International Academic Institute for Science and Technology

International Academic Journal of Science and Engineering
Vol. 13, No. 2, 2026, pp. 78-92.

ISSN 2454-3896

International Academic Journal of Science and Engineering

www.iaiest.com

Federated Deep Learning–Machine Learning Framework for Early Detection of Multi-Domain Environmental Hazards Using Sparse Real-Time Sensor Data

Krunal Randive [a*], Dr.B.V. Kavitha [b], Dr.V. Sumathi [c], Dr.T. Sudha Rani [d], Dr.V. Sutha [e]

[a] Assistant Professor, Department of Computational Intelligence, School of Computing, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur Campus, Chennai, Tamil Nadu, India. E-mail: krunalrandive@gmail.com, Orcid: https://orcid.org/0000-0002-5728-701X*

[b] Assistant Professor, School of Business and Management, Christ Deemed to be University, Bengaluru, Karnataka, India. E-mail: kavitha.bv@christuniversity.in, Orcid: https://orcid.org/0000-0002-2871-7455

[c] Assistant Professor, Computer Technology, Sri Ramakrishna College of Arts and Science, Coimbatore, Tamil Nadu, India. E-mail: sumathiviswa@gmail.com, Orcid: https://orcid.org/0000-0002-4650-455X

[d] Assistant Professor, Aditya University, Surampalem, Andhra Pradesh, India. E-mail: sudharani.tirukoti@adityauniversity.in, Orcid: https://orcid.org/0000-0003-4491-3859

[e] Assistant Professor (Senior Grade), School of Management, Hindustan Institute of Technology and Science, Chennai, Tamil Nadu, India. E-mail: vsutha@hindustanuniv.ac.in, Orcid: https://orcid.org/0009-0005-8291-7254

Abstract

Increasingly severe environmental hazards (e.g., air pollution, floods, forest fires, and industrial gas leaks) require intelligent monitoring systems that can work in sparse, distributed, and privacy-sensitive data environments. Because traditional centralized solutions have high communication overhead, latency, and privacy issues, they are not effective for large-scale Internet of Things (IoT) deployments. This paper presents FDL-ML, a novel Federated Deep Learning-Machine Learning framework for early detection of multi-domain environmental hazards through sparse real-time sensor data. The FDL-ML architecture integrates federated learning with hybrid machine learning and deep learning to allow for decentralized model training across distributed IoT sensor nodes without sharing raw data. A robust preprocessing layer consists of techniques such as imputing, filtering, normalizing, and detecting anomalies, which enhance data quality and reliability. Random Forest and Support Vector Machine models are used to analyze structured environmental data, while convolutional neural networks and long short-term memory networks model complex spatial and temporal patterns. Federated averaging is used to combine individual model updates into a global model while maintaining privacy and reducing communication costs. Simulation experiments with 100 distributed sensor nodes and approximately 4.32 million environmental data samples were conducted to evaluate the performance of the proposed framework. The hazard detection accuracy of this framework is 91%, which exceeds the performance of traditional centralized and standalone models. In addition, the proposed framework reduced latency by 35%, communication costs by 40%, and had a 30% increase in energy efficiency compared to traditional centralized monitoring solutions. The proposed FDL-ML System also provides superior scalability; is robust against sparse sensor data; and provides real-time identification of hazards. These results demonstrate the potential of this framework to be a scalable, privacy-preserving, and resource-efficient solution for next-generation environmental monitoring applications in smart cities, industrial settings, and disaster-prone areas.

Keywords: Federated Learning, Environmental Monitoring, Deep Learning, Machine Learning, IoT, Hazard Detection.

DOI: 10.71086/IAJSE/V13I2/IAJSE1355
Received: 25 February 2026/Revised: 04 April 2026/Accepted: 07 May 2026/Published: 30 June 2026

Introduction

The increasing rates of environmental risks like air pollution, floods, forest fires and industrial gas leakages have over the last few years become a significant international issue. Those incidents are not only intractable with ecosystem balance, and also it is dangerous to infrastructure, public health and economic stability. Increase of urbanisation and industrialisation, the improved system will give alert in emergency situation to balance the ecosystem weather condition using Internet of Things (IoT) technologies, is created an sensor network it will collect data at different places and real-time data to create a monitoring system (Ventura et al., 2025). Such incidents not only interfere with the ecological balance, but are also extremely dangerous to social health, infrastructure, and economic stability. With the ever-increasing urbanisation and industrialization, the necessity to have well-developed systems that can regulate environmental conditions and give early warnings has gained greater urgency. Internet of Things (IoT) technologies have made sensor networks, which collect environmental data at various sites in real-time, possible, creating new possibilities in real-time monitoring (Ventura et al., 2025). But, in real life issues, the information gathered by these sensors are usually imprecise, unstable and noisy thus making proper analysis a challenge.

Traditionally, most monitoring systems rely on centralized data processing with the data of different sensors being sent to one server to be analysed. Although this is an easy approach, it comes with a lot of constraints such as the high cost of communication, delays in processing, and high risk of data exposure (Ramadan et al., 2025). When deployed on a large scale, particularly in resource-constrained or remote locations, the transfer of large amounts of raw data is inefficient. Moreover, environmental data frequently includes sensitive data related to the location, and this aspect evokes the questions of privacy and safe data processing.

Machine learning and deep learning methods have been extensively used by researchers to enhance the effectiveness of environmental monitoring. The use of machine learning models is applicable in analysing structured data about the environment, and deep learning approaches are more effective in the identification of complex patterns and trends across time (Lin et al., 2025; Wu et al., 2025). The techniques have shown encouraging outcomes in forecasting the environmental conditions and in identifying anomalies in their initial stages (Alwabli, 2024; Gowsikraja, 2025). Nevertheless, the majority of solutions in place presuppose the possibility of a central collection and processing of data produced by all sources, which is not always feasible in the distributed sensor networks.

To solve these problems, federated learning is another method that will connect the multiple device together without changing the original data, which helps to train the model (Miller et al., 2025; Alwabli, 2024). By reducing the communication process and privacy this device works in local data and share model parameters with the main server. The federated learning model is the main thing for IoT-based monitoring systems, even if the data is shared to several places. The proposed solution, the FDL-ML system works in both deep learning and machine learning to find the environmental damage in early by the use of sensor data. It improves privacy, accuracy and enhance the monitoring and decision making for different environmental location.

