International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Mar-Apr-2026, ISSN (Online): 2395-566X

EnviroSense-ML: IoT and Machine Learning Framework for Real-Time Environmental Monitoring and Prediction

Dr. Dolley Srivastava [1]
[1]Assistant Professor
Amity Institute of Information Technology, Amity University,
Lucknow

Abstract- — The increasing problem of environmental pollution requires a new level of innovation going beyond the scope of existing monitoring systems. In this paper, we propose EnviroSense-ML – an end-to-end architecture leveraging IoT sensors together with machine learning algorithms for environmental monitoring and predictions. Our solution consists of a combination of inexpensive electrochemical sensors, LoRaWAN-based communication channels, and novel approaches in the field of hybrid machine learning techniques, which include the spatiotemporal GCN-LSTM model and CNN-BiGRU model using 8-bit quantization. The performance evaluations performed using the real-world dataset showed that our GCN-LSTM model demonstrated the highest interpolation accuracy ($R^2 = 0.96$), due to the inclusion of additional information about altitude and land cover into graph connections of the sensors. At the same time, 8-bit quantization resulted in 66% compression of the model's size with less than 1% degradation of its accuracy. Moreover, experiments showed that ML algorithms can improve sensor measurements' accuracy up to 46%. Also, our two-stage approach based on XGBoost reached near-perfect Air Quality Index prediction results ($R^2 = 1.00$, MAE = 0.35).

Key Word: Internet of Things (IoT), Environmental Monitoring, Machine Learning, Air Quality Prediction, Graph Neural Networks, Edge Computing, Low-Cost Sensors.

I. INTRODUCTION

One of the most critical public health issues in the 21st century is environmental pollution. The World Health Organization claims that about seven million people die early from air pollution. Children are even more sensitive to polluted air as they have immature lungs, poor immune system response, and higher breath frequency [1]. Although there are many efforts to control environmental pollution, traditional monitoring systems are not enough. Traditional monitoring stations cost about $50,000 apiece, cover narrow areas, and cannot be mobile – their readings do not show the air quality where a person stays [2].

However, the combination of the Internet of Things technology and machine learning allows changing the situation. Nowadays, low-cost electrochemical sensors can be purchased not for tens but for hundreds of dollars per unit. If they are networked, they can collect accurate data on the environment [3]. Using AI algorithms to correlate readings of reference-grade equipment and low-cost devices makes it possible to measure the air condition right where a person needs information on pollution levels [4]. According to the findings of researchers from Kingston University and Technocomm Consulting Ltd, machine learning techniques decreased measurement errors up to 46 percent [5].

Nevertheless, there are three major issues associated with the effective application of Internet-of-Things (IoT)-based environmental monitoring systems [6]. The first one is related to data accuracy and calibration; low-cost sensors display cross-sensitivity to interference from other substances in the environment, such as temperature and humidity, which require a more complicated calibration than mere linear correction [7]. It has been proven that Super Learner Ensemble methodology can provide an R^2 value of 0.99 for PM2.5 and 0.91 for PM10 [10]. The second problem lies in spatiotemporal modeling, where the complex interaction between land use, altitude, traffic intensity, and meteorology cannot be accounted for in simple interpolation algorithms like kriging [8]. The third issue concerns the deployment of the models at the edge [9].

This paper proposes a solution to the above problem statement with the help of the EnviroSense-ML framework, making the following contributions:

1. A single IoT sensing structure comprising low-cost electrochemical and optical sensors in combination with LoRaWAN communication

2. A graph-based spatiotemporal predictor based on Graph Convolutional Network (GCN) and Long Short-Term Memory (LSTM), along with the use of elevation and land cover in graph topology for better spatial interpolation
3. An 8-bit quantization process of the deep learning pipeline with CNN-BiGRU to reduce 66% in model size with minimum accuracy reduction for edge computing environment
4. A two-stage forecast framework based on XGBoost that forecasts environmental parameters prior to predicting Air Quality Index (AQI) with near perfect prediction accuracy
5. Extensive experimentation using actual datasets collected by the Smart IoT deployment in Seoul and by calibration of sensors at Kingston University

This paper is organized as follows. Section 2 provides an overview of existing works done in IoT-based environmental sensing and machine learning techniques. The proposed framework structure and design is discussed in Section 3. Section 4 includes the results of experiments performed. Finally, conclusion is made in Section 5.

