RESEARCH ARTICLE

EcoPulse-India: An AIoT Framework for Scalable, Privacy-Compliant Urban Air Quality Monitoring on Public Transit Fleets

Satya Prakash[1,*], **Om Prakash Nayak**[1], **Om Prakash Patra**[1], **Manas Ranjan Panigrahy**[1] and **Ashish Kumar Dass**[1]

[1] School of Computer Science and Technology, NIST University, Berhampur, Odisha, India

Abstract

India's stationary air-quality monitoring network remains sparse, particularly in underserved urban and peri-urban areas. EcoPulse-India addresses this limitation by using public transit buses as mobile sensing platforms within an integrated artificial intelligence of things (AIoT) framework. The framework integrates velocity-compensated aerodynamic correction (VCAC), a physics-informed convolutional graph neural network (PI-ConvGNN), multi-agent reinforcement learning (MARL)-based coverage optimisation, and privacy-respecting inference and masking architecture (PRIMA). VCAC compensates for motion-related measurement bias using vehicle kinematic and environmental information, while the quantised int8 PI-ConvGNN enables lightweight edge inference on an ESP32-S3 with a model size of approximately 38 KB. MARL prioritises underserved areas using the vulnerability-weighted coverage index (VWCI), whereas PRIMA applies on-device spatial perturbation and trajectory segmentation to reduce location-privacy risks. Simulation experiments seeded with real-world GPS traces from Delhi DTC and Mumbai BEST fleets showed that VCAC reduced mean absolute error by 22% relative to uncompensated measurements against reference-station data. MARL routing incentives raised city-level mean VWCI to 0.79—a 36.2% relative improvement over the static-routing baseline. Edge inference required 45 ms per pass with average power consumption below 0.8 W. PRIMA retained 94% of grid-cell-level spatial information required for pollution interpolation. These results demonstrate the potential of EcoPulse-India as a lightweight, privacy-aware mobile air-quality monitoring framework for resource-constrained urban environments.

Keywords: air quality monitoring, AIoT, edge computing, physics-informed neural networks, multi-agent reinforcement learning, data privacy, DPDPA 2023, urban sensing, India.

Submitted: 20 April 2026
Revised: 20 May 2026
Accepted: 05 August 2026
Published: 03 September 2026

Vol. 2, **No.** 3, 2026.
10.62762/NGCST.2026.429196

***Corresponding author:**
Satya Prakash
matsya1sm@gmail.com

Citation
Prakash, S., Nayak, O. P., Patra, O. P., Panigrahy, M. R., & Dass, A. K. (2026). EcoPulse-India: An AIoT Framework for Scalable, Privacy-Compliant Urban Air Quality Monitoring on Public Transit Fleets. *Next-Generation Computing Systems and Technologies*, 2(3), 89–98.

1 Introduction

India consistently ranks among the countries with the highest urban $PM_{2.5}$ burden globally—thirteen of the world's twenty most polluted cities were located in India in the most recent global air quality assessment—yet its Continuous Ambient Air Quality Monitoring Station (CAAQMS) network remains thin relative to the population it is meant to serve. Many tier-2 and tier-3 cities have no reference-grade monitors at all. Where stations exist, they are fixed, expensive to maintain, and concentrated in administrative districts rather than in the dense residential or industrial zones where exposure is highest. The result is a data vacuum that makes it difficult for regulators, urban planners, and residents to act on pollution in anything close to real time.

Low-cost sensor (LCS) networks offer an obvious partial remedy. Optical particle counters—particularly Plantower PMS-series devices—have dropped below \$5 USD per unit and can be networked over Wi-Fi or LoRaWAN with commodity microcontrollers. Several Indian cities have piloted static LCS grids [2], and the literature on their calibration is now substantial [3, 4]. What static grids cannot do, however, is reach the geographic periphery. Informal settlements, industrial corridors at city edges, and rapidly urbanizing peri-urban zones are structurally underserved: they lack the physical infrastructure (power, connectivity, security) needed to host fixed nodes, yet their residents often bear the highest pollution burden.

Mobile sensing—mounting LCS units on vehicles—addresses coverage directly. Public transit buses are a natural platform: they move on predictable, high-frequency routes, carry power supplies sufficient for edge computation, and are operated by organizations with existing maintenance infrastructure. DTC in Delhi and BEST in Mumbai together operate several thousand buses covering hundreds of route-kilometres daily across their respective networks. Instrumenting even a fraction of this fleet could produce a continuously updated, city-scale pollution map at a fraction of the cost of a comparable fixed network.

