e-ISSN No.: 2582-4228

Journal of Indian Association for Environmental Management

Journal homepage: www.http://op.niscair.res.in/index/php/JIAEM/index

Applications of Machine Learning in Advanced Pollutant Detection

P.S. Gayathri[1], R. Krishna Sharma[1], S. Nagaveena[2]

[1] Department of Physics, S.T.Hindu College, Nagercoil-629002, Kanyakumari, Affiliated to ManonmaniamSundaranar University, Abishekapatti, Tirunelveli, Tamil Nadu-627012,India.

[2] Department of Nanoscience and Nanotechnology, S.T. Hindu College, Nagercoil - 629002, Kanyakumari.

Email: ps2gayathriraj24@gmail.com

Submitted:12January 2025Revised:19 January 2025 Accepted: 11 March 2025

***Abstract*:Air quality monitoring is essential for safeguarding both human health and the environment, especially as pollution continues to impact millions of lives globally. Detecting harmful pollutants like nitrogen dioxide (NO_2), carbon dioxide (CO_2), ozone (O_3), and ethanol (C_2H_5OH) is a key step toward addressing this critical challenge. This study employs zinc oxide (ZnO)-based sensors combined with machine learning to classify these pollutants effectively. Using decision tree and random forest algorithms in Python, pollutant types are predicted based on parameters such as concentration, temperature, and sensor response time. The dataset, sourced from prior research, underscores the remarkable sensitivity and stability of ZnO-based sensors, even under varying environmental conditions. Rigorous preprocessing ensured data accuracy, enabling reliable predictions and consistent outcomes. With high classification accuracy, this study demonstrates the transformative potential of integrating ZnO-based sensors with machine learning for real-time air quality monitoring, paving the way for actionable solutions to combat pollution and improve lives worldwide.**

***Author Keywords*: Air Quality Monitoring, Pollutant Detection, Machine Learning, Decision Tree, Random Forest, Environmental Health.**

I. Introduction

Air pollution is a serious and growing global challenge, affecting human health, ecosystems, and the climate (Laxmi Kant Bhardwaj et al., 2024; Gary M. Lovett et al., 2009). Pollutants like nitrogen dioxide (NO_2), carbon dioxide (CO_2), and ozone (O_3) are especially concerning due to their harmful effects. NO_2 can cause respiratory issues and contribute to problems like acid rain, CO_2 is a leading driver of climate change, and O_3 poses risks to both human health and agricultural productivity (Nelson Barros et al., 2024). These concerns highlight the need for accurate and real-time air quality monitoring to help mitigate pollution and protect the environment.

Traditional air pollution detection methods, such as gas chromatography and mass spectrometry, are highly reliable but expensive and complex, making them less practical for widespread, real-time use (Andrea Carolina Marcillo Lara et al., 2022). Zinc oxide (ZnO)-based sensors offer a promising alternative. These sensors are cost-effective, highly sensitive, and capable of quickly detecting gases like NO_2, CO_2, and O_3 under various environmental conditions (Bharat Pant et al., 2020). Additionally, the study provides valuable insights into sensor fabrication, further showcasing the capabilities of ZnO-based sensors for practical applications.

Pairing ZnO-based sensors with machine learning can significantly improve their ability to identify pollutants. Machine learning models like decision trees and random forests excel at analyzing complex data and classifying pollutants based on key factors such as gas concentration, temperature, and sensor response time (Anna Glinscaya et al., 2024). By integrating these technologies, this study aims to create practical and scalable solutions for real-time air quality monitoring, contributing to efforts to reduce pollution and its harmful effects.

Sensors for Air Quality Monitoring

Sensors are essential tools for monitoring air pollution, helping us detect and measure harmful pollutants in the environment. They provide real-time data that can guide timely decisions to protect public health and the environment. There are several types of sensors used for air quality monitoring, each designed for specific pollutants and applications based on their sensitivity and operating principles. The following are a few popular kinds of air quality sensors:

Electrochemical Sensors: These sensors work by triggering a chemical reaction with the target gas, which generates an electrical signal (Ali R. Jalalvand et al.,2025). They are highly accurate and are often used to measure gases like nitrogen dioxide (NO_2) and carbon monoxide (CO).

Semiconductor Sensors: Made from materials like metal oxides such as zinc oxide (ZnO), these sensors detect changes in electrical resistance when gases interact with their surface, making them ideal for detecting multiple gases (Shaivalini Singh, 2016).

Optical Sensors: These sensors use light to measure gas concentration, either by absorption or scattering. They are particularly useful for monitoring particulate matter and gases like ozone (O_3) (Sachin Dhawan et al., 2024).