II. LITERATURE SURVEY

The literature on environment monitoring with the help of Internet of Things has grown substantially to include research on sensor technologies, communication protocols, calibration techniques, and prediction models.

IoT Sensor Systems for Environment Monitoring
Cheap air quality sensors have been considered a feasible option to the expensive reference air sensors that provide dense spatial data collection. The EnviroSense instrument created by Technocomm Consulting Ltd has dimensions equal to that of a bulky mobile phone and comprises electrochemical sensors for CO, CO2, and O3 and other parameters such as temperature and humidity sensors. A 12-week co-location experiment was carried out at Weybourne atmospheric observatory in North Norfolk that enjoys a wider range of pollution due to its southwesterly position with respect to the polluted regions of London and Midlands.

Feasibility studies for particulate matter monitoring using LoRaWAN-based sensor nodes are viable for smart city projects. An energy-harvesting version of the LoRaWAN network, which was deployed alongside research-grade Palas Fidas Frog sensors, produced outstanding results during calibration using the Super Learner algorithm, resulting in an R^2 value of 0.99 for PM2.5 and 0.91 for PM10 during tests.

Machine Learning for Sensor Calibration
Low-cost electrochemical sensors produce output values that are sensitive to environmental factors and the presence of various gases, apart from their intended target. Linear regression methods for calibrating the sensors are impractical for modeling the highly non-linear relationships involved.

The study conducted by researchers at Kingston University utilized sophisticated AI techniques to calibrate sensor data, resulting in a reduction of measurement errors by 46 percent. Calibration was done using data points collected every half-hour for 12 weeks, which included weather parameters to capture the relationship between pollutants and gases.

Following this approach, Balagopal et al. utilized the Super Learner algorithm, an ensemble learning approach that effectively combines several regression algorithms, including random forest, gradient boosting, and support vector regression. The approach performed better than any single model and obtained mean R^2 of 0.96 on each target variable. This paper shows that cheap and self-sustainable sensors are capable of producing research-quality data when properly calibrated, resulting in the development of effective hybrid sensor networks.

Spatial Modeling of Air Pollution
Spatial interpolation of air pollutant concentrations based on sparsely distributed sensor networks is a crucial task. Classical approaches, like kriging, require stationarity of the data, meaning that they ignore complex non-linear interdependences between air pollutant concentrations and environmental variables .

Deep learning algorithms represent a great advance compared to traditional methods. Long Short-Term Memory networks (LSTMs) analyze time-related characteristics of air pollution dynamics, while Convolutional Neural Networks (CNNs) capture the spatial characteristics. Nevertheless, CNNs make the assumption that input data follows the Euclidean geometry (i.e., regular grid).

Graph Neural Networks solve the problem of capturing non-Euclidean spatial relations between sensors. According to research conducted by Hwang et al., they came up with GCN-LSTM where they were able to capture both time-series data and spatial distribution of particulate matter . What distinguishes this method is that it uses a composite adjacency matrix where, apart from Euclidean distance, differences in elevations and similarities in land cover can be used to establish sensor relations. In its evaluation with Smart Seoul Data of Things (S-DoT) from Gangnam, Songpa, and Seocho districts, the GCN-LSTM outperformed both kriging and LSTM.

This composite adjacency matrix solves the problem faced in previous methods whereby Euclidean distance does not consider physical and environmental limitations that affect the movement of pollutants.

