IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.141239

A Review On Air Quality Prediction Using Embedded Machine Learning And Deep Learning Models With Quantization Techniques

Anandhu Suresh[1], Lekshmi V[2]

PG Student, MSc Computer Science, Christ Nagar College, Maranalloor, Thiruvananthapuram, Kerala, India[1]

Assistant Professor, PG Department of Computer Science, Christ Nagar College, Maranalloor, Thiruvananthapuram, Kerala, India[2]

Abstract: Air pollution is a major global environmental and health concern, necessitating accurate and timely forecasting of pollutant levels to mitigate adverse effects. Recent advances in machine learning and deep learning enable precise air quality prediction, yet deploying these models in real world resource constrained settings remains challenging. Embedded models augmented with quantization techniques offer an efficient solution by reducing computational costs without significant loss of accuracy. This review synthesizes recent developments in air quality prediction using embedded Machine Learning (ML) and Deep Learning (DL) models with emphasis on 8 bit quantization, hybrid architectures, attention mechanisms, and knowledge distillation. The focal study demonstrates a CNN BiGRU model achieving an R^2 of 0.99 for PM2.5 on IoT devices, balancing high accuracy with model compression. Through analysis of fourteen contemporary works, this paper highlights multi task learning frameworks, spatial temporal modeling over multiple pollutants, and transformer based approaches as emerging trends. Persistent challenges include extending forecast horizons, improving generalizability, and enhancing real time deployment viability. The review concludes by discussing future directions integrating ensemble methods, Bayesian hyperparameter optimization, and quantum learning potentials toward robust, scalable air quality prediction systems.

Keywords: Air Quality Prediction, Deep Learning, Long Short Term Memory (LSTM), Convolutional Neural Network (CNN), Quantization, Edge Computing, Internet Of Things (IoT), Transformer, Knowledge Distillation, Multi Pollutant Forecasting

I. INTRODUCTION

Air quality has critical implications for human health, ecosystems, and overall quality of life. According to global health data, ambient air pollution causes millions of premature deaths each year, primarily due to respiratory and cardiovascular diseases. Pollutants such as particulate matter (PM2.5 and PM10), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), carbon monoxide (CO), and ozone (O_3) pose severe health hazards. Reliable forecasting of these pollutant levels is essential for timely public advisories and effective regulatory actions.

Traditional approaches, including chemical transport and statistical models, often lack the flexibility to capture the complex interactions between meteorological conditions and pollutant emissions. The rise of machine learning and deep learning has transformed air quality forecasting by enabling models to learn nonlinear relationships from large datasets. Architectures such as LSTM networks, CNNs, Gated Recurrent Units (GRUs), and transformers have shown strong performance in modeling temporal sequences and spatial dependencies in air quality data.

However, deploying such sophisticated models on IoT and edge devices presents new challenges due to limited computational and memory resources. Model compression techniques, such as quantization and knowledge distillation, help mitigate these constraints by reducing model size and latency while preserving prediction accuracy. The foundation of this review is an embedded CNN-BiGRU model that employs 8 bit quantization to achieve high performance with low resource requirements, making it suitable for deployment on constrained hardware.

This paper systematically reviews fourteen related studies that investigate a range of methods from hybrid statistical and deep learning models to quantum enhanced and spatial graph based approaches. The objective is to summarize current

DOI: 10.17148/IJARCCE.2025.141239

progress, identify existing limitations, and guide future research toward developing practical, accurate, and scalable air quality prediction systems.

II. BACKGROUND AND RELATED WORKS

2.1 Background

Air pollution poses a significant threat to environmental quality and public health globally, driven by rapid urbanization, industrialization, and increased vehicle emissions, resulting in millions of premature deaths annually. Accurate monitoring and forecasting of key pollutants like PM2.5, PM10, nitrogen dioxide, sulfur dioxide, ozone, and carbon monoxide are crucial for timely health advisories and effective policy decisions. Traditional forecasting methods often fall short in capturing the complex, nonlinear interactions influencing pollutant behavior. Machine learning and deep learning have emerged as powerful tools to model these intricate spatial and temporal patterns from extensive historical data, substantially improving prediction accuracy. However, deploying such models for real time applications on resource limited edge and IoT devices faces challenges in computational demands and energy use. To overcome this, model compression techniques like quantization are increasingly used to reduce model size and complexity without significant loss in performance, enabling efficient, low power embedded air quality forecasting systems suitable for various urban and industrial settings.

