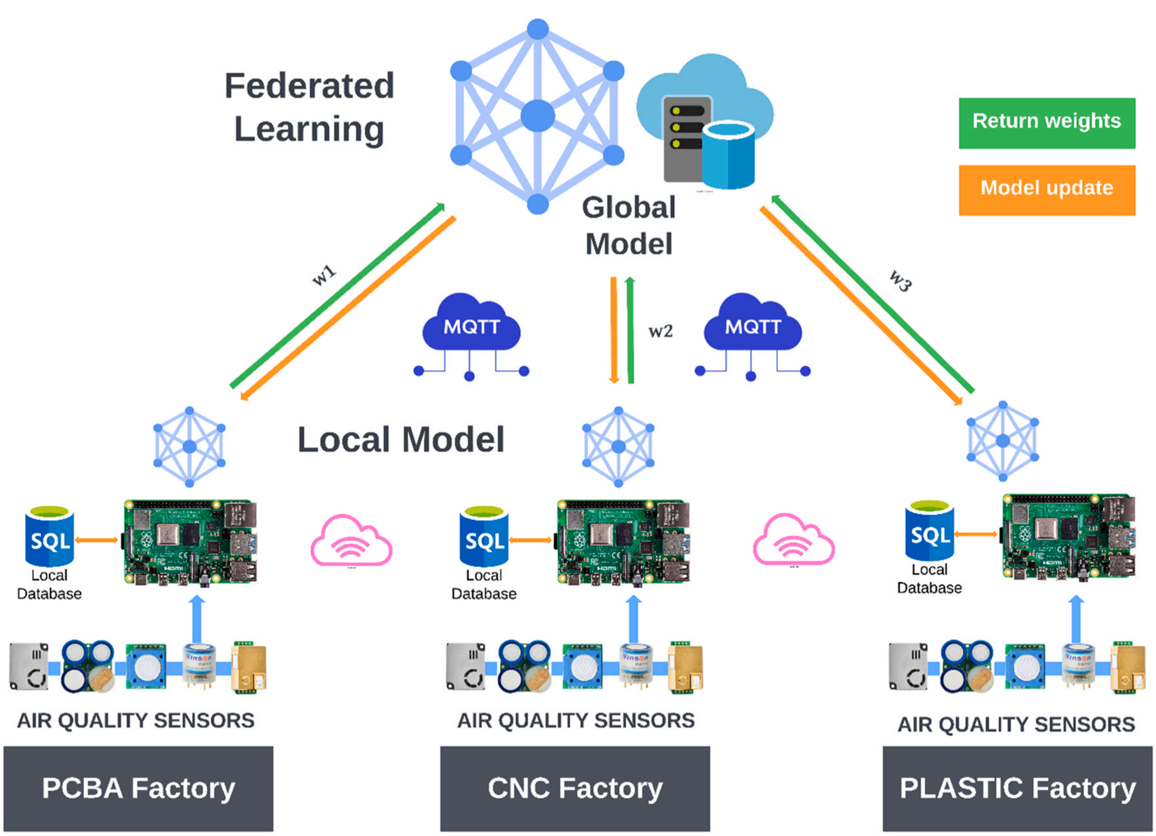


Fig. 1. System Architecture for Real-Time Industrial Environmental Monitoring and Prediction Using the SecureIoT-FL Framework.

capable of monitoring a wide range of pollutants, including PM1.0, PM2.5, PM10, CO_2, VOCs, CH_2O, CO, O_3, NO_2, H_2S, O_2, CH_4, and SO_2. The multi-sensor fusion approach employed in the SecureIoT-FL framework integrates the data from various sensors to improve predictive accuracy significantly. By combining measurements from multiple sensors, the framework can effectively capture the complex interrelationships between different pollutants and environmental factors. This method enhances the robustness of the predictive models, as it allows for a more comprehensive analysis of air quality, leading to better insights and timely interventions in industrial settings. These sensors are connected to a Raspberry Pi 5, which serves as the central processing unit for each node. The data collected by these sensors is continuously stored in a local SQL database, ensuring that all environmental data is captured in real-time and remains securely stored within the local environment. This approach not only ensures data security but also supports continuous and accurate environmental monitoring, essential for proactive industrial management.

A key feature of the proposed system is its innovative SecureIoT-FL framework, which utilizes the Federated Learning (FL) to enhance data privacy while maintaining a high predictive accuracy across all monitored pollutants. Each IoT node is responsible for training a local predictive model using the collected data, though the models designed to forecast the levels of multiple pollutants such as PM2.5, CO_2, O_3, and SO_2. By predicting these pollutant levels, the system provides critical insights into air quality within each industrial environment, enabling for timely interventions and improving of environmental monitoring. Instead of transmitting raw data to a central server—an approach that could pose significant privacy risks—the system transmits only the encrypted model weights from each local node to a global server. This secure transmission is facilitated by the MQTT protocol over AWS, a lightweight and reliable communication method particularly well-suited for IoT environments. Once the global server receives the encrypted model weights from each of the three industrial environments, it aggregates them to update a global predictive model. This updated model is then redistributed back via returning weights to each IoT node, allowing them to refine their local models. This iterative process is repeated over multiple training rounds, conducted biweekly. As more data is collected and more training rounds are completed, the accuracy of both the local and global models improves, benefiting from the diverse data gathered across different industrial settings.

The system's architecture is meticulously designed to ensure not only accurate and real-time monitoring but also the preservation of data privacy throughout the process. By keeping the raw data localized and sharing only encrypted model weights, the system significantly reduces the risk of data breaches—an essential consideration in industrial environments where sensitive operational data is frequently collected. Overall, the SecureIoT-FL framework represents a novel integration of IoT and FL, offering a scalable, secure, and effective solution for environmental monitoring in industrial settings. The system's ability to continuously collect and process data, improving its predictive accuracy over time, ensures adaptability to the specific conditions of different industrial environments, making it a valuable tool for maintaining safe air quality standards and enhancing overall environmental management.

3.2. Design and function of IoT sensing node

This section explores the design and functionality of the IoT sensing node, focusing on the hardware connections and the enclosure design. This discussion provides insights into how the components are integrated to ensure reliable data collection and protection in industrial environments.

3.2.1. Hardware integration and connections

The IoT sensing node in this system is centered around Raspberry Pi 5, which acts as the central processing unit. This node interfaces with a variety of gas sensors and a 4 G Global System for Mobile Communications (GSM) module, all connected through specific communication protocols designed to ensure reliable data collection and transmission in industrial environments. The inclusion of the 4 G GSM module is significant as it enables for real-time data transmission to the cloud, particularly in the remote industrial locations where traditional networking infrastructure may be limited. This capability ensures that the system can operate independently from local network constraints, enhancing its flexibility and scalability for diverse industrial applications. Additionally, the node controls industrial ventilation fans, which can be activated based on the data processed by the system. The sensors integrated into this node are crucial for monitoring various environmental pollutants.

To ensure accurate pollutant measurements, particularly in sensitive pollutants like PM2.5 and volatile organic compounds (VOCs), a straightforward calibration process was implemented for the sensors used in the study. The calibration began with a basic setup utilizing readily available calibration gas mixtures or particulate matter standards sourced from the local suppliers. Each sensor was then placed in a controlled environment, such as a lab chamber, where the concentration of target pollutants could be precisely adjusted and monitored. During the calibration, sensor readings were compared against a reliable reference instrument, such as a high-accuracy PM2.5 monitor or gas chromatograph, to ensure an accurate measurement, with adjusting any discrepancies observed. Additionally, the user-friendly calibration tools, including portable calibration kits containing a well-known concentrations of VOCs, were utilized to facilitate an easy and quick calibration without a need for complex setups.

The ZH09 laser dust sensor is utilized to measure the particulate matter levels in the air, specifically, PM1.0, PM2.5, and PM10 particles. This sensor employs a laser-based detection method that illuminates airborne particles, allowing it to quantify their concentration based on the scattering of light. It communicates with the Raspberry Pi via the UART serial protocol, which has been chosen due to its robustness and reliability—critical factors in industrial environments with stable communication. Similarly, the MICS-2714 gas sensor, used for detecting nitrogen dioxide (NO_2) levels, connects through the UART protocol, ensuring consistent communication even in harsh industrial conditions. For ozone detection, the ZE27-O3 gas sensor is utilized, also relying on the UART protocol to provide reliable data transmission. The system includes the ZE08 gas sensor, which measures formaldehyde (CH_2O) concentrations. This sensor communicates with the Raspberry Pi using the UART serial protocol, a method ideal for short-range, high-speed data exchange, ensuring that critical formaldehyde levels are monitored accurately. Another sensor, the MH-Z19B infrared gas module, is dedicated to measuring carbon dioxide (CO_2) levels. It is connected via the UART protocol, providing crucial data on CO_2 concentrations, which are vital for maintaining safe air quality within the industrial environment. Additionally, the ME3-SO2 electrochemical gas sensor monitors sulfur dioxide (SO_2) levels and is similarly connected through the UART protocol. This sensor plays an essential role in detecting toxic gases that can have significant health impacts. The node also includes the ZCE04B 4-in-1 gas sensor, a versatile multi-gas sensor capable of detecting various gases, including CO, H_2S, O_2 and CH_4. It communicates with the Raspberry Pi using the UART protocol, as shown in Fig. 2, enabling the monitoring of a broad range of pollutants in the industrial setting.

All sensors are connected to the Raspberry Pi through the UART protocol. The only exception is the industrial ventilation system of the factory, which is connected to the Raspberry Pi through the RS485 protocol. This configuration allows the Raspberry Pi to activate the ventilation system when the model predicts that a problem may occur, ensuring proactive management of air quality.

To ensure the collected data is effectively transmitted for further analysis, the node is equipped with a 4 G GSM module, specifically the A7670G. This module is connected to the Raspberry Pi and is responsible for transmitting the collected data to the cloud via the cellular network.

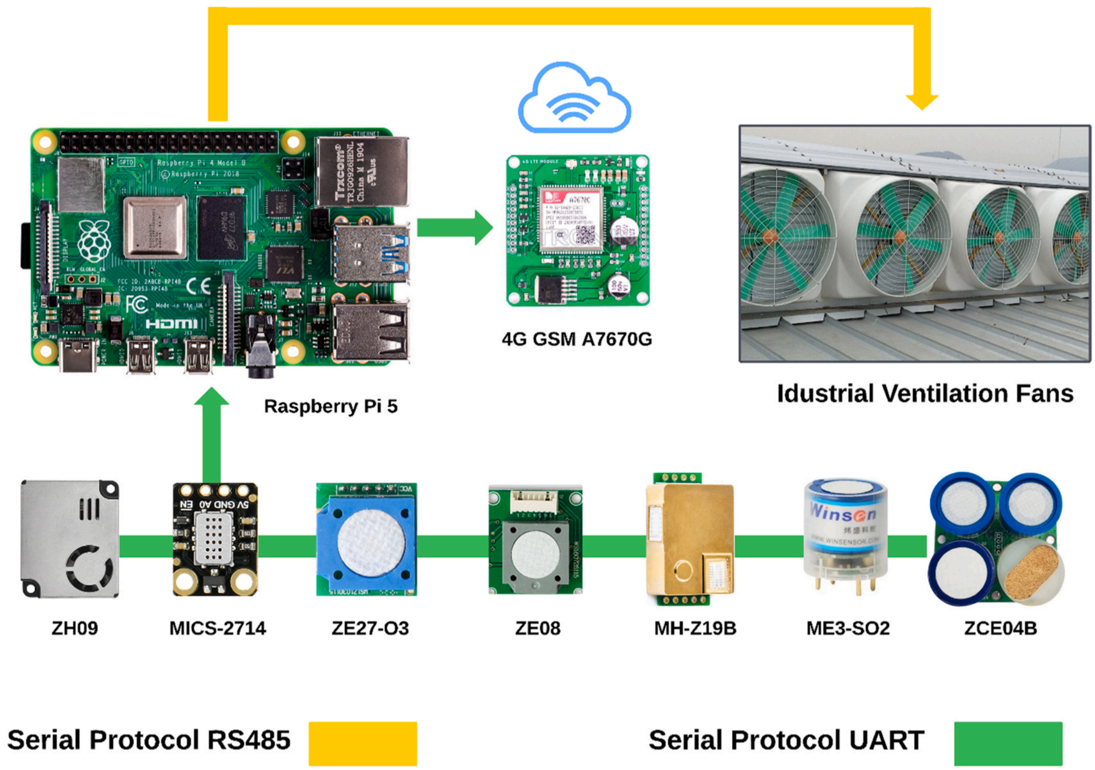


Fig. 2. Hardware Integration and Sensor Connections of the IoT Sensing Node.

This capability is particularly advantageous in remote industrial locations where traditional networking infrastructure may not be available. Furthermore, the Raspberry Pi is integrated with industrial ventilation fans, which it controls based on the real-time data collected from the sensors. When pollutant levels exceed predefined thresholds, the system can automatically activate the fans to ventilate the area, thereby maintaining a safer working environment.

3.2.2. *Energy consumption and power management strategies*

In the design of the IoT sensing node, energy consumption is a critical consideration, particularly for deployments in remote industrial locations where access to a reliable power supply may be limited. Each component's power requirements are directly impacted the overall operational efficiency and sustainability of the system. The central processing unit, a Raspberry Pi 5, typically consumes between 2.5 and 6 watts depending on the workload and processing demands. The power consumption of the individual sensors varies according to their specific functionalities and operational modes. For instance, the ZH09 laser dust sensor has a low power requirement of approximately 0.1 watts, while gas sensors, such as the MICS-2714, consumes around 0.5 watts.

