



# **iJRASET**

International Journal For Research in  
Applied Science and Engineering Technology



---

# **INTERNATIONAL JOURNAL FOR RESEARCH**

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume: 11    Issue: XI    Month of publication: November 2023**

**DOI: <https://doi.org/10.22214/ijraset.2023.56770>**

**[www.ijraset.com](http://www.ijraset.com)**

**Call:  08813907089**

**E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)**



# Survey on Air Pollution in India: Insights, Predictions, and Mixed Gas Analysis for an Improved Future

Somasekhar T<sup>1</sup>, Gagana R<sup>2</sup>, Brunda B<sup>3</sup>, Divya V Chikkamath<sup>4</sup>, Gagana C L Naik<sup>5</sup>

<sup>1</sup>Associate Professor, Computer Science and Engineering, K.S. Institute of Technology, Bengaluru, India

<sup>2, 3, 4, 5</sup> Student, Computer Science and Engineering, K.S. Institute of Technology, Bengaluru, India

**Abstract:** *The sustenance of life on Earth is intricately tied to the quality of our air. But this precious resource is facing mounting threats from the harmful consequences of rapid industrialization, transportation networks, and the everyday practices of modern living. This research paper addresses the pressing issue of rising air pollution in India, with a strong emphasis on environmental well-being. Utilizing data from the Central Pollution Control Board of India (CPCB), our project focuses on air quality visualization, pollution prediction, and mixed gas analysis. We have developed an intuitive web-based platform for visualizing air quality data, offering a valuable tool for the public and policymakers. Additionally, predictive models are employed to anticipate pollution levels, facilitating timely intervention and mitigation. Our mixed gas analysis sheds light on the composition of atmospheric gases, enhancing our understanding of pollution sources. Through this interdisciplinary approach, our research aims to provide a comprehensive solution to combat air pollution, fostering environmental sustainability and public health improvement in India.*

**Keywords:** Air quality, Mixed gas analysis, Machine Learning models, Visualization, Predictions

## I. INTRODUCTION

Air pollution is a pervasive and escalating issue in the modern age, arising from various anthropogenic sources, including industrial operations, transportation, and the combustion of fossil fuels. This widespread contamination of the atmosphere results in the emission of hazardous pollutants, encompassing carbon monoxide (CO), carbon dioxide (CO<sub>2</sub>), particulate matter (PM), nitrogen dioxide (NO<sub>2</sub>), sulphur dioxide (SO<sub>2</sub>), ozone (O<sub>3</sub>), and more. These pollutants not only pose a grave threat to human health but also cast a shadow on the ecological balance, impacting both animals and plants. The far-reaching implications of air pollution range from respiratory diseases like bronchitis and lung cancer to broader environmental concerns such as global warming, acid rain, and climate change. Furthermore, it exacts a heavy economic toll on societies and leads to an array of contemporary environmental issues, including reduced visibility, smog, aerosol formation, and premature deaths. Scientists and researchers have recognized that air pollution extends its adverse effects even to historical monuments. Emissions from vehicles, industrial facilities, agricultural practices, and power plants contribute to the emission of greenhouse gases, which, in turn, exacerbates climate conditions and disrupts plant-soil interactions. These climatic fluctuations have repercussions not only on human and animal populations but also on agricultural productivity, leading to significant economic losses.

Furthermore, contemporary environmental research has expanded its focus to encompass mixed gas analysis, recognizing that the complex cocktail of gases present in the atmosphere has multifaceted effects on both human health and the environment. These mixed gases include volatile organic compounds (VOCs), heavy metals, and other chemical constituents. The monitoring and analysis of mixed gases are essential for gaining a comprehensive understanding of air quality and its impact on public health.

The Air Quality Index (AQI), a critical parameter for assessing air quality, is closely linked to public health. A higher AQI level signifies greater exposure to dangerous pollutants, underscoring the urgency of predicting and monitoring air quality. This is particularly crucial in urban areas undergoing rapid industrial and motorized development. While many air quality studies and research initiatives are directed towards developing countries, where concentrations of deadly pollutants like PM<sub>2.5</sub> are disproportionately high, limited attention has been devoted to air quality prediction in Indian cities. Therefore, there is a compelling need to bridge this gap by analysing and predicting AQI for India.

In response to this pressing issue, our research project takes a comprehensive approach to address the multifaceted problem of air quality.