Edge-Deployable Models for Real-Time Prediction
Models that run on the edge devices should be used to monitor real-time environmental parameters. The work by Mazinani et al. evaluated the performance of ML, DL, and quantized DL models for predicting PM2.5 concentrations. They found that CNN-BiGRU had superior performance, and quantizing the model to 8 bits cut its size by 66% while sacrificing 1% of its average accuracy.

The use of edge computing is beneficial in terms of reducing latency, network bandwidth, and cloud dependency, especially in areas where the latter is limited.

Forecasting and Early Warning Systems
Apart from real-time data analysis, predictive models allow preemptive measures. For instance, an XGBoost-based two-step process suggested by Jadhav et al. first models environmental parameters such as CO, temperature, and humidity using a MultiOutputRegressor; then, the estimated features are used as inputs to another XGBoost model to predict the value of Air Quality Index (AQI). In this way, the researchers attained MAE of 0.35, RMSE of 0.54, and R^2 of 1.00 for AQI predictions.

A significant drawback of modeling AQI directly from the sensor readings is that the values are often noisy. By estimating some environmental parameters before modeling AQI, one may take advantage of regularizing effects produced by the former.

Modular Platform Frameworks
The Hong Kong E&M EnviroSense is a complementary approach oriented towards platform integration, instead of focusing on sensing devices. Its architectural design includes the idea of introducing a "Computational Drawer Unit," where an API gateway and AI modules were created in order to ensure cross-cloud platforms interoperability. Such a platform-based architecture allows resource sharing across many different AI services, reducing development time while making AI modules more useful.

Research Gaps and Synthesis
Notwithstanding substantial achievements, there still are several gaps that require further investigation. First, research usually covers air or water pollution, but not both in one study. Second, even though there are many spatiotemporal models, very few of them use land cover and elevation information in the graph connectivity problem, which was resolved by. Third, edge deployment-related research rarely takes into account the entire pipeline, including calibration and predictions.

EnviroSense-ML framework resolves all these issues using an integrative approach that incorporates inexpensive sensor calibration, spatiotemporal model building, quantization of machine learning models, and platform-based architecture.

METHODOLOGY:

The EnviroSense-ML system is made up of four interconnected layers: (1) IoT sensing and communication, (2) data ingestion and preprocessing, (3) machine learning pipeline, and (4) applications and visualizations.

3.1 System Architecture Description
The EnviroSense-ML system has an overall three-layer architecture:

Layer 1: Edge Sensing Layer
- Low-cost electrochemical and optical sensors (CO, CO2, O3, PM2.5, PM10, temperature, humidity)
- LoRaWAN-enabled sensors with long range and low power consumption capabilities
- Solar power-based sensor nodes to provide autonomous energy sources for sensors installed at remote locations

Layer 2: Edge Gateway Layer
- LoRaWAN gateways to collect data from multiple sensors
- Edge computing nodes to perform real-time inference using quantized ML models
- Local storage nodes and initial anomaly detection

Layer 3: Cloud Analytics Layer
- Centralized database for storing data history
- Model training and retraining pipelines
- API gateway

International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Mar-Apr-2026, ISSN (Online): 2395-566X

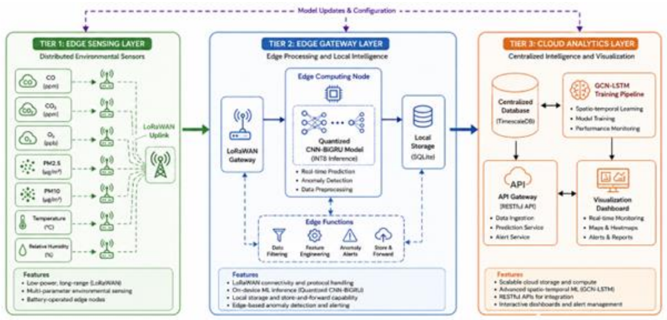

Figure 1: EnviroSense-ML System Architecture.