Infrared Sensors: Commonly used for carbon dioxide (CO_2), these sensors rely on the absorption of specific wavelengths of infrared light by the gas molecules.

One important feature of any sensor is its response time, which indicates how quickly it can detect changes in pollutant levels. A faster response time is crucial for real-time air quality monitoring, especially in areas where pollution levels fluctuate rapidly. The sensor's response time is influenced by factors like the material used, temperature, and the specific gas being measured. The concentration of pollutants also plays a key role in how sensors operate. For instance, in semiconductor sensors, the concentration of the gas being measured is directly correlated with the change in electrical resistance. This relationship allows for accurate measurements of pollutant levels, which is critical for applications that demand precision.

The integration of ZnO-based sensors with advanced technologies, such as machine learning, opens up new possibilities for air quality monitoring systems. These sensors are not only cost-effective but also scalable, making them a practical choice for real-time air quality monitoring across diverse environments. By combining their unique capabilities with smart algorithms, ZnO sensors can help us tackle air pollution more efficiently and effectively.

Machine Learning in Air Quality Monitoring

Monitoring air quality is crucial as polluted air poses serious health risks, including respiratory and heart diseases, and harms the environment through climate change and ecosystem damage (Claudia Banciu et al.,2024). Key pollution sources, like industrial activities and urbanization, highlight the need for real-time monitoring to enable timely warnings and actions to reduce emissions (Guillaume Jean et al., 2024). Traditional systems, though effective, are often expensive and inaccessible, underscoring the demand for affordable, reliable, and scalable solutions to keep communities informed and safe.

Machine learning has made air quality monitoring smarter, faster, and more accessible. Unlike traditional methods that rely on expensive equipment, ML-powered sensors can quickly analyze large amounts of data, detect patterns, and classify air quality based on pollutant levels and environmental conditions. This leads to more accurate and real-time pollution detection, helping to provide timely insights for public health and environmental safety. ML models are also adaptable, adjusting to changing conditions to ensure reliable performance. Plus, their cost-effectiveness makes them a great option for large-scale use in both cities and rural areas. One of the key applications of machine learning in this field is pollutant classification. Models like Decision Trees and Random Forests analyze factors such as gas concentration, temperature, and sensor response time to accurately identify pollutants, making air quality monitoring more efficient and actionable.

Decision Tree

Decision tree algorithms are simple yet powerful tools in machine learning, ideal for tasks like pollutant classification (Jeongwoo Lee et al., 2024). They work by splitting data into branches based on features, creating a tree-like structure that leads to clear predictions (O F Althuwaynee et al., 2020). It splits the data into smaller subsets until a final classification is reached. It is simple, interpretable, and handles non-linear data well. Python's scikit-learn library makes implementation easy, with options to customize tree depth and splitting criteria. They are ideal for monitoring air quality because of their interpretability and simplicity of usage.

A decision tree can be represented mathematically using recursive splitting and conditions. The general mathematical expression for a decision tree involves piecewise functions that divide the feature space into regions based on decision rules. Let $x = (x_1, x_2, \dots, x_p)$ be the input vector of features, and Tbe a decision tree with N leaf nodes. The decision tree partitions the feature space R^P into N disjoint regions R_1, R_2, … ,R_N. The following is an expression for the decision tree's output:

$$T(x) = \sum_{i=1}^{N} c_i \cdot 1(x \in R_i)$$

- c_i: The predicted value (class label or regression output) for region R_i .
- $1(x \in R_i)$: An indicator function that equals 1 if x belongs to region R_i and 0 otherwise.

A decision tree classifies data by selecting features (e.g., pollutant concentration or temperature) and splitting it based on thresholds, using measures like the Gini Index and Entropy to determine the best split. The data is divided step-by-step into smaller groups to improve accuracy, stopping when a group contains only one pollutant, the tree reaches its maximum depth, or further splitting is not possible (Yihang Chen, Shuoyu Chen et al., 2024). New data is classified by following the tree's splits from the root to a leaf node, which provides the predicted pollutant. This method ensures precise and interpretable results by partitioning the feature space into regions, each assigned a predicted class label.