Utilizing data from the Central Pollution Control Board of India (CPCB), our study focuses on three critical aspects: air quality visualization, pollution prediction, and in-depth analysis of mixed gases in the atmosphere. Central to our project is the development of a user-friendly web-based platform for visualizing air quality data, empowering the public and policymakers to monitor air pollution trends and make informed decisions. Notably, our project integrates advanced machine learning models for precise pollution level prediction. These models, informed by historical data, provide a potent tool for timely intervention and data-driven policy formulation.

Our endeavour extends to a meticulous examination of the composition of mixed gases in the atmosphere, shedding light on their origins and contributions to air pollution. By employing machine learning models, we utilize data analysis and predictive modelling to provide stakeholders with vital information for addressing and mitigating the adverse effects of air pollution. This comprehensive approach seeks to bridge the gap between environmental data and actionable solutions, significantly contributing to air pollution mitigation and the enhancement of environmental health in India.

## II. LITERATURE SURVEY

The study conducted in paper [1] comprehensively analyses air pollution data from 23 Indian cities over a six-year period. The dataset undergoes thorough cleaning and preprocessing, encompassing the handling of NAN values, treatment of outliers, and the normalization of data values. A correlation-based feature selection technique is employed to filter pollutants affecting AQI, and skewed features are subjected to logarithmic transformations. Exploratory data analysis methods uncover hidden patterns within the dataset, notably observing a significant decline in pollutants during 2020. Data imbalance is effectively mitigated through SMOTE analysis, leading to the partitioning of the dataset into 75% training and 25% testing subsets. Machine learning-based AQI prediction is conducted with and without SMOTE resampling, with the XGBoost model achieving the highest accuracy for both training and testing sets. Classical statistical error metrics assess and compare model performance. This research contributes significantly to air quality analysis and prediction for India, with potential extensions involving the incorporation of deep learning techniques for AQI prediction.

The study conducted in paper [2] found that the anticipation of the air quality index is achieved by employing various algorithms, including linear regression, Decision Tree, and Random Forest. From the results obtained, it has been inferred that a superior prediction of the air quality index is provided by the Random Forest algorithm. This conclusion is based on the analysis of the data and the comparison of predictive performance, with Random Forest consistently outperforming the other methods in terms of accuracy and reliability.

The enhanced predictive capabilities of the Random Forest algorithm make it the preferred choice for forecasting air quality index, as it offers more robust and dependable results, thus contributing to a more effective and accurate assessment of air quality conditions.

According to a recent study by Reshma J in paper [3], Artificial Neural Networks (ANN) acts as a robust method for forecasting air pollutants. The study consistently demonstrates the precision of ANN in forecasting vital air pollutants, encompassing SO<sub>2</sub>, NO<sub>2</sub>, O<sub>3</sub>, CO, and particulate matter. Moreover, the model's incorporation of meteorological parameters, including temperature, relative humidity, absolute humidity, and wind direction, enhances its predictive capabilities of the model. An additional advantage of the ANN approach is its adaptability to web applications, offering end-users real-time access to air quality information. This empowers individuals to proactively address air pollution's adverse effects, ultimately safeguarding public health and the environment.

In summary, the study in paper [4] highlights the health risks of prolonged exposure to particulate matter and the rising global pollution levels, including in Russia.

Effective environmental measures and predictive monitoring are essential. The research demonstrates the successful use of machine learning methods. However, long-term forecasts are limited due to data quality constraints. More high-quality data is needed, especially in Russian cities, where it remains scarce. This underscores the need for improved air quality monitoring and public awareness efforts, which would expand data resources, enhance machine learning models, and enable more accurate air quality predictions. Increasing the number of air quality monitoring centers in Russia and providing free public access to this information is crucial.

Following the completion of the study in paper [5], the research team achieved success in crafting an Ensemble learning model with the ability to forecast forthcoming air pollution concentrations. This was accomplished by amalgamating basic learning models and employing a meta-learner to generate the ultimate predictions for air pollution concentrations. The researchers also conducted a comprehensive assessment of various machine learning algorithms' performance in forecasting future APC values.



