

Article

Not peer-reviewed version

Smart Monitoring of Air and Waste Using Machine Learning and IoT Integration Approach

Muhammad Khizer and [Rizwan Ayazuddin](#) *

Posted Date: 17 November 2025

doi: 10.20944/preprints202511.1247.v1

Keywords: machine learning; IoT; air quality; waste management



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Smart Monitoring of Air and Waste Using Machine Learning and IoT Integration Approach

Muhammad Khizer ¹ and Rizwan Ayazuddin ^{2,*}

¹ Software Engineering, University of Sialkot, Sialkot, Pakistan
² School of Computer Science, Taylor's University, Subang Jaya, Malaysia
* Correspondence: rizayazuddin@gmail.com

Abstract

This raises serious concerns for public health and environmental sustainability in an increasingly polluted atmosphere. Therefore, advanced monitoring systems must be developed. This research paper presents a novel framework that integrates Machine Learning and Internet of Things (IoT) technologies to monitor and manage air quality and waste in real time. The proposed system utilizes a network of sensors to collect high-resolution data on air pollutants such as PM2.5, PM10, NOx, and CO2, along with waste management parameters such as bin occupancy, using a publicly available dataset from Kaggle. Following rigorous data preprocessing and feature engineering, the framework achieves a peak prediction accuracy of 93.53% using an ANN. The web-based platform enables automated analysis of continuous data, allowing for immediate alerts when pollutant thresholds are exceeded facilitating timely interventions.

Keywords: machine learning; IoT; air quality; waste management

1. Introduction

Air pollution has emerged as one of the most pressing global issues, posing significant challenges to public health, environmental sustainability, and economic development. According to the WHO, millions of premature deaths occur annually due to exposure to air pollutants. Among these, PM2.5, PM10, NOx, and CO2 are considered the most hazardous.

Traditional air quality monitoring systems, typically government-owned and operated, face several limitations: high operating costs, limited spatial coverage, and insufficient capacity for continuous data collection. These challenges hinder their ability to inform policy and enable timely responses to air quality emergencies [14].

Recent advances in the Internet of Things (IoT) and Machine Learning (ML) offer promising solutions to these limitations. IoT facilitates the deployment of interconnected sensors capable of collecting real-time data on air quality and waste management parameters across various geographic regions. This allows for more comprehensive and granular environmental monitoring. Meanwhile, ML algorithms can process large volumes of data, detect patterns, and generate accurate predictions about air quality, thereby enhancing decision-making processes [15].

Several studies have explored the integration of ML with IoT for environmental monitoring. For example, Bellinger et al. conducted a statistical review in air pollution epidemiology, noting a growing number of studies employing ML models such as SVM and ANN for air quality estimation and health impact analysis. Similarly, Hu et al. proposed an ML-based model called HazeEst, which leverages both fixed and mobile sensor data to predict air pollution levels at high spatial resolutions [16].

Despite these advancements, challenges remain particularly in ensuring the reliability and relevance of sensor data, especially from low-cost IoT devices [17-19]. Calibration issues and inconsistent sensor performance can result in poor data quality, which in turn leads to weak ML model performance [20-23]. Moreover, integrating multiple data streams necessitates standardized data consolidation methods to ensure consistency and reliable analysis [24-26].

This paper proposes an integrated framework that combines ML and IoT technologies for real-time air quality monitoring and waste management[27-29]. By utilizing high-resolution sensor data, the system aims to enhance predictive accuracy and provide actionable insights for policymakers and public health officials [30-32]. The following sections present the methodology, results, and implications of this research, contributing to ongoing efforts to improve air quality management and public well-being [33-34].

2. Literature Review

Air pollution has emerged as one of the most pressing global issues, posing significant challenges to public health, environmental sustainability, and economic development. According to the WHO, millions of premature deaths occur annually due to exposure to air pollutants. Among these, PM_{2.5}, PM₁₀, NO_x, and CO₂ are considered the most hazardous.

Traditional air quality monitoring systems, typically government-owned and operated, face several limitations: high operating costs, limited spatial coverage, and insufficient capacity for continuous data collection. These challenges hinder their ability to inform policy and enable timely responses to air quality emergencies [14].

Recent advances in the IoT and machine learning offer promising solutions to these limitations. IoT facilitates the deployment of interconnected sensors capable of collecting real-time data on air quality and waste management parameters across various geographic regions. This allows for more comprehensive and granular environmental monitoring. Meanwhile, ML algorithms can process large volumes of data, detect patterns, and generate accurate predictions about air quality, thereby enhancing decision-making processes [15].