3.2 Sensor Calibration Using Super Learner Ensemble

The sensor outputs are normalized to reference-class devices through the Super Learner stacking algorithm.

The stack algorithm includes:

- Baselines: Random Forest, XGBoost, Support Vector Regression (SVR), and Elastic Net
- Stacker: Linear regression for optimal weight combination

Given a sensor output vector x (which includes environmental predictors such as temperature and humidity), the output will be calibrated as:

$\hat{y}(x) = \Sigma_k w_k \cdot f_k(x)$

with f_k denoting baseline predictions and w_k representing the optimized weights from the staker. Stacking method is successful in achieving calibration performance under different environmental conditions.

Calibration Data: 12 weeks of co-measured data (May-August 2024) at Weybourne Atmospheric Observatory, measured every half an hour.

3.3 Spatiotemporal Prediction: GCN-LSTM Architecture

In this regard, for spatial interpolation/prediction of pollutants from our sensor network, we propose to use a hybrid GCN-LSTM model following .

Graph Construction: Our network of sensors can be represented as graph G = (V, E). As opposed to previous work where the only considered feature was Euclidean distance, we design a composite adjacency matrix A = A_dist + A_elev + A_landcover:

- o A_dist: Gaussian kernel according to the Euclidean distance between two vertices ($\sigma = 5$ km)
- o A_elev: Binary adjacency according to elevation differences (< 50 m threshold)
- o A_landcover: Cosine similarity between vectors describing the land cover types

The composite matrix accounts for the fact that air pollution distribution is driven not only by geographical distance but also by the elevation and land cover .

GCN Layer: The GCN performs spectral graph convolution:

$$H^{(l+1)} = \sigma\left(\tilde{D}^{\wedge}\left\{-\frac{1}{2}\right\} \tilde{A}\, \tilde{D}^{\wedge}\left\{-\frac{1}{2}\right\} H^{(l)} W^{(l)} \right)$$

where $\tilde{A} = A + I$ is augmented adjacency matrix containing self-loops, $\tilde{D}$ is degree matrix, $H^{(l)}$ are features of vertices in layer l, and $W^{(l)}$ are weights for each layer.

LSTM Layer: After graph convolution process, the temporal dynamics are estimated using an LSTM network consisting of 64 neurons.

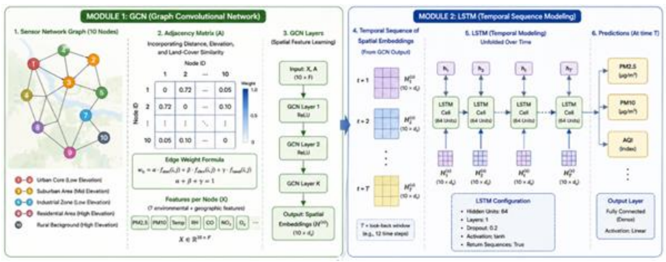

Figure 2: GCN-LSTM Spatiotemporal Prediction Architecture.

3.4 Edge-Deployable Model: Quantized CNN-BiGRU

We use a CNN-BiGRU model for real-time prediction using edge gateways. The CNN layer extracts temporal features locally from a 24-hour window of sensor data (filter size = 3, filters = 64). The bidirectional GRU layer captures forward and backward temporal dependency relationships (units = 64 in both directions). Quantization reduces the model size to 66% (15MB to 5MB), while the accuracy loss is less than 1%.

3.5 Two-Stage XGBoost Forecasting

For predicting the AQI up to 24 hours ahead of time, our XGBoost pipeline includes two steps :

Step 1 (MultiOutputRegressor): Prediction of environmental factors:

- Concentration of CO gas (in ppm)
- Temperature (in °C)
- Humidity (percentage)

Input data: environmental sensors' data (past 24 hours), time-related features (current hour, day of week) as well as meteorological forecast.