The proposed air quality prediction algorithm, which combines Recursive Air Quality Predictor (RAQP) with ensemble learning, exhibited improved results for Nitrogen Dioxide (NO<sub>2</sub>) and Ozone (O<sub>3</sub>), with relatively similar outcomes for Fine Particulate Matter (PM<sub>2.5</sub>) and Carbon Monoxide (CO). Regarding Root Mean Square Error (RMSE), the proposed model outperformed across all APCs.

According to the study in paper [6], a novel technique was introduced for gas sensing and classification using machine learning (ML) algorithms.

This approach accurately classified combinations of three gases with up to 97% accuracy by analyzing autocorrelation Functions (ACFs) and their impedance characteristics. Despite the high dimensionality of the classifier, researchers successfully identified unique features related to gas molecules, such as polarity, molecular weight, and adsorption, allowing precise classification, especially in complex scenarios involving gas mixtures. This cost-effective and lightweight method demonstrates promise for gas mixture classification.

To enhance prediction accuracy for mixed gases, this paper [7] introduces a novel regression prediction model, TCN, known for its robust time series pattern recognition capabilities.

The model is enhanced in two key ways: firstly, by modifying its structure to incorporate additional channels within its residual network for improved information retrieval at each step, and secondly, by optimizing its activation function, which is determined through experimental validation. The results demonstrate the superiority of this approach over LSTM, GRU, and generic TCN in regression prediction tests for two gas mixtures. This method holds promise for effective gas mixture analysis when applied to E-nose systems. In future work, further improvements will be explored to enhance TCN's performance, potentially through synergistic combinations with other neural networks.

In this paper [8], the authors propose enhancements to machine olfactory systems. They introduce a dynamic time warping algorithm (DTW) that significantly improves classification accuracy by 26.87%. Through original feature construction and the PCA method, the classification accuracy rate increases by 25.8%.

Furthermore, the time efficiency of the random forest algorithm is boosted using the extreme random tree algorithm, resulting in a final classification accuracy of 99.28%. Notably, the runtime is substantially reduced to only 103.2568 s, marking a 66.85% decrease from the random forest algorithm.

This paper effectively addresses the classification challenges of mixed gases, enhances the random forest algorithm, and provides a theoretical foundation for simulating the olfactory nervous system.

### III. OBJECTIVES

- 1) *Visualization*: Create a web-based platform that provides visualization of air quality data, including pollutant concentrations and air quality indices for various locations in India.
- 2) *Mixed Gas Analysis*: Develop analytical techniques to assess the composition of mixed gases in the air, emphasizing the identification of pollutants and their sources, enabling more informed decision-making.
- 3) *Environmental Impact Assessment*: Conduct a comprehensive analysis of the environmental impact of air pollution, including its effects on public health, ecosystems, and climate, and present this information in an understandable format on the website.
- 4) *Pollution Prediction*: Implement predictive models that use historical data, meteorological factors, and pollutant trends to forecast air quality, allowing for early warnings and timely intervention.
- 5) *Precision Air Quality Predictions*: Enhance the precision of pollution forecasts by utilizing advanced data analytics and machine learning, ensuring location-specific accuracy.

### IV. METHODOLOGIES

The project delves into a meticulous analysis of the composition of mixed gases in the atmosphere, unravelling their sources and contributions to air pollution.

Through the integrated application of data analysis, machine learning, and web development, this study strives to bridge the gap between environmental data and actionable solutions, making a significant contribution to the mitigation of air pollution and the enhancement of environmental health in India.

The methodological steps adopted in the process are presented in the figure given below.