Several studies have explored the integration of ML with IoT for environmental monitoring. For example, Bellinger et al. conducted a statistical review in air pollution epidemiology, noting a growing number of studies employing ML models such as SVM and ANN for air quality estimation and health impact analysis. Similarly, Hu et al. proposed an ML-based model called HazeEst, which leverages both fixed and mobile sensor data to predict air pollution levels at high spatial resolutions [16].

Despite these advancements, challenges remain particularly in ensuring the reliability and relevance of sensor data, especially from low-cost IoT devices. Calibration issues and inconsistent sensor performance can result in poor data quality, which in turn leads to weak ML model performance. Moreover, integrating multiple data streams necessitates standardized data consolidation methods to ensure consistency and reliable analysis [17].

This paper proposes an integrated framework that combines ML and IoT technologies for real-time air quality monitoring and waste management. By utilizing high-resolution sensor data, the system aims to enhance predictive accuracy and provide actionable insights for policymakers and public health officials. The following sections present the methodology, results, and implications of this research, contributing to ongoing efforts to improve air quality management and public well-being [18].

3. Methodology

The proposed research presents a comprehensive framework that integrates machine learning and IoT technologies for real-time air quality monitoring and waste management. Sensors deployed in urban, industrial, and remote areas collect high-resolution, real-time data on air pollutants such as PM_{2.5}, PM₁₀, NO_x, and CO₂, as well as waste management parameters including bin occupancy and waste categorization. These sensors ensure both spatial and temporal coverage by leveraging fixed monitoring stations and mobile units.

Rigorous preprocessing is conducted on the collected data to handle missing values, remove noise, and eliminate inconsistencies. Error markers are used to fill missing values, and outliers are removed to maintain data integrity. Feature engineering incorporates relevant meteorological parameters temperature, humidity, and wind speed to provide environmental context for air quality predictions. Data normalization ensures that inputs from different sensors are standardized for compatibility.

3.1. About Dataset

The dataset used in this study, publicly available from Kaggle, focuses on air quality assessment across various regions. It consists of 5,000 samples and includes environmental and demographic features that influence pollution levels. Table 1 factors that influence pollution levels.

Table 1. Factors That Influence Pollution Levels.

Temperature (°C)	Humidity (%)	PM2.5 Concentration (µg/m³)	PM10 Concentration (µg/m³)	NO2 Concentration (ppb)
SO2 Concentration (ppb)	CO Concentration (ppm)	Proximity to Industrial Areas (km)	Population Density (people/km²)	Air Quality

The trained ANN model is deployed on a cloud-based IoT platform where sensor data is transmitted securely via protocols such as MQTT or HTTP. The system continuously processes data in real time, performs analysis, and generates predictions. Alerts are triggered when pollutant levels exceed defined thresholds, enabling timely responses from authorities. Results and insights are displayed on an interactive dashboard. The platform also incorporates feedback mechanisms, allowing the ANN model to retrain with updated data and improve predictive accuracy over time as environmental conditions evolve.

3.2. Proposed Framework

The framework addresses key challenges in environmental monitoring, such as sensor calibration, data standardization, and heterogeneous data quality. Periodic calibration ensures accuracy in low-cost IoT sensors. Standardized data collection protocols are adopted to reconcile data from various sources and sensor types. Scalability is a central consideration: the system is designed to adapt to broader geographic areas and diverse environmental conditions, making it suitable for deployment in large urban centers.

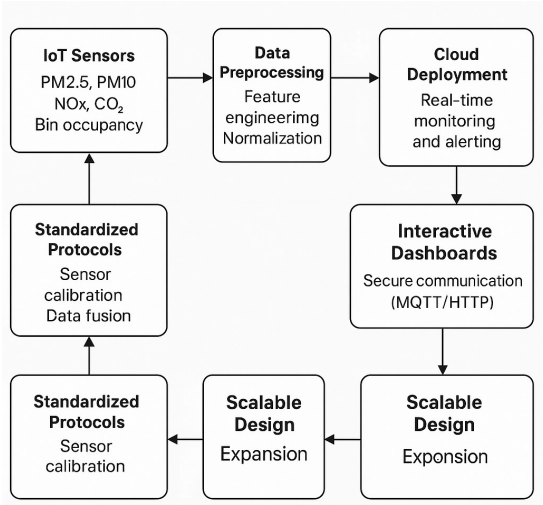


Figure 1. IoT Research Proposes an IoT Integrated ML Driven Framework.