Random Forest

A "forest" of several decision trees is created, which lowers overfitting and improves accuracy. Each tree is trained on a random subset of data, using different features at each step, ensuring diversity among the trees. In this ensemble method, each tree provides a prediction, and the final output is determined by majority vote—the class most frequently predicted by the trees for improved accuracy and stability (Jake Rhodes et al., 2024). Random Forest reduces overfitting and often outperforms single decision trees. Scikit-learn'sRandomForestClassifier and RandomForestRegressor tools provide adjustable parameters like the number and depth of trees. Mathematically:

- N: The forest's decision tree count.
- $h_i(x)$: The prediction of the i-th decision tree for an input x.
- $1\{\cdot\}$: Indicator function, which is 1 if the condition is true and 0 otherwise.
- $\hat{y}$: Final prediction, determined by the class **c** with the highest votes:

$$\hat{y} = argmax_c \sum_{i=1}^{N} 1\{h_i(x) = c\}$$

This collective decision-making process improves reliability and precision compared to a single decision tree. Additionally, Random Forests can rank feature importance, offering insights into which factors, like pollutant concentration or sensor response time, most influence predictions. In air quality monitoring, this approach enhances classification accuracy while uncovering key drivers of sensor performance. It improves classification accuracy by aggregating predictions from all trees, reducing the risk of overfitting.The Random Forest method offers a powerful alternative to single decision trees for classification tasks.

Data Collection

The first step in air quality monitoring involves gathering data from ZnO-based gas sensors, which are highly sensitive to pollutants such as nitrogen dioxide (NO_2), carbon dioxide (CO_2), ozone (O_3), and ethanol (C_2H_5OH). These sensors operate by detecting variations in gas concentration levels and translating them into measurable signals.In addition to gas detection, crucial environmental parameters such as temperature, sensor response time, and pollutant concentration are recorded. These factors play a significant role in determining the accuracy and reliability of pollutant classification, as changes in environmental conditions can affect sensor behavior.

The success of any machine learning model depends heavily on the quality and relevance of the data it uses. In this study, data was gathered from previous research on air quality monitoring systems that used zinc oxide (ZnO) as the primary sensor material. The dataset included important factors like pollutant concentrations (ppm), temperature (°C), and sensor response times (seconds). Data for nitrogen dioxide (NO_2) was obtained from studies conducted by Prasad R. Godse et al. (2024), Nguyen Minh Hieu et al. (2024), Youngho Mun et al. (2013), and Q. Zhang et al. (2018). Carbon dioxide (CO_2) data was drawn from research by Vaibhava Kumar et al. (2023) and Padmanathan Karthick Kannan et al. (2023). For ozone (O_3), references include studies by Tayssir Laribi et al. (2024) and Yina J. Onofre et al. (2019), while ethanol (C_2H_5OH) data came from research by Yi-Hsing Liu et al. (2024) and Sikai Zhao et al. (2021).

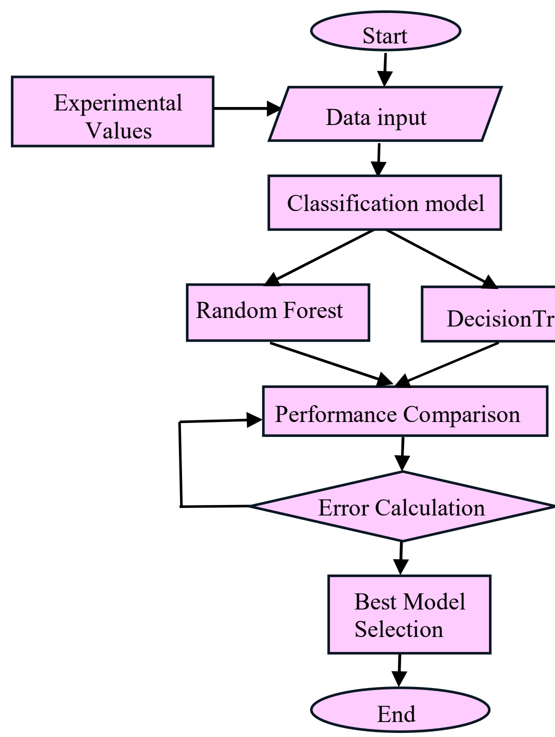

Figure 1 Flowchart representing the entire process

This flowchart (figure 1) provides a clear overview of how ZnO-based sensors and machine learning models work together to monitor air quality. It starts with collecting real-time pollution data from sensors, followed by preprocessing to clean and normalize the information. The data is then fed into decision tree and random forest models, which analyze key factors like gas concentration and temperature to classify pollutants. Once trained, the models are deployed for real-time monitoring, helping detect pollution levels and enabling timely action to protect air quality.

