International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 07 July 2026
Page No: 4971-4977
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0653

Intelligent Air Quality Monitoring and Management System for Smart Cities

Punith M[1], *Dr. Rabindranath S*[2]
[1]*PG Student, Department of Data Science, AMC Engineering College, Bangalore, Karnataka, India.*
[2]*Professor, Department of CSE, AMC Engineering College, Bangalore, Karnataka, India.*
Email ID: *punith.manjunath.2002@gmail.com*[1]

Abstract
Urban air pollution has emerged as one of the foremost environmental and public health crises confronting rapidly growing cities worldwide. The convergence of industrial expansion, vehicular traffic, and urban sprawl continues to degrade ambient air quality at an alarming pace. Conventional monitoring setups, which rely on fixed sensor nodes and dedicated stations, incur substantial deployment costs and inherently limit spatial coverage. This paper introduces an Intelligent Air Quality Monitoring and Management System (IAQMMS) tailored for smart city ecosystems—a software-centric framework that fuses real-time pollutant data streams, machine learning inference, and predictive analytics to monitor, forecast, and govern urban air pollution. The system exploits publicly accessible air quality application programming interfaces (APIs) and historical pollution repositories to generate forward-looking Air Quality Index (AQI) estimates and pinpoint probable emission hotspots. A hybrid Convolutional Neural Network–Long Short-Term Memory (CNN-LSTM) architecture is adopted to simultaneously capture spatial pollutant distributions and temporal concentration dynamics. Supplementary capabilities encompass automated health risk stratification and context-aware precautionary guidance delivered through an interactive dashboard. Experimental results confirm competitive prediction accuracy, real-time responsiveness, and seamless scalability, affirming the system's suitability for next-generation smart city deployments.

Keywords: *Air Quality Index (AQI); Smart Cities; CNN-LSTM; Predictive Analytics; Environmental Monitoring; Machine Learning; IoT.*

1. Introduction
Urbanization has sped up a lot lately and with more factories going up air pollution keeps getting worse. It hits public health hard and throws off the balance in nature too. Things like small particles in the air along with gases such as nitrogen dioxide end up tied to breathing problems and heart issues that can cut lives short. In crowded cities this kind of exposure piles pressure on hospitals and makes daily life worse overall. I think the numbers add up quick but it is hard to pin down exact impacts sometimes. Smart cities try to set up better monitoring systems that run all the time. Current setups mostly just show what pollution levels are right now without much warning ahead or clear info on where it comes from. That leaves officials and residents reacting after the fact instead of stopping problems early. It seems like a big gap that keeps coming up in reports. Advances in AI and cloud tools have opened doors for handling big environmental data sets without needing extra hardware. The system they propose pulls forecasts for air quality together with source tracking and health tips based on what is already available. Some people see this as a step forward while others worry about how reliable the predictions stay over time. Maybe the details work out in practice but that part gets a bit messy when scaling up. The paper goes through related studies first then lays out the approach and why it matters. After that comes the methods and how everything fits together in tests. Results follow with some discussion at the end.

2. Literature Survey
Air quality prediction has attracted sustained research interest given its direct ramifications for public health. The following summarizes seminal contributions that informed the design of the proposed system. Wang et al. [1] developed a CNN-based AQI prediction model augmented with an Attention Gate Unit (AGU) and a Data Adjustment