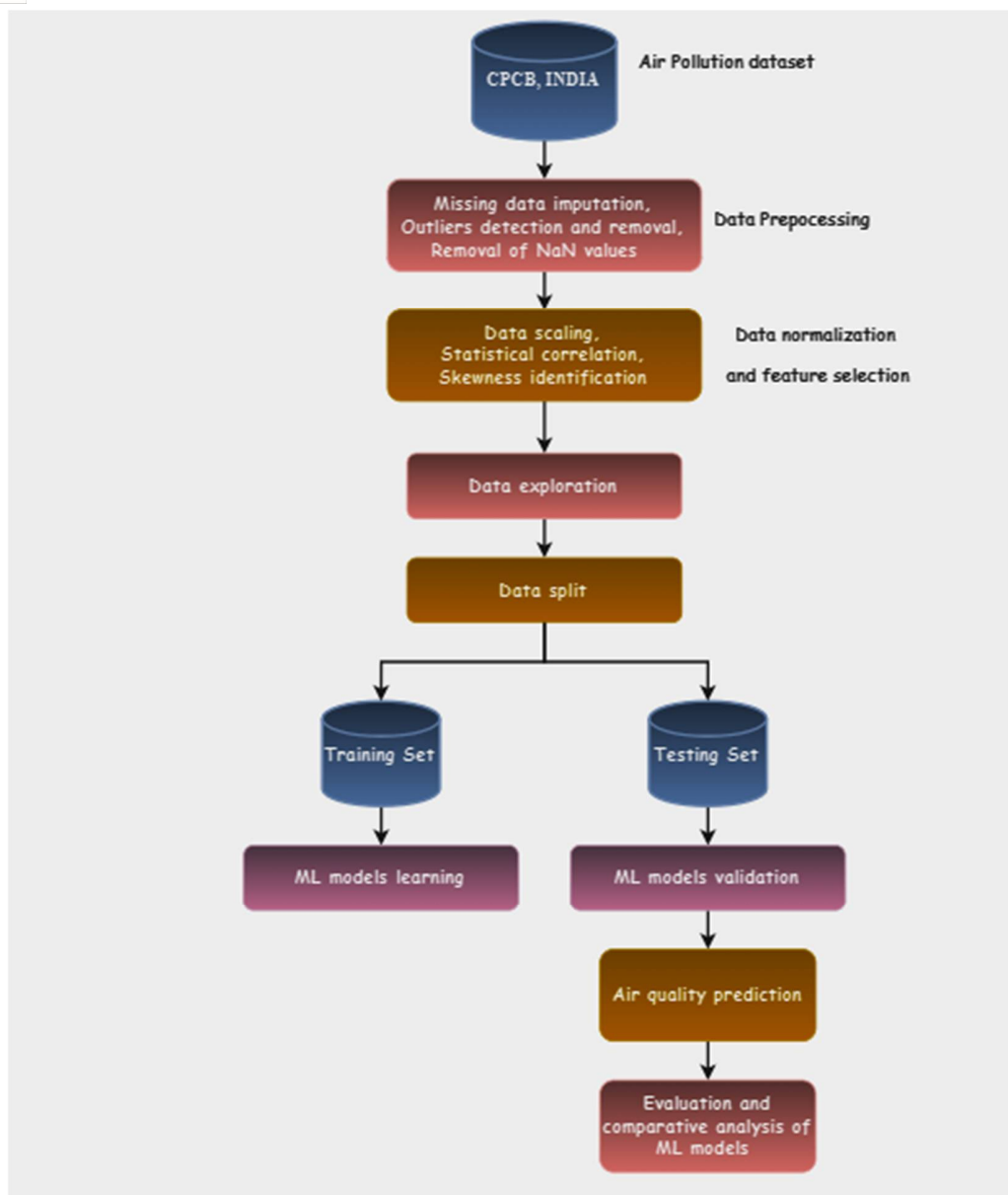


Fig 1: Flowchart

- 1) *Data Preprocessing*: In the data preprocessing phase, we clean and organize the raw air quality and mixed gas data from CPCB. This involves handling missing values, removing outliers, and ensuring data integrity. By preparing the data meticulously, we lay the foundation for accurate analysis and modeling.
- 2) *Data Normalization and Feature Extraction*: Data normalization is crucial for ensuring that all variables are on a consistent scale, enabling fair comparisons and effective modeling. We apply techniques such as Min-Max scaling to standardize the data. Additionally, we perform feature extraction to identify relevant variables that contribute most to our predictive models, reducing dimensionality and enhancing model performance.
- 3) *Exploratory Data Analysis (EDA)*: EDA is a vital step that involves visualizing and exploring the data to gain insights and identify patterns. We employ statistical and graphical techniques to understand the relationships between various features and the air quality parameters. EDA guides us in making informed decisions regarding data preprocessing and model selection.

- 4) **Predictive Modeling:** In this phase, we leverage advanced machine learning models to predict air quality and pollution levels. These models, including regression, neural networks, and ensemble methods, are trained on historical data and are capable of forecasting future air quality conditions. We evaluate and fine-tune these models to ensure their accuracy and reliability. Predictive modeling plays a crucial role in addressing the core objectives of our project, which include visualizing air quality, predicting pollution levels, and conducting mixed gas analysis.