II. Results and Discussion

Data Preprocessing

To improve the accuracy of pollutant classification, the data was carefully processed before training. This involved filtering out any inconsistencies or outliers caused by sensor fluctuations or external disturbances, ensuring the data remained clean and reliable. Next, the recorded values were standardized to maintain a balanced scale, preventing any one factor from having too much influence on the model's predictions. Finally, only the most important features—gas concentration, temperature, and response time—were selected for analysis. By refining the data in this way, the model became more efficient, reducing unnecessary complexity while enhancing overall performance.

Model Selection & Training

To classify pollutants effectively, machine learning models must be carefully chosen and trained. Decision Tree and Random Forest algorithms are selected due to their ability to handle complex, non-linear relationships within the dataset.

The models are trained using 23 collected data points, allowing them to learn patterns in gas detection and environmental factors. The training phase is crucial for enabling the algorithms to make accurate predictions on new data.

Decision Tree and Random Forest Algorithms for Pollutant Classification

In this study, the Decision Tree model was employed as a classification tool to identify key air pollutantsincluding nitrogen dioxide (NO_2), carbon dioxide (CO_2), ozone (O_3), and ethanol (C_2H_5OH), based on sensor data. The algorithm identifies thresholds and patterns within the data, enabling accurate pollutant classification.Using a structured, rule-based approach, the model analyzed input parameters such as gas concentration, temperature, and response time to determine which pollutant was present. The Decision Tree was trained on 23 data samples, learning patterns that allowed it to make predictions on new data. With an accuracy of 74%, the model effectively classified pollutants, though it showed some limitations in handling complex gas mixtures. Despite this, its simplicity and interpretability made it a valuable tool for real-time applications, where quick and transparent decision-making is crucial. While it performed well, further refinements—such as optimizing the tree depth or incorporating additional features—could enhance its accuracy, making it even more effective for air quality monitoring. Figure 2 visualizes the decision tree, showcasing how the input parameters guide the classification process step by step.

Then the Random Forest model was employed in this study to improve pollutant classification by leveraging the power of multiple decision trees. Instead of relying on a single tree, Random Forest built an ensemble of trees, each trained on different subsets of the data. This approach reduced the risk of overfitting and improved the model's ability to handle complex gas interactions. Trained on the same 23 data samples, the Random Forest model achieved an accuracy of 84%, outperforming the Decision Tree model. When tested on eight new samples, it correctly identified seven pollutants, demonstrating its robustness and reliability. By averaging the predictions from multiple trees, the model produced more stable and accurate classifications, making it a stronger choice for real-world air quality monitoring. Its ability to generalize well across different conditions makes Random Forest a practical and efficient tool for continuous air pollution tracking, ensuring timely and accurate pollutant detection.

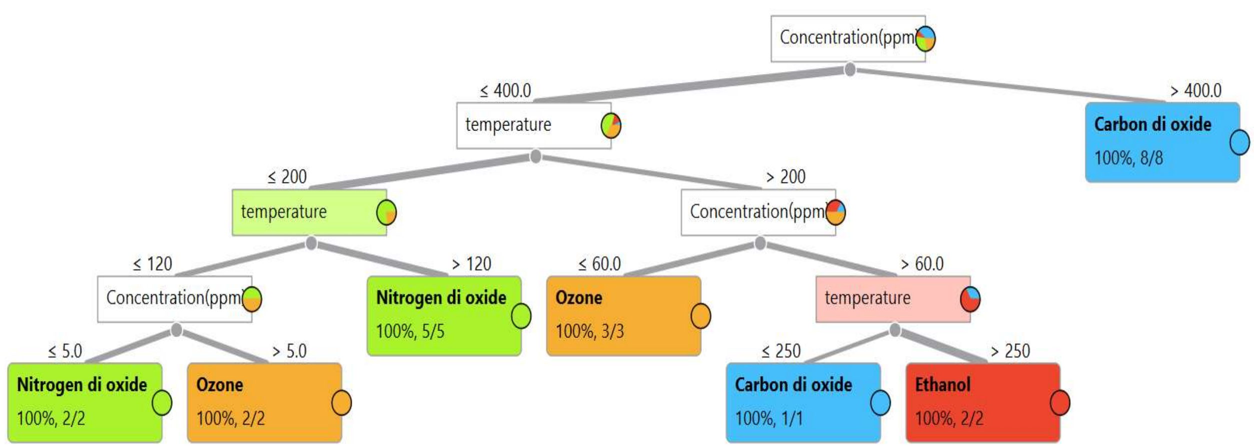

Figure 2: Decision Tree

Evaluating & Validating the Model

Once the machine learning models were trained, the next step was to see how well they performed in real-world conditions. It wasn't just about memorizing the training data but proving that the models could accurately identify pollutants when faced with new, unseen samples. To further validate the models, they were tested on, eight fresh air samples were , each containing key details like gas concentration, temperature, and sensor response time—important factors in classifying air pollutants.