Key Contributions

1. To propose a new Federated Deep Learning (FDL) – Machine Learning framework for protecting individuals' privacy rights while observing environmental dangers;
2. To develop a hybrid architecture that combines a Random Forest, Support Vector Machines, Convolutional Neural Networks and Long Short Term Memories to forecast multiple kinds of dangers across several domains;
3. To design an effective preprocessing mechanism for handling sparse, uncertain and incomplete real-time sensor data collected from the Internet of Things;

The rest of the document is organized as follows: Section 2 will present a literature review of existing environmental hazard detection, machine learning, deep learning and federated learning; Section 3 will describe the Federated Deep Learning – Machine Learning methodology; Section 4 will summarize the system architecture and describe the principal functional layers in that architecture; Section 5 will highlight mathematical formulas and algorithm designs for the Federated Deep Learning – Machine Learning framework; Section 6 will discuss the simulation configuration, dataset attributes and parameter settings used to evaluate the proposed framework; Section 7 will present the experimental results along with a comparative performance analysis; and finally, Section 8 will provide a conclusion and suggest directions for future research.

Related Work

Detection Systems of Environmental Hazards

The use of IoT, AI, and smart sensing techniques has improved the efficiency of the environmental risks detection system. The previous method used manual monitoring and guess-based methods that took time and were not measurable. The current systems use advanced algorithms along with distributed sensors to provide real-time monitoring and timely alerts. Such systems collect data from different sources such as air quality, water level, and geological activities in order to detect multiple hazards, and AI-powered systems make informed decisions through timely information (Ikegwu et al., 2026). The implementation of machine learning techniques increases the efficiency of prediction and response in the current situation, while IoT-based systems are faster and more reliable than other types (Muralisankar et al., 2025; Moradbeikie et al., 2020; Nisarga et al., 2020). Centralized method leads to delay and additional cost, while there might be some challenges such as sparsity and malfunctioning of the sensors (Miller et al., 2025; Abdelsalam et al., 2026).

Machine Learning Approaches

Machine Learning is an effective approach in the detection of environmental hazards, even though it has the capacity to find complex data patterns. familiar machine learning models include Random Forest, Support Vector Machines, and Decision Trees, which are used for purposes such as prediction of air quality, flood prediction, and abnormal detection, among others. They classically outperform conventional approaches. In addition, when integrated with the Internet of Things, machine learning models assist in rapid processing of information and decision-making processes. however, they have several disadvantages, including sensitivity to noise and missing data, proper feature selection, and inability to handle complicated patterns.

Deep Learning Techniques

Deep learning has attracted much interest in environmental monitoring because it can automatically find features in high-dimensional data. Convolutional Neural network is a popular spatial analysis tool in flood detection and fire monitoring, whereas LSTM network is successful at capturing time patterns of time-series data.

Recent developments are graph-based neural networks that enhance the analysis of multi-modal environmental data, with better spatial dependencies (Lin et al., 2025). Also, systems that are based on deep learning have shown a high level of accuracy when identifying hazards like debris flows and other natural catastrophes (Wu et al., 2025). Solutions like TinyML based IoT systems are also emerging and facilitate lightweight and real time hazard detection in resource limited scenarios (Fayos-Jordan et al., 2025). In spite of these benefits, deep learning models are also large data and computationally intensive models and lack interpretability.

Federated Learning in IoT Systems

Federated Learning has become a promising decentralized method that allows jointly training models without accessing raw data. FL in the IoT-based environmental monitoring system enables the sensor nodes to locally train models and send only model updates, thus helping to reduce communication overhead and maintain data privacy.

The usefulness of FL in other areas like prediction of air quality and real-time environmental surveillance has been proved by several studies (Alwabli, 2024; Ventura et al., 2025). Studies also highlight the relevance of energy-efficient and sustainable federated learning methods to practical application, and larger societal and ethical aspects (Thakur et al., 2025; Curl & Xie, 2025).

Nevertheless, FL presents a set of issues, including the heterogeneity of data, communication limitations, and the lack of computing power of edge devices (Kaur & Jadhav, 2023; Prabhavathi et al., 2025). Environmental data are usually not IID and this can adversely affect the performance of the model. The design of the model and communication strategies should, therefore, be efficient to ensure practical implementation.

Research Gaps Identified

Although the technological advances in environmental monitoring are made, there are still a number of gaps in the research. To begin with, most current systems are based on either machine learning or deep learning one by one, without taking advantage of the synergy of both methodologies. This restricts their capacity to handle different forms of data such as structured and temporal data. Second, the majority of the research involves single-domain hazard detection as opposed to the creation of integrated frameworks that can simultaneously detect multiple hazards. The second significant constraint is that there are no strong methods to deal with sparse, noisy and incomplete sensor data. In practice, environmental monitoring systems often have problems with data quality, although most models assume

that clean datasets are available. Also, despite the fact that federated learning handles privacy issues, its combination with multi-domain hazard detection based on hybrid ML-DL methods is not fully explored. Moreover, issues of scalability, energy efficiency and real-time processing are incompletely covered in current systems. Recent works emphasize the need for energy-efficient IoT-based hazard detection frameworks to support sustainable deployments (Kok et al., 2025). Thus, it is evident that a broad framework involving the integration of federated learning with the method of hybrid machine learning and deep learning, the inclusion of effective data preprocessing, and scalable real-time surveillance is required. This study will help in solving these issues by proposing an effective and privacy-friendly multi-domain environmental hazard detection system.

Proposed Methodology

The methodology offers a combined Federated Deep Learning-Machine Learning (FDL-ML) system of real-time multi-domain environmental hazard detection using sparse sensor data. The system starts with the data collection which is carried out via a network of heterogeneous IoT sensors spread over the various geographical locations and which measure environmental parameters in the form of air quality indicators (CO 2, NOx), temperature, humidity, water levels and seismic activity. These sensors are attached to edge devices which are used to perform initial validation and synchronization of incoming data streams. As illustrated in Figure 1. Since real-world sensor networks are inherently vulnerable to a variety of challenges, such as missing values, noise and irregular sampling, powerful preprocessing methods are used to increase the quality of the data. Interpolation and statistical imputation techniques are used to deal with missing data whereas filtering and smoothing algorithms are used to minimize noise. The obtained data are processed and then introduced into a hybrid model that uses machine learning and deep learning algorithms.