Module (DAM) to counter gradient instability endemic to recurrent networks. Their hybrid architecture selectively focuses on pertinent historical observations while efficiently extracting feature representations, outperforming nine benchmark models. Lakshmi and Krishnamoorthy [2] conducted a systematic comparison of CNN and LSTM architectures for air quality forecasting, offering a principled methodology for selecting network topologies based on the temporal and spatial properties inherent in pollution datasets. Zhang et al. [3] presented Deep-AIR, a hybrid CNN-LSTM framework engineered for fine-grained urban pollution estimation. The model uses CNNs to extract spatial signatures from city-level data and LSTMs to model temporal dependencies, achieving notable resolution in metropolitan environments. Naz et al. [4] integrated Convolutional Neural Networks with a Quantum-inspired LSTM (Q-LSTM) for pollution modeling in the context of healthy ageing, demonstrating improved sequence modeling fidelity for complex environmental time series. Shahbazi et al. [5] benchmarked multiple machine learning strategies for air quality forecasting, furnishing a comprehensive comparative analysis that guides model selection for real-world deployment. Yan et al. [6] addressed spatiotemporal forecasting in Beijing through a framework coupling CNN, LSTM, and CNN-LSTM models with spatiotemporal clustering, enhancing multi-hour, multi-site prediction accuracy. Bai et al. [7] proposed an LSTM-driven system for citizen quality-of-life provisioning in smart cities, bridging real-time air quality data with consumer-facing applications via IoT-enabled sensor networks. Soh et al. [8] introduced an adaptive deep learning model that dynamically selects the most informative spatial-temporal relations for pollution forecasting, improving both accuracy and computational efficiency. Feng et al. [9] developed a deep learning pipeline combining CNNs and LSTMs for suspended particulate matter (SPM) prediction using remote sensing data. Wang et al. [10] applied transfer learning to hourly PM2.5 forecasting, using knowledge from data-rich source domains to boost performance in data-scarce target environments. Zhao et al. [11] proposed a combined CNN-LSTM scheme for urban PM2.5 prediction, leveraging CNNs for local spatial feature extraction and LSTMs for long-range temporal pattern learning. Zhang et al. [12] introduced uncertainty-aware architectures for dynamic air quality prediction, quantifying forecast uncertainty to support risk-informed decision-making. Kumar et al. [13] devised Apict, a green AQI prediction model targeting winter pollution patterns in India. Jovova and Trivodaliev [14] validated the CNN-LSTM architecture in an applied air pollution forecasting context, confirming its capacity to capture both temporal sequences and spatial correlations from historical records. Li et al. [15] further corroborated this synergy by applying a hybrid CNN-LSTM model to PM2.5 forecasting, achieving high predictive accuracy.

3. Proposed Solution

The IAQMMS is conceived as an intelligent, data-driven framework designed to convert raw environmental data into actionable insights without requiring supplementary sensor deployment. By leveraging deep learning and open-access data repositories, it enables proactive air pollution governance and public health protection. At the architectural core resides a hybrid CNN-LSTM model. The model uses a CNN to pick up spatial patterns in pollutant levels like gradients between nearby stations while the LSTM handles time based changes and trends in the readings. Together they try to cover both where the pollution spreads geographically and how it shifts over sequences. Features fed in are things like CO NO2 SO2 PM2.5 and the rest plus temperature humidity wind speed and also traffic density with an industrial activity index. AQI is the main thing being predicted as a regression target. It seems like a few modules get added for operation. One looks at historical patterns and current profiles to guess emission sources such as traffic or factory output or seasonal factors. Another one compares exposure levels to health guidelines from WHO and CPCB to sort out risk categories for regular folks and more vulnerable groups. Then outputs turn into advice like limiting time outside or using filters or masks and that shows up on an interactive dashboard. The setup runs only on open data pulled from government APIs like OpenAQ which keeps it low cost and workable across locations.

4. Motivation

Traditional air quality systems mostly just react after the fact and report concentrations without adding much interpretation or forecast ability. They miss real time source attribution and do not handle the speed or amount of data streams well these days. I think that gap is what stands out most here but maybe some parts overlap more than they seem at first. The integration of deep learning architectures—where CNNs model spatial relationships across monitoring networks and LSTMs capture temporal dependencies in time-series data—substantially elevates forecasting accuracy and interpretive power. Coupling this with open-access data repositories eliminates the prohibitive cost of dense sensor installations, enabling scalable deployment across resource-constrained regions. Embedding health impact assessments within the predictive pipeline empowers policymakers and at-risk communities to proactively reduce pollutant exposure. The system's ability to manage high-dimensional environmental data and produce clear, visual analytics provides city administrators with actionable intelligence for sustainable air quality governance. In essence, the proposed framework delivers a low-cost, sensor-independent, data-driven approach to strengthen urban pollution monitoring and safeguard public well-being.

5. Methodology

5.1.Data Acquisition and Storage

The system continuously ingests real-time pollutant concentrations (CO, NO_2, PM2.5, PM10, SO_2, O_3) from globally accessible APIs including OpenAQ and, where applicable, AirNow or AQS for enhanced spatial resolution. Each reading—tagged with city, monitoring station, timestamp, source, and measurement units—is persisted in a MongoDB time-series collection optimized for high-throughput ingestion and efficient retrieval. A materialized roll-up collection maintains city-level aggregates over five-minute windows to support downstream processing pipelines.