Three technical problems, however, have prevented mobile LCS deployments from reaching production scale.

Aerodynamic bias. Optical particle counters measure particle concentration in the air that enters their intake port. On a moving vehicle, that intake volume changes with bus speed and wind direction. A bus accelerating into a headwind pulls in more air per second than a stationary sensor in still conditions; a bus traveling with a tailwind pulls in less. Standard calibration procedures conducted under static lab conditions do not account for this, and the resulting concentration errors are not random—they correlate with speed and heading in ways that systematically distort the data [5].

Sparse-sensor interpolation. Reconstructing a continuous city-scale pollution field from a moving set of point measurements is an underdetermined problem. Data-driven spatiotemporal models (e.g., standard GNNs) can fit training data well but produce physically implausible outputs under distribution shift—pollution gradients running against prevailing wind, or concentration peaks appearing in locations insulated from emission sources [6, 7]. In a deployment context, implausible outputs erode trust and can trigger inappropriate regulatory responses.

Data privacy. GPS-timestamped sensor readings trace the movement of individual buses and, by extension, identifiable drivers or commuters. India's DPDPA 2023 establishes strict purpose-limitation and data-minimisation obligations for processors of personal data. Raw trajectory uploads would expose fleet operators to legal liability and create a surveillance record that is both unnecessary for pollution monitoring and difficult to delete once stored in the cloud [8].

EcoPulse-India integrates solutions to all three problems—plus a fourth concerning equitable spatial coverage—into a single deployable framework. This paper describes the architecture, the design decisions behind each subsystem, and the simulation results obtained using real Delhi and Mumbai fleet traces.

2 Related Work

2.1 Low-Cost Sensors and Aerodynamic Bias

The Plantower PMS5003 and PMS7003 have been deployed in dozens of urban LCS studies since 2017. Their performance under controlled conditions is well-documented: uncorrected readings track reference-method $PM_{2.5}$ with correlation coefficients above 0.9 in dry, low-humidity conditions, degrading in humid environments due to hygroscopic growth of particles [4]. Most published calibration approaches use linear or polynomial regression against reference instruments, sometimes augmented with humidity and temperature corrections.

What the literature addresses less thoroughly is intake-dynamic bias on moving platforms. A handful of studies using mobile low-cost sensors have noted measurement variability attributable to vehicle motion [5], and attributed it qualitatively to changes in dynamic pressure at the intake port, but quantitative fluid-dynamic models of the effect are rare. Standard machine learning calibration pipelines treat each measurement as independently drawn from the sensor's operating distribution; they do not model the physical relationship between vehicle kinematics and intake air volume. The VCAC subsystem described in Section 3.1 attempts to close this gap.

2.2 Spatiotemporal Pollution Modelling

Graph neural networks have become the dominant method for spatiotemporal interpolation of urban environmental data. Architectures such as DCRNN [6] and STGCN [7], originally developed for traffic forecasting, have been widely adopted for this purpose owing to their ability to learn spatial correlations across sensor graphs and temporal dynamics through recurrent or convolutional layers. Their limitation is that they are purely data-driven: they can reproduce patterns present in training data but have no mechanism for enforcing physical consistency in novel conditions.

Physics-informed machine learning (PIML) inserts differential equation constraints—typically from fluid dynamics or atmospheric transport—into the training objective, penalising outputs that violate known physics [9]. This has been demonstrated on pollution interpolation problems using full-precision models on server hardware [10], but deploying PIML on microcontrollers introduces a second constraint: the model must fit within the memory and compute budget of the target device. Quantisation (reducing weight precision from float32 to int8) reduces model size by roughly $4\times$ and inference latency proportionally, at some accuracy cost. The int8 quantised PI-ConvGNN described in Section 3.2 is, to our knowledge, the first physics-informed pollution model validated for deployment on an ESP32-class processor.

2.3 Coverage Optimisation and Environmental Justice

Classical sensor placement optimisation uses information-theoretic criteria (entropy, mutual information) to maximise coverage given a fixed sensor budget [11]. These approaches treat all spatial cells as equally valuable. Environmental justice literature argues otherwise: communities with higher demographic vulnerability—lower income, pre-existing health burdens, proximity to industrial emission sources—have both greater need for pollution data and, historically, lower access to it [12]. Incorporating vulnerability weighting into sensor placement has been proposed conceptually but its integration into a dynamic, MARL-driven mobile routing framework has not been demonstrated at fleet scale. Recent work by Ji et al. [13] has addressed combinatorial optimisation of mobile sensor deployment on bus fleets using trip-based scheduling, but without reinforcement learning or demographic vulnerability weighting in the objective.