2.2 Related Works

Numerous studies have explored machine learning (ML) and deep learning (DL) techniques for various aspects of air quality forecasting:

Embedded and Quantized Models: Mazinani et al. developed an embedded hybrid CNN-BiGRU model optimized with 8 bit quantization, demonstrating exceptional accuracy ($R^2 = 0.99$) for PM2.5 forecasting and practical deployment on resource constrained IoT devices. This highlights the model's potential for efficient real time monitoring in urban environments.

Multi Step and Multi Pollutant Forecasting: Chatterjee et al. introduced a Hyper Heuristic Multi Chain Model (H2MCM) composed of multiple interconnected LSTM units, capable of accurate predictions up to 12 hours ahead for several pollutants. Their approach improved robustness via transformer based heuristic error correction mechanisms.

Hybrid Statistical and Machine Learning Approaches: Cao et al. combined Empirical Mode Decomposition (EMD) with ARIMA and Support Vector Regression (SVR) to effectively characterize both linear and nonlinear pollutant dynamics in multiple Chinese cities, showing improved short term forecasting performance.

Quantum Enhanced Learning: Naz et al. proposed a quantum enhanced CNN-LSTM architecture incorporating Variational Mode Decomposition to reduce noise and improve prediction accuracy, marking a novel integration of quantum computing paradigms into air quality modeling.

Deep Sequential and Attention Models: Jayanth et al. presented a hybrid model integrating Holt Winters smoothing with Transformer and BiGRU networks, fine tuned by Bayesian optimization to capture seasonal and trend elements of pollutant time series data effectively.

Knowledge Distillation and Graph Based Networks: Kumar et al. combined Graph Convolutional Networks with Transformer GRUs under a knowledge distillation framework, resulting in lightweight yet accurate models suited for multi city Indian metropolitan pollutant forecasting.

Metaheuristic Optimization Techniques: Ahmed et al. applied Particle Swarm Optimization (PSO) for optimizing neural network parameters in predicting industrial CO emissions with remarkable precision.

Multi Task and Edge Ready Architectures: Xie et al. designed a knowledge distilled multi task Transformer named Bidirectional Mamba2, excelling in fast, accurate multi pollutant predictions optimized for edge computing deployment.

Hybrid Filtering and Energy Efficient Models: Chatterjee et al. utilized a hybrid Kalman Filter and Artificial Neural Network (ANN) framework for indoor air quality management, emphasizing energy efficiency and real time control.

Health Impact Prediction: Ramesh et al. employed multi scale CNNs enhanced with attention mechanisms to correlate pollutant levels with associated health risks, advancing understanding of pollution driven health outcomes.

III. LITERATURE SURVEY

- In this paper Air Quality Prediction via Embedded Machine Learning and Deep Learning Models with Quantization Techniques [1], Mazinani et al. propose an embedded hybrid CNN BiGRU model optimized with 8 bit quantization. The model achieves outstanding prediction accuracy ($R^2 = 0.99$) for PM2.5 and demonstrates real time deployment feasibility on IoT edge devices, highlighting its practical utility for low power air quality monitoring.

- In this paper Future Air Quality Prediction Using Long Short-Term Memory Based on Hyper Heuristic Multi-Chain Model [2], Chatterjee et al. introduce a Hyper Heuristic Multi Chain Model (H2MCM) consisting of interconnected

DOI: 10.17148/IJARCCE.2025.141239

LSTM units, facilitating up to 12 hour ahead multi pollutant forecasting. The approach incorporates transformer based heuristic error correction, substantially improving prediction robustness.

- In this paper A Hybrid Air Quality Prediction Model Based on Empirical Mode Decomposition [3], Cao et al. combine Empirical Mode Decomposition with ARIMA and SVR to capture linear and nonlinear pollutant dynamics effectively in urban China. This hybrid method excels in short term pollutant forecasting.

- In this paper Air Quality and Healthy Ageing: Predictive Modeling of Pollutants Using CNN Quantum-LSTM [4], Naz et al. present a quantum enhanced CNN-LSTM framework integrating Variational Mode Decomposition to denoise inputs and boost prediction accuracy. While promising, the method requires quantum hardware support.