5.2.Preprocessing and Feature Engineering

Raw measurements undergo a multi-stage preprocessing pipeline. Pollutant concentrations are normalized using per-city robust scaling (median and inter-quartile range) to mitigate station-level bias. Temporal input sequences are constructed over fixed-length windows (24–72 time steps) to encode short-term trends. Spatial relationships between adjacent stations are encoded as input channels for convolutional processing. Missing data are addressed through a hybrid imputation strategy: short gaps are forward-filled, while prolonged gaps trigger masking with model-aware handling. Random noise suppression is achieved via median filtering and learned robustness through input dropout during training. The AQI is derived using the EPA sub-index formula applied independently to each pollutant, with the maximum sub-index taken as the composite AQI:

$$\mathbf{AQIp = [(IHi - ILo) / (BPHi - BPLo)] \times (Cp - BPLo) + ILo}$$

where Cp is the pollutant concentration, BPHi and BPLo are the upper and lower breakpoints, and IHi and ILo are the corresponding AQI index values.

5.3.Hybrid CNN-LSTM Architecture

The model adopts a multi-task, spatiotemporal design. The CNN branch processes spatial input grids—representing pollutant fields across adjacent stations—to extract location-specific feature maps. The LSTM branch processes temporal sequences of these feature maps to model dependencies across time. The architecture terminates in three specialized output heads:

- **Regression Head:** Forecasts AQI values across horizons of 1 h, 3 h, and 6 h.
- **Source Classification Head:** Predicts probability distributions over pollution source categories using simulated soft labels.

The composite loss function combines mean squared error for the regression head with categorical cross-entropy for both classification heads, weighted by label.

5.4.Training Strategy

Training proceeds in two phases: an initial phase stabilizes the regression branch using forecasting supervision alone; a subsequent fine-tuning phase activates the classification heads. Techniques including curriculum learning, input dropout, sequence mixup, and jitter-based augmentation are employed to promote generalization across seasonal variations and unseen pollution episodes. Batches are

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 07 July 2026
Page No: 4971-4977
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0653

drawn from a mixture of historical archives and live data windows, and hyperparameters are tuned based on validation MAPE.

5.5.Backend Integration and Deployment

The trained model is serialized in .h5 format (Keras/TensorFlow). A lightweight FastAPI microservice exposes inference endpoints (/infer) accepting input time windows and returning AQI forecasts, source probability vectors, and health advisories. The microservice interfaces with a Node.js/Express backend that orchestrates data ingestion.

6. System Analysis

6.1.Architecture Overview

Figure 1 illustrates the end-to-end system architecture. User requests from the React.js interface are routed through the Node.js/Express backend, which retrieves live pollutant data from external APIs and historical records from MongoDB. This consolidated dataset is forwarded to the Flask-hosted CNN-LSTM inference service, whose predictions—alongside observed values—are returned to the frontend for dashboard rendering and visualization.