To effectively manage the power consumption and ensure the longevity of the IoT nodes, it is essential to adopt strategies such as duty cycling, where components are powered down during the periods of inactivity, and they are energy-efficient whenever possible. Additionally, an optimizing data transmission intervals and processing tasks can significantly reduce the overall energy footprint of the system. The total power consumption of the IoT sensing unit with all components, ranges from 6.6 to 10.1 watts, depending on the operational state. This total power requirement is crucial for recognizing the energy needs of the system and guiding future enhancements for energy management. (Table 1)

The Table 2 summarizes the power consumption of each component within the IoT sensing node, providing insights into the total energy requirements and future strategies for optimization.

3.2.3. *Enclosure design*

The enclosure for the IoT sensing node is meticulously designed to ensure optimal airflow, allowing the air inside the factory to easily enter and exit the enclosure, as illustrated in Fig. 3. This airflow is critical for accurate sensor readings, ensuring that the sensors measure real-time environmental conditions within the factory. The enclosure is composed of five layers, each engineered to facilitate the entry of air from various angles, enabling it to reach the internal sensors effectively. This design not only protects the sensors, controllers, and GSM module from external damage but also promotes efficient air circulation within the enclosure. By allowing air to flow freely through multiple openings, the enclosure ensures that the sensors are exposed to the actual air quality conditions in the factory, resulting in accurate and reliable measurements.

Furthermore, this design safeguards the internal components while maintaining a continuous flow of air, which is essential for the sensors to detect precise pollutant levels. The strategic placement of ventilation holes allows air to circulate within the enclosure, preventing any potential stagnation and ensuring that the sensors capture true environmental readings.

3.3. *Deployment of IoT nodes*

The IoT nodes were strategically deployed across three distinct factories: a PCBA factory, a CNC machining factory, and a plastic injection factory as shown in Fig. 4, all located in the industrial zone of Ikitelli, Istanbul, and Turkey respectively. These factories were specifically chosen due to the variety of pollutants they generate, making them ideal candidates for comprehensive environmental monitoring. The IoT nodes were installed inside each factory on March 1, 2024, and remained in an operational mode until August 1, 2024. During this period, the nodes are continuously collecting the data on various pollutants in each factory, providing valuable insights into the air quality within these industrial environments. The deployment in these specific factories was crucial for capturing a diverse range of pollutants, thereby allowing for a

Table 1
Comparison of AI Models, Datasets, Applications, and Data Privacy in Industrial and Environmental IoT Systems.

Ref.	AI Model used for Prediction	Dataset	Real Application	Data Privacy
[30]	Stacked bidirectional Long Short-Term Memory (Bi-LSTM)	Publicly available datasets	IoT edge computing system	No, traditional centralized learning
[31]	hybrid model called BO-HyTS	Publicly available datasets	IoT systems deployed in urban areas of India for monitoring and predicting air pollution level	No, traditional centralized learning
[32]	TCN-GCN	Publicly available datasets	Prediction for detection and maintenance in IIoT networks	No, traditional centralized learning
[33]	LSTM	Publicly available datasets	Forecasting pollutant concentrations in a Region Of Interest for various forecast horizons	No, traditional centralized learning
[34]	SPAM	Publicly available datasets	Smart City framework for predictive air quality management	No, traditional centralized learning
[35]	MLADCF	Datasets generated from real-world IoT deployments	Agricultural management	No, traditional centralized learning
[36]	EEMD-WOA-LSTM	Datasets generated from real-world IoT deployments	Remote monitoring capabilities for environmental conditions	No, traditional centralized learning
[37]	Long Short-Term Memory (LSTM)	Publicly available datasets	Forecasting in smart buildings	No, traditional centralized learning
[38]	LSTM	Datasets generated from real-world IoT deployments	Sustainable Air Pollution Monitoring System for Environmental Protection	No, traditional centralized learning
Proposed System	LSTM and CNN	Datasets collected from three different industrial environments over a 5-month period	Industrial environmental monitoring and prediction, with integrated control for ventilation systems	Yes, Federated Learning Framework is used to ensure data privacy by preventing the sharing of datasets

comprehensive analysis of environmental conditions across different industrial processes.

3.3.1. PCBA factory

The PCBA factory was selected because it is a significant source of particulate matter (PM2.5 and PM10) and volatile organic compounds (VOCs), both of which are byproducts of the soldering and assembly processes involved in printed circuit board manufacturing [50]. Monitoring these pollutants is critical for ensuring both worker safety and compliance with environmental regulations. According to international safety standards, the occupational exposure limit for PM2.5 is typically set at 35 μg/m^3 over a 24-hour period, while PM10 is limited to 50 μg/m^3. For VOCs, the permissible exposure limit (PEL) varies depending on the specific compound, but commonly falls within 500 ppb to 1 ppm for compounds like formaldehyde (CH_2O). Elevated VOC levels can lead to a range of health issues, including respiratory problems, headaches, and long-term effects such as organ damage and increasing of cancer risk. Therefore, monitoring VOC concentrations is essential for protecting worker health and maintaining compliance with safety regulations.

Additionally, the monitoring of gases such as hydrogen sulfide (H_2S) and methane (CH_4) is also critical in the PCBA factory environment. The PEL for H_2S is typically set at 10 ppm for a 15-minute exposure, as it poses severe health risks, including respiratory distress and neurological effects. On other side, Methane, even-though it is less toxic, is a potent greenhouse gas with regulatory guidelines to minimize its environmental impact. There is a need for establishing alert thresholds for H_2S and CH_4 gases when the levels approaching to 5 ppm in the former and exceeding 1000 ppm in the latter which ensure proactive management of air quality and worker safety. The alert has been set for PM2.5 PM10 and VOCs at 30 μg/m^3, 45 μg/m^3 and 400–900 ppb, respectively. These thresholds ensure early detection and corrective measures to maintain a safe working environment, highlighting the importance of the continuous monitoring and proactive management of air quality in the PCBA factory.

Table 2
Power Consumption of IoT Sensing Node Components.

Component	Power Consumption (Watts)	Estimated Daily Consumption (Wh)
Raspberry Pi 5	4.25	102
ZH09 Laser Dust Sensor	0.1	2.4
MICS−2714 Gas Sensor	0.5	12
ZE27-O3 Gas Sensor	0.5	12
ZE08 Formaldehyde Sensor	0.5	12
MH-Z19B CO2 Sensor	0.5	12
ME3-SO2 Sensor	0.5	12
ZCE04B 4-in−1 Gas Sensor	0.5	12
4 G GSM Module (A7670G)	1.0	24
Total Consumption	**6.6**	**180**

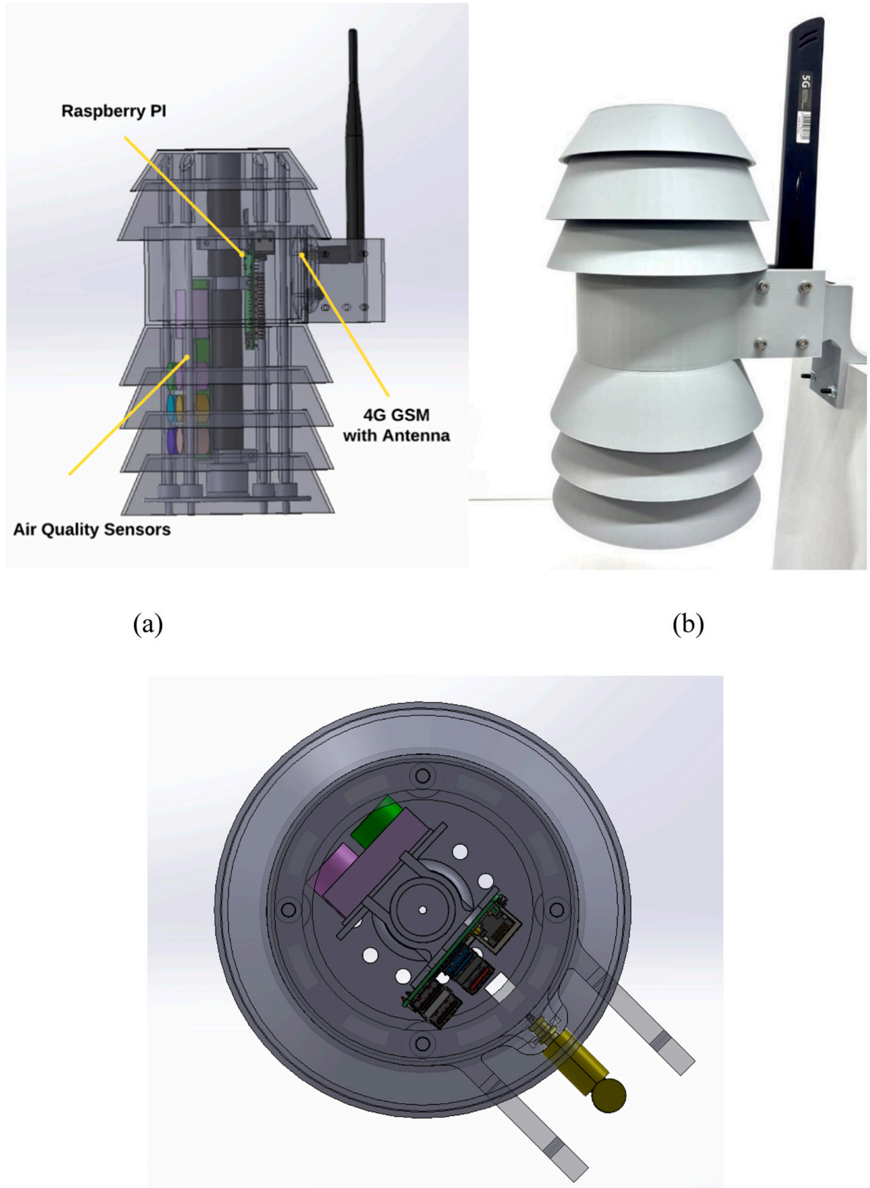
3.3.2. CNC machin factory

The CNC machining factory, on the other hand, was chosen due to its emission of metal particulates (PM10), nitrogen dioxide (NO_2), and ozone (O_3) [51]. These pollutants arise from the cutting, grinding, and finishing operations typical of CNC machining, where high temperatures and material abrasion generate hazardous airborne particles and gases. The safety threshold of PM10 is generally set at 50 μg/m^3 over a 24-hour period, with an alerting level established at 45 μg/m^3 to ensure an early intervention. The Nitrogen dioxide (NO_2) exposure is limited to 200 μg/m^3 as a 1-hour mean, with alerts triggered at 180 μg/m^3. High concentrations of NO_2 can cause respiratory issues and aggravate conditions such as asthma. On the other hand, the ozone (O_3) safety levels are typically set at 100 μg/m^3 in the mean of 8-hour, while alerting thresholds are set at 90 μg/m^3 to provide warnings before unsafe exposure levels are reached. The elevated ozone levels can lead to lung irritation and reduce lung function.

3.3.3. Plastic injection factory

Lastly, the plastic injection factory was included in the study because of its release of sulfur dioxide (SO_2), carbon dioxide (CO_2), and VOCs during the plastic molding and curing processes [52]. These emissions can pose significant health risks and contribute to environmental pollution, making their monitoring essential.

The occupational exposure limit of carbon dioxide (CO_2) is generally



(c)

Fig. 3. IoT Sensing Node Enclosure: (a) Transparent Section Revealing Internal Components, (b) 3D Printed Model, (c) broken out view.

5000 ppm over an 8-hour period, with an alert set at 4500 ppm to ensure an early detection before unsafe levels are reached. The high concentrations of CO_2 can lead to drowsiness and impaired cognitive function. The permissible exposure limit VOCs depends on the specific compound, but a common threshold for many VOCs is around 500 ppb to 1 ppm. Alerting thresholds are typically set between 400 ppb and 900 ppb depending on the specific VOC. These compounds can contribute to a range of health issues, from headaches to more serious long-term effects such as liver or kidney damage.

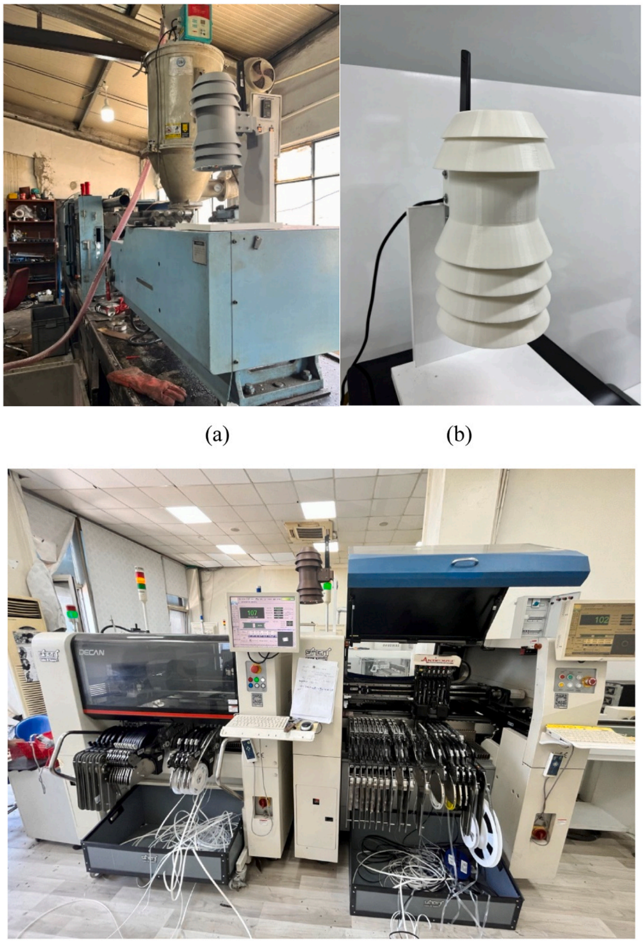
By implementing these alert levels, the monitoring system ensures that any rise in pollutant levels is detected early, allowing for timely precautionary measures to protect both worker health and the environment.

