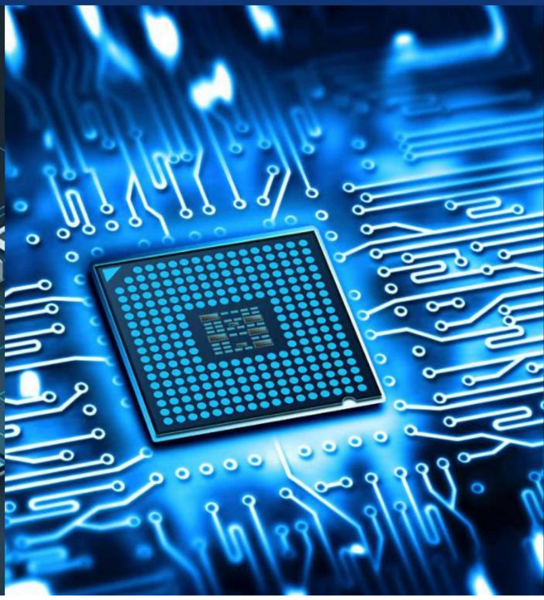
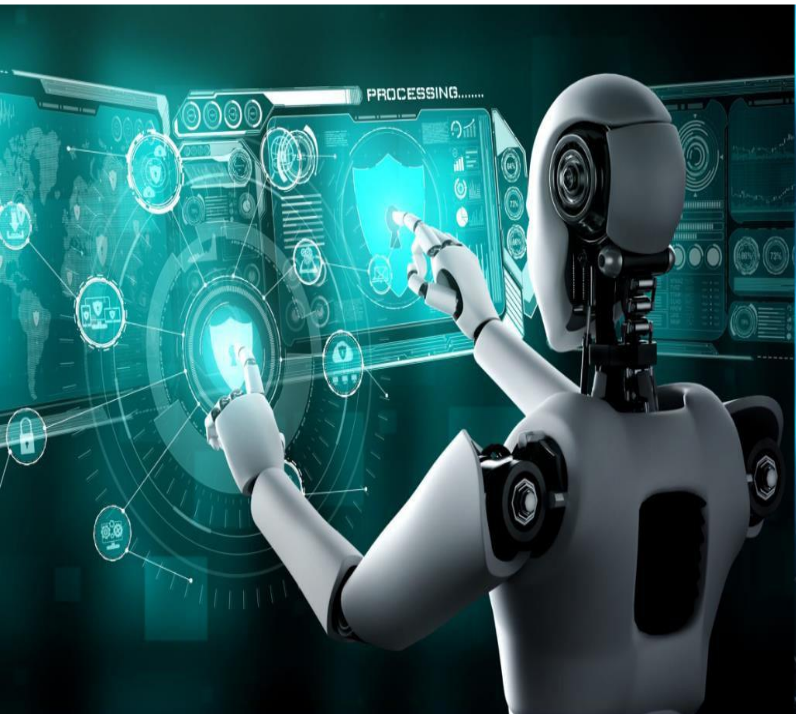
ISSN(O): 2320-9801
ISSN(P): 2320-9798


International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771

Volume 14, Issue 4, April 2026

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Intelligent Visual Analytics for Monitoring Air Quality with AI

Ramya V, Sujith.S, Sujithkumar.V, Swaraj.S, Sethupathy.D

Assistant Professor, Department of Computer Science and Engineering, The Kavery Engineering College, Mecheri, Salem, India

Department of Computer Science and Engineering, The Kavery Engineering College, Mecheri, Salem, India