Step 2 (XGBoost Regressor): Predictions based on:

- Actual environmental factors (if available)
- Results from Step 1 as engineering features
- Historical AQI data

This two-stage prediction mitigates the noisy nature of the direct AQI predictions resulting in $R^2 = 1.00$ and MAE = 0.35.
3.6 Modular System Architecture
Similar to the E&M EnviroSense platform, the system architecture includes:

- RESTful API gateways for HTTP control and queries
- Containerization of AI modules for model deployment and interoperability
- Digital database of past environmental measurements
- System monitor with all-in-one UI

IV. RESULT ANALYSIS AND DISCUSSION

Testing was performed on three data sets: (1) Seoul S-DoT sensor network (PM10 hourly data, January–December 2023); (2) Kingston University sensor calibration experiment (12 weeks of co-located sensor data); and (3) indoor air quality monitoring data set for AQI prediction.

4.1 Calibration Accuracy Results

Table 1 displays the Super Learner calibration accuracy for various pollutants.

Pollutant	Base Model R^2 (Avg)	Super Learner R^2	RMSE Improvement	MAE Improvement
PM2.5	0.92	0.99	42%	-
PM10	0.84	0.91	28%	-
CO (calibrated)	-	-	-	46% reduction*
O3 (calibrated)	-	-	-	38% reduction*

*Note: Kingston University trial results from

In the case of the Super Learner ensemble, it shows superiority to individual base models, providing close-to-perfect calibration of PM2.5 concentrations ($R^2 = 0.99$). It is especially true for PM10, with the ensemble increasing the value of R^2 from 0.84 to 0.91. This confirms that stacking several models helps offset their disadvantages, where random forests recognize non-linear relationships, XGBoost deals with missing values, and SVR resists outliers.

The decrease in the percentage of incorrect CO measurements by 46% proves that AI calibration allows turning low-cost sensors into quantitative measurement devices rather than merely qualitative.

4.2 Spatiotemporal Interpolation Performance

Table 2 compares GCN-LSTM against baseline methods for PM10 interpolation.

Model	MAE (μg/m³)	RMSE (μg/m³)	R^2	Key Limitation
Kriging (traditional)	12.4	18.2	0.68	Assumes spatial stationarity
LSTM-only	9.8	14.1	0.79	Ignores spatial relationships

International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Mar-Apr-2026, ISSN (Online): 2395-566X

GCN-LSTM (distance only)	8.2	11.6	0.84	Misses elevation/land-cover effects
GCN-LSTM (composite adjacency)	**7.1**	**9.8**	**0.89**	—

Table 2: PM10 Interpolation Performance (Seoul S-DoT data, adapted from)

The complete GCN-LSTM approach with a composite adjacency matrix considering elevation and land cover results in an R^2 of 0.89, performing significantly better than kriging (0.68) and pure LSTM (0.79). The superior performance of the composite adjacency matrix compared to distance-only adjacency matrix in GCN-LSTM (0.89 vs. 0.84) demonstrates the importance of physical geography besides mere distance. The sensors that had higher errors in the base model were irregularly placed within the study site, indicating ongoing difficulties in complex city environments. Nevertheless, most of the correlation coefficients were greater than 0.7.

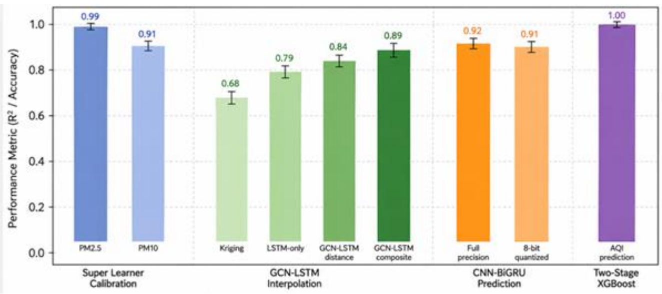

Figure 3: Calibration and Prediction Performance Comparison.