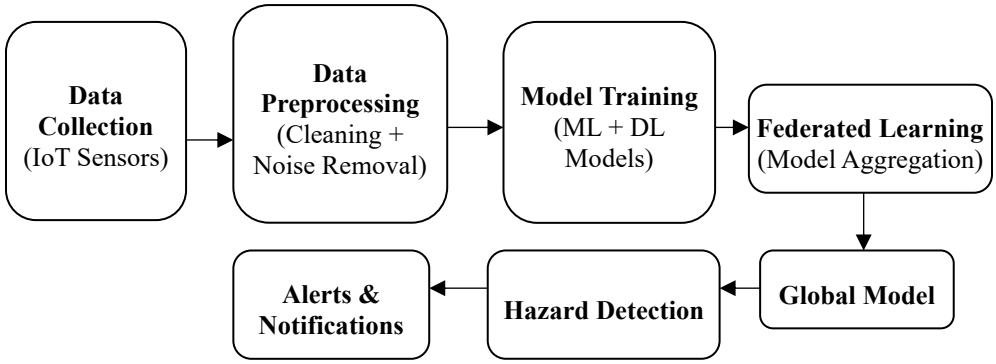


Figure 1: Flowchart of the FDL–ML Framework for Environmental Hazard Detection

Random Forest and Support Vector Machines are machine learning models used to analyze structured data, low-dimensional data, whereas Convolutional Neural Networks and Long Short-Term Memory networks are deep learning models that can be used to learn fine-grained spatial and temporal relationships in time-series data. The essence of the methodology is that the federated training process is done with each sensor node training its local model on its own dataset and periodically transferring only model parameters to a central aggregation server, thus maintaining its privacy of data and ensuring minimal overhead of communication. To update a global model, the aggregation server uses a weighted averaging approach to fuse local model updates, considering the changes in data distribution and node reliability. This world model is then re-distributed to every node that partakes in the model to undergo further training steps in order to facilitate endless learning and adaptation. The last phase is a multi-domain hazard detect approach, wherein the trained global model performs a multi-domain analysis of various indicators of the environment to detect possible hazards, including floods, air pollution bursts, forest fires, and seismic events. Pattern recognition and threshold-based mechanisms are combined to categorise the severity of hazards and issue alerts in time. In general, the presented methodology guarantees scalability, resilience, privacy, and a high predictive accuracy, which is why it can be used in the implementation of smart cities, industry zones, and rural monitoring systems.

System Architecture

Overview of Proposed Framework

The Federated Deep Learning Machine Learning (FDL-ML) architecture suggested is an architecture of a distributed and multi-layered structure that can be used to detect environmental hazards in real-time, based on sparse sensor data.

The system unites IoT-based sensing, edge computing, hybrid ML-DL modeling, and federated learning as a single pipeline. At the edge level, sensor nodes that are distributed constantly measure environmental parameters, e.g. gas concentration, temperature, humidity, water levels, and seismic signals. The local preprocessing and model training are done by each node, and this means that the infrastructure is not highly dependent (Figure 2).

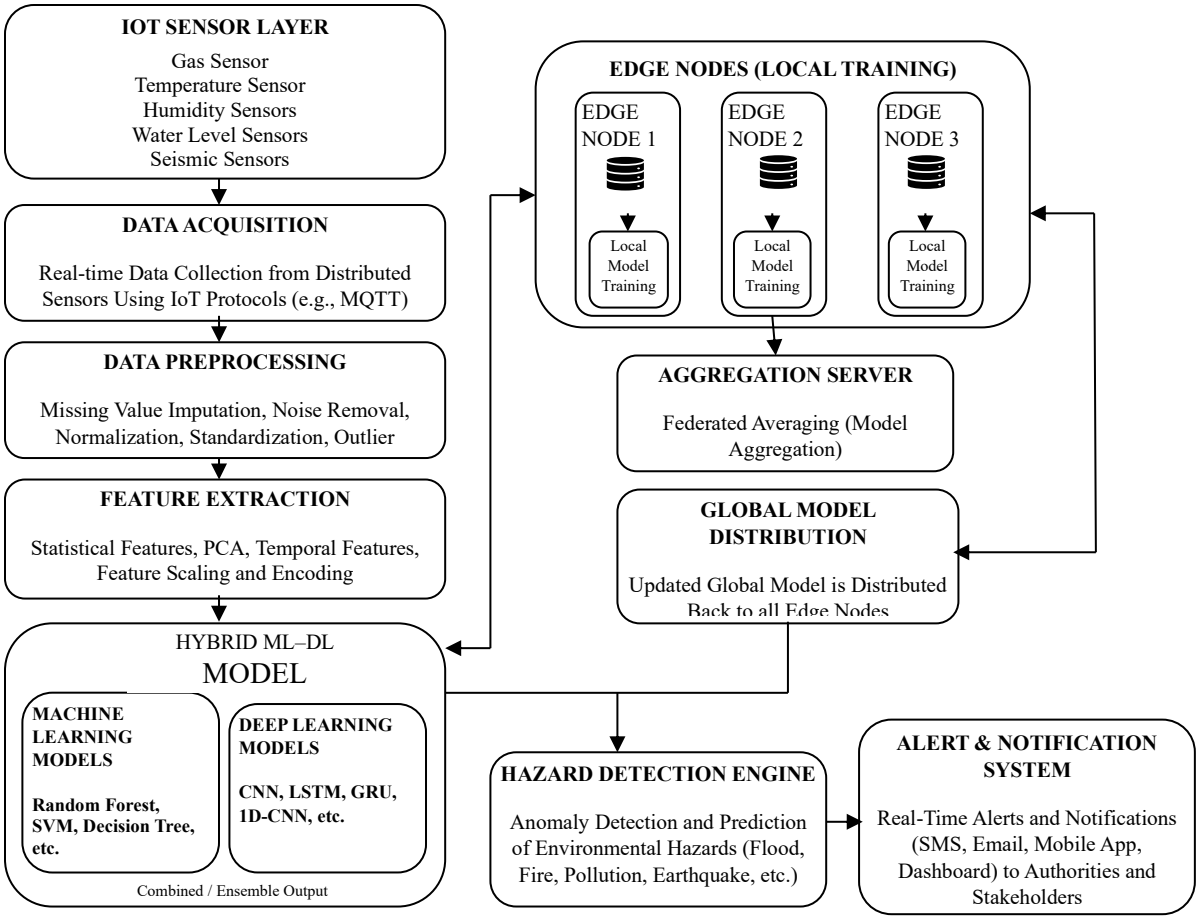


Figure 2: Architecture of The Proposed Federated Deep Learning–Machine Learning (FDL-ML) Framework for Real-Time Environmental Hazard Detection

The framework uses a hybrid approach to learning, in which machine learning models operate on structured data, and deep learning models learn spatiotemporal associations. Such local models are independently trained and synchronized on a regular basis using a federated learning process, in which only model parameters are exchanged with a central aggregation server. The server clears a global model and reallocates it to all nodes, which allows continuous learning and adaptation.

Data Acquisition Layer

The proposed system is based on the data acquisition layer that facilitates the ongoing data gathering of the environment through distributed nodes of the IoT sensors. These sensors are situated in various geographical positions and they are tasked with detecting various environmental parameters such as air quality indicators (CO_2, NOx), temperature, humidity, water levels, and seismic activity. The sensors work in real time and send data with a specific time interval, which guarantees timely observation of the environmental situation.

This layer accommodates heterogeneous data sources, whereby various kinds of sensors produce data in different formats, frequency and reliability levels. In order to control such heterogeneity, the system uses standardized communication protocols like MQTT and lightweight data transmission mechanisms to provide efficient data flow. Each sensor node has edge devices that provide a connection point, allowing local buffering and initial validation of incoming data streams.