ABSTRACT: The IoT-based device was integrated with the thinger.io platform, which has proven stability and effectiveness. This platform enabled stable, real-time data storage and transmission, providing cost-effective and reliable environmental monitoring. The system's robustness and affordability present significant advantages, making it scalable for wider use in similar regions with budget constraints. Monthly data collection at one-minute intervals provided overall insight into the pollution levels, alongside humidity and temperature data. The dataset was classified using five machine learning algorithms, leading to important findings. The Gradient Boosting (GD) model reached the best precision, closely followed by the Random Forest (RF) model. The Support Vector Classifier (SVC) has proven to have a strong performance, while K-Nearest neighbors (KNN) achieved an acceptable accuracy. Logistic Regression (LR) exhibited relatively lower accuracy. Notably, the region recorded high pollution levels, with predominant classifications falling into the higher risk categories, from unhealthy and very unhealthy to hazardous. The results highlight the ability of machine learning (ML) techniques to accurately classify AQ data. The successful integration of IoT platforms with ML models demonstrates the potential for developing low-cost, scalable, and stable systems for real-time AQ monitoring. This hybrid approach is crucial, especially in developing countries with limited research on pollution and budget constraints, as it may help improve environmental management in highly populated and industrialized regions.

KEYWORDS: Air pollution,Artificial intelligence, Deep learning, Monitoring tools, Health PRISMA guidelines

I. INTRODUCTION

The economic and urban development of cities has led to increased air pollution, which has become a major problem for human health. Worldwide, an estimated 4.2 million premature deaths are related to air pollution: 29% due to lung cancer, 17% to acute lower respiratory tract infections, 24% to stroke, 25% to heart disease, and 43% to chronic obstructive pulmonary disease [1]. According to the World Health Organization (WHO), the main air pollutants are CO, SO2, and NO2. In addition to these gases, volatile organic compounds (VOC) and particulate matter (PM) also pose a serious threat [2]. Exposure to these contaminants can cause minor problems, i.e., nose and eye irritation. In the long term, however, it can even cause deadly diseases, such as cancer. High levels of pollution increase the risk of acid rain, which can cause disease in humans [3]. Furthermore, the ecosystem is adversely affected, influencing the growth of trees and plants. The technology used in industrialized countries leads to an increase in industries and vehicles and, therefore, higher levels of air pollution.

Air pollution is harmful to the environment, with some emissions contributing to global warming. Therefore, reducing air pollution is a crucial—albeit time-intensive—endeavor. One of the most important tasks in the fight against air pollution is the accurate measurement and prediction of air quality (AQ). This would enable the design of essential measures to prevent and minimize the effects of air pollution [4]. To gain knowledge about AQ levels in real time, technology is required that provides information and warnings and allows monitoring on the web. One of these technologies is the Internet of Things (IoT), which can transmit data through a network without human interaction. It is an internet-based system that facilitates links and interactions between the environment and users through the internet. The arrangement of a net of sensors, satellite information, and IoT devices has demonstrated the potential to acquire extensive AQ data over time.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