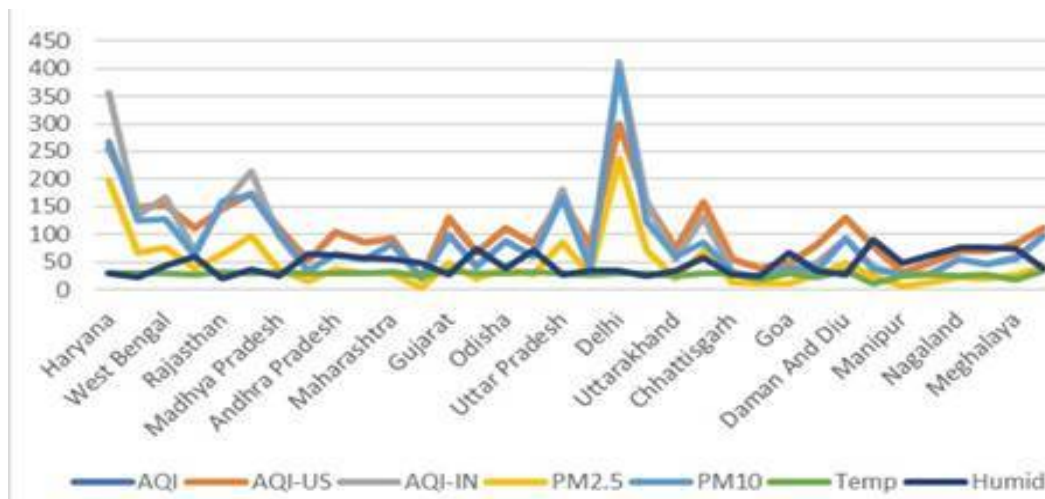


Fig 2: Analysis of Air parameters

By following this methodology, our research project aims to provide a robust framework for understanding and addressing air pollution in India, ultimately contributing to environmental sustainability and improved public health.

## V. REQUIREMENTS

### A. Programming Language

Python: Python is a versatile language commonly used in data analysis, machine learning, and web development.

Data Processing and Analysis:

Pandas: A Python library for data manipulation and analysis.

NumPy: It is used for numerical operations and array processing.

Jupyter Notebook: An interactive environment for data exploration and analysis.

### B. Data Visualization

Matplotlib: A popular library for creating static, animated, and interactive visualizations.

Seaborn: A data visualization library built on Matplotlib for creating attractive statistical graphics.

Plotly: Offers interactive and dynamic data visualizations.

Power BI: For more advanced, interactive data visualization, dashboards, and reporting.

### C. Machine Learning and Predictive Modeling

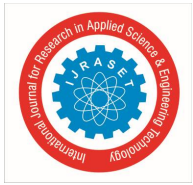
Scikit-learn: It is a popular machine learning library in Python that provides a wide range of tools and algorithms for tasks such as classification, regression, clustering, dimensionality reduction, and model evaluation.

XGBoost: High-performance gradient boosting frameworks.

### D. Web Development

HTML, CSS, and JavaScript: Fundamental web technologies for designing and creating web pages and interactivity.

XAMPP (Cross-platform, Apache, MySQL, PHP, Perl) is an open-source web server package that simplifies the setup of a local web development environment by combining Apache, MySQL, PHP, and Perl.



#### E. Database and Storage

MySQL: Relational databases for storing structured data.

JDBC (Java Database Connectivity) is a Java-based API for connecting and interacting with relational databases.

#### F. Deployment and Hosting

Netlify is a cloud-based web hosting and automation platform that allows developers to build, deploy, and manage modern web projects with features like continuous integration, serverless functions, and content delivery.

### VI. CONCLUSION

In conclusion, our survey paper address the escalating air pollution crisis in India, and offer a comprehensive solution. Utilizing data from the Central Pollution Control Board of India (CPCB), our user-friendly web platform provides air quality visualization, empowering the public and policymakers alike. Informed by a thorough literature survey, our mixed gas analysis, facilitated by advanced machine learning models, seeks to unravel the intricate composition of atmospheric gases, crucial for identifying pollution sources. Moving forward, our commitment includes refining predictive models, utilizing historical data to forecast air quality levels with precision. This approach not only strives to contribute to the enhancement of environmental sustainability and public health in India but also seeks to set a benchmark on a global scale for fostering cleaner, healthier futures.