The results showed a clear difference in performance between the models. The Decision Tree model correctly identified six out of eight pollutants, achieving an accuracy of 74%. Meanwhile, the Random Forest model performed even better, classifying seven out of eight pollutants correctly with an improved accuracy of 84%. This reinforced the strength of Random Forest, which relies on multiple decision trees working together to reduce errors and improve classification accuracy, making it a more dependable choice for complex pollutant classification.

	Random Forest	Tree	temperature	oncentration(ppm	response time (s)
1	Nitrogen di oxide	Nitrogen di oxide	27	5.0	0.62
2	Carbon di oxide	Carbon di oxide	300	2000.0	10.00
3	Nitrogen di oxide	Nitrogen di oxide	27	0.5	1.55
4	Ozone	Nitrogen di oxide	27	20.0	5.28
5	Ethanol	Ethanol	300	100.0	28.00
6	Carbon di oxide	Carbon di oxide	27	1000.0	19.50
7	Nitrogen di oxide	Carbon di oxide	120	120.0	160.00
8	Carbon di oxide	Carbon di oxide	300	10000.0	20.00

Figure 3: Predicted by decision tree and random forest algorithm respectively

Evaluation results for target (None, show average over classes)

Model	AUC	CA	F1	Prec	Recall	MCC
Random Forest	0.928	0.840	0.842	0.847	0.840	0.752
Tree	0.841	0.740	0.743	0.753	0.740	0.599

Figure 4: Classification Accuracy (CA) of decision tree and random forest algorithm respectively

To ensure these models were genuinely reliable, additional checks were carried out. A confusion matrix helped analyze where the models made correct or incorrect predictions, while cross-validation techniques confirmed that their performance remained consistent across different datasets. These evaluations highlighted that the models—particularly Random Forest—were not only accurate but also stable and reliable, making them well-suited for real-time air quality monitoring.While these models provide promising results, certain challenges remain. The dataset used in this study was relatively small, which may limit the models' ability to generalize across different environmental conditions. Additionally, pollutants often exist in complex mixtures, which can make classification more challenging, as interactions between gases can influence sensor responses. Other environmental factors, such as temperature, humidity, and sensor degradation over time, may also impact accuracy, requiring periodic recalibration and optimization.

Real-Time Implementation & Monitoring

Once trained, the models can be integrated into a real-time air quality monitoring system using ZnO-based sensors. These sensors can be deployed in different locations, continuously collecting pollution data. The collected data can then be processed instantly using either a local computer or a cloud-based system. The machine learning model will analyze the data and classify pollutants, providing real-time insights into air quality. The results can be displayed on a live dashboard, and alerts can be triggered if pollutant levels exceeded safe limits, enabling timely action.

Real-time monitoring is essential for minimizing exposure to harmful pollutants. For example, if NO_2 levels peak during rush hour traffic, authorities can implement traffic control measures to reduce emissions. A sudden rise in CO_2 levels may indicate poor ventilation, prompting necessary adjustments in indoor spaces. Similarly, detecting high ozone levels can signal the formation of photochemical smog, allowing for preventive measures to be taken. Beyond immediate alerts, continuous monitoring also helps track pollution trends over time, aiding in policy-making and urban planning.

Despite its benefits, real-time monitoring presents certain challenges. Sensors may require regular calibration to maintain accuracy, as factors like temperature and humidity can affect their performance. Additionally, processing large volumes of real-time data must be efficient to ensure timely responses. Overcoming these challenges is crucial for maintaining a reliable and scalable air quality monitoring system.

Future Scope

This study demonstrates that machine learning, combined with ZnO-based sensors, is a powerful tool for air quality monitoring Expanding the dataset by collecting more data from various locations and climates would help improve model accuracy and generalizability. Additionally, exploring more advanced machine learning techniques, such as deep learning models like Long Short-Term Memory (LSTM) networks, could enable better predictions of pollution trends over time.

Enhancing sensor sensitivity is another key area for future research. By incorporating hybrid materials, such as ZnO combined with SnO_2, sensors could become even more responsive to pollutants, improving detection capabilities. Furthermore, integrating this technology into mobile applications could allow individuals to check real-time air quality levels directly from their smartphones, increasing public awareness and engagement in pollution control efforts.