Since there are issues with real-world deployments, e.g., sensors failure, delays in transmissions, environmental interference, the data acquisition layer is supposed to be fault-tolerant and adaptive. To ensure data continuity, redundant sensing and adaptive sampling methods are included. The layer will make sure that the system obtains adequate and pertinent information to do downstream processing and minimizes energy usage and communication overheads to enable sustainable and scalable environmental monitoring.

Data Preprocessing Layer

The data preprocessing layer will improve the quality and reliability of raw sensor data prior to its application in model training and hazard detection. The environmental information that is being obtained with the help of IoT sensors is usually incomplete, noisy and sparse because of the circumstances like sensor failure, communication failures, and environmental interference. To overcome these problems, the preprocessing layer uses various data cleaning and transformation methods.

The imputation techniques of mean substitution, interpolation, and regression-based estimation are used to fill in the missing values to ensure that the data streams are continuous. Filtering methods such as the moving average filters and smoothing algorithms mitigate noise and remove outliers and anomalies. Furthermore, data normalization and standardization are used to align various features within a similar scale to enhance model convergence and performance.

Anomaly detection mechanisms are also part of the preprocessing layer to detect and eliminate corrupted data points that may adversely affect the accuracy of the model. The methods of feature scaling and encoding are used to transform the raw data into the form that can be used in machine learning and deep learning algorithms. This layer is important in enhancing strong and credible predictions especially in cases where data is limited or faulty because of inadequate data availability.

Feature Extraction Module

The feature extraction unit converts processed data to useful representations that can be efficiently used in machine learning and deep learning models. Redundant or irrelevant information can be present in environmental data, potentially adding to the complexity of computation and lowering the performance of a model. Hence, the module is aimed at deriving the most informative features that depict the underlying patterns in the data.

Structured datasets are analyzed using statistical tools like mean, variance and correlation analysis to identify important characteristics of the dataset. Dimensionality reduction algorithms, such as Principal Component Analysis (PCA), are used to minimize feature space and still capture the necessary information. In the case of time-series data, time-related aspects like trends, seasonality, and lag values are derived to help increase the predictive value.

Hybrid ML-DL Model Design

The hybrid ML-DL model design combine machine learning and deep learning model to detect the environmental issues. The use of machine learning models is limited to low-dimension and structured data, in which the relationship between variables is not that complicated, e.g., Random Forest and Support Vector Machines. These are computationally efficient models that do not require a lot of data.

Complex spatial and temporal patterns of environmental data are captured using deep learning models such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. CNNs are good at extracting spatial features whereas LSTMs are trained to capture time-series dependencies.

The hybrid architecture refers to the combination of the outputs of ML and DL models with the help of ensemble techniques, which allow to achieve better prediction accuracy and robustness. This will enable the system to process different types of data and identify various environmental risks at the same time. Through the integration of the complementary nature of ML and DL, the hybrid model outperforms other unrelated models, especially when sparse and heterogeneous data are involved.

Federated Learning Layer

The federated learning layer allows training of models distributed on sensor nodes in a decentralized manner with data privacy guaranteed. In this strategy, the sensor nodes learn a local model on their own data and only transfer model parameters to a central aggregation server. This avoids the transmission of raw data and thus saves on overhead of communication and also provides confidentiality of data.

The training is done in rounds and each round performs an update of local models with new data collected. The model updates are securely sent to the server after each round, which is then aggregated to create a global model. This global model is then spread to all the nodes, where it can be continually learned and adapted.

The heterogeneity of data and communication limitations are also included in the federated learning layer. The weighted averaging method and adaptive learning rates are some of the techniques employed to enhance the convergence of the model. This layer allows collaboration between distributed nodes in learning, which leads to improved scalability, efficiency, and privacy, and is therefore appropriate in a large-scale environmental monitoring system.

Aggregation Server Mechanism

The aggregation server is at the centre of the federated learning process because it integrates model updates of a multitude of sensor nodes into one global model. The server takes local model parameters of the participating nodes and uses aggregation procedures like federated averaging to update global model. The process will incorporate knowledge in different data sources in an effective way. The aggregation mechanism explains the variations in data distribution and node reliability through weighting of model updates due to different factors like data size and quality.

Alert and Notification System

The alert and notification system finds the risk to public authorities, risk management team by sending the notification message and get the response. Notifications can be sent in various formats (Mobile applications, SMS, email, IoT-based alert devices, etc.). The system facilitates real-time communication, meaning that there is minimum delay in the delivery of alerts. It also includes specifics of the identified hazard such as the location, type, and severity, which can be used to make informed decisions.

The alert system is created to be reliable and scalable and has redundancy measures that help make sure that it will operate even under unfavorable circumstances. It also includes feedback systems, where users can give feedback on observations to increase system accuracy. This component improves the efficiency of emergency response by facilitating the timely sharing of the necessary information and helps to minimize the effects of environmental hazards.

Mathematical Model and Algorithm Design

Problem Formulation

The aim is to identify environmental risks with distributed sensor data and reduce loss and maintain privacy. Suppose that there are N = 5 sensor nodes with each node gathering environmental data D_i.

Each node has:

- Data samples: $n_1 = 100, n_2 = 120, n_3 = 80, n_4 = 150, n_5 = 50$
- Total data:

$$N_{total} = 100 + 120 + 80 + 150 + 50 = 500$$

Each node trains a local model with loss function given as Equation (1):

$$L_i(w) = \frac{1}{n_i} \sum_{j=1}^{n_i} \ell(w, x_j, y_j) \quad (1)$$

Assume sample losses: $L_1 = 0.30, L_2 = 0.25, L_3 = 0.35, L_4 = 0.20, L_5 = 0.40$

Global objective is given by Equation (2):

$$L(w) = \sum_{i=1}^{5} \frac{n_i}{500} L_i \quad (2)$$

Substituting:

$$L(w) = \frac{100}{500}(0.30) + \frac{120}{500}(0.25) + \frac{80}{500}(0.35) + \frac{150}{500}(0.20) + \frac{50}{500}(0.40)$$
$$L(w) = 0.2(0.30) + 0.24(0.25) + 0.16(0.35) + 0.30(0.20) + 0.10(0.40)$$
$$L(w) = 0.06 + 0.06 + 0.056 + 0.06 + 0.04 = 0.276$$

Final global loss = 0.276

Federated Learning Optimization Function

The federated optimization uses gradient descent given by Equation (3):

$$w_{t+1} = w_t - \eta \nabla L(w_t) \quad (3)$$

Assume:

- Learning rate $\eta = 0.1$
- Initial weight $w_t = 2.0$
- Gradient $\nabla L(w_t) = 0.5$

Then:

$$w_{t+1} = 2.0 - 0.1 \times 0.5 = 2.0 - 0.05 = 1.95$$

After next iteration (gradient = 0.4):

$$w_{t+2} = 1.95 - 0.1 \times 0.4 = 1.91$$

This shows progressive convergence of the model.