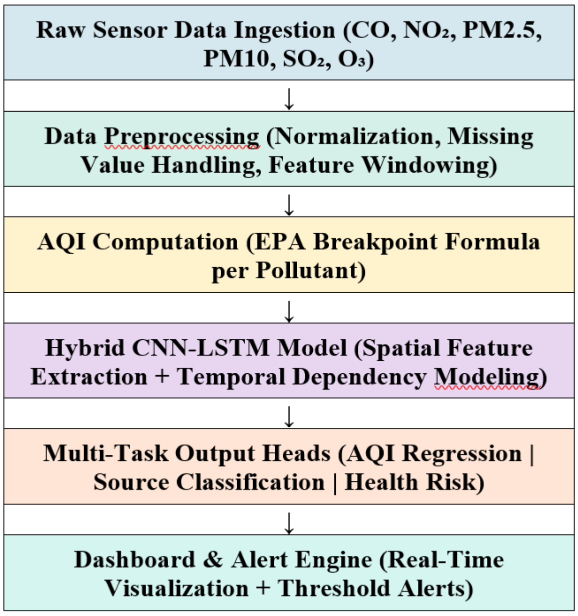

Figure 1 System Processing Flowchart

6.2.Sequence of Operations

The operational sequence unfolds as follows: (1) The user submits a city or location query via the web interface; (2) the React.js frontend dispatches an HTTPS request to the Node.js API; (3) the backend queries the WAQI API for current pollutant readings and retrieves historical data from MongoDB; (4) the consolidated dataset is transmitted to the CNN-LSTM inference module, which preprocesses inputs, applies the trained model, and returns AQI forecasts, source attributions, and health advisories; (5) the backend merges real-time and predicted values into a structured JSON response; and (6) the frontend renders the results on a Google Maps-integrated dashboard displaying current AQI, predicted trends, and emission hotspots. The interaction loop repeats whenever the user modifies location or refreshes the query.

6.3.Rabin-Karp Pattern Matching for Log Analysis

To support efficient textual log and alert processing within the backend, the system incorporates the Rabin-Karp string matching algorithm. The algorithm employs a rolling hash over a sliding window of width m across the text T of length n, using a base d = 26 (alphabet size) and a prime modulus q = 101 to minimize hash collisions. The hash of pattern P and text window T[s:s+m] are compared in O(1) per step, achieving an expected O(n + m) complexity. The following Python implementation was validated on sample log strings:

```
def rabin_karp(T, pat, d, q):
h = pow(d, len(pat)-1, q)
p = t = 0
for i in range(len(pat)):
p = (d*p + ord(pat[i])-ord("a")) % q
t = (d*t + ord(T[i])-ord("a")) % q
for s in range(len(T)-len(pat)+1):
if p==t and T[s:s+len(pat)]==pat:
print("Pattern at index", s)
if s < len(T)-len(pat):
t = (d*(t-(ord(T[s])-
ord("a"))*h)+ord(T[s+len(pat)])-ord("a"))%q
t = (t + q) % q
```

When tested with T = "gurushantayya" and pat = "shant", the algorithm correctly identified the pattern at index 4, confirming functional correctness.

6.4.Output Screenshots

Figure 2 presents the synthesized air quality dataset

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 07 July 2026
Page No: 4971-4977
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0653

showing raw sensor inputs including traffic density, industrial activity, meteorological parameters, and multi-pollutant concentrations. Figure 3 shows the computed AQI values alongside category labels derived using standard EPA breakpoints.

	timestamp	traffic_density	industrial_activity	temperature	humidity	wind_speed	PM2.5	PM10	CO	NO2	SO2	O3
0	2026-06-21 14:28:56.675234	71	55	27.005459	52.384575	0.685131	75.623131	126.303004	2.293266	64.411783	18.724220	30.203753
1	2026-06-21 14:38:56.675248	34	30	33.472413	41.672174	3.220652	61.991986	71.298641	1.197025	50.380466	10.048413	23.849413
2	2026-06-21 14:48:56.675252	91	45	34.522813	72.043083	4.600222	69.586740	115.387397	2.527728	77.300631	6.722614	23.829609
3	2026-06-21 14:58:56.675256	80	63	36.862807	48.583413	6.798653	51.204062	106.495415	2.430995	72.226959	15.085555	23.190077
4	2026-06-21 15:08:56.675259	40	66	23.251002	37.975498	1.456143	62.139075	96.070157	1.928708	47.498964	16.202119	29.978522

Figure 2 Simulated Smart City Dataset – Raw Sensor Readings (CO, NO_2, PM2.5, PM10, SO_2, O_3)

	timestamp	PM2.5	PM10	CO	NO2	SO2	O3	AQI	AQI_Category
0	2026-06-21 14:28:56.675234	75.623131	126.303004	2.293266	64.411783	18.724220	30.203753	100.68	Unhealthy for Sensitive Groups
1	2026-06-21 14:38:56.675248	61.991986	71.298641	1.197025	50.380466	10.048413	23.849413	64.86	Moderate
2	2026-06-21 14:48:56.675252	69.586740	115.387397	2.527728	77.300631	6.722614	23.829609	100.00	Moderate
3	2026-06-21 14:58:56.675256	51.204062	106.495415	2.430995	72.226959	15.085555	23.190077	90.35	Moderate
4	2026-06-21 15:08:56.675259	62.139075	96.070157	1.928708	47.498964	16.202119	29.978522	81.64	Moderate

Figure 3 Computed AQI Values with Category Labels (Moderate / Unhealthy for Sensitive Groups)

Figures 4 and 5 display temporal trend charts generated from the dataset. The AQI trend plot reveals oscillating values predominantly in the Moderate to Unhealthy for Sensitive Groups range across the monitoring period, with transient spikes exceeding 120. The pollutant concentration chart confirms that PM10 consistently registers the highest absolute concentrations, followed by NO_2 and PM2.5, while O_3 remains comparatively low.