### VII. ACKNOWLEDGMENT

We extend our heartfelt gratitude to Dr. Rekha B Venkatapur, Professor and HOD of CSE, and Mr. Somasekhar T, Associate Professor, Dept of CSE for their invaluable and insightful contributions throughout the planning and execution of this project. Their generous commitment of time and expertise have been sincerely appreciated. We would also like to acknowledge the unwavering support and encouragement provided by the esteemed faculty members at KSIT.

### REFERENCES

- [1] K. Kumar, B.P Pande, "Air pollution prediction with machine learning: a case study of Indian cities", Iranian Society of Environmentalists (IRSEN) and Science and Research Branch, Islamic Azad University.  
International Journal of Environmental Science and Technology (2023) 20:5333–5348  
Available at: <https://doi.org/10.1007/s13762-022-04241-5>
- [2] Ritik Sharma, Gaurav Shilimkar, Shivam Pisal, "Air Quality Prediction by Machine Learning", Vishwakarma Institute of Technology, Pune, Maharashtra, India.  
International Journal of Scientific Research in Science and Technology.  
<https://doi.org/10.32628/IJSRST218396>
- [3] Reshma J, "Analysis and Prediction of Air Quality", Assistant Professor, Department of Computer Science and Engineering, BNM Institute of Technology, Bengaluru, India.  
International Research Journal of Engineering and Technology (IRJET) Volume: 07 Issue: 01 | Jan 2020 e-ISSN: 2395-0056 p-ISSN: 2395-0072  
Available at: [www.irjet.net](http://www.irjet.net)
- [4] Ekaterina Gladkova, Liliya Saychenko, "Applying machine learning techniques in air quality prediction", Saint Petersburg Mining University, St Petersburg, Russia.  
X International Scientific Siberian Transport Forum.  
Available at: [www.sciencedirect.com](http://www.sciencedirect.com)
- [5] Dionis A. Padilla, Glenn V. Magwili, Luis Benjamin Z. Mercado, Jean Tristan L. Reyes, "Air Quality Prediction using Recurrent Air Quality Predictor with Ensemble Learning ", School of Electrical, Electronics, and Computer Engineering, Mapua University Philippines.  
<https://ieeexplore.ieee.org/Xplore/home.jsp>
- [6] Kookjin Lee(1), Sangjin Nam(2), Hyojun Kim(3)\*, Dae-Young Jeon(4), Dongha Shin(3), Hyeong-Gyun Lim(3), Chulmin Kim(1), Doyoon Kim(1), Yeonsu Kim(1), Sang-Hoon Byeon(3), and Gyu-Tae Kim(1), "Detection and Accurate Classification of Mixed Gases Using Machine Learning with Impedance Data",  
1. School of Electrical Engineering Korea University, Republic of Korea  
2. Dept of Computer Science and Engineering Korea University, Republic of Korea  
3. Dept of Environmental Health College of Health Sciences Korea University, Republic of Korea. \* Samsung Electronics Co. Ltd Samsung-ro Yongin-si, Republic of Korea  
4. Institute of Advanced Composite Materials Korea Institute of Science and Technology, Republic of Korea  
Available at: [www.advtheorysimul.com](http://www.advtheorysimul.com)



[7] Liwen Zeng (1), Yang Xu (2), Sen Ni (1) Min Xu (3), Pengfei Jia (1), "A mixed gas concentration regression prediction method for electronic nose based on two-channel TCN".

1. School of Electrical Engineering, Guangxi University, Nanning, China.

2. Guangxi Key Laboratory of Intelligent Control and Maintenance of Power Equipment, Guangxi University, Nanning, China.

3. College of Food Science and Bioengineering, Xihua University, Chengdu, China. <https://www.sciencedirect.com/journal/sensors-and-actuators-b-chemical>

[8] Yonghui Xu (1), Xi Zhao (1), Yinsheng Chen (2) and Zixuan Yang (1), "Research on a Mixed Gas Classification Algorithm Based on Extreme Random Tree",

1. School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin 150001, China.

2. School of Measurement and Control Technology and Communication Engineering, Harbin University of Science and Technology, Harbin 150001, China.

Available at: [www.mdpi.com/journal/applsci](http://www.mdpi.com/journal/applsci)





10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)