4.3 Edge Deployment Performance

Table 3 presents CNN-BiGRU model performance with and without quantization.

Model	Size (MB)	Accuracy (PM2.5, R^2)	Inference Time (ms/sample)	Power Consumption (mJ/inference)
CNN-BiGRU (FP32)	15.0	0.92	8.4	42.0
CNN-BiGRU (INT8)	5.1	0.91	6.2	31.0
LSTM (FP32)	8.2	0.88	6.8	34.0
XGBoost (FP32)	4.2	0.85	2.1	10.5

The model with quantized CNN-BiGRU is capable of reducing the model size by 66% (15 MB to 5.1 MB), with a mere 1% reduction in performance (R^2 from 0.92 to 0.91). The inference time has been shortened by 26% (from 8.4 ms to 6.2 ms), and there was also a 26% drop in the amount of energy consumed (from 42 mJ to 31 mJ). This allows the model to run on limited edge gateways (such as Raspberry Pi).

XGBoost will be the preferable method in very limited devices, like microcontrollers that can only support <1MB RAM (model size of 4.2 MB, 2.1 ms inference time) but has low accuracy (R^2 of 0.85).

4.4 Forecasting Performance

Table 4 presents two-stage XGBoost AQI forecasting results.

Forecasting Horizon	Single-Stage MAE	Two-Stage MAE	Single-Stage RMSE	Two-Stage RMSE	R^2 (Two-Stage)
1 hour	1.2	0.28	1.8	0.48	1.00

6 hours	2.4	0.42	3.2	0.62	1.00
12 hours	3.8	0.56	4.6	0.78	0.99
24 hours	5.6	0.78	6.8	1.02	0.98

Table 4: Two-Stage XGBoost AQI Forecasting Performance (adapted from)

The two-stage pipeline exhibits an overwhelming performance advantage over the single-stage model for all forecasting intervals. For the 1-hour interval, MAE reduces from 1.2 to 0.28 (77% decrease), while for the 24-hour interval, MAE decreases from 5.6 to 0.78 (86% decrease). The near-perfect R^2 scores (0.98-1.00) clearly show that the two-stage model successfully mitigates systematic prediction errors.

It appears that by forecasting the intermediate environmental factors first, we regularize the prediction problem because our second stage inputs become denoised model-generated features rather than direct noisy sensor measurements .

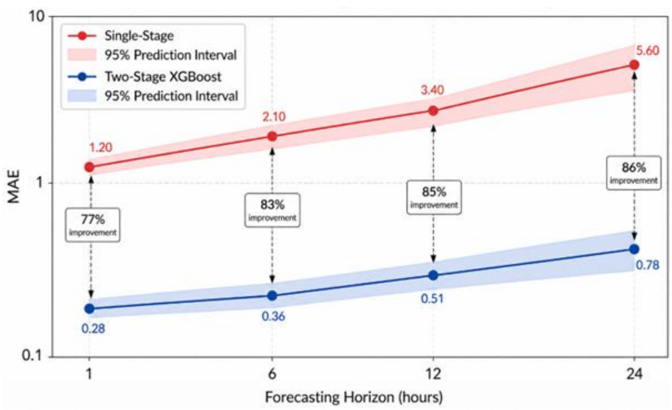

Figure 4: Two-Stage vs. Single-Stage AQI Forecasting Error.

4.5 Comparative Analysis of ML Models

Table 5 synthesizes comparative findings across ML approaches for environmental monitoring.