2.4 Edge AI and Data Privacy in India

DPDPA 2023 is India's first comprehensive data protection statute [8]. It applies to any processing of digital personal data, including location trajectories generated by GPS-equipped devices. Existing LCS deployments in India have generally treated privacy as an afterthought: analyses of urban air quality data governance in Indian smart cities highlight that data-minimisation and purpose-limitation considerations are not yet consistently integrated into deployment practice [14], leaving raw location traces exposed to secondary use. On-device redaction—processing location data locally before transmission—is the technically sound alternative, but it requires lightweight inference capable of running within the power and memory constraints of embedded processors [15]. The broader challenge of deploying intelligence on resource-constrained IoT nodes, including the standardisation and interoperability constraints that govern such deployments, has been documented extensively in the IoT systems literature [16]. The PRIMA module described in Section 3.4 implements this approach.

3 Methodology

Figure 1 shows the three-layer EcoPulse-India AIoT architecture, comprising the physical, edge, and virtual/cloud layers. The physical layer integrates mobile sensing platforms, fixed CPCB reference stations, and the embedded sensing hardware. The edge layer performs local processing, including VCAC aerodynamic correction, PRIMA privacy-preserving preprocessing, lightweight PI-ConvGNN inference, data buffering, communication, positioning correction, and sensor lifecycle management. The virtual/cloud layer performs city-scale pollution-field reconstruction, MARL-based coverage optimisation, CPCB data fusion, compliance reporting, AQI visualisation, and hotspot alerting. The following subsections describe the principal computational components of the framework.

3.1 Velocity-Compensated Aerodynamic Correction (VCAC)

The core intuition behind VCAC is straightforward: a PMS sensor mounted on a bus does not sample a fixed volume of ambient air per second. It samples whatever volume is pushed or pulled through the intake port by the combination of the bus's motion and the ambient wind. If the bus is moving at 40 km/h into a 10 km/h headwind, the effective intake velocity is 50 km/h; if it is moving with the same wind, the effective intake velocity is 30 km/h. The raw particle count per unit time therefore reflects the particle count *per unit intake volume*, which is not constant.

VCAC models this effect using a computational fluid dynamics (CFD) proxy trained on simulated intake flows across the geometric envelope of a standard DTC-class city bus. The proxy takes two inputs available in real time on every instrumented bus:

- Vehicle speed and heading from the onboard GPS module (sampled at 1 Hz).
- Ambient wind speed and direction, obtained from the nearest meteorological station and transformed into the vehicle coordinate frame using GPS-derived vehicle velocity and IMU-derived orientation.

The proxy outputs a scalar correction factor $\kappa(v, \theta)$, where v is vehicle speed and θ is the wind-relative heading angle. Corrected concentration C_{corr} is computed as:

$$C_{\text{corr}} = \frac{C_{\text{raw}}}{\kappa(v, \theta)} \tag{1}$$

The CFD proxy was developed using OpenFOAM simulations covering vehicle speeds from 0 to 70 km/h and wind angles from 0 to 360 in 15 increments, then distilled into a lookup table compact enough to reside in ESP32 flash memory. Wind angle is updated at 10 Hz using IMU heading integration with GPS-derived correction to prevent drift.

VCAC is applied continuously in the bus sensor unit's interrupt service routine, before any data is passed downstream. This means all measurements transmitted off-device are already aerodynamically corrected; the cloud layer never receives raw counts.

3.2 Physics-Informed Convolutional Graph Neural Network (PI-ConvGNN)

The PI-ConvGNN operates in two deployment modes. At the edge, a quantised lightweight variant runs on the ESP32-S3 for anomaly detection and short-horizon interpolation within the local sensing neighbourhood. In the cloud, a full-precision city-scale variant aggregates measurements from mobile and fixed sensing nodes to reconstruct the spatial distribution of $PM_{2.5}$. The model combines graph-based spatiotemporal learning with a physics-informed regularisation term derived from atmospheric advection and diffusion.

3.2.1 Graph construction

Mobile bus-mounted sensors and fixed CAAQMS reference stations are represented as nodes in a dynamic spatial graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ denotes the set of sensing nodes and $\mathcal{E}$ denotes their spatial connections. Because the locations of mobile sensors vary over time, the graph topology is updated every 5 min using a k-nearest-neighbour (kNN) procedure with $k = 5$.