On a broader scale, linking air quality data with smart city infrastructure could lead to more proactive environmental management. By connecting real-time monitoring systems with urban planning and public health policies, cities can make informed decisions to reduce pollution exposure and create healthier environments. Implementing adaptive learning techniques in the models would also allow them to adjust dynamically to changing environmental conditions, ensuring long-term reliability.

TABLE 1
Experimental Vs. Predicted

Sl. No.	Ref.	Material	Actual \gas	prediction made by	
				Random forest	Decision Tree
1	[26]	Zinc oxixe and Gold based sensor	Nitrogen di oxide	Nitrogen di oxide	Nitrogen di oxide
2	[15]	Zinc Oxide thinfilm based sensor	Carbon di oxide	Carbon di oxide	Carbon di oxide
3	[17]	Zinc Oxide and Silver based sensor	Nitrogen di oxide	Nitrogen di oxide	Nitrogen di oxide
4	[25]	Zinc oxide and Tin oxide based sensor	Ozone	Ozone	Nitrogen di oxide
5	[20]	Zinc Oxide and Copper Oxide Nanowire based sensor	Ethanol	Ethanol	Ethanol
6	[22]	Zinc Oxide and Nickel based sensor	Carbon di oxide	Carbon di oxide	Carbon di oxide
7	[21]	Zinc Sulfide and Zinc Oxide with Gallium doping based sensor	Ozone	Nitrogen di oxide	Carbon di oxide
8	[15]	Zinc Oxide thinfilm based sensor	Carbon di oxide	Carbon di oxide	Carbon di oxide

III. CONCLUSION

This study highlights the potential of combining ZnO-based sensors with machine learning for effective air quality monitoring. By employing decision tree and random forest algorithms, the models successfully classified key pollutants—nitrogen dioxide (NO_2), carbon dioxide (CO_2), ozone (O_3), and ethanol (C_2H_5OH)—with accuracy rates of 74% and 84%, respectively. The results demonstrate the reliability of these models in real-world applications, offering a cost-effective and scalable alternative to traditional monitoring methods.

In addition to pollutant detection, this research underscores the significance of ZnO-based sensor fabrication. The development of these sensors plays a crucial role in enhancing detection sensitivity and ensuring accurate pollutant classification. Further advancements in sensor materials, design, and calibration could improve long-term performance and adaptability across different environmental conditions.While the findings are promising, expanding the dataset, incorporating advanced machine learning models, and refining sensor fabrication techniques could further enhance accuracy and scalability. By addressing these areas, future research can contribute to the development of more sophisticated, real-time air quality monitoring systems, ultimately supporting environmental sustainability and public health initiatives.

Acknowledgment

We sincerely appreciate the contributions of various researchers whose work provided the essential data for this study. We extend our gratitude to Prasad R. Godse et al. (2024), Nguyen Minh Hieu et al. (2024), Youngho Mun et al. (2013), and Q. Zhang et al. (2018) for their valuable research on nitrogen dioxide (NO_2). We also acknowledge Vaibhava Kumar et al. (2023) and Padmanathan Karthick Kannan et al. (2023) for their studies on carbon dioxide (CO_2), which greatly supported our analysis. Our appreciation goes to Tayssir Laribi et al. (2024) and Yina J. Onofre et al. (2019) for their insightful work on ozone (O_3), as well as Yi-Hsing Liu et al. (2024) and Sikai Zhao et al. (2021) for their research on ethanol (C_2H_5OH). Their contributions have been instrumental in developing and validating our machine learning models for air quality monitoring.

IV. References

A`li R. Jalalvand , Mohammad Mehdi Karami, Roles of nanotechnology in electrochemical sensors for medical diagnostic purposes: A review(January,2025). https://doi.org/10.1016/j.sbsr.2024.100733.

Andrea Carolina Marcillo Lara, Juan C. Baca Cabrera, Anja Widdig, Claudia Birkemeyer, A comparison between mobile and stationary gas chromatography–mass spectrometry devices for analysis of complex volatile profiles (November 2022). DOI:10.1007/s00216-022-04391-y

Anna Glinscaya, Ilia Panfilov, Anatoliy Kukartsev, Elena Suprun, Andrey Boyko, Use of decision trees for water quality assessment: Analysis of key parameters (October 2024). DOI:10.1051/bioconf/202413003002

Bharat Pant, Ahalapitiya Jayatissa, Gas Sensor Application of Zinc Oxide (November 2020). DOI:10.1201/9780429023187-13. pp.228-240

Claudia Banciu, Adrian Florea, Razvan Bogdan, Monitoring and Predicting Air Quality with IoT Devices (September 2024). DOI:10.3390/pr12091961

Gary M. Lovett, Timothy Tear, David C Evers, Stuart E.G. Findlay, Bernard Jackson Cosby Jr , Judy K. Dunscomb, Charles T. Driscoll, Kathleen C Weathers, Effects of Air Pollution on Ecosystems and Biological Diversity in the Eastern United States(April 2009) DOI:10.1111/j.1749-6632.2009.04153.x pp.99-135.