- In this paper Enhancing Air Quality Prediction Through Holt-Winters Smoothing and Transformer BiGRU with Bayesian Optimization [5], Jayanth et al. design a hybrid model capturing seasonal and trend components of pollutant data, optimized via Bayesian methods for enhanced forecasting precision.

- In this paper Optimizing Air Pollution Forecasting Models Through Knowledge Distillation: A Novel GCN and TRANS-GRU Methodology for Indian Cities [6], Kumar et al. combine Graph Convolutional Networks with Transformer GRU models under knowledge distillation to develop lightweight but highly accurate air pollution forecasting systems.

- In this paper MetaForecaster: A PSO-Driven Neural Model for Sustainable Industrial Air Quality Management [7], Ahmed et al. utilize Particle Swarm Optimization to fine tune neural networks for accurate industrial CO emission predictions.

- In this paper Knowledge-Distilled Multi-Task Model with Enhanced Transformer and Bidirectional Mamba2 for Air Quality Forecasting [8], Xie et al. introduce a fast, accurate multi task Transformer model optimized for edge device deployment.

- In this paper Indoor Air Wellness: A Predictive Model for Pollution Control Using Advanced AI Techniques [9], Chatterjee et al. combine Kalman filtering with neural networks to design energy efficient indoor air quality management solutions.

- In this paper AirQuaNet: A Convolutional Neural Network Model With Multi-Scale Feature Learning and Attention Mechanisms for Air Quality-Based Health Impact Prediction [10], Ramesh et al. leverage multi scale CNNs with attention to link pollutant exposure with health risks.

DOI: 10.17148/IJARCCE.2025.141239

Reference no.	Methodology	Dataset(s)	Accuracy	Merits	Demerits
A. Mazinani [1]	CNN BiGRU with 8 bit quantization	Brindisi airport (Italy), PM2.5	$R^2 = 0.99$, RMSE = 2.03	Real time embedded IoT deployment; 66% model size reduction	Limited pollutant scope; short term forecasts only
K. Chatterjee [2]	Multi-chain LSTM Hyper Heuristic Model	Hyderabad, multiple pollutants	PLCC ~ 1 (Pearson corr.)	Handles long range dependencies; improves error correction	Complex architecture; high computational cost
Y. Cao [3]	Empirical Mode Decomposition + ARIMA + SVR	Various Chinese cities, multi pollutants	Superior short-term accuracy	Combines linear and nonlinear modeling benefits	Performance decreases after 3 hours; lacks attention mechanisms
F. Naz [4]	CNN Quantum LSTM with Variational Mode Decomposition	Belfast dataset, PM2.5	$R^2 = 0.965$, RMSE = 9.85	Quantum enhanced accuracy gains; noise reduction	Requires quantum hardware; computationally demanding
T. Jayanth [5]	Holt-Winters + Transformer BiGRU with Bayesian Optimization	Amaravati, India, multiple pollutants	$R^2 = 0.99$	Captures seasonal and trend components; thorough hyperparameter tuning	Computationally expensive; complex model
S. Kumar [6]	Graph Convolutional Networks + Transformer GRU with Knowledge Distillation	Delhi, Mumbai, Kolkata, Bangalore datasets	~+2.36% R^2 improvement; 6.5x parameter reduction	Efficient spatial modeling; lighter inference models	Dataset limited to Indian cities; graph construction required
M. Ahmed [7]	Particle Swarm Optimized Neural Network	Industrial CO emission data, Italy	R^2 ~0.99999	Optimized weights; temporal pattern differentiation	Single pollutant (CO); limited domain applicability
Z. Xie [8]	Multi Task Enhanced Transformer + Bidirectional Mamba2 with Knowledge Distillation	Guangzhou, Chengdu, Beijing multi pollutant datasets	Maintains accuracy within 5% of teacher; 98% faster inference	Multi-pollutant prediction, edge deployment	Complex architecture; steep learning curve

K. Chatterjee [9]	Kalman Filter + ANN Hybrid	Indoor scenarios in Hyderabad	Improved energy efficiency by 48-60%	Practical deployment in smart buildings; energy saving	Limited to indoor environment; less generalizable outdoors
K. Ramesh [10]	AirQuaNet CNN with Multi Scale Feature Learning and Attention	Health impact linked pollutant datasets	Low RMSE, high R^2	Relates pollutants to health outcomes; multi scale feature extraction	No real world deployment validation

IV. CHALLENGES AND LIMITATIONS

4.1 Dataset Limitations

- Limited Data Diversity and Size: Many air quality datasets suffer from limited geographic coverage, short time spans, and insufficient variability, which restrict the generalizability of trained models across different regions and climates.
- Sensor Quality and Missing Values: Variability in sensor accuracy, calibration, and occasional missing data points introduce noise and inconsistencies that challenge model robustness.
- Lack of Multi Pollutant and Multi Modal Data: A majority of datasets focus on single pollutants or lack integration with meteorological and traffic related data, limiting comprehensive forecasting performance.