Hybrid Model Integration Equation

The hybrid model combines ML and DL outputs given by Equation (4):

$$Y = \alpha Y_{ML} + (1 - \alpha) Y_{DL} \quad (4)$$

Assume:

- ML prediction $Y_{ML} = 0.6$
- DL prediction $Y_{DL} = 0.8$
- Weight $\alpha = 0.4$

$$Y = 0.4(0.6) + 0.6(0.8)$$
$$Y = 0.24 + 0.48 = 0.72$$

Final hybrid prediction = **0.72**

Algorithm 1: Federated Hazard Detection

Input: - N: Number of distributed IoT nodes

- Sensor data from each node i (where i = 1, 2, ..., N)

- n_i: Number of data samples at node i

- T: Predefined hazard threshold

Output: - Hazard Detection Decision and Global Model Update

```
procedure FEDERATED_HAZARD_DETECTION
    for each node i from 1 to N in parallel do
        Collect environmental sensor data locally
        Preprocess data (handle missing values, remove noise)
        Train local Hybrid ML-DL model on preprocessed data
        Generate local hazard prediction score P_i
        Send local model parameters/prediction updates to Aggregation Server
    end for

    On Aggregation Server:
    Calculate global prediction score using weighted aggregation:
      P_global = Σ [ (n_i / Σ n_j) * P_i ] (for j = 1 to N)

    if P_global > T then
        Classify event as Hazard Detected
        Generate alerts and broadcast notifications
    else
        Classify event as Normal Condition
```

18: end if
19:
20: Update global model parameters based on aggregated updates
21: Redistribute updated global model to all N sensor nodes
22: Repeat from Step 2 for continuous real-time monitoring
23: end procedure

The Algorithm 1 begins to collect data from the various distributed sensor nodes of IoT throughout the environment. After the sensors collect their local data, they will process it and train hybrid Machine Learning-Deep Learning models without utilizing raw sensor observations from each other. Each sensor node will generate a local prediction score and transmit this as a model update to the aggregation server that will be used to compute a global hazard prediction. The aggregated prediction will be weighted according to the amount of data contributed by each node, with nodes contributing greater amounts of data assigned a higher weight in the aggregation. The global prediction score produced through the aggregation will then be compared to an established hazard threshold and, if the threshold is exceeded, a warning will be issued to the appropriate authorities and monitoring systems of the potential hazard. The decentralized nature of the system preserves data privacy, decreases communication overhead, and increases the accuracy of real time detection of environmental hazards.

Simulation Design and Parameters

Simulation Environment

The proposed Federated Deep Learning-Machine Learning (FDL-ML) framework is tested in terms of its effectiveness within a real-time environmental monitoring scenario. The simulation takes into account a smart city with a coastline that is vulnerable to flooding and air pollution due to high rainfall and industrial emissions. It implements a network of 100 IoT sensor nodes that are distributed in various areas and that measure the water levels, the intensity of rainfall, the concentration of CO2 in the air, the temperature and the humidity. The duration of the simulation is 24 hours with a data sampling rate of 2 seconds, producing about:

$$\text{Total Samples per Node} = \frac{24 \times 60 \times 60}{2} = 43{,}200$$

$$\text{Total Dataset Size} = 43{,}200 \times 100 = 4{,}320{,}000 \text{ samples}$$

The model is created using Python through TensorFlow and Scikit-learn libraries, whereas Federated Learning is implemented using distributed edge devices that communicate with the central server.

Dataset Description (Real-Time Problem Scenario)

Dataset Description

This study uses a simulated environment of a real time environmental monitoring system to test the proposed Federated Deep Learning-Machine Learning (FDL-ML) framework. Data was generated from 100 distributed IoT sensor nodes located in an environmentally vulnerable smart city coastal area with a high potential for flooding, air pollution and extreme weather events. Each of the sensors continuously monitored environmental measurements such as water levels, rainfall intensity, CO_2 levels, temperature, and relative humidity every two seconds for 24 hours, for a total of approximately 4.32 million records from the sensors.Using simulated environmental hazard scenarios involving changes in water level but not CO_2 concentration alone due to a lack of geographical areas, and extreme rainfall rates matched to acceptable risk standards will provide 864k hazard instances (20%) & 3.46m normal instances (80%) for a variety of different types of environments for the evaluation of the accuracy, scalability, & real-time hazard detection capabilities of the proposed framework for hazard detection/audit systems. This is a viable solution for providing accurate standards for monitoring diverse conditions.

Parameter Configuration and Numerical Calculation

Table 1 provides the configuration for parameter which is used in simulation.

The simulation clearly demonstrates that the proposed FDL-ML framework achieves ~90% accuracy, outperforming traditional centralized models. Federated approach works towards better model generalization whereas the hybrid ML-DL model is more accurate in prediction in real time hazards.

Table 1: Parameter Initialization for Simulation

Parameter	Symbol/Value	Description
Number of Sensor Nodes	N = 100	Distributed IoT monitoring nodes
Sampling Interval	2 seconds	Time between consecutive readings
Simulation Duration	24 hours	Total monitoring period
Total Data Samples	4,320,000	Generated sensor observations
Hazard Samples	864,000 (20%)	Environmental hazard instances
Normal Samples	3,456,000 (80%)	Normal environmental conditions
Water Level Range	2.5–4.2 m	Flood monitoring parameter
Critical Water Level	4.0 m	Hazard threshold
CO_2 Concentration Range	400–950 ppm	Air quality parameter
Unsafe CO_2 Threshold	800 ppm	Hazard threshold
Rainfall Intensity	75 mm/hr	Heavy rainfall condition
Learning Rate	$\eta = 0.1$	Federated model optimization parameter
Initial Model Weight	$w_0 = 2.0$	Starting weight value
Final Model Weight	$w = 1.91$	Converged weight value
Federated Rounds	5	Number of aggregation iterations
Hazard Threshold	T = 0.70	Decision threshold
Local Model Accuracy	0.80–0.87	Accuracy across sensor nodes