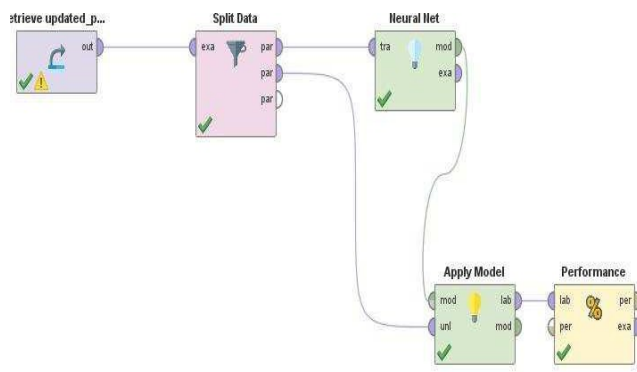


Figure 2. Rapid Miner Model.

4. Results

The proposed framework demonstrates a robust integration of machine learning and IoT technologies for efficient air quality monitoring and waste management. A preprocessed dataset containing key air quality parameters PM2.5, PM10, NOx, CO₂ and waste management metrics such as bin occupancy was used to train and test various machine learning models.

The dataset was split into training (70%) and testing (30%) subsets. K-fold cross-validation was employed to ensure model validation and prevent overfitting. Among the tested models, the ANN achieved the highest accuracy of 93.53%, outperforming other models as shown in table

Table 2. Final Accuracy Achieved.

Sr. No	Selected Algorithm	Accuracy
i.	K-Nearest Neighbors	79.07%
ii.	Navie Bayes	91.73
iii.	Decision Tree	86.0%,
iv.	Random Forest	89.0%
v.	Artificial Neural Network	93.53%
vi.	Deep Learning	80.33%
vii.	Ensemble Vote	89.73%
viii.	Bagging (NB)	91.67%

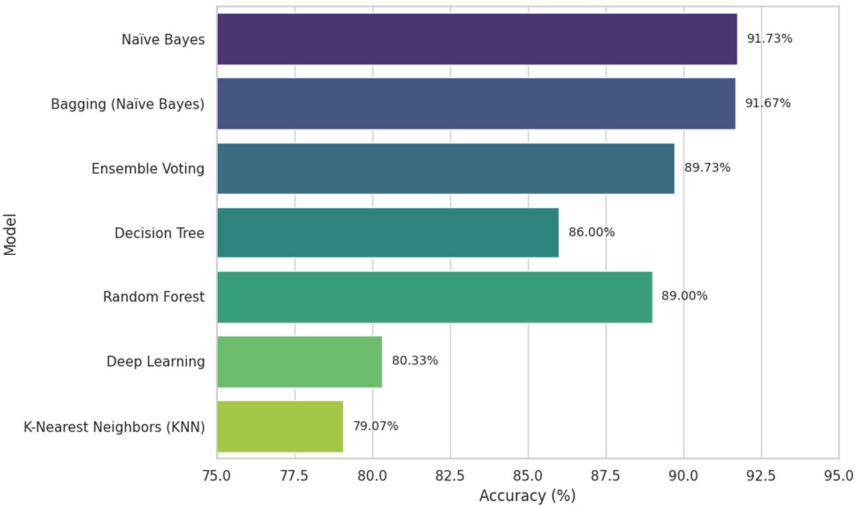


Figure 3. Accuracy Comparison of Machine Learning Models.

Compared to conventional systems, which typically rely on limited data and simpler algorithms, this intelligent framework presents significant advantages in scalability, accuracy, and real-time responsiveness. The ANN model's capacity to uncover nonlinear and multi-dimensional patterns is especially beneficial, given the complexity of environmental systems.

5. Conclusions

This research highlights the promising potential of integrating ML and IoT technologies in air quality monitoring, real-time observation, and waste management. The proposed model, which employs a network of high-resolution sensors and an ANN, achieves a peak prediction accuracy of 93.53% for waste composition. The cloud-based platform enables continuous data analysis and automatic alerts, facilitating timely interventions to address air quality issues and safeguard public health. Despite these advancements, challenges such as inconsistent data quality, sensor calibration issues, and the lack of standardized data collection protocols continue to hinder optimal system performance. To address these limitations and enhance the proposed framework, future work will focus on standardizing sensor calibration and data acquisition protocols to ensure data reliability and consistency. The exploration of advanced ML techniques, including deep learning and ensemble methods, will be pursued to further improve model accuracy and robustness. The incorporation of geo-spatial and meteorological data will enable more detailed analysis of pollutant distribution and their environmental impacts. Furthermore, scaling the framework to larger and more diverse urban areas will assess its adaptability and effectiveness in varied environmental conditions. These future efforts aim to contribute meaningfully toward the development of more effective and sustainable environmental monitoring systems that protect public health and promote long-term ecological balance.