3.4. SecureIoT-FL framework

In this study, the SecureIoT-FL framework was implemented to facilitate secure, decentralized training of machine learning models across multiple industrial environments. This approach was specifically chosen to address the strict data privacy policies of industrial factories, which prohibit sharing datasets with a central server. The framework effectively preserves data privacy while still enabling collaborative model training across diverse industrial settings.

3.4.1. Overview of federated learning

FL is a decentralized machine learning technique that allows individual IoT nodes or devices to train models locally on their respective datasets. Instead of transmitting raw data to a central server, only the model parameters (e.g., weights and biases) are shared [53,54,55]. These parameters are aggregated by the central server to update a global model, which is then redistributed back to the nodes for further training. This iterative process continues over multiple rounds, allowing the model to learn from diverse data across different environments while keeping the underlying data private and secure.



(a) (b)

(c)

Fig. 4. Installation Sites of IoT Nodes: (a) Plastic Injection Factory, (b) CNC Machining Factory, (c) PCBA Factory.

Each IoT node, deployed in distinct industrial environments—specifically a PCBA factory, a CNC machining factory, and a plastic injection factory—was equipped with sensors to monitor various air pollutants such as PM2.5, PM10, CO_2, VOCs and SO_2. These nodes used Long Short-Term Memory (LSTM) and to predict pollutant levels based on locally collected data. In this study, each round of the FL process was set to two weeks. This time frame was chosen to allow sufficient data accumulation at each IoT node before model training. Over these two weeks, each IoT node continuously collected data from its sensors, leading to an increasingly rich dataset for each subsequent round. After every two weeks, the training, aggregation, and redistribution processes described above were executed, resulting in a new global model iteration. The predictive models developed through this FL process were integrated with the industrial ventilation systems. When the model predicted a rise in pollutant levels beyond a certain threshold, the system automatically activated the ventilation to maintain air quality, ensuring a safer environment for workers.

3.4.2. Mathematical modeling of the secureIoT-FL framework

This study implements the SecureIoT-FL framework, specifically designed for industrial IoT environments to enable secure, decentralized model training across multiple factories. Due to the stringent data privacy requirements in these settings, traditional centralized learning methods are insufficient. This approach utilizes the N-FedAvg algorithm, secure communication via MQTT over AWS, and Transport Layer Security (TLS) encryption to ensure data confidentiality while facilitating collaborative learning. The following mathematical formulation details the entire process, from data accumulation to global model updates, within this secure and robust framework.

3.4.2.1. Factory online status check. First, the function $O_i(t)$ that represents whether factory i is online at time t, is shown in Eq. (1).

$$O_i(t) = \begin{cases} 1 \textit{if factoryiis online at timet}, \\ 0 \textit{if factoryiis offline at timet}. \end{cases} \tag{1}$$

For three factories, the overall system status, $S(t)$ is given by:

$$S(t) = O_1(t) \times O_2(t) \times O_3(t) \tag{2}$$

3.4.2.2. Dataset accumulation. The amount of data $n_f(t)$ collected at each factory f increases over time as the IoT nodes continuously gather data from various sensors. The data collection is conducted every 5 minutes, which results in a significant accumulation of data overtime.

Data Collection Per Day:

For each factory, the total number of data points collected per day is calculated as follows:

$$\Delta n_f = 288 \quad \times 12 \quad = \quad 3456 \textit{data points per day per factory} \tag{3}$$

Where:

- 288 represents the number of 5-minute intervals in a 24-hour period ($12 \times 24 = 288$).
- 12 is the number of parameters that were obtained from the 7 sensors in each IoT node.

Cumulative Dataset Size Over Time:

The cumulative dataset size $n_f(t_d)$ at factory f by day d can be expressed as:

$$n_f(t_d) = \quad n_f(t_{d-1}) + \Delta n_f \tag{4}$$

Where:

- $n_f(t_{d-1})$ is the dataset size at the end of the previous day.
- Δn_f is the data accumulated per day, which is 3744 data points.

After 14 days (which corresponds to one round of training in the **SecureIoT-FL** framework), the total accumulated data $n_f(t_{train})$ at factory f is given by:

$$n_f(t_{train}) = n_f(t_0) + 14 \times \quad \Delta n_f \tag{5}$$

Where:

- $n_f(t_0)$ is the initial dataset size at the start of the study (usually 0 if no prior data exists).
- $14 \times \Delta n_f$ represents the total data accumulated over 14 days.

3.4.2.3. Local model training. Each IoT node i trains its local model using the accumulated dataset $D_i(t_{train})$:

$$L_i(w) = \frac{1}{|D_i(t_{train})|} \quad \sum_{(x_j, y_j) \in D_i(t_{train})} l\left(w; x_j, y_j\right) \tag{6}$$

Where:

- $L_i(w)$: This represents the loss function for the local model at IoT node i. The loss function $L_i(w)$ is used to measure how well the model, with weight parameters w, predicts the target values based on the input data.
- $D_i(t_{train})$: This denotes the dataset collected by IoT node i up to the time t_{train}, which is the time of the current training round.

- $|D_i(t_{train})|$: This represents the size of the dataset $D_i(t_{train})$, i.e., the number of data points collected by node i by time t_{train}.
- $l\left(w; x_j, y_j\right)$: This is the loss (or error) for a single data point (x_j, y_j), where x_j is the input data and y_j is the ground-truth label or target value. The function l measures the difference between the model's prediction and the actual value, typically using a loss function like Mean Squared Error (MSE) [56,57].
- Summation $\sum_{(x_j,y_j)\in D_i(t_{train})}$: This indicates that the loss function is summed over all data points (x_j, y_j) in the dataset $D_i(t_{train})$, collected by node i up to the training time t_{train}.

3.4.2.4. Communication and encryption using MQTT over AWS. When the model parameter $w_f(t+t_{train})$ are ready to be sent to the global server, communication is carried out using the MQTT protocol over AWS, and the model parameters are encrypted using TLS. TLS ensures a secure communication by encrypting the data, protecting it from interception or tampering during transmission. It uses a combination of symmetric encryption for data confidentiality and asymmetric encryption (public and private keys) for the secured key exchange. By employing TLS, the transmitted model parameters remain private and unchanged, complying with the industry-standard security protocols [58].

Let Encrypt(w_i, TLS) represents the encryption function:

$$\widehat{w}_f(t+t_{train}) = \mathrm{Encrypt}(w_f(t+t_{train}), TLS) \quad (7)$$

The encrypted model parameters $\widehat{w}_f(t+t_{train})$ are then transmitted to the global server:

$$\widehat{w_f}(t+t_{train}) \rightarrow \text{MQTT over AWS} \rightarrow \text{Global Server} \quad (8)$$

3.4.2.5. Global model aggregation (N-FedAvg) with encrypted weights. Within the SecureIoT-FL framework, the N-FedAvg (Normalized Federated Averaging) algorithm is employed to perform a secure and balance global model aggregation across all participating factories. This enhanced approach built upon the conventional FedAvg by incorporating normalization to assure balancing of contributions from clients with varying data volumes, thereby improving stability and accuracy.

After receiving encrypted weights from each factory, the global server decrypts them using the TLS protocol, which ensures secure transmission. The decryption function can be represented as follows:

$$w_f(t+t_{train}) = \mathrm{Decrypt}\left(\widehat{w}_f(t+t_{train}), TLS\right) \quad (9)$$

Where $w_f(t+t_{train})$ denotes the local model weights for factory f after decryption.

The N-FedAvg algorithm then updates the global model within the SecureIoT-FL framework using the following formula:

$$w_G(t+t_{round}) = \frac{1}{N}\sum_{f=1}^{N}\frac{n_f(t_{train})}{n_{total}(t_{train})} w_f(t+t_{train}) \quad (10)$$

Where:

- f: Represents each factory (client) in the SecureIoT-FL system.
- w_f: The local weights of the model trained at factory f.
- $n_f(t_{train})$: The number of data points collected by factory f at time t_{train}.
- $n_{total}(t_{train})$: The cumulative number of data points collected across all factories by time t_{train}.
- N: The total number of participating factories.

The Key Advantages of N-FedAvg within SecureIoT-FL are discussed as follows:

1. **Balanced aggregation across clients:**
   By dividing the sum of all weighted local models by N, the N-FedAvg algorithm ensures that the update of global model accurately reflects the contributions of all clients, providing a balanced aggregation as core part in the **SecureIoT-FL** framework's reliability.
2. **Normalization for data imbalance:**
   The N-FedAvg algorithm applies a normalization to each client's contribution based on the data volume, maintaining consistent model performance within **SecureIoT-FL.** This feature enables to deal well with client data imbalances and stabilizes the global model update, making it less sensitive to fluctuations in individual client data volumes.
3. **Consistency in model weights:**

By scaling each local model's contribution, N-FedAvg ensures that the global model weights remain within a stable range which is crucial for maintaining the model accuracy across rounds. Without such normalization, the global model in the SecureIoT-FL framework may become skewed if some factories contribute significantly with larger datasets than others, leading to potential inaccuracies.

3.4.2.6. Global model update and redistribution. The updated global model $w_G(t+t_{round})$ is then encrypted before being sent back to the factories:

$$\widehat{w}_G(t+t_{round}) = \mathrm{Encrypt}(w_G(t+t_{round}), TLS) \quad (11)$$

The encrypted global model is then distributed to all factories using MQTT:

$$\widehat{w}_G(t+T_{round}) \rightarrow \text{MQTT over AWS} \quad (12)$$

Upon receiving the global model, each factory decrypts it:

$$w_G(t+t_{round}) = \mathrm{Decrypt}(\widehat{w}_G(t+t_{round}), TLS) \quad (13)$$

3.4.2.7. Continuous data accumulation and training. For subsequent kth rounds, the accumulated dataset size increases:

$$n_f(t_{round,k}) = n_f(t_0) + 14 \times k \times \Delta n_f \quad (14)$$

The global model after the kth round is updated as:

$$w_G\left(t_{round,k+1}\right) = \frac{1}{N}\sum_{f=1}^{N}\frac{n_f\left(t_{round,k}\right)}{n_{total}\left(t_{round,k}\right)} w_f\left(t_{round,k}\right) \quad (15)$$

3.4.2.8. Condition for global model update. The model is only updated if all factories are successfully participated:

$$w_G\left(t_{round,k+1}\right) = \begin{cases} \mathrm{Average}(w_f | \mathscr{O}_f(t_{round,k}) = 1) \; if S(t_{round,k}) = 1, \\ w_G\left(t_{round,k}\right) otherwise. \end{cases} \quad (16)$$

Where:

- $w_G\left(t_{round,k+1}\right)$: Represents the global model's weights after the completion of round $k+1$. This is the updated set of weights that will be used by all participating IoT nodes in the next round.
- $w_G\left(t_{round,k}\right)$: Denotes the global model's weights before the start of round $k+1$. If the conditions for updating the model are not met, these weights are carried over to the next round without any modification.
- Average($w_f|\mathscr{O}_f\left(t_{round,k}\right) = 1$): This term calculates the average of the local model weights w_f from all IoT nodes (factories) that successfully completed their local training during round k.
- $\mathscr{O}_f\left(t_{round,k}\right) = 1$ indicates that factory f is online during the rounding of k.

- $S(t_{round,k})$: A status indicator function that evaluates whether all factories successfully participated in the training round k.
- If $S(t_{round,k}) = 1$, it indicates that all factories are online and have successfully completed their training for that round.

3.4.3. Key benefits of the secureIoT-FL framework

- Data Privacy Compliance: The SecureIoT-FL framework was essential in maintaining data privacy, especially given the industrial factories' reluctance to share their datasets. By keeping data local and sharing only encrypted model parameters, the framework effectively addressed privacy concerns while enabling collaborative learning.
- Enhanced Predictive Accuracy: Collaborative learning across multiple industrial environments allowed the global model to improve its predictive accuracy with each round, leveraging the diverse datasets from different industrial contexts.
- Scalability and Flexibility: The SecureIoT-FL framework demonstrated scalability, allowing for the seamless integration of additional IoT nodes without compromising privacy or performance. New nodes could easily join the existing FL process.
- Real-Time Environmental Control: By integrating predictive models with ventilation systems, the framework enabled real-time environmental control. The system could predict potential air quality issues and activate ventilation preemptively, ensuring a safer industrial workspace.

3.5. Predictive modeling

This study employed Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN) to forecast pollutant levels in industrial environments. Both models were integrated to the SecureIoT-FL framework, providing accurate predictions based on real-time data collected by the IoT nodes.