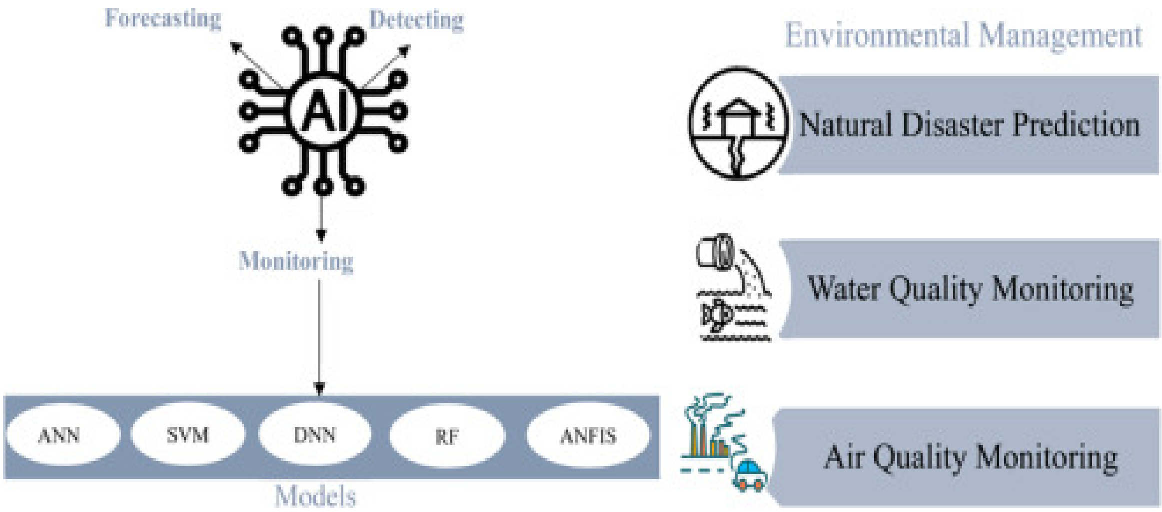


Fig 1: Artificial intelligence in environmental monitoring

This enables scientists to study tendencies, classify pollution sources, and evaluate control measure feasibility. Complete air information is analyzed in the IoT cloud. With respect to IoT design, the system involves sensing, network, and application layers. With such extensive environmental data, accurate predictions can only be made through detailed evaluation, which can be successfully carried out by machine learning (ML) algorithms [5]. A standard IoT system typically involves four factors [6]. These are IoT devices, IoT gateway and connectivity, cloud and data processing, and a user interface.

Several authors have assessed pollution levels using IoT technology. Witczak et al. provided an overview of IoT-based monitoring and control systems, discussing various aspects of these systems, including their architecture, applications, and challenges [7]. Edupuganti et al. designed an IoT-based methodology to measure, in real time, temperature, humidity, various gases, microbes, and light intensity. It included sensor nodes (MQ2, MQ135, DTH11, and LDR), a gateway (Arduino Uno), a Wi-Fi module (ESP8266) (Espressif Systems, Shanghai, China), an LCD, and a publicly accessible cloud server (ThingSpeak platform from The MathWorks, Inc., Natick, MA, USA). Graphs were built based on sensor information collected from the webpage, and were connected to Secure Digital (SD) cards for data storage. Pollution was monitored by connecting a Global Positioning System (GPS) module to the selected location before publishing the information on a public webpage [8]. To improve existing systems, Mohan et al. proposed a three-stage air pollution monitoring system.

Using gas sensors (DHT-22, MG811, MQ-7) and Raspberry Pi 4, an IoT package was designed. To efficiently monitor air pollution (humidity, temperature, noise), data were stored in a public cloud for pattern analysis [9]. Ardebili et al. highlighted the role of IoT in enhancing the resilience of cyber-physical systems using real-time monitoring. A smart solar panel supported by digital twin technologies was used to demonstrate the effectiveness of this approach in assessing the resilience of systems during disturbances [10]. Finally, Bai et al. demonstrated that continuous checking was effective in a remote detection system, alongside rapid web association. Likewise, wireless sensor network (WSN)-checking frameworks were executed considering different types of contamination, i.e., water, soil, or radioactive pollution [11]. To meet monitoring air pollution challenges in resource-limited environments, Zhu et al. used low-cost, energy-independent sensor systems. One example is based on energy harvesting, which is a promising option for ensuring environmental monitoring sustainability [12].

Samples were collected in a residential area in a suburb of Baghdad, Iraq, called the Dora area, which is located south of the city center. This industrialized and densely populated area has become a significant residential neighborhood. Following urban expansion, it now falls within the boundaries of the Dora oil refinery, established in 1955 and previously located in an uninhabited area. The coordinates of the Dora area are approximately 33.3035° N latitude and

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

44.3928° E longitude. Currently, residents suffer from the refinery emissions, which contribute to various health issues, including skin diseases and cancer, as well as visual distortions due to volatile compounds [16]. A thermal power plant, built in 1976 and operated by the Iraqi Ministry of Electricity, is located approximately 10 km away from the refinery. These two facilities (Figure 3) emit a large volume of pollutants, adversely affecting AQ in the surrounding area. According to the Environmental Reality Report issued in 2017 by the Iraqi Ministry of Environment, the Al-Jadriya area, adjacent to these facilities, also suffers from elevated levels of SO2 and CO. Local residents have expressed concerns about the environmental and health effects stemming from both the refinery and the power plant, demanding their relocation to outside Baghdad border [17].