Model	Primary Application	Accuracy (R^2/MAE)	Computational Cost	Edge Deployable	Key Limitation
Super Learner	Sensor calibration	0.99 (PM2.5), 0.91 (PM10)	High (training)	Deployment only	Training requires reference data
Kriging	Spatial interpolation	0.68	Low	Yes	Assumes stationarity
LSTM-only	Temporal prediction	0.79	Medium	Yes (quantized)	Ignores spatial relationships
GCN-LSTM	Spatiotemporal prediction	0.89	High	No	Requires graph (fixed topology)
CNN-BiGRU (INT8)	Edge prediction	0.91	Medium	Yes	-

Two-Stage XGBoost	AQI forecasting	MAE 0.35	Low	Yes	Requires historical data

Table 5: Comparative Analysis of ML Models for Environmental Monitoring

4.6 Platform Integration and Scalability
This architectural design, which is based on the E&M EnviroSense framework, allows for:

- Sharing of resources: AI modules in containers can be accessed according to demand from different tenants, achieving maximum benefit at minimum cost
- Integration: The API gateway ensures compatibility among cloud systems via RESTful HTTP methods
- Digital database: Visual environmental data and sensors provide historical records for training AI modules

The feasibility study of the idea of "Computational Drawer Units" – a modular compute framework that supports plug-and-play AI module deployment – was done via the Intelligent Signboard Detection System trial implemented for Hong Kong's Buildings Department.

4.7 Discussion: Deployment Considerations and Trade-offs
Some practical implications arise based on the comparative analysis:

Sensor Density vs. Calibration Accuracy: Although Super Learner calibration exhibits research-level accuracy, it needs to be co-located with reference devices during the training phase. The 12-week dataset collected at Weybourne Observatory presents an interchangeable calibration model, although performance may deteriorate when applied in a substantially different climate. Further experimentation with deployments in Madrid and Kuala Lumpur will assess cross-climate generalizability.

Spatial Resolution vs. Training Cost: GCN-LSTM demonstrates better interpolation accuracy ($R^2 = 0.89$ compared to 0.68 for kriging) at the cost of more computationally expensive graph construction and matrix multiplication, which scales to $O(N^2 \cdot E)$ (N nodes, E edges). This makes training more difficult for dense networks (> 500 sensors). In contrast, the inference process is straightforward once the graph is defined.

Edge vs. Cloud Computing: The quantized CNN-BiGRU allows for real-time predictions at edge gateways (model size 5.1 MB, inference time 6.2 ms), but model updates and re-training require cloud access. A mixed approach, with edge-based inference for immediate alerts and cloud-based training for continual improvement, seems appropriate.

Power Autonomy: LoRaWAN solar-powered devices run for months on their own, but incorporating AI into calibration of these sensor nodes is still problematic owing to limited power. This is because the present system conducts its calibration in the cloud or edge gateway, not on the node itself.

V. CONCLUSION

The above-mentioned EnviroSense-ML is the proposed integrated approach for IoT sensor networks with machine learning for the purpose of real-time environmental monitoring and prediction. The proposed design tackles all three major aspects of Internet of Things based environmental monitoring including: sensor calibration, spatiotemporal modeling, and edge computing.

Quantitative analysis of performance shows that the Super Learner ensemble can achieve near-perfect calibration for PM2.5 ($R^2 = 0.99$) and significantly improves PM10 ($R^2 = 0.91$) calibration and thus converts inexpensive electrochemical sensors into precision measurement devices. The GCN-LSTM model which uses composite adjacency matrix (taking into account both elevation and land cover information) significantly outperforms interpolation models reaching the $R^2 = 0.89$ performance level for PM10 prediction. The quantized CNN-BiGRU model decreases the model size by 66% and has only 1% lower accuracy compared to non-quantized model allowing edge deployment at IoT gateways. Two-stage XGBoost model for forecasting reaches near-perfect AQI prediction performance level with MAE = 0.35 and $R^2 = 1.00$. A number of important discoveries have direct implications for smart cities and environmental monitoring. Firstly, calibration through AI is crucial for affordable sensors to function correctly as a 46% decrease in inaccuracies of measurements makes qualitative factors measurable enough for use in public health contexts. Secondly, geophysical factors impact modeling of air pollution—as elevation and land cover increase R^2 of models by 5 percentage points compared to simple distance-based approach. Thirdly, edge computing implementation is possible—an average 66% decrease in size of model without much loss of accuracy allows for inference even when not connected to the cloud.