3.5.1. Long short-term memory (LSTM) model

LSTM networks are a type of recurrent neural network (RNN) particularly well-suited for time series forecasting due to their ability to capture long-term dependencies in data. The LSTM model in this study was designed to predict future pollutant levels based on historical data.

- Input Features: The input to the LSTM model included time-series data of pollutant concentrations such as PM2.5, CO_2, O_3, and SO_2.

- Model Architecture: The LSTM model consisted of an input layer, one or more LSTM layers, and a fully connected output layer. The core equations governing LSTM cell operations are:

➤ Forget Gate: Determines what information to discard from the cell state.

$$f_t = \sigma(W_f \bullet [h_{t-1}, x_t] + b_f) \tag{17}$$

Where:

- f_t: Forget gate activation at time t.
- σ: Sigmoid activation function.
- W_f: Weight matrix of the forget gate.
- h_{t-1}: Hidden state from the previous time step.
- x_t: Input data at time t.
- b_f: Bias of the forget gate.

➤ Input Gate: Decides which values from the input will update the cell state.

$$i_t = \sigma(W_i \bullet [h_{t-1}, x_t] + b_i) \tag{18}$$

$$\widetilde{C}_t = \tanh(W_C \bullet [h_{t-1}, x_t] + b_C) \tag{19}$$

Where:

- i_t: Input gate activation at time t.
- W_i: Weight matrix of the input gate.
- W_C: Weight matrix of candidate cell state.
- b_i, b_C: Biases of the input gate and candidate cell state, respectively.
- tanh: Hyperbolic tangent activation function.

➤ Cell State Update: Updates the cell state C_t with new information.

$$C_t = f_t \bullet C_{t-1} + i_t \bullet \widetilde{C}_t \tag{20}$$

Where:

- C_t: Updated cell state at time t.
- C_{t-1}: Cell state from the previous time step.

➤ Output Gate: Determines the output based on the cell state.

$$o_t = \sigma(W_o \bullet [h_{t-1}, x_t] + b_o) \tag{21}$$

$$h_t = o_t \bullet \tanh(C_t) \tag{22}$$

Where:

- o_t: Output gate activation at time t.
- W_o: Weight matrix of the output gate.
- h_t: Hidden state at time t, which is also the output of the LSTM cell.
- b_o: Bias of the output gate.

- **Training process:** The LSTM model was trained using the Adam optimizer and mean squared error (MSE) as the loss function [59]. The training was performed locally at each IoT node, with model weights being updated and aggregated across the network through the **SecureIoT-FL** framework.

3.5.2. Convolutional neural network (CNN) model

CNNs are typically used for spatial data but have also been applied effectively to time-series data by treating the temporal sequence as a spatial structure. The CNN model in this study was used to capture the spatio-temporal relationships in pollutant concentration data.

- **Input features:** The input to the CNN model included the same pollutant concentrations and environmental variables used in the LSTM model. However, the data was reshaped into a format suitable for convolutional layers.
- **Model architecture:** The CNN model consisted of an input layer, several convolutional layers, pooling layers, and a fully connected output layer. The key operation in the CNN is the convolution, defined as:

$$S(i,j) = (X^*W)(i,j) = \sum_m \sum_n X(i+m, j+n) \bullet W(m,n) \tag{23}$$

where $S(i,j)$ is the feature map, X is the input matrix, and W is the filter (or kernel) applied to the input.

- **Pooling layers:** These layers were used to reduce the spatial dimensions of the feature maps, summarized by:

$$P(i,j) = \max(S(i,j)) \tag{24}$$

The final layers consisted of fully connected layers to interpret the spatial features extracted by the convolutions.

- **Training process:** Similar to the LSTM model, the CNN was trained using the Adam optimizer and MSE loss function. Training was conducted locally, with model weights shared and aggregated using the FL process.

3.5.3. Uncertainty quantification in the predictive modeling

To improve the reliability of pollutant forecasts and manage the prediction uncertainties inheritance in the real-world industrial environments, the uncertainty quantification methods were incorporated into the SecureIoT-FL framework. An approach, called Monte Carlo dropout, is used in this study which introduces a randomness during the model inference, enabling the model to generate multiple predictions for the same input and quantify the uncertainty by observing the variance among these predictions [60]. By applying the dropout at the inference stage, each forward pass can be considered as a sample from a Bayesian posterior distribution, allowing the framework to estimate the uncertainty in the forecasted values.

Bayesian inference methods also contribute to the model's robustness by integrating the prior knowledge about pollutant patterns and updating predictions based on the observed data [61]. These methods provide probabilistic outputs that allow stakeholders to assess confidence levels for each prediction, particularly when the pollutant levels approach critical thresholds. This probabilistic approach enhances the decision-making process, ensuring that the safety interventions (e.g., ventilation activation) are reliable and interpretable predictions. In other words, Such uncertainty quantification techniques supports the goal of providing confidence intervals for each prediction, where the stakeholders can use them to gauge the accuracy and dependability of pollutant forecasts in the real-time industrial settings.

3.6. Accuracy and performance metrics

To evaluate the predictive accuracy and performance of the LSTM and CNN models within the **SecureIoT-FL** framework, several key metrics were employed: Accuracy, Mean Absolute Percentage Error (MAPE), Coefficient of Determination (R^2), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). These metrics provided a comprehensive assessment of how well the models predicted pollutant levels in real-time.

3.6.1. Accuracy

Accuracy is defined as the degree to which the model's predictions are close to the actual values. It measures the proportion of predictions that fall within a specified error range, indicating the model's reliability for practical use.

The formula for Accuracy is as follows:

$$Accuracy = \frac{Number\ of\ Accurate\ Predictions}{Total\ Number\ of\ Predictions} \times \quad 100\% \tag{25}$$

In this study, a prediction is considered accurate if it falls within a predefined acceptable error margin from the actual value, i. e., a prediction is called "accurate" if the absolute percentage error is within ± 10 % of the actual pollutant level. This threshold provides a reasonable balancing between the precision and practical usability, ensuring that the model predictions are sufficiently close to the real-world conditions.

3.6.2. Mean absolute percentage error (MAPE)

MAPE measures the accuracy of the model by calculating the percentage error between the predicted and actual values. It is defined as:

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y}_i}{y_i}\right| \times 100\% \tag{26}$$

where y_i is the actual value, $\widehat{y}_i$ is the predicted value, and n is the number of observations. MAPE provides an intuitive understanding of the prediction error relative to the magnitude of the actual values.

3.6.3. Coefficient of determination (R^2)

R^2 measures the proportion of variance in the dependent variable that is predictable from the independent variables. It is defined as:

$$R^2 = 1 - \quad \frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{27}$$

where $\overline{y}$ is the mean of the actual values. R^2ranges from 0 to 1, with higher values indicating better model performance. An R^2 value closer to 1 suggests that the model explains a large proportion of the variance in the data.

3.6.4. Root mean squared error (RMSE)

RMSE is a standard measure of the differences between predicted and actual values. It is particularly sensitive to outliers, making it useful for assessing the overall error magnitude:

$$RMSE = \sqrt{\quad \frac{1}{n} \quad \sum_{i=1}^{n}(y_i - \widehat{y}_i)^2} \tag{28}$$

RMSE provides a direct interpretation of the model's error in the same units as the original data, making it easier to assess the practical impact of the error.

3.6.5. Mean absolute error (MAE)

MAE measures the average magnitude of errors in a set of predictions, without considering their direction. It is defined as:

$$MAE = \frac{1}{n} \quad \sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{29}$$

MAE gives a straightforward indication of the average error in the predictions, providing a measure of how far the predicted values are from the actual values.

4. Results and discussion

This section presents and analyzes the outcomes of the study, focusing on the effectiveness of the SecureIoT-FL framework in monitoring and predicting environmental conditions across diverse industrial settings. This section includes a detailed examination of the collected data, visualization of key pollutant trends, and an in-depth analysis of the performance of the predictive models employed—namely, Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN). The performance of these models is validated against several metrics to assess their accuracy and predictive capabilities in forecasting pollutant levels. Through this comprehensive evaluation, the framework's ability to provide reliable, real-time environmental monitoring while maintaining data privacy and security is demonstrated.

4.1. Data collection and storage

The IoT sensing nodes, installed across three distinct factories—PCBA, CNC machining, and plastic injections, were tasked with collecting environmental data every five minutes. Each node was equipped with a suite of 7 sensors designed to monitor a variety of pollutants, including PM2.5, PM10, CO_2, VOCs, CH_2O, CO, O_3, NO_2, H_2S, O_2, CH_4 and SO_2. The data collected by these sensors was stored locally at each factory, ensuring that sensitive information remained secure and was not shared externally.

The data collection began on March 1, 2024, and continued uninterrupted for a period of five months until August 1, 2024. Over this duration, the IoT nodes gathered extensive data for each pollutant, with each pollutant contributing 288 data points per day, resulting in a significant accumulation over time. Given the frequency of data collection and the number of pollutants being monitored, each pollutant generated approximately 44,352 samples over the five-month period. When considering all 12 pollutants together, the total number of samples collected by each IoT node amounted to 576,576 samples, as shown in Table 3. This extensive dataset provided a comprehensive view of the environmental conditions within each factory.

To facilitate continuous model improvement, the predictive models were retrained every two weeks using the newly accumulated data. This biweekly training cycle allowed the models to adapt to changing environmental conditions and progressively enhance their predictive accuracy. The table below illustrates the accumulated data collected at each two-week interval, highlighting the steady increase in data volume and its integration into the model training process:

4.1.1. Data quality management and handling missing data

During the data collection period, missing or corrupted data points can arise due to sensor malfunctions or intermittent network connectivity. To mitigate the impact of these issues on model performance, particularly during federated learning rounds, the framework implements data pre-processing steps to identify and address anomalies. Specifically, missing data is imputed using temporal interpolation methods when feasible, or alternatively, affected data segments are excluded from model training if the data gaps are too large. Additionally, data integrity checks are performed at each IoT node before data is shared with the central server, ensuring that only reliable data informs the federated learning process. This approach helps maintain data quality and model robustness, even in the presence of occasional data inconsistencies.

4.2. Data analysis

The analysis of the datasets collected from the PCBA, CNC machining, and plastic injection factories provides insights into the environmental parameters over a period of five months. Here is a consolidated table summarizing the key statistical measures—Count, Mean, Standard Deviation (STD), Minimum (Min), 25th Percentile (25 %), Median (50 %), 75th Percentile (75 %), and Maximum (Max)—for each factory as shown in Table 4.

From the combined data, it is observed that the particulate matter (PM1.0, PM2.5, and PM10) levels vary across the factories, with PCBA showing a mean PM1.0 of 19.8 μg/m^3, while the plastic injection and CNC machining factories display slightly different profiles in their PM levels. The variability in particulate matter, reflected in the standard

Table 3
Accumulated Data Collected by IoT Sensors Every Two Weeks Over a Five-Month Period.

Round	Time Period	Days	Samples per Pollutant	Total Samples Across All Sensors
1	Weeks 1–2 (1/3–14/3)	14	4032	48,384
2	Weeks 3–4 (15/3–28/3)	28	8064	96,768
3	Weeks 5–6 (29/3–11/4)	42	12,096	145,152
4	Weeks 7–8 (12/4–25/4)	56	16,128	193,536
5	Weeks 9–10 (26/4–9/5)	70	20,160	241,920
6	Weeks 11–12 (10/5–23/5)	84	24,192	290,304
7	Weeks 13–14 (24/5–6/6)	98	28,224	338,688
8	Weeks 15–16 (7/6–20/6)	112	32,256	387,072
9	Weeks 17–18 (21/6–4/7)	126	36,288	435,456
10	Weeks 19–20 (5/7–18/7)	140	40,320	483,840
11	**Weeks 21-22 (19/7 - 1/8)**	**154**	**44,352**	**532,224**

Table 4
Combined Factory Environmental Statistics for PCBA, CNC Machining, and Plastic Injection Factories.