II. RELATED REVIEWS

In recent years, there has been a surge in literature documenting efforts to harness the potential of ML and DL techniques for air pollution forecasting and monitoring. These efforts have resulted in notable improvements in prediction accuracy. This review paper explores the latest advancements in air pollution forecasting and real-time monitoring. To emphasize the unique and innovative aspects of our approach, we have summarized six existing review papers, facilitating a comparative analysis.

Our comprehensive review of AI applications in air quality research allows for an in-depth examination of the complex, multifaceted factors influencing pollution. We assess a spectrum of pollutants and their health impacts using ML and DL models. This allows the identification of the most suitable techniques for specific pollutants based on spatial and temporal characteristics.

Our review offers researchers a comprehensive and holistic perspective on AI techniques across air pollution analysis domains. We assess the evolution of monitoring and prediction capabilities based on ML versus DL methods. highlights our expansive scope compared to existing literature, spanning forecasting, monitoring, drivers, health impacts, ML, and DL approaches. This comprehensive vantage point empowers impactful solutions to this critical environmental and public health challenge.

The selection and review of articles on air pollution forecasting and monitoring using AI techniques in this paper comply with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines. To ensure a comprehensive article search, we have used the following databases: PubMed, ScienceDirect, and Web of Science. The search was carried out based on the following keywords: ”Air Pollution”, "Human Health,” "Machine Learning,” "Artificial Intelligence,” and "Deep Learning,”. This article's examination encompasses the most recently accepted papers until June 20, 2023.

Following PRISMA guidelines, the initial literature search identified 412 research papers. After removing 193 duplicate papers, 219 papers with unique titles were screened further. Papers were excluded at this stage if they did not meet the inclusion criteria of 1) being peer-reviewed research articles, 2) being published in English, 3) being relevant to the application of AI techniques for air pollution analysis, forecasting, or prediction, and 4) accessible in full text. Based on these criteria, 99 papers were excluded. The remaining 120 articles underwent a detailed assessment of quality and relevance to identify papers published in Q1 journals per Scimago rankings.

III. METHODS

Data quality is an essential prerequisite for improving ML models. Data preprocessing plays a crucial role in generalization capabilities, as it reduces noise. This improves the speed and implementation of ML algorithms, especially in the case of large datasets. Data mining and supervision techniques include feature selection, outlier removal, and missing value imputation. In the case of accurately estimating AQI using ML, effective preprocessing is essential for refining datasets, thereby improving the reliability and accuracy of models. Techniques, i.e., imputation and normalization, are critical to preparing input data for ML algorithms, ultimately contributing to more robust and accurate AQ predictions. By rigorously addressing outliers and biases, preprocessing facilitates the extraction of clearer information, which is essential for reliable AQI forecasting shows the percentage distribution of pollutants at the study location.