Guillaume Jean, Diana Ailyn, Climate Change and Air Quality: Interrelationship between climate change and air pollution levels(August 2024)

Jake Rhodes, Adam G. Rustad, Random Forest-Supervised Manifold Alignment(November 2024). DOI:10.48550/arXiv.2411.15179

Jeongwoo Lee, Caryl Anne M. Barquilla ,Kitae Park, Andy Hong, Urban form and seasonal PM2.5 dynamics: Enhancing air quality prediction using interpretable machine learning and IoT sensor data(November 2024). DOI:10.1016/j.scs.2024.105976.

Laxmi Kant Bhardwaj, Vinay Vikram Singh, Air Pollution And Its Effect On Human Health(April 2024) DOI:10.58532/V3BCAG4P1CH1 pp.1-6.

Nelson Barros, Tania Fontes, Impact of Kitchen Natural Gas Use on Indoor NO2Levels and Human Health: A Case Study in two European Cities (September 2024) DOI:10.3390/app14188461

Nguyen Minh Hieu, Anh D. Phan, Phuoc Cao Van, Viet Anh Cao, Nguyen Manh Hung, Nguyen Duc Chinh, Sutripto Majumder, Truong Hong Cuong, Hoang Gia Chuc, Do Van Minh, Do Quang Trung, Tu Nguyen, NguyenVan Du, Manh Trung Tran, Chunjoong Kim, Dojin Kim, Pham Thanh Huy, Jong-Ryul Jeong, The kinetics of a light irradiation enhanced room temperature NO2 gas sensor using hybrid ZnO/ ZnTe nanorod structures(December 2024). DOI:10.1039/d4ra06969b. pp. 39418-39428.

Nilesh Kumar, Ruhina Prasad, POLLUTION (May 2024)(Poster)

O F Althuwaynee, A L Balogun , A Aydda , T Gumbo, Classification of air pollutants API Inter-Correlation using decision tree algorithms(February 2020). DOI:10.1088/1755-1315/419/1/012022

Padmanathan Karthick Kannan, Saraswathi Ramiah, John Bosco Balaguru Rayappan, CO2 gas sensing properties of DC reactive magnetron sputtered ZnO thin film(May 2014) DOI:10.1016/j.ceramint.2014.05.011.pp.13115–13122.

Prasad R. Godse, Sujit A. Kadam, Yuvraj B. Jadhao, Yuan-Ron Ma, Tanaji M. Nimbalkar, Yogesh M. Jadhav, Vikas B. Patil, Ultra-responsive and highly sensitive 1D ZnO nanotubes for detecting perilous low levels of NO2 gas (February 2024). DOI:10.1039/D3MA00962A. pp. 2826–2840

Q. Zhang, G. Xie, M. Xu, Y. Su, H. Tai, H. Du and Y. Jiang, Sens. Actuators, B, 2018, 259, 269–281.

Sachin Dhawan , Anand Kumar , M. Vijaya Kumar , Mukesh Khare , S.M. Shiva Nagendra , Dalip Singh Mehta , Field calibration protocol for optical-based air quality sensors in Indian conditions (March 2024). https://doi.org/10.1016/j.matpr.2024.03.020.

Shaivalini Singh, Al doped ZnO based metal–semiconductor–metal and metal–insulator–semiconductor–insulator–metal UV sensors (April, 2016). https://doi.org/10.1016/j.ijleo.2016.01.012.

Sikai Zhao, Yanbai Shen, Fulai Hao, Changke Kang, Baoyu Cui, Dezhou Wei, Fanli Meng, P-n junctions based on CuO-decorated ZnO nanowires for ethanol sensing application(February 2021). DOI:10.1016/j.apsusc.2020.148140

Tayssir Laribi, Riadh Souissi, Sandrine Bernardini, M. Marc Bendahan, Bouguila Noureddine, S. Alaya, Highly responsive and selective ozone sensor based on Ga doped ZnS–ZnO composite sprayed films(January 2024). DOI:10.1039/d3ra06959a. pp.413-423

Vaibhava Kumar, Ajeet Singh, Bal Chandra Yadav, Hemant Kumar Singh, Deep Prakash Singh, Sandip Kumar Singh, Navin Chaurasiya, Environment-sensitive and fast room temperature CO2 gas sensor based on ZnO, NiO and Ni-ZnO nanocomposite materials (December 2023). DOI:10.1016/j.efmat.2023.12.002.