Factories	Statistic	PM1.0 (μg/m^3)	PM2.5 (μg/m^3)	PM10 (μg/m^3)	CO_2 (ppm)	VOC (ppb)	CH_2O (ppb)	CO (ppm)	NO_2 (ppb)	H_2S (ppm)	O_2 (%)	CH_4 (ppm)	SO_2 (ppm)
PCBA	count	44352	44352	44352	44352	44352	44352	44352	44352	44352	44352	44352	44352
	mean	19.8	24.6	27.0	541.8	2.9	0.2	0.5	0.2	0	20.9	0	0
	std	2.8	3.3	3.7	145.5	0.4	0.1	0	0.1	0	0	0	0
	min	8.1	11.5	9.3	178.8	0	0	0.5	0	0	20.8	0	0
	25 %	17.9	22.3	24.5	444.5	3.0	0.1	0.5	0.2	0	20.9	0	0
	50 %	19.8	24.6	27.0	542.2	3.0	0.2	0.5	0.2	0	20.9	0	0
	75 %	21.7	26.8	29.5	639.9	3.0	0.2	0.5	0.3	0	20.9	0	0
	max	30.8	37.6	41.4	1235.7	3.0	0.5	0.5	0.5	0	21.0	0	0
CNC Machining	count	44352	44352	44352	44352	44352	44352	44352	44352	44352	44352	44352	44352
	mean	50.0	60.1	70.2	500.4	1.0	0.1	1.0	0.1	0	20.0	0.5	12.0
	std	20.0	24.7	29.8	50.2	0.5	0	0.5	0	0	0.2	0.1	1.0
	min	0	0	0	292.1	0	0	0	0	0	19.2	0.1	7.7
	25 %	36.4	43.2	49.9	466.5	0.7	0.1	0.7	0.1	0	19.9	0.4	11.3
	50 %	50.0	59.9	70.2	500.7	1.0	0.1	1.0	0.1	0	20.0	0.5	12.0
	75 %	63.4	76.8	90.3	534.4	1.3	0.1	1.3	0.1	0	20.1	0.6	12.7
	max	131.9	164.3	193.1	715.7	3.0	0.4	2.9	0.3	0	20.8	1.0	15.9
Plastic Injection	count	44352	44352	44352	44352	44352	44352	44352	44352	44352	44352	44352	44352
	mean	95.3	112.3	123.0	528.9	0.8	0.5	0	0.2	0	20.0	0	15.3
	std	48.3	53.3	56.9	144.6	1.2	0.4	0.4	0.7	0	0	0	174.9
	min	0	0	0	0	0	0	0	0	0	20.0	0	10.0
	25 %	56.0	70.0	78.0	433.0	0	0.4	0	0	0	20.0	0	11.8
	50 %	88.0	109.0	121.0	477.0	0	0.5	0	0	0	20.0	0	11.8
	75 %	124.0	139.0	150.0	578.0	2.0	0.6	0	0.2	0	20.0	0	11.9
	max	703.0	888.0	1024.0	1611.0	3.0	6.2	8.0	10.0	0	21.0	0	15.1

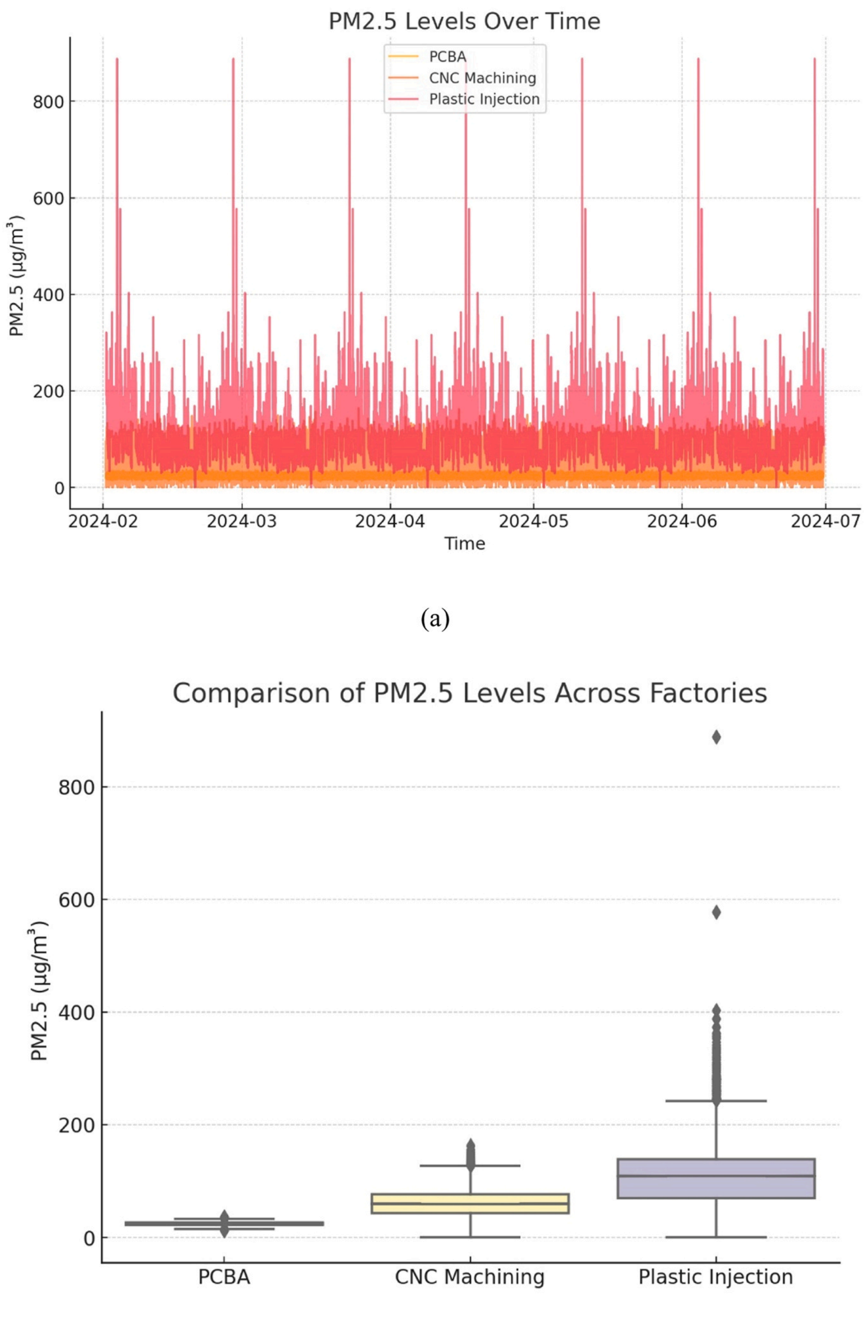
deviations, suggests differences in processes or environmental controls between the factories. The CO_2 levels, another key parameter, average around 541 ppm in the PCBA factory, indicate a typical range for indoor environments with active industrial processes. These values are within standard comfort ranges for industrial settings, suggesting effective climate control measures are in place. However, the temperature shows a slight variability (STD ~1.3), which could be attributed to differences in operation times or ventilation effectiveness.

Volatile Organic Compounds (VOC) grades and concentrations of gases such as CO, O_3, and NO_2 also highlight the differences in emissions between the factories. For instance, the PCBA factory recorded a VOC grade average of 2.9 ppb (parts per billion), indicating higher organic emissions, which could be due to the nature of materials and processes used in PCB assembly compared to CNC machining or plastic injection molding. The gases like H_2S and CH_4 were consistently low or negligible across all datasets, aligning with expected safety and environmental standards for such industrial operations.

Overall, the data reflects the specific environmental conditions prevalent in each factory type. Differences in particulate matter, gas concentrations, and climate conditions underscore the importance of tailored environmental controls to manage air quality and maintain safe working conditions in diverse manufacturing environments. This analysis will help in identifying areas for improvement, such as enhancing ventilation or adjusting processes to minimize harmful emissions.

4.3. Data visualization

The Time Series Plot of PM2.5 Levels provides a detailed



(a)

(b)

Fig. 5. data visualization of PM2.5: (a) Time Series Plot of PM2.5 Levels, (b) Box Plot Comparing PM2.5 Levels Across Factories.

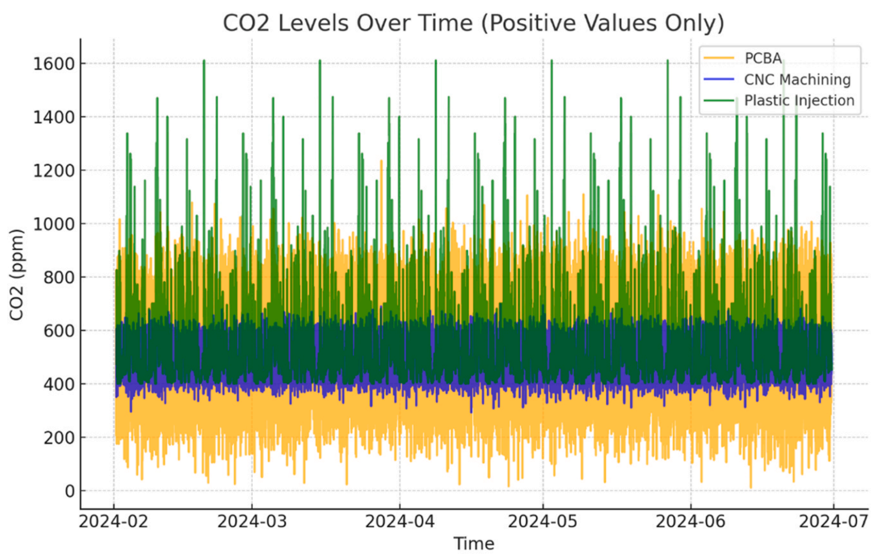
visualization of the PM2.5 concentrations over time for the PCBA, CNC machining, and plastic injection factories as shown in Fig. 5. This plot reveals the dynamic nature of PM2.5 levels in each factory, with visible fluctuations that reflect daily operational changes, shifts in production intensity, or variations in environmental controls. Over the five-month period, each factory exhibits distinct patterns of PM2.5 emissions, suggesting that the processes or control measures in place impact air quality differently across the three types of manufacturing. For example, the CNC machining factory shows relatively consistent PM2.5 levels with minor variations, indicating stable operations or effective pollutant management strategies. In contrast, the plastic injection factory demonstrates more pronounced peaks and troughs, which could be attributed to changes in production cycles or specific manufacturing processes that temporarily elevate particulate emissions. The PCBA factory displays a moderate level of fluctuation, potentially indicating intermediate levels of production changes or varying effectiveness of emission control measures.

The Time Series Plot of CO_2 Levels reveals trends in emissions over time for each factory by focusing exclusively on positive CO_2 values, which ensures a more precise reflection of the environmental conditions as shown in Fig. 6. The PCBA factory is characterized by significant variability and occasional peaks in CO_2 levels, suggesting larger fluctuations in emissions. On the other hand, the CNC machining and plastic injection factories maintain relatively steady and lower CO_2 levels, indicating a more consistent impact on air quality.

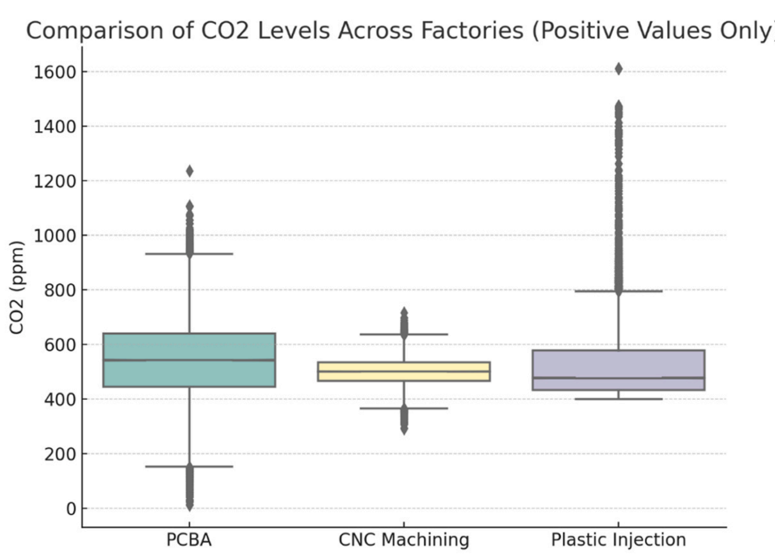
In the Box Plot comparing CO_2 levels across factories, the representation of positive values highlights these observations. The PCBA factory demonstrates a higher median CO_2 level along with a wider interquartile range, signifying greater variability in its emissions. In contrast, the CNC machining and plastic injection factories have narrower distributions and lower median CO_2 levels, which reflect more effective emission control and more consistent operational performance.

4.4. Performance validation discussion

The framework analysis in this section paves the stage for evaluating



(a)



(b)

Fig. 6. data visualization of CO2: (a) Time Series Plot of CO2 Levels, (b) Box Plot Comparing CO_2 Levels Across Factories.