For two connected nodes i and j, the spatial distance is defined as d_{ij}. Edge weights are assigned according to inverse distance as

$$w_{ij} = \frac{1}{d_{ij} + \delta}, \tag{2}$$

where δ is a small positive constant introduced to avoid numerical instability when two sensing locations are very close. The resulting weighted graph allows the model to aggregate information from spatially neighbouring measurements while adapting to the movement of instrumented buses.

3.2.2 Physics-informed training objective

The data-driven component of PI-ConvGNN minimises the mean squared error between predicted and observed pollutant concentrations at the supervised sensing nodes. The corresponding loss is defined as

$$\mathcal{L}_{\text{data}} = \frac{1}{N} \sum_{i=1}^{N} \left(\hat{C}_i - C_i \right)^2, \tag{3}$$

where C_i and $\hat{C}_i$ denote the observed and predicted $PM_{2.5}$ concentrations, respectively, and N is the number of supervised observations.

Figure 1. EcoPulse-India three-layer AIoT architecture. The physical layer integrates public-transit and ride-sharing fleets, CPCB reference stations, and embedded sensing nodes. The edge layer provides local calibration, privacy-preserving preprocessing, lightweight neural inference, positioning correction, data buffering and communication, and sensor lifecycle management. The virtual/cloud layer performs city-scale PI-ConvGNN inference, MARL-based coverage optimisation, CPCB data fusion, compliance reporting, AQI visualisation, and hotspot alerting.

To improve physical consistency, a physics-informed regularisation term is introduced based on a simplified advection–diffusion equation:

$$\frac{\partial C}{\partial t} + \mathbf{u} \cdot \nabla C = D\nabla^2 C, \tag{4}$$

where C denotes pollutant concentration, $\mathbf{u}$ is the ambient wind-velocity field obtained from the nearest available meteorological station, and D is the effective eddy diffusivity coefficient. Local emission and removal processes are not explicitly represented because spatially and temporally resolved source–sink inventories were unavailable for the simulation. Their unresolved effects are therefore learned implicitly by the data-driven component of the model.

For the dynamic sensor graph, the diffusion component is approximated using the weighted graph Laplacian. Let $\hat{\mathbf{C}}$ denote the vector of predicted concentrations at all graph nodes. The graph-based Laplacian is defined as

$$\mathbf{L} = \mathbf{D}_g - \mathbf{W}, \tag{5}$$

where $\mathbf{W} = [w_{ij}]$ is the weighted adjacency matrix and $\mathbf{D}_g$ is the diagonal degree matrix with $D_{g,ii} = \sum_j w_{ij}$. The spatial diffusion term is then approximated as

$$\nabla^2 \hat{C} \approx -\mathbf{L}\hat{\mathbf{C}}, \tag{6}$$

with the sign convention determined by the adopted graph-Laplacian definition. Temporal derivatives are approximated using consecutive model outputs:

$$\frac{\partial \hat{C}}{\partial t} \approx \frac{\hat{C}^{\,t} - \hat{C}^{\,t-\Delta t}}{\Delta t}. \tag{7}$$

The resulting physics residual is expressed as

$$R_{\text{phys}} = \frac{\partial \hat{C}}{\partial t} + \mathbf{u} \cdot \nabla \hat{C} - D\nabla^2 \hat{C}, \tag{8}$$

and the physics-informed loss is

$$\mathcal{L}_{\text{phys}} = \frac{1}{N} \sum_{i=1}^{N} \left(R_{\text{phys},i}\right)^2. \tag{9}$$

The overall training objective combines the data-fitting and physical-consistency terms:

$$\mathcal{L} = \mathcal{L}_{\text{data}} + \lambda \mathcal{L}_{\text{phys}}, \tag{10}$$

where λ controls the contribution of the physics-informed regularisation. In the reported experiments, $\lambda = 0.3$ was selected based on validation performance on held-out CAAQMS reference stations.

3.2.3 *Edge quantisation*

After training the full-precision model, post-training int8 quantisation was applied using TensorFlow Lite to obtain the lightweight edge variant. A representative calibration dataset containing 10,000 inference samples from the Delhi simulation corpus was used to determine the quantisation ranges. The resulting serialised int8 model occupies approximately 38 KB, compared with approximately 150 KB for its float32 counterpart.