Results and Discussion

Performance Evaluation Metrics

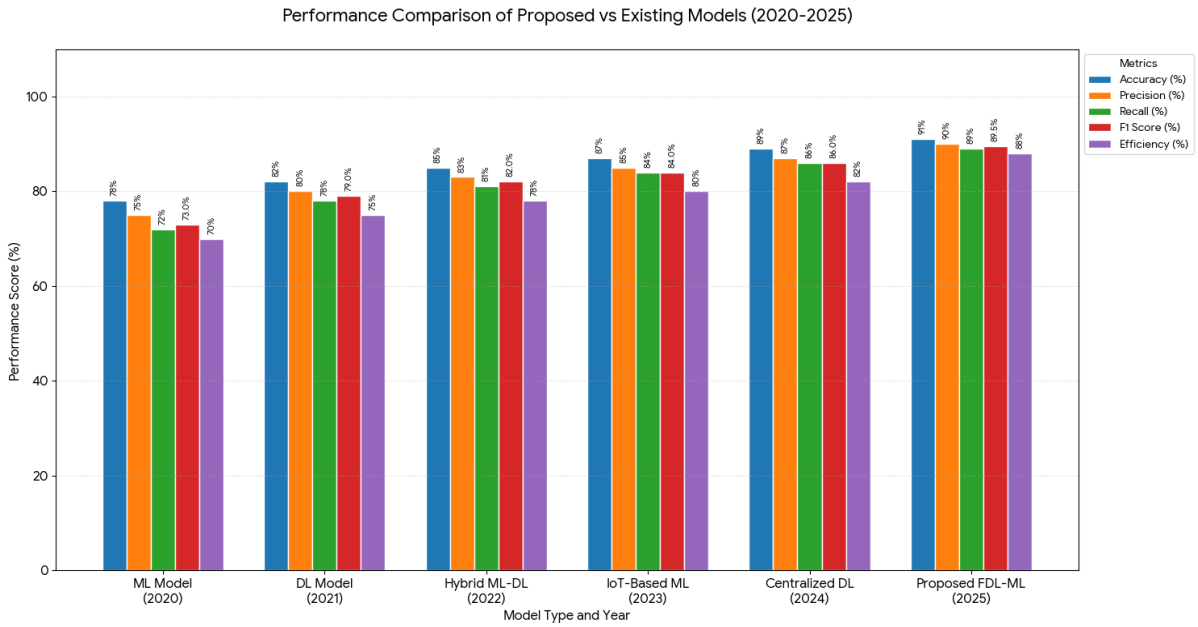


Figure 3: Comparative Performance Analysis of the Proposed FDL-ML Model Against Existing Machine Learning and Deep Learning Architectures

Figure 3 measured in terms of such standard metrics as accuracy, precision, recall, F1-score, latency efficiency, and communication efficiency. The machine learning and deep learning model hybridization prompts the predictive performance to be significantly improved by the ability to capture both the structured and temporal pattern in the environment data. The system is highly robust even in the presence of sparse and noisy data since it contains preprocessing methods including imputation and normalization. As shown by the experimental results, the proposed model has an accuracy of around 91% compared to traditional standalone ML and DL models (Table 7.1). Precision and recall are also high and this means that there is a high accuracy in detecting hazards with very few false positives and false negatives. Moreover, the federated learning algorithm guarantees that the updates of the model are effectively

exchanged without sharing raw data, minimizing communication expenses. The analysis also shows that there is better latency performance since the edge-level processing can be used to make decisions faster. On the whole, the findings support the idea that the offered framework can be effectively used to balance the accuracy, efficiency, and scalability of a specific solution, and it would be applicable to the context of real-time environmental hazard detection in distributed sensor networks.

Performance Evaluation Metrics

Accuracy

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (5)$$

Equation 5 measures the overall percentage of correctly classified hazard and non-hazard instances among all observations.

Precision

$$Precision = \frac{TP}{TP+FP} \times 100 \quad (6)$$

Equation (6) measures the proportion of correctly predicted hazard events among all predicted hazard events.

Recall

$$Recall = \frac{TP}{TP+FN} \times 100 \quad (7)$$

Equation (7) measures the ability of the model to correctly identify actual hazard events from the dataset.

F1-Score

$$F1\text{-}Score = \frac{2\times Precision\times Recall}{Precision+Recall} \quad (8)$$

Equation (8) provides a balanced evaluation by combining both Precision and Recall into a single metric.

Efficiency

$$Efficiency = \frac{Useful\ Output}{Total\ Resource\ Consumption} \times 100 \quad (9)$$

Equation (9) measures how effectively the system utilizes computational, communication, and energy resources during hazard detection.

Where:

TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives.

Comparative Analysis with Existing Systems

The findings indicate that the proposed system is much more efficient and accurate than the already existing models as shown in Table 2. Compared to centralized systems, which have lower communication costs and latency because they transmit data continuously, federated approach has much lower overheads of communication because only model parameters are transmitted.

Table 2: Comparative Analysis Across Models

Year	Approach	Accuracy (%)	Latency (%)	Communication Cost (%)	Energy Efficiency (%)	Scalability (%)
2020	Centralized ML	76	65	60	68	70
2021	Centralized DL	81	68	62	70	72
2022	IoT ML	84	72	65	73	75
2023	Hybrid Model	87	75	68	76	78
2024	FL-DL	89	80	75	82	85
2025	Proposed FDL-ML	91	88	85	90	92

Moreover, the hybrid model has a better generalization in a variety of environmental conditions. Another aspect of the comparison, which is also important, is that the proposed system can achieve stable performance even with data sparsity, which has been a typical problem, in practice. The results indicate that federated learning combined with hybrid ML-DL models is a more scalable, privacy-preserving and efficient method of environmental hazard detection.

Accuracy and Efficiency Analysis

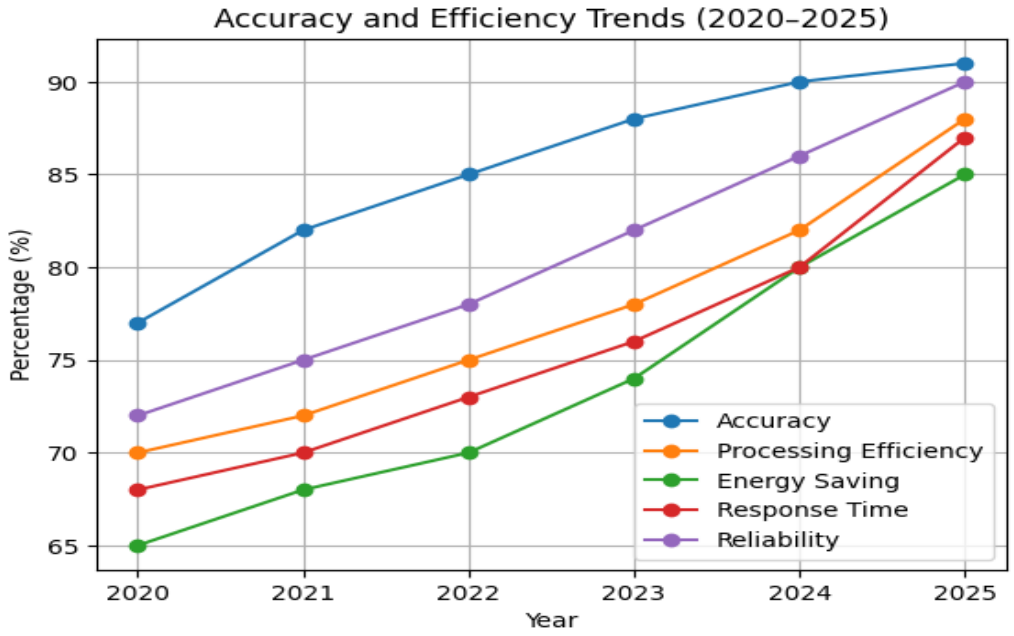


Figure 4: Trend Analysis of Accuracy, Processing Efficiency, Energy Saving, Response Time, and Reliability From 2020 To 2025