References

1. S. Kabir, R. Ul Islam, M. S. Hossain, and K. Andersson, "An integrated approach of belief rule base and deep learning to predict air pollution," *Sensors (Switzerland)*, vol. 20, no. 7, pp. 1–25, Apr. 2020, doi: 10.3390/s20071956.
2. Essamlali, H. Nhaila, and M. el Khaili, "Supervised Machine Learning Approaches for Predicting Key Pollutants and for the Sustainable Enhancement of Urban Air Quality: A Systematic Review," *Sustainability (Switzerland)*, vol. 16, no. 3. Multidisciplinary Digital Publishing Institute (MDPI), Feb. 01, 2024. doi: 10.3390/su16030976.
3. K. Hu, A. Rahman, H. Bhugubanda, and V. Sivaraman, "HazeEst: Machine Learning Based Metropolitan Air Pollution Estimation from Fixed and Mobile Sensors," *IEEE Sensors Journal*, vol. 17, no. 11, pp. 3517–3525, Jun. 2017, doi: 10.1109/JSEN.2017.2690975.
4. M. A. Haq, "Smotednn: A novel model for air pollution forecasting and aqi classification," *Computers, Materials and Continua*, vol. 71, no. 1, pp. 1403–1425, 2022, doi: 10.32604/cmc.2022.021968.
5. K. S. Rautela and M. K. Goyal, "Transforming air pollution management in India with AI and machine learning technologies," *Scientific Reports*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-71269-7.
6. Bekkar, B. Hssina, S. Douzi, and K. Douzi, "Air-pollution prediction in smart city, deep learning approach," *Journal of Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021- 00548-1.
7. Y. Xu et al., "Evaluation of machine learning techniques with multiple remote sensing datasets 2 in estimating monthly concentrations of ground-level PM," 2018.
8. C. Bellinger, M. S. Mohamed Jabbar, O. Zaiane, and A. Osornio-Vargas, "A systematic review of data mining and machine learning for air pollution epidemiology," *BMC Public Health*, vol. 17, no. 1. BioMed Central Ltd., Nov. 28, 2017. doi: 10.1186/s12889-017-4914-3.
9. S. Al-Janabi, M. Mohammad, and A. Al-Sultan, "A new method for prediction of air pollution based on intelligent computation," *Soft Computing*, vol. 24, no. 1, pp. 661–680, Jan. 2020, doi: 10.1007/s00500-019-04495-1.
10. U. Rehman et al., "A Machine Learning-Based Framework for Accurate and Early Diagnosis of Liver Diseases: A Comprehensive Study on Feature Selection, Data Imbalance, and Algorithmic Performance,"