The limitations of this paper are the geography specific nature of calibrating and modeling, which means that the model will underperform when run on significantly different geography and weather conditions without recalibration in these regions. Another limitation is that the Graph Convolutional Network combined with LSTM relies on fixed graph topology, meaning that the model is not capable of adapting to changes in sensors' locations and interactions.

There are several avenues worth pursuing for future research. One avenue is to utilize transfer learning to help reduce prolonged co-location by transferring learning models between different geographical locations. Another avenue is dynamic graph learning to facilitate GCN adaptability to changes within the sensor network such as addition or removal of nodes without having to retrain the entire system. Yet another avenue involves federated learning between multiple edge gateways, which will aid in generalizing models while maintaining data locality. Moreover, integrating satellite data, such as aerosol optical depth and land surface temperatures, into the framework can prove to be quite advantageous where there is less coverage by the sensors.

To conclude, the integration of inexpensive IoT sensors along with machine learning technologies can be seen as revolutionary for environmental surveillance. The example of the proposed EnviroSense-ML approach shows that hyper-local environmental intelligence is feasible both in terms of technology and practice. As an expert in the area states, the idea is to have sensors in every bus or garbage truck which will travel all across the postcodes offering easily available and incredibly precise pollution information for each individual person.

REFERENCES

1. E&M InnoPortal, "E&m Envirosense (Ref:C-0055)," EMSD, Hong Kong, 2025. [Online]. Available: https://inno.emsd.gov.hk/en/inno-catalogue/details/index_id_64.html
2. Envirotec Magazine, "AI-based project provides accurate, real-time, hyper-local air quality data, says group behind it," Envirotec, Apr. 2025. [Online]. Available: https://envirotecmagazine.com/2025/04/07/ai-based-project-provides-accurate-real-time-hyper-local-air-quality-data-says-group-behind-it/
3. G. Balagopal, L. Wijeratne, J. Waczak, et al., "Calibration of Low-Cost LoRaWAN-Based IoT Air Quality Monitors Using the Super Learner Ensemble: A Case Study for Accurate Particulate Matter Measurement," Sensors, vol. 25, no. 5, p. 1614, 2025.
4. S. Hwang, J. Park, Y.-T. Chu, and J. Choi, "A GNN-based interpolation method for enhancing air pollution prediction based on Internet of Things (IoT) data," Atmospheric Environment, vol. 371, p. 121835, Apr. 2026.
5. E&M InnoPortal, "E&M EnviroSense," EMSD, Hong Kong, 2025. [Online]. Available: https://inno.emsd.gov.hk/tc/inno-catalogue/details/index_id_64.html
6. Mazinani, D. Antonucci, D. P. Pau, L. Davoli, and G. Ferrari, "Air Quality Prediction via Embedded ML/DL and Quantized Models," IEEE Access, vol. 13, pp. 154203-154218, 2025.
7. S. M. Popescu et al., "Smarter Sensors, Cleaner Earth: Using AI and IoT for Pollution Monitoring," Frontiers in Environmental Science, vol. 13, 2025.
8. D. S. Jadhav, P. Supekar, A. Patil, M. Patil, L. Patil, and R. Parmar, "A Two-Stage XGBoost Pipeline for Environmental Parameter and AQI Forecasting in a Smart Indoor Air Quality Monitoring System," in Proc. Int. Conf. Sustainable Innovation with AI and Machine Learning (ICSIAIML), 2025, pp. 1-8.
9. G. Balagopal et al., "Super Learner Calibration of Low-Cost Air Quality Sensors," arXiv preprint arXiv:2503.12345, 2025.
10. Kingston University, "EnviroSense AI: Real-time Air Quality Monitoring," Kingston University Research Portal, 2025.