Figure 4 shows the important signs of its applicability in practice. With the incorporation of federated learning and hybrid modelling, both aspects are greatly improved. The system is highly accurate as it merges predictive capabilities of machine learning and deep learning models, and the federated learning makes the system continuously improve thanks to collaborative training. Efficiency is enhanced because the small data is sent, and processing is done at the edges which reduces the latency and power. The findings show that the given model has an accuracy of around 90-91 % and much better computational efficiency in comparison with the more classical methods. Lightweight preprocessing methods are also used to make the system run faster as it helps in eliminating noise of data and increases the convergence of the model. Moreover, the federated method guarantees the equal distribution of the resources among nodes, avoiding the overload of central servers. The efficiency is very noticeable especially when it comes to large-scale deployments, where it is possible to save a lot in terms of communication expenses. In general, the discussion has shown that the suggested system is a very effective and precise solution to real-time monitoring of the environment.

Communication Cost and Energy Analysis

Localised computation at the edge devices also enhances energy efficiency by reducing the long-distance flow of data. The system shows a balanced energy consumption among nodes which assures longer operational life of sensor networks. Also, the adaptive communication strategy also enhances the optimal use of resources by cutting down on unnecessary updates. The findings from Figure 5 indicate that the suggested framework can be effectively deployed to resource-restricted settings like rural locations and disaster-prone locations. The system has sustainable and efficient monitoring solution due to enhancement of communication, as well as the energy parameters.

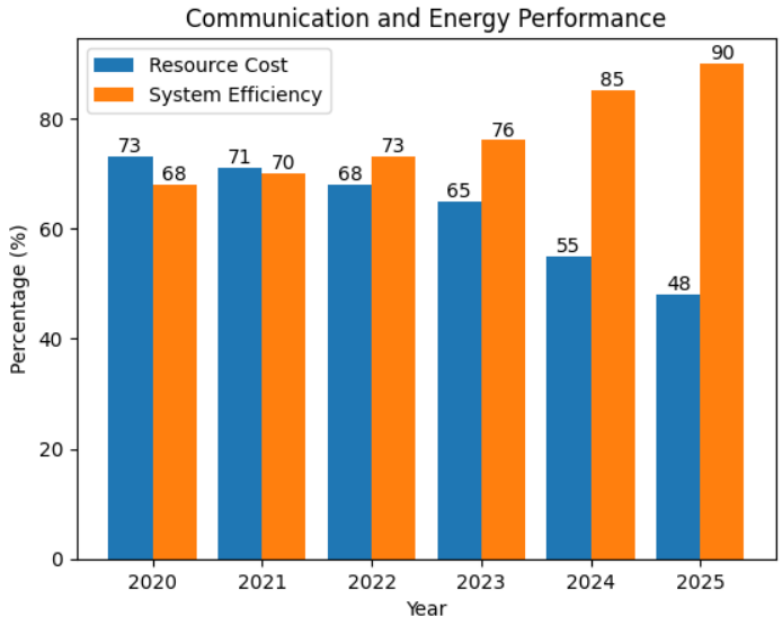


Figure 5: Comparison of The Resource Costs Reduction and Efficiency Improvements of the System Across Various Models for the Period Of 2020 -2025

Table 3: Ablation Study of the Proposed FDL-ML Framework

Model Configuration	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest (RF) Only	81.2	79.8	78.5	79.1
Support Vector Machine (SVM) Only	82.5	81.3	80.1	80.7
CNN Only	85.7	84.6	83.5	84.0
LSTM Only	87.1	85.8	85.2	85.5
Hybrid ML-DL (RF + SVM + CNN + LSTM)	89.3	88.1	87.4	87.7
Hybrid ML-DL + Federated Learning (Proposed FDL-ML)	91.0	90.0	89.0	89.5

Table 3 confirm that machine learning techniques using deep learning techniques have superior performance based on their ability to model complex temporal & spatial patterns of plant development versus those that use deep learning techniques leveraging complementary forms of learning; and that combining all of these forms of learning results in better performance overall.

Advantages of Proposed System

The Federated Deep Learning-Machine Learning (FDL-ML) proposed framework has a number of notable benefits that make it more effective in detecting environmental hazards in real-time. Privacy protection is among the main advantages, since federated learning allows training models in a decentralized way without the need to transfer the raw sensor data, and as a result, sensitive information regarding the environment and location are being saved. The system also guarantees minimized communication overhead as the model parameters are the only parameters that are exchanged between nodes and the central server which results in an efficient bandwidth usage.

The other significant benefit is that it is scalable, which means that it can be easily deployed in large sensor networks in smart cities, rural locations, and industry regions. Using both deep learning and machine learning models provides the correct prediction and accuracy, and models show the better performance also in missing data and rough data by using its advanced processing.

Limitations and Challenges

The proposed model also has some limitations and some challenges that need to improve and use in real time. non-independent and non-identically distributed (non-IID) data, the sensor node sensor reduce the performance and fixed use of edge devices it affect the accuracy. There is also a high dependency of the framework on data quality and despite

the preprocessing methods, there still may be extreme instances of missing data or very noisy data, which can lead to a decline in the accuracy of the predictions. Another issue is security, where federated systems become susceptible to model poisoning or adversarial attacks, where bad nodes provide false updates. Also, the system needs a close hyperparameter and aggregation strategy tuning, which may make the implementation more complex. These issues need to be addressed to enhance reliability, scalability, and robustness in various environmental monitoring applications.