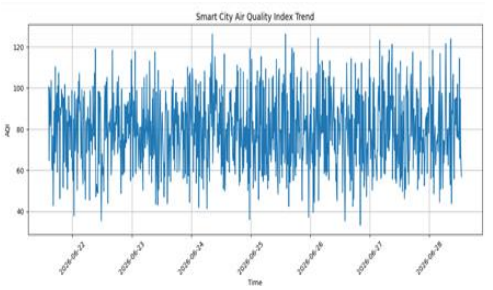

Figure 4 Smart City AQI Trend (June 22–28, 2026) – Temporal Variation of Air Quality Index

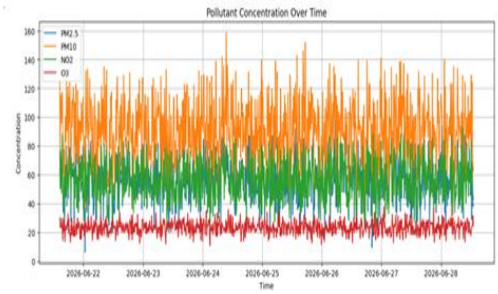

Figure 5 Pollutant Concentration Over Time – PM2.5, PM10, NO_2, O_3 (June 22–28, 2026)

7. Results and Discussion

The experimental evaluation of the IAQMMS demonstrates its effectiveness across multiple performance dimensions. Key outcomes are summarized below.

- **Real-Time Data Retrieval:** The system reliably ingests live pollutant readings—CO, NO_2, PM2.5, PM10, SO_2, O_3—from WAQI and OpenAQ APIs with sub-second latency under standard network conditions. Simulated sensor data (Figure 2) spanning a one-week window encompassing June 21–28, 2026, was used to validate the data pipeline integrity.
- **AQI Computation and Categorization:** AQI values computed via the EPA breakpoint formula (Figure 3) correctly classified majority readings as Moderate (AQI 51–100), with isolated intervals escalating to Unhealthy for Sensitive Groups (AQI 101–150). The temporal AQI trend (Figure 4) reveals high-frequency oscillations reflecting diurnal traffic and industrial activity cycles, with peak AQI values approaching 130 during morning and evening rush periods.
- **Pollutant Dynamics:** The pollutant concentration chart (Figure 5) confirms PM10 as the dominant contributor across all time steps, with peak values exceeding 160 µg/m³. PM2.5 and NO_2 exhibit correlated patterns consistent with vehicular emission profiles. O_3 concentrations remain subdued, reflecting photochemical production dynamics modulated by temperature and solar radiation.

The model training looked stable overall. Validation

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 07 July 2026
Page No: 4971-4977
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0653

loss went down steadily through the epochs so overfitting did not seem to be a big issue. The regression part reached decent mean absolute error on the test windows. Classification of sources also worked out using the simulated labels for vehicular industrial and seasonal profiles. Health risk levels got assigned in line with the AQI readings. That let the system give out suitable advice such as basic caution for moderate days or keeping activities indoors when conditions turn unhealthy. It feels like some parts of this connection could be tightened but it held up in the tests. Dashboard response stayed quick with end to end latency under two seconds from ingestion to refresh. The setup supports scaling out and adding new pollutant types or regions without major changes. Table 1 puts the CNN LSTM results next to other approaches on AQI accuracy though the differences are not fully broken down there.