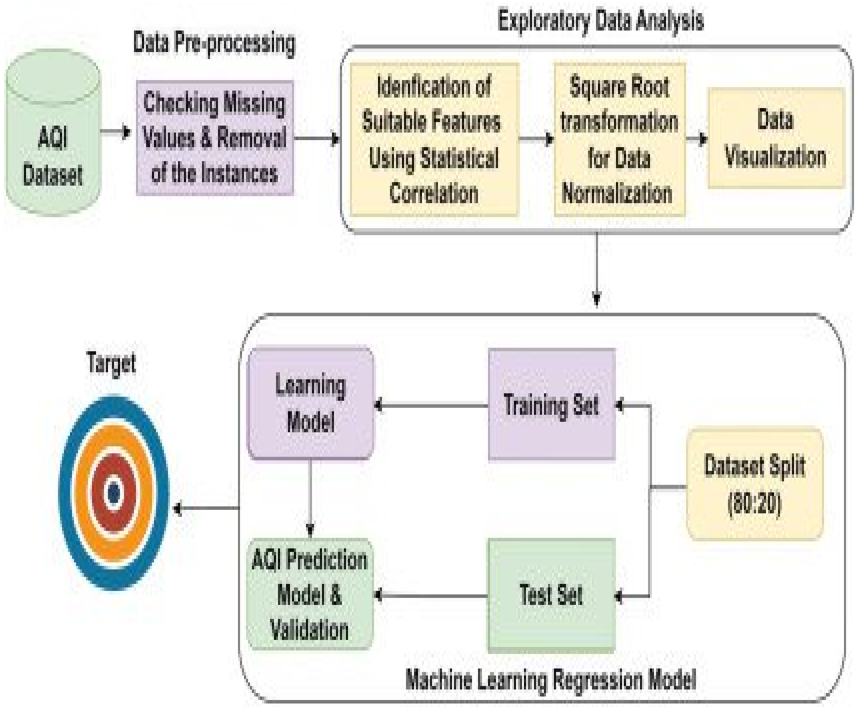


Fig 2: Air quality prediction by machine learning models

Data preprocessing began with inspection of the dataset for null values. As evidenced, the dataset did not contain any missing (null) values for any of the features. This indicates that data were fully recorded without gaps or potential issues, i.e., without data collection errors or equipment failures, so there were no missing entries. Therefore, unlike scenarios in which a significant amount of missing data must be addressed using methods like median imputation, these data did not require such measures.

A dataset may include extreme values that differ from other data and fall outside the estimated range. These are considered outliers. Understanding and even removing outliers can often improve ML model quality. Outliers include noise and distortion, which lead to biased model training, reducing generalization ability. These abnormalities can disrupt the learning process, resulting in a model that may fail to provide consistent and accurate AQI forecasts, affecting the overall quality and reliability of predictions. The presence of outliers in the data can seriously compromise the accuracy of AQI predictions using ML models.

IV. RESULT ANALYSIS

Governmental organizations rely on AQI, a specific metric within the dataset being investigated. This metric is used both to transfer AQ data to the population and for the training of analysts. Industrial emissions from activities, i.e., refining and fossil-based power generation, significantly contribute to air pollution. To enhance the precision of AQ assessments, each pollutant is categorized as one of six classes, depending on its concentration, ranging from “good” to “hazardous”. Numerical values are also assigned to pollutants according to a classification system, ranging from (0–good) to (5–hazardous), to facilitate ML algorithms for real-time analysis and prediction.

The categorization of each pollutant according to its health risks is essential. CO is considered dangerous in concentrations above 30.5 ppm, according to Environmental Protection Agency (EPA) standards, due to the severe risk of poisoning, which can affect the body’s oxygen supply. On the contrary, although CO_2 is not toxic at normal levels, it becomes hazardous when its concentration exceeds 10,000 ppm, leading to symptoms such as unconsciousness and dizziness, particularly in indoor spaces. Similarly, NH_3 is classified as hazardous when its concentration exceeds 200 ppm, which may cause severe respiratory irritation and other health complications. CH_4 is categorized as hazardous at

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

levels exceeding 40,000 ppm, at which there is risk of ignition, as outlined by Occupational Safety and Health Administration guidelines (OSHA). Additionally, LPG and H2 are not inherently toxic. They are classified as hazardous at concentrations above 5000 ppm and 200 ppm, respectively, due to their potential for explosive reactions.