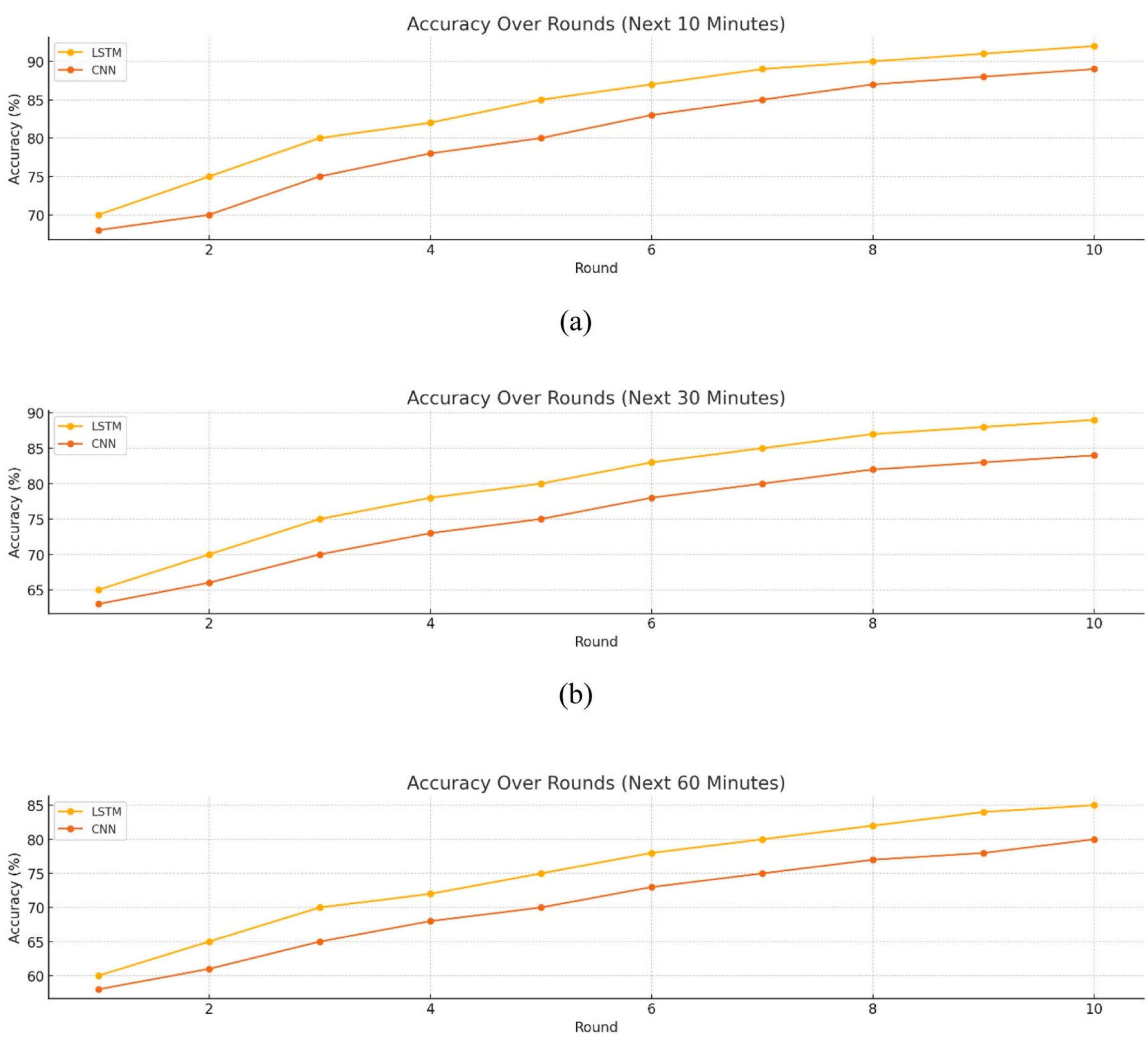
the predictive performance of both LSTM and CNN models over 10 rounds of training, covering predictions for the next 10, 30, and 60 minutes. The key performance metrics, including Accuracy, Mean Absolute Percentage Error (MAPE), R^2, Root Mean Square Error (RMSE), and Mean Absolute Error (MAE), are used to assess the effectiveness of the SecureIoT-FL framework in improving the model performance through FL. Additionally, as the number of IoT nodes increases and datasets expand, the computational efficiency and scalability of the framework is critically analyzed to assure a sustained performance across diverse industrial environments.

4.4.1. Accuracy improvements across prediction horizons

This section evaluates the performance improvements of the LSTM and CNN models across different prediction horizons: 10, 30, and 60 minutes. As shown in Fig. 7, which illustrates the accuracy line chart changes during each round, the SecureIoT-FL framework's effectiveness in enhancing predictive accuracy is highlighted by the progressive gains observed over 10 rounds of training. These improvements are driven by the accumulation of new data with each round, allowing the models to refine their understanding of the temporal dynamics of pollutant levels in diverse industrial environments. The following analysis provides insights into how the models adapt and improve as they are exposed to increasingly comprehensive datasets, demonstrating the power of FL in a real-world industrial IoT context.

As demonstrated in Fig. 7, the accuracy trends of LSTM and CNN models show a progressive improvement over 10 rounds, with increasing accuracy for predictions at 10, 30, and 60-minute intervals. The accuracy in this context refers to the percentage of the corrected predictions compared to the actual pollutant levels. This metric, along with the others, helps to validate the model's predictive capabilities as the rounds of training are continuously performed.

Next 10 Minutes: For the 10-minute prediction horizon, both LSTM and CNN models exhibited notable improvements in key performance metrics, including MAPE, RMSE, and R^2, across the 10 rounds of training, as shown in Table 5. The LSTM model's MAPE decreased significantly from 10.0 % in Round 1–4.2 % in Round 10, reflecting the enhanced accuracy in short-term predictions. Similarly, the CNN model's MAPE dropped from 10.5 % to 4.5 % over the same period, highlighting the models' growing accuracy in estimating pollutant levels, which is crucial for timely interventions in industrial environments. As illustrated in Fig. 8(b), the loss (MAPE) for both models shows a consistent decline over the rounds, demonstrating the effectiveness of the iterative training process in refining model performance. RMSE values followed a similar downward trend, with the LSTM model's RMSE reducing from 0.15 to 0.04, and the CNN model's RMSE from 0.16 to 0.05. The consistent decline in RMSE underscores the models' ability



(c)

Fig. 7. Accuracy Trends Across Rounds: (a) Predictions for next 10 minutes, (b) Predictions for next 30 minutes, (c) Predictions for next one hour.

Table 5
Prediction for the Next 10 Minutes.

Round	Model	Accuracy (%)	MAPE (%)	R^2	RMSE	MAE
Round 1	LSTM	70.00	10.00	0.70	0.15	0.12
	CNN	68.00	10.50	0.68	0.16	0.13
Round 2	LSTM	75.00	8.50	0.75	0.12	0.10
	CNN	70.00	9.00	0.72	0.13	0.11
Round 3	LSTM	80.00	7.00	0.80	0.10	0.08
	CNN	75.00	8.00	0.76	0.11	0.09
Round 4	LSTM	82.00	6.00	0.83	0.09	0.07
	CNN	78.00	7.00	0.78	0.10	0.08
Round 5	LSTM	85.00	5.00	0.86	0.08	0.06
	CNN	80.00	6.00	0.81	0.09	0.07
Round 6	LSTM	87.00	4.80	0.88	0.07	0.06
	CNN	83.00	5.50	0.83	0.08	0.06
Round 7	LSTM	89.00	4.60	0.90	0.06	0.05
	CNN	85.00	5.20	0.85	0.07	0.06
Round 8	LSTM	90.00	4.50	0.92	0.05	0.05
	CNN	87.00	5.00	0.87	0.06	0.05
Round 9	LSTM	91.00	4.40	0.93	0.04	0.04
	CNN	88.00	4.80	0.88	0.05	0.04
Round 10	LSTM	**92.00**	**4.20**	**0.94**	**0.04**	**0.04**
	CNN	**89.00**	**4.50**	**0.90**	**0.05**	**0.04**

to minimize forecast errors if more training data can be included, contributing to a more reliable short-term prediction.

In addition to reductions in MAPE and RMSE, the Coefficient of Determination (R^2), which indicates the proportion of variance explained by the models, showed significant improvement for both LSTM and CNN across the rounds. LSTM's R^2 increased from 0.70 in Round 1–0.94 in Round 10, demonstrating a substantial enhancement in the model's capacity to capture the underlying patterns in pollutant data. Similarly, the CNN model showed an increase in R^2 from 0.68 to 0.90, reflecting improved explanatory power and a deeper understanding of pollutant behavior in the short term. The results for the 10-minute prediction horizon clearly demonstrate the effectiveness of the SecureIoT-FL framework in boosting predictive performance through continuous data accumulation and model refinement, as evidenced by the declining loss values depicted in Fig. 8.

Next 30 Minutes: For the 30-minute prediction horizon, both LSTM and CNN models demonstrated significant advancements in their performance metrics throughout the 10 training rounds. The LSTM model's accuracy increased from 65 % in Round 1–89 % in Round 10, while CNN's accuracy rose from 63 % to 84 %, indicating that the models became progressively better at capturing medium-term pollutant variations. This improvement is closely tied to the continuous accumulation of new data in each round, allowing the models to refine their predictive capabilities with each iteration.

As the models adapted to the increased data, notable reductions in MAPE and RMSE were observed, underscoring the enhanced accuracy of their medium-term forecasts. The LSTM model's MAPE decreased from 12.0 % in Round 1–5.2 % by Round 10, while CNN's MAPE reduced from 12.5 % to 6.0 %, as shown in Table 6 and illustrated in Fig. 8. This reduction reflects the models' growing ability to make accurate predictions for the next 30 minutes. Correspondingly, RMSE values also showed a downward trend, with LSTM's RMSE dropping from 0.20 to 0.06, and CNN's RMSE decreasing from 0.22 to 0.09, indicating a marked improvement in the models' capacity to minimize forecast errors over medium-term intervals.

The enhancement in the Coefficient of Determination (R^2) further illustrates the models' improved performance. LSTM's R^2 value increased from 0.60 in Round 1–0.89 in Round 10, and CNN's R^2 rose from 0.58 to 0.85, suggesting that the models became more adept at explaining the variance in pollutant levels as the rounds progressed. This trend signifies that the FL framework effectively facilitated the integration of new data from various industrial environments, thus enhancing the models' predictive accuracy and robustness. Overall, the performance improvements observed in the 30-minute predictions affirm the value of the SecureIoT-FL framework in supporting medium-term forecasting. The combination of data accumulation and FL enabled the models to better capture the temporal dynamics of pollutants, leading to more reliable predictions.

Next 60 Minutes: For the 60-minute prediction horizon, the LSTM and CNN models showed consistent improvements in performance metrics over the 10 training rounds, demonstrating the SecureIoT-FL framework's ability to enhance long-term predictive capabilities, as shown in Table 7 and illustrated in Fig. 8. The LSTM model's accuracy rose from 60 % in Round 1–85 % in Round 10, while the CNN model's accuracy increased from 58 % to 80 %. Although these gains were slightly lower than those observed for shorter prediction horizons, they underscore the framework's capacity to refine long-term forecasts through iterative learning and data accumulation. Both models exhibited notable reductions in MAPE and RMSE, highlighting their enhanced accuracy for longer-term predictions. The LSTM model's MAPE decreased from 15.0 % in Round 1–6.5 % in Round 10, and the CNN model's MAPE reduced from 16.0 % to 7.5 %. These reductions reflect the models' growing ability to adapt to the complexities of forecasting pollutant levels an hour in advance. Similarly, RMSE values declined significantly, with LSTM's RMSE dropping from 0.25 to 0.10 and CNN's RMSE reducing from 0.27 to 0.12, indicating that the models became increasingly effective at minimizing forecast errors over extended periods.

The improvements in Coefficient of Determination (R^2) further validate the models' enhanced performance for the 60-minute horizon. LSTM's R^2 increased from 0.50 in Round 1–0.82 in Round 10, while CNN's R^2 rose from 0.48 to 0.80. These enhancements suggest that, despite the inherent challenges of long-term forecasting, the models were able to learn and generalize effectively from the FL framework. The ongoing accumulation of data from diverse industrial settings played a crucial role in these improvements, allowing the models to better capture the temporal patterns associated with pollutant dynamics.

The results for the 60-minute prediction horizon highlight the robustness and scalability of the SecureIoT-FL framework in supporting long-term environmental forecasting. The steady gains in accuracy, MAPE, RMSE, and R^2 demonstrate that the models are capable of progressively refining their predictions, even in the face of the more complex and variable conditions associated with longer-term forecasting.

Throughout the training rounds, the LSTM model consistently outperformed the CNN model in terms of accuracy, MAPE, and RMSE. This superior performance is likely due to LSTM's ability to capture temporal dependencies more effectively, which is crucial for time series data like pollutant levels. While CNN models excel in spatial feature extraction, their relative limitations in temporal sequence modeling likely contributed to the observed differences in performance.

To explore the maximum prediction limits of the LSTM model, the model's performance is analyzed across longer prediction horizons and various training rounds. The training continuation beyond of 10 rounds shows that the model's accuracy has a potential to extend beyond of the 60-minute mark. However, based on the current trend, to achieve a high accuracy for predictions beyond of 60 minutes, it would likely require to get more training rounds.

Through the current evaluation, it is estimated that an approximately of 15–20 training rounds are required to maintain a satisfactory accuracy (>70 %) for predictions up to 90 minutes. For the predictions beyond of 90 minutes, more substantial training or enhancements to the model architecture might be necessary to maintain the predictive accuracy, highlighting the current limitations of the technology and opportunities for future improvement.