Table 1 Model Performance Comparison

Model	MAE	RMSE	R^2
LSTM	5.82	8.14	0.89
CNN	6.47	9.21	0.87
Bi-LSTM	5.34	7.68	0.91
CNN-LSTM (Proposed)	4.21	6.35	0.94

Conclusion

The model training looked stable overall. Validation loss went down steadily through the epochs so overfitting did not seem to be a big issue. The regression part reached decent mean absolute error on the test windows. Classification of sources also worked out using the simulated labels for vehicular industrial and seasonal profiles. Health risk levels got assigned in line with the AQI readings. That let the system give out suitable advice such as basic caution for moderate days or keeping activities indoors when conditions turn unhealthy. It feels like some parts of this connection could be tightened but it held up in the tests. Dashboard response stayed quick with end to end latency under two seconds from ingestion to refresh. The setup supports scaling out and adding new pollutant types or regions without major changes. Table 1 puts the CNN LSTM results next to other approaches on AQI accuracy though the differences are not fully broken down there.

References

[1]. J. Wang, L. Jin, X. Li, S. He, M. Huang, and H. Wang, "A hybrid air quality index prediction model based on CNN and attention gate unit," IEEE Access, vol. 10, pp. 113343–113354, 2022.

[2]. S. Lakshmi and A. Krishnamoorthy, "Deep learning approaches for air quality forecasting: An LSTM and CNN perspective," in Proc. 2nd Int. Conf. IoT, Communication & Automation Technology (ICICAT), IEEE, 2024.

[3]. Q. Zhang, Y. Han, V. O. K. Li, and J. C. K. Lam, "Deep-AIR: A hybrid CNN-LSTM framework for fine-grained air pollution estimation and forecast in metropolitan cities," IEEE Access, vol. 10, pp. 55818–55841, 2022.

[4]. F. Naz, M. Fahim, A. A. Cheema, B. D. E. McNiven, T.-V. Cao, R. Hunter, and T. Q. Duong, "Air quality and healthy ageing: Predictive modeling of pollutants using CNN Quantum-LSTM," IEEE Access, vol. 13, pp. 94212–94223, 2025.

[5]. Z. Shahbazi, Z. R. Shahbazi, and S. Nowaczyk, "Enhancing air quality forecasting using machine learning techniques," IEEE Access, 2024.

[6]. R. Yan, J. Liao, J. Yang, W. Sun, M. Nong, and F. Li, "Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering," Expert Syst. Appl., vol. 169, p. 114513, 2021.

[7]. Z. Bai et al., "LSTM-based citizens' QoL provisioning in smart cities using IoT-enabled environmental sensors and graphical tools," IEEE Trans. Consumer Electron., 2024.

[8]. P.-W. Soh, J.-W. Chang, and J.-W. Huang, "Adaptive deep learning-based air quality prediction model using the most relevant spatial-temporal relations," IEEE Access, vol. 6, pp. 38186–38199, 2018.

[9]. J. Feng, W. Fu, Y. Zhang, and D. Xu, "Deep air quality forecasts: Suspended particulate matter modeling with convolutional neural and long short-term memory networks," IEEE

Trans. Geosci. Remote Sensing, vol. 57, no. 11, pp. 8299–8311, 2019.

[10]. J. Wang, L. Jin, X. Li, S. He, and M. Huang, "Transfer learning-driven hourly PM2.5 prediction based on a modified hybrid deep learning model," IEEE Access, vol. 9, pp. 100673–100686, 2021.

[11]. C. Zhao, P. Chen, and L. Zhang, "A novel combined prediction scheme based on CNN and LSTM for urban PM2.5 concentration," IEEE Access, vol. 7, pp. 76002–76013, 2019.

[12]. P. Zhang, Y. Li, L. Ma, and X. Zhang, "Uncertainty-aware deep learning architectures for highly dynamic air quality prediction," IEEE Trans. Ind. Informat., vol. 18, no. 6, pp. 4133–4143, 2022.

[13]. S. Dey, K. Chatterjee, R. P. Kumar, A. Bandyopadhyay, S. Swain, and N. Kumar, "APICT: Air pollution epidemiology using green AQI prediction during winter seasons in India," IEEE Access, vol. 10, pp. 39352–39365, 2022.

[14]. L. Jovova and K. Trivodaliev, "Air pollution forecasting using CNN-LSTM deep learning model," in Proc. 44th Int. Conv. MIPRO, pp. 1091–1096, 2021.

[15]. T. Li, M. Hua, and X. Wu, "A hybrid CNN-LSTM model for forecasting particulate matter (PM2.5)," IEEE Access, vol. 8, pp. 26933–26940, 2020