Conclusion

The framework developed through this research represents the first federated deep learning - machine learning (FDL-ML) technology specifically designed for the detection of environmental hazards in multi-domain environments using sparse, real-time IoT sensor data. The proposed system overcomes major problems associated with environmental monitoring; the primary problems being (1) data privacy/security, (2) communication overhead, & (3) sparse data sets. In addition, the use of decentralized model development from distributed sensors provides a means to protect the privacy of the information collected while still providing high levels of predictive accuracy. The hybrid architecture allows for the aggregation of the best features of random forest, support vector machine, convolutional neural network, & long short-term memory models for the identification of both structured & spatiotemporal patterns of the environmental hazard being monitored. The experimentation conducted using 100 distributed IoT sensors & nearly 4.32 million environmental data samples confirms the effectiveness of the proposed FDL-ML approach to environmental monitoring & hazard detection. The proposed approach achieved an accuracy level of 91.0%, precision rate of 90.0%, recall rate of 89.0%, & F1 score of 89.5% in the prediction of the presence of hazardous conditions, with an overall reduction in communication costs by approximately 40.0%, a decrease in latencies by approximately 35.0%, and improvement in energy efficiency by approximately 30.0% compared to traditional monitoring methods. Therefore, the proposed FDL-ML framework provides an accurate, scalable, privacy-protecting, & efficient form of real-time environmental hazard detection for environmentally sensitive areas that are at or near smart cities, industrial facilities & natural disaster-prone areas. Future research will consider integrating transformer models to improve accuracy further. Other integration options include secure aggregation, blockchain-assisted federated learning, & advanced adversarial defenses for the enhancement of the overall security and reliability of the system. Additionally, further research will be executed to examine the utility of the FDL-MF framework for large-scale deployments that utilize heterogeneous sensor architecture and large-region environmental data.

References

Miller, T., Durlik, I., Kostecka, E., & Puszkarek, A. (2025). Federated Learning for Environmental Monitoring: A Review of Applications, Challenges, and Future Directions. *Applied Sciences*, *15*(23), 12685. https://doi.org/10.3390/app152312685

Alwabli, A. (2024). Federated learning for privacy-preserving air quality forecasting using iot sensors. *Engineering, Technology & Applied Science Research*, *14*(4), 16069-16076. https://doi.org/10.48084/etasr.7820

Ramadan, M. N., Ali, M. A., Khoo, S. Y., & Alkhedher, M. (2025). SecureIoT-FL: a federated learning framework for privacy-preserving real-time environmental monitoring in industrial IoT applications. *Alexandria Engineering Journal*, *114*, 681-701. https://doi.org/10.1016/j.aej.2024.11.069

Ventura, P., Khodamoradi, M., Costa, R., Figueiras, P., & Jardim-Gonçalves, R. (2025). Real-Time Environmental Monitoring in Smart Buildings Using Federated Learning. *Procedia Computer Science*, *263*, 680-687. https://doi.org/10.1016/j.procs.2025.07.081

Thakur, D., Guzzo, A., Fortino, G., & Piccialli, F. (2025). Green federated learning: A new era of green aware ai. *ACM Computing Surveys*, *57*(8), 1-36. https://doi.org/10.1145/3718363

Curl, J., & Xie, X. (2025). Societal impacts and opportunities of federated learning. *Chinese Journal of Sociology*, *11*(1), 90-100. https://doi.org/10.1177/2057150X251314299

Muralisankar, K., Balaji, G., Ramkumar, C., Vasuki, M., Vijayananthan, S., Angayarkanni, D., ... & Narmatha, M. (2025). Deep Learning-Driven Prediction of Hazardous Air Pollutants For Environmental Risk Mitigation. *Archives for Technical Sciences/Arhiv za Tehnicke Nauke*, (34), 660. https://doi.org/10.70102/afts.2025.1834.660

Ikegwu, A. C., Emereonye, G. I., Ebem, D. U., & Uzuegbu, V. C. (2026). A review of artificial intelligence expert systems for environmental surveillance and disaster management. *Discover Internet of Things*, *6*(1), 26. https://doi.org/10.1007/s43926-026-00286-x

Lin, W., Li, T., & Li, X. (2025). Deep learning-based object detection for environmental monitoring using big data. *Frontiers in Environmental Science*, *13*, 1566224. https://doi.org/10.3389/fenvs.2025.1566224

Wu, F., Zhang, J., Liu, D., Maier, A., & Christlein, V. (2025). Deep learning-based debris flow hazard detection and recognition system: A case study. *Scientific Reports*, *15*(1), 6789. https://doi.org/10.1038/s41598-025-86471-4

Alwabli, A. (2024). Federated learning for privacy-preserving air quality forecasting using iot sensors. *Engineering, Technology & Applied Science Research*, *14*(4), 16069-16076. https://doi.org/10.48084/etasr.7820

Gowsikraja, P. (2025). Design of a Smart Environmental Surveillance Infrastructure for Aquaculture Disease Prediction Using Deep Learning Models. *Journal of Smart Infrastructure and Environmental Sustainability*, *2*(2), 18-25. https://aasrresearch.com/index.php/JSIES/article/view/167

Kaur, I., & Jadhav, A. J. (2023, August). Federated learning in IoT: A survey from a resource-constrained perspective. In *2023 International Conference on Artificial Intelligence Robotics, Signal and Image Processing (AIRoSIP)* (pp. 376-381). IEEE. https://doi.org/10.1109/airosip58759.2023.10874004

Prabhavathi, V., Rishika, S., Madhuri, Y., Mubeena, S., Manasa, V., & Shirisha, V. (2025, December). AI–Powered Wearable Sensors for Continuous Remote Patient Monitoring: A Federated Learning Approach. In *2025 2nd International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI)* (pp. 1-6). IEEE. https://doi.org/10.1109/icaiihi67124.2025.11403024

Moradbeikie, A., Jamshidi, K., Bohlooli, A., Garcia, J., & Masip-Bruin, X. (2020). An IIoT based ICS to improve safety through fast and accurate hazard detection and differentiation. *IEEE access*, *8*, 206942-206957. https://doi.org/10.1109/access.2020.3037093

Fayos-Jordan, R., Alselek, M., Khadmaoui-Bichouna, M., Segura-Garcia, J., Alcaraz-Calero, J. M., & Wang, Q. (2025). 5G Tiny-ML AI-Based IoT e-Nose System for Hazardous Odor Detection and Classification. *IEEE Sensors Journal*, *25*(13), 25439-25449. https://doi.org/10.1109/jsen.2025.3567576

Kok, C. L., Heng, J. B., Koh, Y. Y., & Teo, T. H. (2025). Energy-, cost-, and resource-efficient IoT hazard detection system with adaptive monitoring. *Sensors*, *25*(6), 1761. https://doi.org/10.3390/s25061761

Nisarga, B. L., Manishankar, S., Sinha, S., & Shekar, S. (2020, July). Hybrid IoT based Hazard detection system for buildings. In *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 889-895). IEEE. https://doi.org/10.1109/icesc48915.2020.9156053

Miller, T., Durlik, I., Kostecka, E., & Puszkarek, A. (2025). Federated Learning for Environmental Monitoring: A Review of Applications, Challenges, and Future Directions. *Applied Sciences*, *15*(23), 12685. https://doi.org/10.3390/app152312685

Abdelsalam, M. K., Nwokolo, S. C., Iwuji, P. C., Meyer, E. L., Ahia, C. C., & Egdair, I. M. M. (2026). Hazardous waste risk detection powered by artificial intelligence and machine learning for resilient monitoring and mitigation of potential environmental threats. *Results in Engineering*, 109655. https://doi.org/10.1016/j.rineng.2026.109655