The significant performance improvements observed for both models underscore the potential of the SecureIoT-FL framework for real-time environmental monitoring in industrial contexts. Accurate pollutant forecasts up to 60 minutes ahead enable proactive interventions, such as activating ventilation systems, thereby enhancing workplace safety and compliance with environmental regulations. Additionally, the use of FL

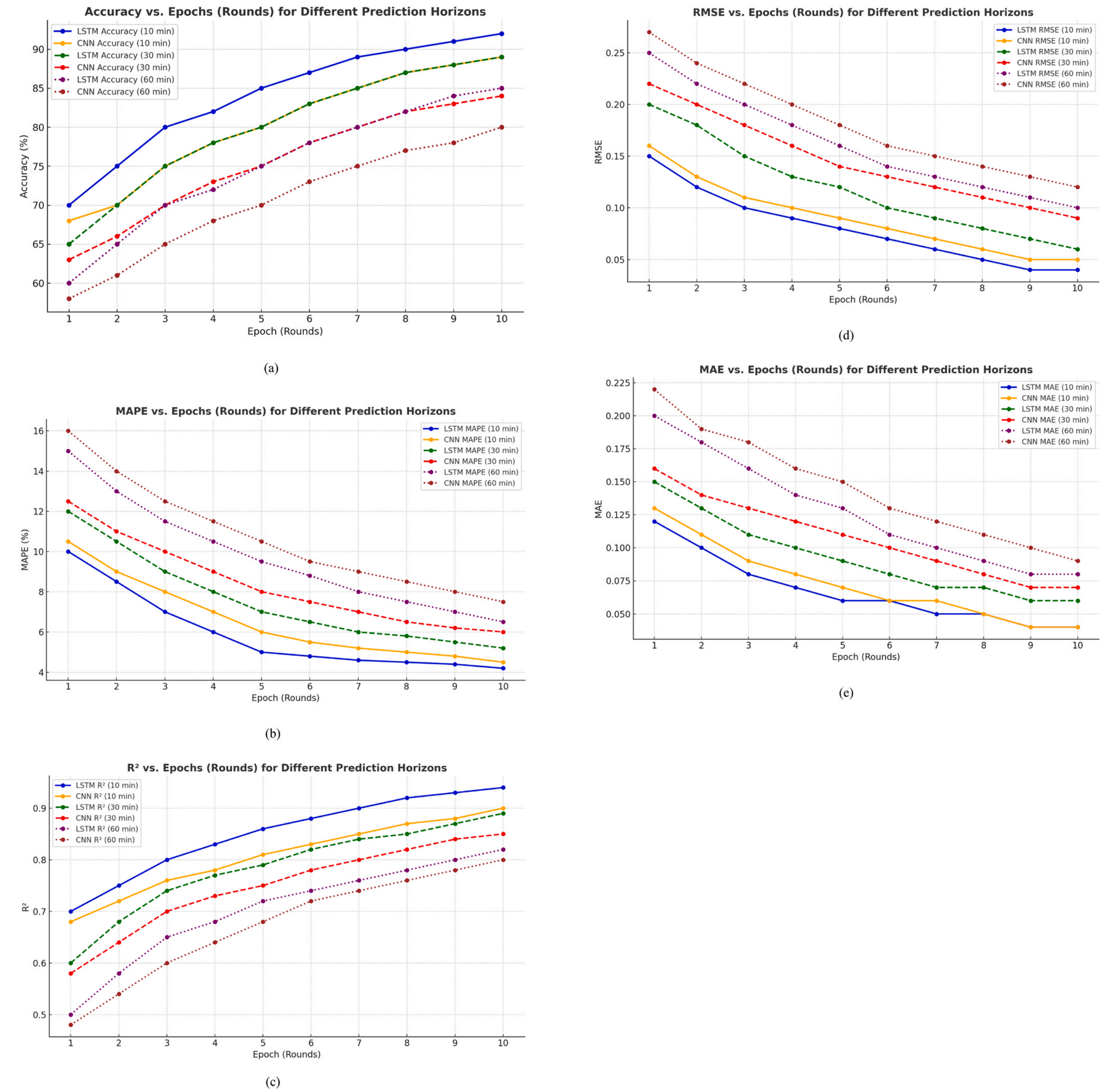


Fig. 8. Various Performance Metrics for the LSTM and CNN Predictions Over Epochs (Rounds) for Different Prediction Horizons: (a) Accuracy, (b) MAPE, (c) R^2, (d) RMSE, (e) MAE.

ensures that sensitive data remains on-site, addressing privacy concerns while leveraging collective intelligence across different industrial nodes. This balanced approach to accuracy, privacy, and scalability positions the SecureIoT-FL framework as a leading solution for predictive maintenance and environmental safety in industrial IoT applications.

These findings reinforce the value of the SecureIoT-FL framework and highlight opportunities for future research to optimize FL strategies, model architectures, and training protocols, thereby further enhancing predictive performance in industrial environments. The continuous accumulation and utilization of new data with each training round are crucial to these advancements, demonstrating the importance of iterative learning in improving model performance over time.

4.4.2. Comparison with centralized models

To illustrate the advantages of the federated learning (FL) approach in the SecureIoT-FL framework, the performance of the proposed federated model is compared with a centralized model trained on pooled data from all factories. As shown in Table 8, the centralized approach yields slight improvements in prediction accuracy, Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE) across different prediction horizons. The centralized model has specifically achieved 2–3 % higher accuracy in the short-term (10-minute) forecasts and approximately 1–2 % higher accuracy in the medium (30-minute) forecasts. MAPE and RMSE metrics are also marginally better in the centralized model, with improving the value of MAPE by 1 % in the short-term forecasts.

Table 6
Prediction for the Next 30 Minutes.

Round	Model	Accuracy (%)	MAPE (%)	R^2	RMSE	MAE
Round 1	LSTM	65.00	12.00	0.60	0.20	0.15
	CNN	63.00	12.50	0.58	0.22	0.16
Round 2	LSTM	70.00	10.50	0.68	0.18	0.13
	CNN	66.00	11.00	0.64	0.20	0.14
Round 3	LSTM	75.00	9.00	0.74	0.15	0.11
	CNN	70.00	10.00	0.70	0.18	0.13
Round 4	LSTM	78.00	8.00	0.77	0.13	0.10
	CNN	73.00	9.00	0.73	0.16	0.12
Round 5	LSTM	80.00	7.00	0.79	0.12	0.09
	CNN	75.00	8.00	0.75	0.14	0.11
Round 6	LSTM	83.00	6.50	0.82	0.10	0.08
	CNN	78.00	7.50	0.78	0.13	0.10
Round 7	LSTM	85.00	6.00	0.84	0.09	0.07
	CNN	80.00	7.00	0.80	0.12	0.09
Round 8	LSTM	87.00	5.80	0.85	0.08	0.07
	CNN	82.00	6.50	0.82	0.11	0.08
Round 9	LSTM	88.00	5.50	0.87	0.07	0.06
	CNN	83.00	6.20	0.84	0.10	0.07
Round 10	LSTM	**89.00**	**5.20**	**0.89**	**0.06**	**0.06**
	CNN	**84.00**	**6.00**	**0.85**	**0.09**	**0.07**

Table 7
Prediction for the Next 60 Minutes.

Round	Model	Accuracy (%)	MAPE (%)	R^2	RMSE	MAE
Round 1	LSTM	60.00	15.00	0.50	0.25	0.20
	CNN	58.00	16.00	0.48	0.27	0.22
Round 2	LSTM	65.00	13.00	0.58	0.22	0.18
	CNN	61.00	14.00	0.54	0.24	0.19
Round 3	LSTM	70.00	11.50	0.65	0.20	0.16
	CNN	65.00	12.50	0.60	0.22	0.18
Round 4	LSTM	72.00	10.50	0.68	0.18	0.14
	CNN	68.00	11.50	0.64	0.20	0.16
Round 5	LSTM	75.00	9.50	0.72	0.16	0.13
	CNN	70.00	10.50	0.68	0.18	0.15
Round 6	LSTM	78.00	8.80	0.74	0.14	0.11
	CNN	73.00	9.50	0.72	0.16	0.13
Round 7	LSTM	80.00	8.00	0.76	0.13	0.10
	CNN	75.00	9.00	0.74	0.15	0.12
Round 8	LSTM	82.00	7.50	0.78	0.12	0.09
	CNN	77.00	8.50	0.76	0.14	0.11
Round 9	LSTM	84.00	7.00	0.80	0.11	0.08
	CNN	78.00	8.00	0.78	0.13	0.10
Round 10	LSTM	**85.00**	**6.50**	**0.82**	**0.10**	**0.08**
	CNN	**80.00**	**7.50**	**0.80**	**0.12**	**0.09**

Table 8
Comparative performance metrics of federated and centralized models across different prediction horizons.

Prediction Horizon	Model	Accuracy (%)	MAPE (%)	R^2	RMSE
10 Minutes	FL	92	4.2	0.94	0.04
Centralized	94–95	3.2	0.96	0.03	
30 Minutes	FL	89	5.2	0.89	0.06
Centralized	90–91	5.0	0.90	0.05	
60 Minutes	FL	85	6.5	0.82	0.10
Centralized	86	6.3	0.83	0.09	

However, despite these modest gains, the centralized approach requires the aggregation of all factory data at a central server, which increases the transmission costs and compromises the data privacy. The federated approach, by contrast, enables for real-time environmental monitoring across the industrial nodes while maintaining data privacy, making it a scalable and secure solution for diverse industrial applications.

4.4.3. Scalability and computational efficiency

As the number of IoT nodes within the SecureIoT-FL framework expands, the computational efficiency and performance of the federated learning system become increasingly critical. The decentralized architecture of the federated learning facilitates the distributed computation, significantly alleviating the computational burden on individual nodes. Each IoT device is responsible for processing its local data and only transmitting model updates to the central server. This approach minimizes the data transfer requirements and reduces the latency, making it especially advantageous when the dataset grows with the addition of new nodes. Consequently, the system can maintain efficient training processes without overwhelming its computational resources.

Furthermore, the iterative nature of the federated learning process is effective in managing larger datasets. As models undergo training across multiple rounds, they continuously adapt and refine their predictions based on the cumulative data that are contributed by all participating nodes. This mechanism not only enhances predictive accuracy but also ensures that the framework can effectively accommodate larger and more complex datasets.

By leveraging the strengths of the federated learning, the proposed SecureIoT-FL framework positions itself as a robust solution capable of efficiently scaling and ultimately supporting the real-time environmental monitoring in diverse industrial contexts.

4.5. Ethical considerations

In the development and deployment of the SecureIoT-FL framework for industrial environmental monitoring, several ethical considerations were carefully addressed to ensure responsible use of technology and protection of stakeholder interests. A primary ethical concern involved data privacy and security, as industrial data often contains sensitive information that, if exposed, could compromise operational integrity and competitive advantage. To address this, the framework was designed using FL, which allows data to remain localized at each factory site. Only model parameters are shared with the global server, thereby maintaining strict confidentiality and minimizing the risk of data breaches. Additionally, the system's implementation aligns with regulatory standards for data protection, ensuring compliance with relevant laws and guidelines such as the General Data Protection Regulation (GDPR). This commitment to privacy and data security not only safeguards the information of participating factories but also builds trust among stakeholders, encouraging broader adoption of the technology.

Another ethical consideration pertains to the potential impact on workers' health and safety. The framework aims to improve air quality and environmental conditions within industrial settings, thereby enhancing workplace safety. However, it is essential that these interventions are implemented transparently and with the cooperation of factory management and workers to ensure that the technology is used to support, rather than replace, existing safety protocols. Stakeholder engagement and clear communication were prioritized throughout the project's deployment to align the system's operations with the needs and rights of the workforce. Furthermore, the responsible use of AI and machine learning algorithms was emphasized, avoiding biases in the models that could lead to unequal or unjust outcomes. Regular audits and validations of the models were conducted to ensure fair and accurate predictions across different industrial environments, maintaining ethical integrity in predictive maintenance and environmental management.

Overall, the ethical framework guiding this study emphasizes data privacy, compliance with regulatory standards, worker safety, and the responsible use of AI, ensuring that the SecureIoT-FL framework is not only effective but also ethically sound in its application.

5. Limitations and future directions

While the SecureIoT-FL framework demonstrates a significant promise in enhancing the predictive accuracy and data privacy for industrial environmental monitoring, it is essential to acknowledge the presence of some limitations. A key challenge is the potential for biases in the sensor data, which may rise from the calibration errors, environmental factors, or sensor degradation. These biases can affect to the reliability of the data and, consequently, the accuracy of the model's predictions. Moreover, the generalizability of the results may be limited, as the framework was primarily tested in PCBA, CNC machining, and plastic injection factories. This specificity means that the findings might not directly apply to the other industrial settings with differing pollutant profiles or operational conditions, necessitating a further adaptation or retraining of the models. The reliance on the federated learning also presents challenges, as it requires stable communication between the global server and local nodes. Any disruptions in connectivity could affect on the efficiency of model updates and real-time performance.

Looking ahead, future research should focus on addressing these limitations by developing an advanced sensor calibration techniques and enhancing data quality management to minimize such above-mentioned biases. The expanding the framework test across a broader range of industrial applications will help validate its effectiveness in diverse operational contexts. This needs to conduct many case studies in various environments can provide valuable insights into the framework's adaptability and scalability. Furthermore, the advancing of predictive analytics tools that leverage sophisticated machine learning algorithms could improve the system's forecasting capabilities, facilitating more informed decision-making and timely interventions. The user experience through intuitive interface design will also enhanced, allowing the stakeholders to monitor and respond to the environmental conditions in real time. Lastly, the investigation of the SecureIoT-FL integration with other smart systems, such as automated manufacturing processes, could create synergies that enhance the operational efficiency and environmental safety.

Finally, future work should consider the development of adaptive learning mechanisms that enable the SecureIoT-FL framework to be accounted for the evolving industrial environments, including shifts in pollutant levels or the introduction of new pollutants. Such mechanisms could allow the model to recalibrate in real-time, and improve its accuracy and resilience upon the environmental conditions change.

6. Conclusion

This study introduced SecureIoT-FL, an innovative IoT-based framework that integrates Federated Learning (FL) for real-time environmental monitoring and forecasting in industrial settings while preserving data privacy. The framework was deployed across three distinct factories—PCBA, CNC machining, and plastic injection—equipped with IoT nodes and multiple gas sensors. Over a five-month period, the system successfully collected more than 516,000 data points. To ensure data privacy, SecureIoT-FL utilized FL, allowing local model training with encrypted model weights shared via MQTT over AWS. The key outcomes of the proposed work include significant improvements in the model performance over 10 biweekly training rounds, with increasing the accuracy from 70 % to 92 % and 68–89 % for LSTM and CNN models, respectively. This approach enabled effective forecasting of pollutant concentrations over multiple timeframes (10, 30, and 60 minutes), supporting timely interventions to reduce worker exposure to harmful pollutants, including proactive measures such as ventilation system activation. While SecureIoT-FL sets a new standard for secure and scalable industrial environmental monitoring, future work should address potential sensor data biases and explore broader industrial applications to further validate the framework's effectiveness. Overall, SecureIoT-FL demonstrates the significant potential of FL in the IIoT, offering a robust solution for real-time environmental monitoring and predictive maintenance.