On the ESP32-S3 operating at 240 MHz, the quantised model completed a single inference pass in approximately 45 ms. The average measured power consumption during inference was 0.74 W, while idle power consumption between inference cycles at a 1 Hz update rate was approximately 0.12 W. These results indicate that the quantised PI-ConvGNN can support lightweight edge inference within the computational and power constraints of the proposed mobile sensing platform.

3.3 MARL and Vulnerability-Weighted Coverage Index (VWCI)

3.3.1 *VWCI formulation*

The VWCI scores each 500 m $\times$ 500 m grid cell g in the city by combining spatial data gap and demographic vulnerability:

$$\text{VWCI}(g) = \alpha \cdot \text{DataGap}(g) + (1 - \alpha) \cdot \text{Vuln}(g) \tag{11}$$

where $\text{DataGap}(g)$ is the fraction of the past 24 hours with no sensor measurement within 500 m of cell g, $\text{Vuln}(g)$ is a composite index combining census poverty rates, population density, proximity to emission sources (industrial areas, major roads, landfills), and distance to healthcare facilities. $\alpha = 0.5$ in baseline experiments.

3.3.2 *MARL routing*

Each instrumented bus operates as an independent agent. At each decision point (bus terminus or major junction), the agent selects a route from the set of feasible routes using a policy gradient algorithm. The reward function is:

$$r_t = \sum_{g \in \text{route}} \text{VWCI}(g) - \beta \cdot \Delta\text{schedule} \tag{12}$$

where Δschedule is the deviation from the published timetable and β penalises service disruption. Agents share a global VWCI map updated every 30 minutes but act independently, avoiding the coordination overhead of centralised multi-agent approaches.

The MARL dispatch app (Figure 1) communicates routing recommendations to drivers via a lightweight Android application, providing INR 2–5 micro-incentives per completed high-VWCI route segment, funded through municipal air quality programme budgets. Drivers are not required to follow recommendations; the incentive structure is designed to increase voluntary compliance without mandating route changes that could affect transit reliability.

3.4 PRIMA: Privacy-Respecting Inference and Masking Architecture

PRIMA operates as a preprocessing stage that runs on the ESP32-S3 before any data leaves the bus unit. It performs two operations:

Table 1. EcoPulse-India performance summary.

Metric	Value	Baseline
MAE reduction (VCAC)	22%	uncompensated LCS
VWCI improvement (MARL)	+36.2% (0.58 → 0.79)	static routing
Edge inference latency	45 ms	—
Average edge power draw	< 0.8 W	—
Int8 model size	38 KB	150 KB (float32)
PII retention after PRIMA	< 6%	100% (raw trace)
Spatial resolution retention	94%	—

Spatial fuzzing. Each GPS coordinate is perturbed by a random offset drawn from a Laplace distribution with scale parameter $b = 150$ m—sufficient to prevent precise route reconstruction while preserving enough spatial information for 500 m-resolution pollution mapping. The perturbation is consistent within a 30-second window (preventing temporal re-identification from perturbation patterns) and refreshed independently at each window boundary.

Trajectory segmentation. Continuous GPS traces are broken into independent 30-second segments before transmission. Each segment is tagged with an anonymised session identifier that changes every 30 minutes. This prevents a cloud-side observer from linking segments into a continuous route, since no persistent identifier connects them across sessions.

PRIMA was designed to support the principles of purpose-limited processing and data minimisation under India's Digital Personal Data Protection (DPDP) framework. By performing spatial perturbation and trajectory segmentation locally before transmission, PRIMA reduces the amount and granularity of potentially linkable location information leaving the sensing device. However, these technical safeguards should not be interpreted as establishing legal compliance by themselves. Compliance ultimately depends on the specific processing context, lawful basis, notice, consent where applicable, retention practices, and other obligations imposed by the DPDP Act and Rules.

4 Experiments

All experiments used simulated environments seeded with real GPS traces from the DTC Delhi fleet (312 buses, routes spanning 2,800 km of road network) and the BEST Mumbai fleet (274 buses, 1,900 km). CAAQMS ground truth was drawn from the 40 active Delhi CPCB stations and 22 active Mumbai stations. Simulation covered a 72-hour period under three representative meteorological regimes: winter inversion (low wind, high $PM_{2.5}$), monsoon onset (variable wind, moderate $PM_{2.5}$), and dry summer (moderate wind, moderate $PM_{2.5}$). Table 1 summarizes the headline performance metrics across all four subsystems, each of which is analysed in detail in the following subsections.