4.2 Model Interpretability

- Opaque Deep Learning Models: The complexity of deep neural networks, especially hybrid architectures combining CNNs, RNNs, and transformers, results in low explainability.
- Requirement for Explainable AI (XAI): For regulatory acceptance and public trust, air quality forecasting models must integrate explainability methods to make predictions transparent and actionable for stakeholders.

4.3 Computational and Deployment Constraints

- High Computational Resources: Many state of the art DL models demand substantial training time and computational power, hindering their deployment in real time or resource constrained edge devices.
- Model Compression Trade offs: Though quantization and distillation reduce model size, they can sometimes lead to accuracy degradation, necessitating careful balance between efficiency and prediction fidelity.
- Scalability Challenges: Deploying forecasting models across large sensor networks with diverse hardware capabilities remains a practical hurdle.

V. CONCLUSION AND FUTURE WORK

5.1 CONCLUSION

The reviewed studies demonstrate the transformative impact of machine learning (ML) and deep learning (DL) technologies in advancing air quality forecasting. Hybrid models combining CNNs and RNNs, especially BiGRU architectures, have proven highly effective in capturing spatial temporal pollutant patterns, achieving impressive accuracy with embedded, quantized implementations optimized for real time deployment on edge devices. Multi step and multi pollutant forecasting frameworks leveraging LSTM chains and transformers further improve robustness and extend forecasting horizons. Additionally, hybrid statistical ML techniques and quantum enhanced models open new avenues for handling complex air quality dynamics. Despite these successes, challenges remain, including limitations imposed by dataset scarcity and variability, limited interpretability of deep models, and the substantial computational demands restricting widespread adoption in resource constrained environments.

5.2 FUTURE WORK

To overcome current obstacles and further progress air quality forecasting, future research should focus on:

1. Expanding Dataset Diversity and Coverage: Coordinated efforts for generating large scale, multi city, multi seasonal datasets encompassing diverse pollutant types, meteorological conditions, and urban settings to improve model generalizability.
2. Integrating Multi Modal and Multi Source Data: Developing models that seamlessly fuse data from sensor networks, satellite observations, traffic, and meteorology to comprehensively model pollution dynamics.
3. Enhancing Model Explainability: Incorporating Explainable AI (XAI) tools to increase transparency of DL model

decisions, enabling better trust and adoption by policymakers and environmental agencies.

4. Optimizing Real Time Edge Deployment: Advancing lightweight, quantized architectures and adaptive online learning methods for low latency and energy efficient deployment in IoT and edge computing platforms.
5. Improving Forecasting Horizons: Developing novel architectures that maintain accuracy over extended forecast windows beyond current multi hour limits using transformer and attention mechanisms.
6. Mitigating Bias and Ensuring Fairness: Creating representative datasets addressing diverse urban, climatic, and socio economic contexts to minimize algorithmic biases and support equitable environmental policies.

Addressing these priorities will be essential for realizing scalable, interpretable, and accurate air quality prediction systems that can effectively guide public health interventions and environmental management worldwide.

REFERENCES

[1] A. 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. 123678-123695, 2025, doi: 10.1109/ACCESS.2025.3603920.

[2] K. Chatterjee, S. S. Kumar, R. P. Kumar, et al., "Future Air Quality Prediction Using Long Short-Term Memory Based on Hyper Heuristic Multi Chain Model," IEEE Access, vol. 12, pp. 123678-123705, 2024, doi: 10.1109/ACCESS.2024.3441109.

[3] Y. Cao, D. Zhang, S. Ding, W. Zhong, and C. Yan, "A Hybrid Air Quality Prediction Model Based on Empirical Mode Decomposition," Tsinghua Science and Technology, vol. 29, no. 1, pp. 99-111, Feb. 2024, doi: 10.26599/TST.2022.9010060.