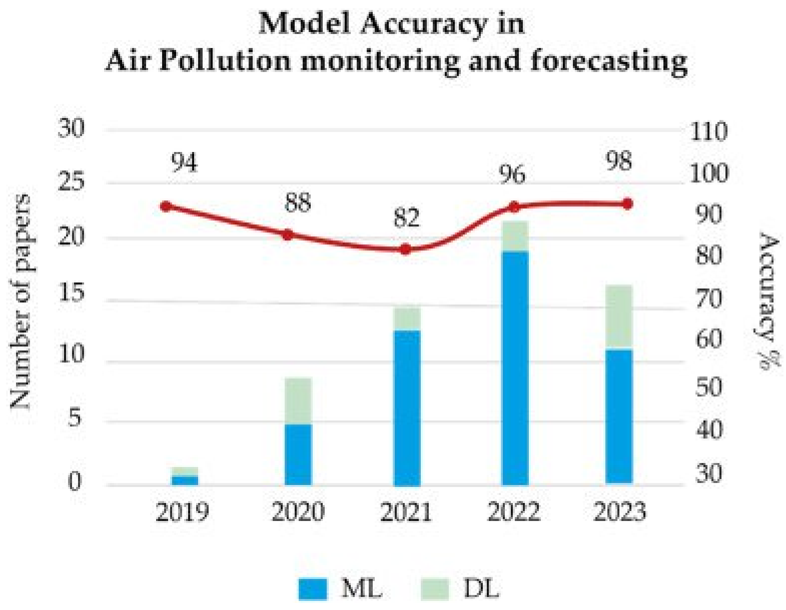


Fig 3: Application of artificial intelligence in air pollution monitoring and forecasting

Classification algorithms function as supervised learning (SL) methods that require labeled input data (pollutant concentration), while the output corresponds to predefined AQI categories. This process allows the model to learn from observations and accurately classify new instances. Unlike regression models, which produce numerical results, classification models generate categorical results. The methodologies used in this study utilize labeled datasets, in which each observation is aligned with a known AQI category established during the feature engineering phase, facilitating the training of ML models to consistently assess AQ.

To categorize the resulting AQI, supervised dedicated algorithms, containing SVC, RFC, Gradient boosting classifier, K-Nearest neighbors (KNN), and LR, were used. When designing ML models, several architecture options are often offered. Defining the optimal architecture for a particular model may be difficult, as it requires exploring a range of options using different hyperparameters. This manuscript focuses on exploring and selecting the optimal model architecture through hyperparameter tuning.

V. CONCLUSIONS AND FUTURE WORK

The direct impact of contamination on life quality emphasizes the crucial demand for comprehensive air quality (AQ) supervision, specifically in densely populated urban areas, like Baghdad (Iraq). With over 9 million inhabitants and 4 million vehicles, pollution levels in the city require thorough investigation. However, due to the lack of attention and limited research on pollution, many regions suffer from a lack of adequate data, hindering efforts to address environmental health concerns. This work aims to fill this gap by developing a low-cost, internet of things (IoT)-based device for the real-time detection of harmful gases and particulate matter. It was installed in the neighborhood of Dora (Baghdad's most polluted area) due to its proximity to an oil refinery and a power plant, coupled with vehicle emissions and dust accumulation. The main innovation of this project lies in the hybrid nature of the proposed monitoring and predicting system, combining both hardware and artificial intelligence (AI) components.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