4.1 Aerodynamic Correction (VCAC)

Across both fleet datasets, VCAC-corrected $PM_{2.5}$ readings showed 22% lower mean absolute error against the nearest CAAQMS reference compared with uncompensated readings from the same sensor units. The improvement was largest during periods of high bus speed (above 35 km/h on arterial roads) and strong crosswinds—conditions where uncorrected readings showed systematic overestimates of 30–45% relative to reference. Under low-speed urban-core conditions (below 15 km/h), the correction effect was smaller (8–12% MAE reduction), consistent with reduced dynamic pressure variation at low speeds.

The VCAC correction factor surface $\kappa(v, \theta)$, derived from the OpenFOAM simulation corpus, is non-linear: intake bias is most severe at intermediate speeds and at heading angles near 45° and 315° (quartering winds), reflecting the geometric asymmetry of the bus intake port placement relative to the prevailing airstream.

4.2 Edge Inference (PI-ConvGNN)

The int8 quantised PI-ConvGNN completed a single inference pass in 45 ms on the ESP32-S3 at 240 MHz. Average power consumption during inference was 0.74 W; idle power between inference cycles (at 1 Hz update rate) was 0.12 W. The model fits in 38 KB of SRAM with no flash-based weight streaming required, leaving sufficient headroom for VCAC lookup tables and PRIMA buffers within the 512 KB available.

Physics constraint violations (instances where model output exceeded the advection-diffusion residual threshold ϵ) occurred in 2.1% of outputs on the held-out test set for the quantised edge model, versus 0.8% for the float32 cloud model. This difference is attributable to quantisation rounding in the gradient penalty terms during training and is considered acceptable for the anomaly-detection function of the edge model; city-scale heatmap production uses the cloud model.

4.3 Coverage Optimisation (MARL + VWCI)

Starting from a baseline VWCI of 0.58 (representing the coverage distribution of the unmodified bus routes), MARL routing incentives raised city-level mean VWCI to 0.79—a 36.2% relative improvement over the static-routing baseline. The gains were concentrated in peri-urban residential zones and near industrial corridors that existing routes did not serve during peak measurement hours. Schedule deviation across the fleet averaged 3.2 minutes per trip, below the 5-minute threshold set by DTC operational requirements.

Driver uptake of routing recommendations averaged 61% of eligible route-choice decisions during the simulation period (modelled as probabilistic compliance based on prior survey data on driver response to similar navigation incentive schemes). Sensitivity analysis showed that VWCI gains remained above 25% at compliance rates as low as 40%, suggesting the system is robust to partial adoption.

4.4 Privacy Compliance (PRIMA)

After PRIMA redaction, a linkage attack using the Munkres optimal assignment algorithm recovered less than 6% of original GPS fix locations to within 200 m. Trajectory reconstruction—linking redacted segments into continuous routes—succeeded for fewer than 3% of sessions when tested against a 30-day trace corpus. Spatial information useful for pollution interpolation (cell-level presence, not precise location) was retained at 94% fidelity relative to unredacted data, as measured by the fraction of 500 m grid cells correctly flagged as visited.

5 Discussion

The VCAC results confirm what the fluid dynamics literature predicts but what operational LCS deployments typically ignore: the intake dynamics of a sensor mounted on a moving vehicle are not equivalent to a static sensor in ambient air. The 22% MAE reduction is likely a lower bound on the benefit; the simulation corpus covers standardised DTC bus geometry, and real-world installations on heterogeneous vehicle types (Ola/Uber cars, two-wheelers) would require geometry-specific CFD calibration. Extending VCAC to ride-sharing fleets is technically feasible but would require manufacturer cooperation to obtain intake geometry specifications.

The PI-ConvGNN results demonstrate the feasibility of deploying a compact physics-informed model on resource-constrained edge hardware. The quantised model exhibited a physics-residual violation rate of 2.1% on the held-out test set, compared with 0.8% for the float32 cloud model. The modest increase in physics-residual violations after quantisation likely reflects approximation errors introduced when the trained float32 weights and activations are mapped to int8 representations during inference. Because post-training quantisation is applied only after optimisation of the physics-informed training objective, it does not directly alter the gradient-based physics regularisation during training. Nevertheless, the observed difference indicates a trade-off between numerical precision and physical consistency. Further evaluation across alternative quantisation strategies, including quantisation-aware training and mixed-precision inference, would be useful to determine whether this gap can be further reduced.