- International Journal of Intelligent Systems, vol. 2024, no. 1, Jan. 2024, doi: <https://doi.org/10.1155/2024/6111312>.
11. T. M. Ali et al., "A Sequential Machine Learning-cum- Attention Mechanism for Effective Segmentation of Brain Tumor," *Frontiers in Oncology*, vol. 12, Jun. 2022, doi: <https://doi.org/10.3389/fonc.2022.873268>.
 12. Mir et al., "A novel approach for the effective prediction of cardiovascular disease using applied artificial intelligence techniques," *ESC heart failure*, Jul. 2024, doi: <https://doi.org/10.1002/ehf2.14942>.
 13. Nawaz et al., "A Comprehensive Literature Review of Application of Artificial Intelligence in Functional Magnetic Resonance Imaging for Disease Diagnosis," *Applied Artificial Intelligence*, pp. 1–19, Oct. 2021, doi: <https://doi.org/10.1080/08839514.2021.1982185>.
 14. Muzafar, S., & Jhanjhi, N. Z. (2020). Success stories of ICT implementation in Saudi Arabia. In *Employing Recent Technologies for Improved Digital Governance* (pp. 151-163). IGI Global Scientific Publishing.
 15. Jabeen, T., Jabeen, I., Ashraf, H., Jhanjhi, N. Z., Yassine, A., & Hossain, M. S. (2023). An intelligent healthcare system using IoT in wireless sensor network. *Sensors*, 23(11), 5055.
 16. Shah, I. A., Jhanjhi, N. Z., & Laraib, A. (2023). Cybersecurity and blockchain usage in contemporary business. In *Handbook of Research on Cybersecurity Issues and Challenges for Business and FinTech Applications* (pp. 49-64). IGI Global.
 17. Hanif, M., Ashraf, H., Jalil, Z., Jhanjhi, N. Z., Humayun, M., Saeed, S., & Almuhaideb, A. M. (2022). AI-based wormhole attack detection techniques in wireless sensor networks. *Electronics*, 11(15), 2324.
 18. Rautela, K. S., Goyal, M. K., & Surampalli, R. Y. (2025). AI and Machine Learning for Optimizing Waste Management and Reducing Air Pollution. *Journal of Hazardous, Toxic, and Radioactive Waste*, 29(3), 04025014.
 19. Madan, B., Nair, S., Katariya, N., Mehta, A., & Gogte, P. (2025). Smart waste management and air pollution forecasting: Harnessing Internet of things and fully Elman neural network. *Waste Management & Research*, 0734242X241313286.
 20. Abbood, M. M. (2025). Investigation of IoT and Deep Learning Techniques Integration for Smart City Applications. *American Journal of Computing and Engineering*, 8(1), 57-68.
 21. Shah, I. A., Jhanjhi, N. Z., Amsaad, F., & Razaque, A. (2022). The role of cutting-edge technologies in industry 4.0. In *Cyber Security Applications for Industry 4.0* (pp. 97-109). Chapman and Hall/CRC.
 22. Humayun, M., Almufareh, M. F., & Jhanjhi, N. Z. (2022). Autonomous traffic system for emergency vehicles. *Electronics*, 11(4), 510.
 23. Muzammal, S. M., Murugesan, R. K., Jhanjhi, N. Z., & Jung, L. T. (2020, October). SMTrust: Proposing trust-based secure routing protocol for RPL attacks for IoT applications. In *2020 International Conference on Computational Intelligence (ICCI)* (pp. 305-310). IEEE.
 24. Brohi, S. N., Jhanjhi, N. Z., Brohi, N. N., & Brohi, M. N. (2023). Key applications of state-of-the-art technologies to mitigate and eliminate COVID-19. *Authorea Preprints*.
 25. William, P., Kuppusamy, S., Samundeeswari, A., Vijayabharathi, R., Hajiyeva, R. J., Thangavel, K., & Alagarsamy, M. (2025). Integrating IoT and Machine Learning for Real-Time Monitoring of Infectious Waste Management. In *Hospital Waste Management and Toxicity Evaluation* (pp. 299-320). IGI Global Scientific Publishing.
 26. Lakhouit, A. (2025). Revolutionizing urban solid waste management with AI and IoT: a review of smart solutions for waste collection, sorting, and recycling. *Results in Engineering*, 104018.
 27. Khalil, M. I., Humayun, M., Jhanjhi, N. Z., Talib, M. N., & Tabbakh, T. A. (2021). Multi-class segmentation of organ at risk from abdominal ct images: A deep learning approach. In *Intelligent Computing and Innovation on Data Science: Proceedings of ICTIDS 2021* (pp. 425-434). Singapore: Springer Nature Singapore.
 28. Humayun, M., Jhanjhi, N. Z., Niazi, M., Amsaad, F., & Masood, I. (2022). Securing drug distribution systems from tampering using blockchain. *Electronics*, 11(8), 1195.
 29. Lim M., Abdullah A., Jhanjhi N.Z. (2021). Performance optimization of criminal network hidden link prediction model with deep reinforcement learning. *Journal of King Saud University - Computer and Information Sciences*, 33(10), 1202-1210. <https://doi.org/jksuci.2019.07.010>

30. Ahmed Q.W., Garg S., Rai A., Ramachandran M., Jhanjhi N.Z., Masud M., Baz M. (2022). AI-Based Resource Allocation Techniques in Wireless Sensor Internet of Things Networks in Energy Efficiency with Data Optimization. *Electronics* (Switzerland), 11(13), . <https://doi.org/electronics11132071>
31. Aldughayfiq B., Ashfaq F., Jhanjhi N.Z., Humayun M. (2023). YOLO-Based Deep Learning Model for Pressure Ulcer Detection and Classification. *Healthcare* (Switzerland), 11(9), . <https://doi.org/healthcare11091222>
32. Ogbolumani, O. A., & Adekoya, M. (2025). Intelligent waste management optimization through machine learning analytics. *Journal of Science Research and Reviews*, 2(1), 7-26.
33. Arun, M. (2025). Investigation of a deep learning-based waste recovery framework for sustainability and a clean environment using IoT. *Sustainable food technology*, 3(2), 599-611.
34. Kumar T., Pandey B., Mussavi S.H.A., Zaman N. (2015). CTHS Based Energy Efficient Thermal Aware Image ALU Design on FPGA. *Wireless Personal Communications*, 85(3), 671-696. <https://doi.org/s11277-015-2801-8>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.