CRediT authorship contribution statement

Montaser Ramadan: Writing – original draft, Software, Methodology, Conceptualization, Project administration, Validation. **Mohammed A. H. Al:** Writing – review & editing, Supervision. **SHIN YEE Khoo:** Writing – review & editing. **Mohammad Alkhedher:** Writing – review & editing.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Montaser Ramadan reports writing assistance was provided by University of Malaya.

Acknowledgment

ADU author acknowledges financial support from Abu Dhabi University's Office of Research and Sponsored Programs. Grant ID: 19300889.

References

[1] S. Mozaffari, et al., Effects of occupational exposures on respiratory health in steel factory workers, Front Public Health 11 (2023) 1082874.
[2] J. Susilo, F. Tunjungsari, Resiko gangguan pernapasan pada industri besi baja, CoMPHI J. Community Med. Public Health Indones. J. 3 (2) (2022) 110–121.
[3] J. Liu, Y. Yang, Controlling industrial air-pollutant emissions under multi-factor interactions based on a developed hybrid-factorial environmental input–output model, Sustainability 15 (9) (2023) 7717.
[4] W. Dargie, J. Wen, L.A. Panes-Ruiz, L. Riemenschneider, B. Ibarlucea, G. Cuniberti, Monitoring toxic gases using nanotechnology and wireless sensor networks, IEEE Sens J. 23 (11) (2023) 12274–12283.
[5] Y. Mu, C. Duan, X. Li, Y. Wu, A monitoring method for corporate environmental performance based on data fusion in china under the double carbon target, Sustainability 15 (12) (2023) 9391.
[6] H. Wang, W. Zhang, D. Yang, Y. Xiang, Deep-learning-enabled predictive maintenance in industrial internet of things: methods, applications, and challenges, IEEE Syst. J. 17 (2) (2022) 2602–2615.
[7] M.N.A. Ramadan, M.A.H. Ali, S.Y. Khoo, L. Hamad, M. Alkhedher, Revolutionizing agri-food technology: development and validation of the Portable Intelligent Oil Recognition System (PIORS), Smart Agric. Technol. 9 (2024) 100624, https://doi.org/10.1016/j.atech.2024.100624.
[8] M. Ansari, M. Alam, An intelligent IoT-cloud-based air pollution forecasting model using univariate time-series analysis, Arab J. Sci. Eng. 49 (3) (2024) 3135–3162.
[9] H. Wang, W. Zhang, D. Yang, Y. Xiang, Deep-learning-enabled predictive maintenance in industrial internet of things: methods, applications, and challenges, IEEE Syst. J. 17 (2) (2022) 2602–2615.
[10] C. De Capua, G. Fulco, M. Lugarà, F. Ruffa, An improvement strategy for indoor air quality monitoring systems, Sensors 23 (8) (2023) 3999.
[11] M.N.A. Ramadan, M.A.H. Ali, S.Y. Khoo, M. Alkhedher, M. Alherbawi, Real-time IoT-powered AI system for monitoring and forecasting of air pollution in industrial environment, Ecotoxicol. Environ. Saf. 283 (2024) 116856.
[12] V. Hura, L. Monastyrskii, IOT-based solution for detection of air quality using ESP32, Artif. Intell. (3) (2023) 86–93.
[13] M.N.A. Ramadan, M.A.H. Ali, S.Y. Khoo, M. Alkhedher, M. Alherbawi, Real-time IoT-powered AI system for monitoring and forecasting of air pollution in industrial environment, Ecotoxicol. Environ. Saf. 283 (2024) 116856.
[14] P.K. Srivastava, A.K. Yadav, Machine Learning Techniques and Industry Applications, IGI Global, 2024.
[15] M.N.A. Ramadan, M.A.H. Ali, S.Y. Khoo, M. Alkhedher, AI-powered IoT and UAV systems for real-time detection and prevention of illegal logging, Results Eng. (2024) 103277.
[16] M.N.A. Ramadan, et al., Towards early forest fire detection and prevention using AI-powered drones and the IoT, Internet Things (2024) 101248.
[17] S.M. Popescu, et al., Artificial intelligence and IoT driven technologies for environmental pollution monitoring and management, Front Environ. Sci. 12 (2024) 1336088.
[18] P. Hembert, C. Ghnatios, J. Cotton, F. Chinesta, Assessing Sensor Integrity for Nuclear Waste Monitoring Using Graph Neural Networks, Sensors 24 (5) (2024) 1580.
[19] R. Arabelli, E. Boddepalli, M. Buradkar, N.V.S. Goriparti, M.K. Chakravarthi, IoT-Enabled Environmental Monitoring System Using AI. 2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), IEEE, 2024, pp. 1–6.
[20] T. Filho, et al., A standard-based internet of things platform and data flow modeling for smart environmental monitoring, Sensors 21 (12) (2021) 4228.
[21] L. García, A.-J. Garcia-Sanchez, R. Asorey-Cacheda, J. Garcia-Haro, C.-L. Zúñiga-Cañón, Smart air quality monitoring IoT-based infrastructure for industrial environments, Sensors 22 (23) (2022) 9221.
[22] S.M. Popescu, et al., Artificial intelligence and IoT driven technologies for environmental pollution monitoring and management, Front Environ. Sci. 12 (2024) 1336088.

[23] A. Khorramifar, et al., Environmental engineering applications of electronic nose systems based on MOX gas sensors, sensors 23 (12) (2023) 5716.
[24] H. Wu, T. Yang, H. Li, Z. Zhou, Air quality prediction model based on mRMR–RF feature selection and ISSA–LSTM, Sci. Rep. 13 (1) (2023) 12825.
[25] Z. Wu, Y. Tian, M. Li, B. Wang, Y. Quan, J. Liu, Prediction of air pollutant concentrations based on the long short-term memory neural network, J. Hazard Mater. 465 (2024) 133099.
[26] Z. Zhang, S. Zhang, C. Chen, J. Yuan, A systematic survey of air quality prediction based on deep learning, Alex. Eng. J. 93 (2024) 128–141.
[27] K. Liu, et al., Time series prediction of the chemical components of PM2. 5 based on a deep learning model, Chemosphere 342 (2023) 140153.
[28] B. Zhang, et al., Air pollutant diffusion trend prediction based on deep learning for targeted season—North China as an example, Expert Syst. Appl. 232 (2023) 120718.
[29] S.M. Cabaneros, B. Hughes, Methods used for handling and quantifying model uncertainty of artificial neural network models for air pollution forecasting, Environ. Model. Softw. 158 (2022) 105529.
[30] J. Han, G.H. Lee, J. Lee, T.Y. Kim, J.K. Choi, A novel deep-learning-based robust data transmission period control framework in iot edge computing system, IEEE Internet Things J. 9 (23) (2022) 23486–23505.
[31] M. Ansari, M. Alam, An intelligent IoT-cloud-based air pollution forecasting model using univariate time-series analysis, Arab J. Sci. Eng. 49 (3) (2024) 3135–3162.
[32] Y. Zhang, H. An, Y. Xing, Y. Liu, T. Zhang, Learning temporal and spatial features jointly: A unified framework for space-time data prediction in industrial IoT networks, IEEE Sens J. 23 (16) (2023) 18752–18764.
[33] I.-I. Prado-Rujas, A. García-Dopico, E. Serrano, M.L. Córdoba, M.S. Pérez, A multivariable sensor-agnostic framework for spatio-temporal air quality forecasting based on Deep Learning, Eng. Appl. Artif. Intell. 127 (2024) 107271.
[34] K. Chatterjee, et al., Towards cleaner industries: smart cities' impact on predictive air quality management, IEEE Access (2024).
[35] O. Farooq, P. Singh, M. Hedabou, W. Boulila, B. Benjdira, Machine learning analytic-based two-staged data management framework for internet of things, Sensors 23 (5) (2023) 2427.
[36] U.R. Babu, N. Sharma, K.S. Kumar, D.D. Madhukar, P.R. Sonawane, N. Nishant, IoT Based Environmental Monitoring System using LSTM-WAO-EEMD. 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), IEEE, 2023, pp. 1616–1621.
[37] L. Closson, C. Cérin, D. Donsez, J.-L. Baudouin, Design of a meaningful framework for time series forecasting in smart buildings, Information 15 (2) (2024) 94.
[38] A. Ramachandran, K. Gayathri, R.H. Aswathy, S.Zulaikha Beevi, Modeling of internet of things enabled sustainable environment air pollution monitoring system, Glob. NEST J. 25 (4) (2023) 172–179.
[39] M. Imam, S. Adam, S. Dev, N. Nesa, Air quality monitoring using statistical learning models for sustainable environment, Intell. Syst. Appl. 22 (2024) 200333.
[40] A. Coiduras-Sanagustín, E. Manchado-Pérez, C. García-Hernández, Understanding perspectives on personal data privacy in Internet of Things (IoT): a systematic literature review (SLR), Heliyon (2024).
[41] A.H. Eyeleko, T. Feng, A critical overview of industrial internet of things security and privacy issues using a layer-based hacking scenario, IEEE Internet Things J. (2023).
[42] X. Wu, Y. Liu, J. Tian, Y. Li, Privacy-preserving trust management method based on blockchain for cross-domain industrial IoT, Knowl. Based Syst. 283 (2024) 111166.
[43] Q. Liu, et al., Privacy and integrity protection for IoT multimodal data using machine learning and blockchain, ACM Trans. Multimed. Comput., Commun. Appl. 20 (6) (2024) 1–18.
[44] N. He, Z. Zhang, X. Wang, T. Gao, Efficient privacy-preserving federated deep learning for network intrusion of industrial IoT, Int. J. Intell. Syst. 2023 (1) (2023) 2956990.
[45] L. Zhang, X. Lei, Y. Shi, H. Huang, C. Chen, Federated learning for iot devices with domain generalization, IEEE Internet Things J. 10 (11) (2023) 9622–9633.
[46] P. Zeng, A. Liu, N.N. Xiong, S. Zhang, M. Dong, TD-MDB: a truth discovery based multi-dimensional bidding strategy for federated learning in industrial IoT systems, IEEE Internet Things J. (2023).
[47] J. Lin, F. Li, H. Han, Federated Temporal Learning Based Cyber Attack Detection for Distributed Industrial IoT Systems. 2023 CAA Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS), IEEE, 2023, pp. 1–5.
[48] D. Patil, S. Goel, S.K. Garud, M.D. Kokate, A. Nashte, P. Rane, Federated learning in real-time medical iot: optimizing privacy and accuracy for chronic disease monitoring, J. Electr. Syst. 19 (3) (2023).
[49] A.A. Sinaci, et al., Privacy-preserving federated machine learning on FAIR health data: a real-world application, Comput. Struct. Biotechnol. J. 24 (2024) 136–145.
[50] P. Zhou, et al., PM2. 5, PM10 and health risk assessment of heavy metals in a typical printed circuit noards manufacturing workshop, J. Environ. Sci. 26 (10) (2014) 2018–2026.
[51] Y. Li, S. Zong, H. Jing, N. Gao, H. Ye, J. Chen, The influence of ventilation modes on oil mist particles diffusion in a machining workshop, Heliyon 10 (5) (2024).
[52] Y. Ren, et al., Characterization of VOC emissions and health risk assessment in the plastic manufacturing industry, J. Environ. Manag. 357 (2024) 120730.
[53] P.M. Mammen, Federated learning: Opportunities and challenges, *arXiv preprint arXiv*:2101.05428, 2021.
[54] Q. Yang, Y. Liu, T. Chen, Y. Tong, Federated machine learning: Concept and applications, ACM Trans. Intell. Syst. Technol. (TIST) 10 (2) (2019) 1–19.
[55] J. Konečný, H.B. McMahan, F.X. Yu, P. Richtárik, A.T. Suresh, and D. Bacon, Federated learning: Strategies for improving communication efficiency, *arXiv preprint arXiv:1610.05492*, 2016.
[56] J. Gibson, Minimum mean squared error estimation and mutual information gain, Information 15 (8) (2024) 497.
[57] X. Wang, L. Zhou, H. Lin, Deep regression learning with optimal loss function (no. just-accepted), J. Am. Stat. Assoc. (2024) 1–20.
[58] R. Oppliger, SSL and TLS: Theory and Practice, Artech House, 2023.
[59] Z. Chang, Y. Zhang, W. Chen, Electricity price prediction based on hybrid model of adam optimized LSTM neural network and wavelet transform, Energy 187 (2019) 115804.
[60] D. Bethell, S. Gerasimou, R. Calinescu, Robust uncertainty quantification using conformalised monte carlo prediction (in), Proc. AAAI Conf. Artif. Intell. (2024) 20939–20948.
[61] T. Putri, I. Jaya, T. Toharudin, F. Kristiani, Sensitivity analysis of the PC hyperprior for range and standard deviation components in Bayesian Spatiotemporal high-resolution prediction: an application to PM2. 5 prediction in Jakarta, Indonesia, Int. J. Data Netw. Sci. 8 (2) (2024) 871–880.