The MARL coverage results have a caveat that deserves direct acknowledgment: the 61% compliance rate is modelled, not measured. Real driver behaviour under a micro-incentive scheme may differ substantially, particularly for BEST's unionised workforce, where routing changes require negotiation with the Brihanmumbai Electric Supply and Transport Workers' Union. This limitation does not preclude the feasibility of the approach, but it indicates that real-world VWCI gains should be expected to vary considerably by city and fleet operator.

On privacy, PRIMA's design makes one deliberate trade-off: the 150 m spatial fuzz radius means that EcoPulse cannot reliably attribute pollution readings to specific road segments narrower than $\sim$300 m. In dense urban cores where emissions from a single road are of regulatory interest, this limits spatial resolution below what the sensor hardware could otherwise achieve. A city that needs street-level accuracy and is prepared to accept higher privacy risk could reduce the fuzz radius; PRIMA's parameters are configurable.

The framework does not solve sensor degradation. PMS-series optical particle counters in India's high-humidity, high-dust environments show meaningful sensitivity drift within 3–6 months of unprotected outdoor exposure. EcoPulse's architecture includes a 90-day maintenance schedule (Figure 1, "Sensor lifecycle"), but drift detection between maintenance intervals relies on cross-calibration against adjacent sensors using BLE k-hop rendezvous. This works when sensors are sufficiently dense; in sparse rural deployments, cross-calibration is not always feasible.

6 Conclusion

EcoPulse-India demonstrates that the gap between India's air quality data needs and its monitoring infrastructure can be partially closed using hardware that already exists—city buses—combined with edge AI tight enough to run on a $5 processor. The four technical contributions (VCAC aerodynamic correction, int8 PI-ConvGNN, MARL-VWCI routing, PRIMA on-device redaction) address distinct failure modes of prior mobile LCS approaches and are individually deployable; operators who cannot adopt the full stack can still benefit from VCAC alone or PRIMA alone.

The simulation results are encouraging, but two validation steps remain necessary before production deployment: a physical pilot on a subset of DTC or BEST buses with concurrent comparison against reference monitoring stations, and a legal review of the PRIMA implementation against the applicable provisions of the Digital Personal Data Protection Act, 2023 and the Digital Personal Data Protection Rules, 2025, taking into account their phased commencement.

The broader point is that the constraints that define Global South urban environments—thin infrastructure, low hardware budgets, fast-growing informal settlements, strong privacy regulation—are not primarily obstacles to good air quality monitoring. They are design constraints that favour distributed, lightweight, privacy-first architectures over the centralised dense-sensor models developed in European and North American cities. EcoPulse-India is one approach; others will follow from the same starting conditions.

Data Availability Statement

The data used to support the findings of this study are available from the corresponding author upon request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] IQAir. (2025). *2024 World Air Quality Report*. IQAir Group, Goldach, Switzerland. Available at https://www.iqair.com/world-most-polluted-cities/world-air-quality-report-2024-en.pdf

[2] Guttikunda, S. K., Nishadh, K. A., & Jawahar, P. (2019). Air pollution knowledge assessments (APnA) for 20 Indian cities. *Urban Climate*, *27*, 124-141. [CrossRef]

[3] Kumar, P., Morawska, L., Martani, C., Biskos, G., Neophytou, M., Di Sabatino, S., ... & Britter, R. (2015). The rise of low-cost sensing for managing air pollution in cities. *Environment international*, *75*, 199-205. [CrossRef]

[4] Jayaratne, R., Liu, X., Thai, P., Dunbabin, M., & Morawska, L. (2018). The influence of humidity on the performance of a low-cost air particle mass sensor and the effect of atmospheric fog. *Atmospheric Measurement Techniques*, *11*(8), 4883-4890. [CrossRef]

[5] Munir, M. M., Adrian, M., Saputra, C., & Lestari, P. (2022). Utilizing low-cost mobile monitoring to estimate the PM2. 5 inhaled dose in urban environment. *Aerosol and Air Quality Research*, *22*(6), 220079. [CrossRef]

[6] Li, Y., Yu, R., Shahabi, C., & Liu, Y. (2017). Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv preprint arXiv:1707.01926*. [CrossRef]

[7] Yu, B., Yin, H., & Zhu, Z. (2017). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv preprint arXiv:1709.04875*. [CrossRef]

[8] Government of India. (2023). *The Digital Personal Data Protection Act, 2023*. Ministry of Electronics and Information Technology, New Delhi. Available at https://www.meity.gov.in/static/uploads/2024/06/2bf1f0e9f04e6fb4f8fef35e82c42aa5.pdf