Yihang Chen, Shuoyu Chen, Yicheng Yang, Siming Lu, Comparison of decision tree and ensemble algorithms(July 2024). DOI:10.54254/2755-2721/55/20241535. pp. 241-248.

Yi-Hsing Liu, Sheng-JoueYoungb, Liang-Wen Ji, Yen-Lin Chub, Shoou-Jinn Chang, High-sensitive ethanol gas sensor using Ag modifedZnO nanosheets (November 2024). DOI:10.1016/j.talo.2024.100386

Yina J. Onofre, Ariadne C. Catto, Sandrine Bernardini, Tomas Fiorido, Khalifa Aguir, Elson Longo, Valmor R. Mastelaro, Luís F. da Silva, Marcio P.F. de Godoy, Highly selective ozone gas sensor based on nanocrystalline Zn0.95Co0.05O thin film obtained via spray pyrolysis technique(June 2019). DOI:10.1016/j.apsusc.2019.01.197

Youngho Mun, Sunghoon Park, Soyeon An, Chongmu Lee, Hyoun Woo Kim, NO2 gas sensing properties of Au-functionalized porous ZnO nanosheets enhanced by UV irradiation(December, 2013). DOI:10.1016/j.ceramint.2013.04.035 .pp. 8615–8622.

ANNEXURE I
Data Collected

Pollutant Type	Temp.	Conc.	Response Time
Nitrogen di oxide	170	50	36
Nitrogen di oxide	200	200	41
Nitrogen di oxide	200	5	13
Nitrogen di oxide	200	1	1
Nitrogen di oxide	175	10	2.83
Nitrogen di oxide	27	5	18
Ozone	27	20	5.28
Ozone	250	60	9
Ozone	250	20	46
Ozone	260	30	160
Carbon di oxide	27	500	14.32
Carbon di oxide	27	1500	23.12
Carbon di oxide	27	2000	27.25
Carbon di oxide	400	5000	90
Carbon di oxide	450	10000	10
Carbon di oxide	250	200	9
Carbon di oxide	250	1025	17
Carbon di oxide	300	10000	20
Carbon di oxide	300	5000	120
Ethanol	350	400	20.3
Ethanol	300	100	28

ANNEXURE II

Algorithm for Pollutant Classification Using Decision Tree and Random Forest

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier,
RandomForestRegressor
from sklearn.metrics import accuracy_score, classification_report,
mean_squared_error

# Load dataset (Replace with actual dataset path)
data = pd.read_csv("ZnO based sensors CSV.csv ")

# Define feature columns and target variable
features = ["Concentration", "Temperature", "Response_Time"]
target = "Pollutant_Type"

# Splitting data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(data[features],
data[target], test_size=0.2, random_state=42)

# Standardizing the feature values
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

# Decision Tree Model
dt_model = DecisionTreeClassifier()
dt_model.fit(X_train, y_train)
y_pred_dt = dt_model.predict(X_test)

# Random Forest Classifier Model
rf_classifier = RandomForestClassifier(n_estimators=100,
random_state=42)
rf_classifier.fit(X_train, y_train)
y_pred_rf = rf_classifier.predict(X_test)

# Evaluate classification models
print("Decision Tree Accuracy:", accuracy_score(y_test,
y_pred_dt))
print("Random Forest Accuracy:", accuracy_score(y_test,
y_pred_rf))
print("Classification Report for Decision Tree:\n",
classification_report(y_test, y_pred_dt))
print("Classification Report for Random Forest:\n",
classification_report(y_test, y_pred_rf))

# Random Forest Regressor for Predicting Pollutant Concentration
if "Concentration" in data.columns:
X_reg = data.drop(columns=["Concentration"])
y_reg = data["Concentration"]

X_train_reg, X_test_reg, y_train_reg, y_test_reg =
train_test_split(X_reg, y_reg, test_size=0.2, random_state=42)

rf_regressor = RandomForestRegressor(n_estimators=100,
random_state=42)
rf_regressor.fit(X_train_reg, y_train_reg)
y_pred_reg = rf_regressor.predict(X_test_reg)

print("Random Forest Regressor Mean Squared Error:",
mean_squared_error(y_test_reg, y_pred_reg))
```