[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. 94212-94227, 2025, doi: 10.1109/ACCESS.2025.3570526.

[5] T. Jayanth, A. Manimaran, V. R. K. Reddy, and R. N., "Enhancing Air Quality Prediction Through Holt–Winters Smoothing and Transformer-BiGRU With Bayesian Optimization," IEEE Access, vol. 13, pp. 180756–180780, 2025, doi: 10.1109/ACCESS.2025.3621231.

[6] S. Kumar, V. Kour, A. Raj, et al., "Optimizing Air Pollution Forecasting Models Through Knowledge Distillation: A Novel GCN and TRANS_GRU Methodology for Indian Cities," IEEE Access, vol. 13, 2025, doi: 10.1109/ACCESS.2025.3546504.

[7] M. Ahmed, S. Islam, M. H. Sulaiman, M. M. Hassan, and T. Bhuiyan, "MetaForecaster: A PSO-Driven Neural Model for Sustainable Industrial Air Quality Management," IEEE Access, vol. 13, pp. 121670-121681, 2025, doi: 10.1109/ACCESS.2025.3587716.

[8] Z. A. Xie, C. O. Chow, J. H. Chuah, and W. J. K. R., "Knowledge-Distilled Multi-Task Model With Enhanced Transformer and Bidirectional Mamba2 for Air Quality Forecasting," IEEE Access, vol. 13, pp. 158870-158880, 2025, doi: 10.1109/ACCESS.2025.3595679.

[9] K. Chatterjee, M. Raju, B. Bala, et al., "Indoor Air Wellness: A Predictive Model for Pollution Control Using Advanced AI Techniques," IEEE Access, vol. 13, pp. 42099-42114, 2025, doi: 10.1109/ACCESS.2025.3547808.

[10] S. Chadalavada, S. Yaman, A. Sengur, et al., "AirQuaNet: A Convolutional Neural Network Model With Multi-Scale Feature Learning and Attention Mechanisms for Air Quality-Based Health Impact Prediction," IEEE Access, vol. 13, pp. 96261-96276, 2025, doi: 10.1109/ACCESS.2025.3574722.

[11] R. Faldo, S. Mandala, R. P. Astuti, et al., "APD-BayNet: Jakarta Air Quality Index Prediction Using Bayesian Optimized TabNet," IEEE Access, vol. 13, pp. 57734-57748, 2025, doi: 10.1109/ACCESS.2025.3555961.

[12] B. C. M. Tam, S.-K. Tang, and A. Cardoso, "A Multi-Seasonal SS-MSTL-DR Approach With Efficient Training Using Deep Learning and LLaMA: A Case Study of 6-Element Air Quality Prediction in a Subtropical City," IEEE Access, vol. 13, pp. 172542–172546, 2025, doi: 10.1109/ACCESS.2025.3615518.

[13] W. Cao, R. Zhang, and W. Cao, "Multi-Site Air Quality Index Forecasting Based on Spatiotemporal Distribution and PatchTST: Enhanced Evidence From Hebei Province in China," IEEE Access, vol. 12, 2024, doi: 10.1109/ACCESS.2024.3460187.

[14] A. Abalo-García, S. Hernández-García, I. Ramírez, and E. Schiavi, "MPD: A Meteorological and Pollution Dataset: A Comprehensive Study of Machine and Deep Learning Methods for Air Pollution Forecasting," IEEE Access, vol. 13, pp. 41282-41283, 2025, doi: 10.1109/ACCESS.2025.3547038.

[15] Y. Boddu, A. Manimaran, B. Arunkumar, M. Sucharitha, and J. Suresh Babu, "Advanced Air Quality Forecasting Using an Enhanced Temporal Attention-Driven Graph Convolutional Long Short-Term Memory Model With Seasonal-Trend Decomposition," IEEE Access, vol. 13, 2024, doi: 10.1109/ACCESS.2024.3515095.

[16] O. S. Gomes, M. O. Binelo, M. F. B. Binelo, J. P. C. Oliveira, E. Galvani, and R. R. Alves, "An Artificial Neural Network for Short Time Air Temperature Prediction," IEEE Access, vol. 13, pp. 77593–77597, 2025, doi: 10.1109/ACCESS.2025.3565731.