[9] Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse

problems involving nonlinear partial differential equations. *Journal of Computational physics*, *378*, 686-707. [CrossRef]

[10] Li, L., Wang, J., Franklin, M., Yin, Q., Wu, J., Camps-Valls, G., ... & Reichstein, M. (2023). Improving air quality assessment using physics-inspired deep graph learning. *npj Climate and Atmospheric Science*, *6*(1), 152. [CrossRef]

[11] Krause, A., Singh, A., & Guestrin, C. (2008). Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, *9*(2). Available at https://www.jmlr.org/papers/volume9/krause08a/krause08a.pdf

[12] Morello-Frosch, R., Zuk, M., Jerrett, M., Shamasunder, B., & Kyle, A. D. (2011). Understanding the cumulative impacts of inequalities in environmental health: implications for policy. *Health affairs*, *30*(5), 879-887. [CrossRef]

[13] Ji, W., Han, K., & Liu, T. (2023). Trip-based mobile sensor deployment for drive-by sensing with bus fleets. *Transportation Research Part C: Emerging Technologies*, *157*, 104404. [CrossRef]

[14] Kaginalkar, A., Kumar, S., Gargava, P., & Niyogi, D. (2023). Stakeholder analysis for designing an urban air quality data governance ecosystem in smart cities. *Urban Climate*, *48*, 101403. [CrossRef]

[15] Warden, E., & Situnayake, D. (2019). *TinyML: Machine Learning with TensorFlow Lite on Arduino and Ultra-Low-Power Microcontrollers*. O'Reilly Media. Retrieved from https://www.semanticscholar.org/paper/TinyML%3A-Machine-Learning-with-TensorFlow-Lite-on-Warden-Situnayake/9d312f454068259015e38608f7ebb65947978f65

[16] Bandyopadhyay, D., & Sen, J. (2011). Internet of things: Applications and challenges in technology and standardization. *Wireless personal communications*, *58*(1), 49-69. [CrossRef]

Satya Prakash is a final-year Computer Science student at NIST University, Berhampur, Odisha, India, specializing in high-performance full-stack development and systems programming across the MERN stack, Java, and Go. He is the creator of IntelliRide, an AI framework for smart mobility that earned 1st Runner-Up at the GFI 2026 International Conference, and of EcoPulse-India, an AIoT framework for real-time environmental sensing. His work spans multi-algorithm pipelines, federated learning, and low-level operating systems, with a focus on lean, high-impact AI solutions tailored to the infrastructure realities of emerging economies. (Email: matsya1sm@gmail.com)

Om Prakash Nayak is a graduate in Computer Science and Technology from NIST University, Berhampur, Odisha, India, with a strong foundation in programming, software development, and core computer science principles. His interests include artificial intelligence, computer networks, and web development, and his hands-on experience covers the development of IoT-based applications and smart systems. (Email: omnayak2903@gmail.com)

Om Prakash Patra is a B.Tech. student in Computer Science and Technology at NIST University, Berhampur, Odisha, India, with hands-on experience in machine learning, computer vision, and IoT systems. He has completed internships at NIST University in CVML and at IISER Berhampur in IoT platform development using FastAPI, gaining practical exposure to deep learning models, backend APIs, authentication, and real-time databases. His interests span artificial intelligence, IoT, backend development, and computer vision. He is a core member of the NIST Counselling Service. (Email: omprakashpatra437@gmail.com)

Manas Ranjan Panigrahy is a graduate in Computer Science and Engineering from NIST University, Berhampur, Odisha, India, with a foundation in software engineering, complex programming, and core computational theory. Driven by an interest in artificial intelligence, computer networking, and web development, he has worked on the development of IoT-integrated applications and smart systems. (Email: manasranjanpanigrahy047@gmail.com)

Ashish Kumar Dass holds an M.Tech. degree in Computer Science and Engineering from the National Institute of Science and Technology (NIST), Berhampur, Odisha, India, and is currently pursuing his Ph.D. in Computer Science and Engineering under BPUT, Odisha. With over fifteen years of teaching experience, he serves as an Assistant Professor at the School of Computer Science and Engineering, NIST University, Berhampur, Odisha, India. His research interests include underwater wireless sensor networks, ad-hoc and sensor networks, computer security, artificial intelligence, and mathematical modeling. (Email: ashishkumardass@nist.edu)