REFERENCES

1. Méndez, M.; Merayo, M.G.; Núñez, M. Machine learning algorithms to forecast air quality: A survey. Artif. Intell. Rev. 2023, 56, 10031–10066. [Google Scholar] [CrossRef] [PubMed]
2. Rajasekar, D.; Sekar, A.; Rajasekar, M. Air Quality Monitoring and Disease Prediction Using IoT and Machine Learning. Int. J. Innov. Res. Comput. Sci. Technol. 2020, 8, 389–395. [Google Scholar] [CrossRef]
3. Nilesh, N.; Patwardhan, I.; Narang, J.; Chaudhari, S. IoT-based AQI Estimation using Image Processing and Learning Methods. In Proceedings of the 2022 IEEE 8th World Forum on Internet of Things, WF-IoT 2022, Yokohama, Japan, 26 October–11 November 2022. [Google Scholar] [CrossRef]
4. Vasantha, S.V.; Mageswari, R.U.; Ramesh, K.; Vaishnavi, J. Air Quality Prediction System using ML and DL Techniques. In Proceedings of the 2022 IEEE North Karnataka Subsection Flagship International Conference, NKCon 2022, Karnataka, India, 16–17 December 2022. [Google Scholar] [CrossRef]
5. Idrees, Z.; Zheng, L. Low cost air pollution monitoring systems: A review of protocols and enabling technologies. J. Ind. Inf. Integr. 2020, 17, 100123. [Google Scholar] [CrossRef]
6. Rajashekar, R.C. IoT-based Air Pollution Monitoring: Algorithms and Implementation. Master's Thesis, International Institute of Information Technology, Hyderabad, India, 2021. [Google Scholar]
7. Witczak, D.; Szymoniak, S. Review of Monitoring and Control Systems Based on Internet of Things. Appl. Sci. 2024, 14, 8943. [Google Scholar] [CrossRef]
8. Edupuganti, S.; Tenneti, N.S.S.; Iqbal, M.M.; Rajaram, G. An IoT Implemented Dynamic Air Pollution Monitoring System. EAI Endorsed Trans. Internet Things 2023, 9, e4. [Google Scholar] [CrossRef]
9. Mohan, A.M.; George, A.M.; Baby, A.; Gopi, S. Real-time Air Quality Index Monitoring and Alert System using IoT Technology. Int. J. Emerg. Res. Areas 2023, 3, 76–80. [Google Scholar] [CrossRef]
10. Ardebili, A.A.; Martella, C.; Longo, A.; Rucco, C.; Izzi, F.; Ficarella, A. IoT-Driven Resilience Monitoring: Case Study of a Cyber-Physical System. Appl. Sci. 2025, 15, 2092. [Google Scholar] [CrossRef]
11. Bai, Z.; Hu, Z.; Bian, K.; Song, L. Real-time Prediction for Fine-grained Air Quality Monitoring System with Asynchronous Sensing. In Proceedings of the ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing—Proceedings, Brighton, UK, 12–17 May 2019. [Google Scholar] [CrossRef]
12. Zhu, Q.; Zhu, L.; Wang, Z.; Zhang, X.; Li, Q.; Han, Q.; Yang, Z.; Qin, Z. Hybrid triboelectric-piezoelectric nanogenerator assisted intelligent condition monitoring for aero-engine pipeline system. Chem. Eng. J. 2025, 519, 165121. [Google Scholar] [CrossRef]
13. Rosca, C.-M.; Stancu, A. Integration of AI in Self-Powered IoT Sensor Systems. Appl. Sci. 2025, 15, 7008. [Google Scholar] [CrossRef]
14. Imam, M.; Adam, S.; Dev, S.; Nesa, N. Air quality monitoring using statistical learning models for sustainable environment. Intell. Syst. Appl. 2024, 22, 200333. [Google Scholar] [CrossRef]
15. Ghosh, H.; Tusher, M.A.; Rahat, I.S.; Khasim, S.; Mohanty, S.N. Water Quality Assessment Through Predictive Machine Learning. In Lecture Notes in Networks and Systems; Springer: Singapore, 2023. [Google Scholar] [CrossRef]
16. Environmental concerns surround Dora refinery and power station in Baghdad » 964media. Available online: https://en.964media.com/7756/ (accessed on 5 July 2024).
17. Smoke Al-Doura Chokes Baghdad's Skies: An Investigation—Al-Aalem. Available online: https://al-aalem.com/%D8%A3%D8%AF%D8%AE%D9%8 6 % D 8 % A 9 - % D 8 % A 7 % D 9 % 8 4 % D 8 % A F % D 9 % 8 8 % D 8 % B 1 % D 8 % A 9 - % D 8 % A A % D 8 % A E % D 9 % 8 6 % D 9 % 8 2 - % D 8 % B 3 % D 9 % 8 5%D8%A7%D8 % A 1 - % D 8 % A 8 % D 8 % B A % D 8 % A F % D 8 % A 7 % D8%AF-%D8%AA%D8%AD%D9%82%D9%8A%D9%82/#_ftn1 (accessed on 10 September 2024).
18. Tella, A.; Balogun, A.L.; Adebisi, N.; Abdullah, S. Spatial assessment of PM10 hotspots using Random Forest, K-Nearest Neighbour and Naïve Bayes. Atmos. Pollut. Res. 2021, 12, 101202. [Google Scholar] [CrossRef]
19. Anitha, M.; Kumar, L.S. Development of an IoT-Enabled Air Pollution Monitoring and Air Purifier System. Mapan—J. Metrol. Soc. India 2023, 38, 669–688. [Google Scholar] [CrossRef]
20. Micro AL2O3 Ceramic Tube, Tin Dioxide. Available online: https://www.hwsensor.com (accessed on 21 April 2024).