



INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Comprehensive Study On Secondary Organic Aerosols And Pm2.5 Source Profiling

¹Dr.Santosh Kumar Singh., ²Mr.Amit Kumar Pandey, ³Madhu Pandey, ⁴Kajal Mistry

¹HOD (Information Technology), ²Assistant Professor ^{1,2,3,4}

Department of IT ,Thakur College of Science and Commerce
Thakur Village, Kandivali (East), Mumbai-400101, Maharashtra, India |

Abstract

Fine particulate matter (PM_{2.5}) is a critical pollutant with far-reaching impacts on human health, climate, and environmental quality. This study focuses on two primary objectives: source identification and characterization, and health impact assessment. Utilizing advanced source apportionment techniques and chemical characterization methods, the research identifies and quantifies the contributions of key PM_{2.5} sources such as vehicular emissions, industrial activities, and natural events like wildfires and dust storms. Accurate source profiling is essential for formulating targeted mitigation strategies tailored to regional and local conditions. The findings provide a scientific basis for evidence-based policymaking, emphasizing cleaner transportation systems, stricter industrial emission controls, and sustainable land-use practices.

Keywords

PM_{2.5}, Source Profiling, SOAs, Chemical Characterization, Air Pollution, Public Health.

1. Introduction

Air pollution, particularly fine particulate matter (PM_{2.5}), is one of the most pressing environmental and public health challenges of the 21st century. PM_{2.5} refers to airborne particles with a diameter of less than 2.5 micrometers, small enough to penetrate deep into the respiratory system and bloodstream. Exposure to these particles is linked to a wide range of health issues, including respiratory diseases, cardiovascular disorders, and premature mortality, particularly in vulnerable groups such as children, the elderly, and individuals with pre-existing health conditions.

In addition to its health impacts, PM_{2.5} also contributes to climate change and environmental degradation, affecting ecosystems and biodiversity. The sources of PM_{2.5} pollution are diverse and multifaceted, encompassing anthropogenic activities such as vehicular emissions, industrial processes, fossil fuel combustion, and agricultural practices, as well as natural phenomena like wildfires and dust storms. This complexity underscores the need for precise source identification and characterization to understand the contributions of different sources to overall pollution levels. Such insights are essential for designing effective mitigation strategies that are both targeted and region-specific. This project focuses on two key objectives: source identification and health impact assessment. Using advanced techniques such as source apportionment and

chemical characterization, the research will identify and quantify the contributions of major PM_{2.5} sources. By profiling sources, including transportation systems, industrial operations, and natural contributors, the study aims to provide a clear understanding of their relative contributions to pollution levels. Simultaneously, the project evaluates the health impacts of PM_{2.5} exposure, focusing on vulnerable populations. By analyzing exposure levels and their associated health outcomes, the study seeks to quantify the societal and economic costs of air pollution, providing actionable insights into the impacts on public health. The findings of this study will form the basis for evidence-based policymaking, aimed at reducing PM_{2.5} concentrations and mitigating its adverse impacts. Proposed strategies include promoting cleaner transportation technologies, implementing stricter industrial emission controls, and adopting sustainable land-use practices. By integrating source profiling with health impact assessments, the project aspires to contribute to global efforts in improving air quality, protecting public health, and fostering environmental sustainability, addressing one of the most significant environmental and societal challenges of our time

focuses on identifying PM_{2.5} sources and assessing their health impacts to develop effective mitigation strategies. Using Google Colab, you implement three key algorithms: SETA (SOA Estimation and Trend Analysis), which analyzes trends in Secondary Organic Aerosols (SOAs) to understand their contribution to PM_{2.5}; PSPA (PM_{2.5} Source Profiling and Apportionment), which identifies and quantifies pollution sources such as vehicular emissions, industrial activities, and natural contributors; and ACAC (Aerosol Chemical Analysis and Classification), which characterizes aerosol chemical composition to determine their origin and environmental impact. By integrating these techniques, your study provides data-driven insights to support evidence-based policymaking, advocating for cleaner technologies, stricter emission controls, and sustainable land-use practices to improve air quality and public health.

2. SETA Technique

The SETA (SOA Estimation and Trend Analysis) Technique plays a crucial role in your research on Secondary Organic Aerosols (SOAs), Source Profiling of PM_{2.5}, and Chemical Characterization of Aerosols by providing a comprehensive approach to estimating SOA levels, identifying sources, and analyzing their trends. This technique integrates receptor modeling methods such as Positive Matrix Factorization (PMF) and Chemical Mass

Balance (CMB) to apportion SOA contributions

within PM_{2.5}, helping to differentiate between anthropogenic sources like vehicular and industrial emissions and biogenic sources from natural organic compounds. Additionally, SETA incorporates chemical marker-based analysis using compounds such as levoglucosan, hopanes, and polycyclic aromatic hydrocarbons (PAHs) to chemically characterize SOAs and their formation pathways. Advanced spectrometry techniques like Gas Chromatography-Mass Spectrometry (GC-MS) and High-Resolution Mass Spectrometry (HR-MS) enable precise identification of SOA components, improving the understanding of PM_{2.5} composition. Furthermore, SETA utilizes trend analysis techniques, including time-series modeling and machine learning, to study the seasonal and long-term variations of SOA concentrations, linking them to meteorological conditions and emission patterns. This allows for predictive assessments of SOA trends, contributing to effective air pollution mitigation strategies. In addition to source characterization and trend analysis, SETA assesses the health impact of SOAs by evaluating their oxidative potential, which is directly linked to respiratory and cardiovascular diseases. By integrating these methodologies, SETA not only enhances the understanding of SOA formation and transformation processes but also supports evidence-based policymaking aimed at reducing PM_{2.5} levels and improving air quality, aligning perfectly with the objectives of your research project.

3. PM_{2.5} Source Profiling and Apportionment (PSPA)

Application of PSPA Algorithm for PM_{2.5} Source Profiling Using the Dataset The provided dataset contains PM_{2.5}, PM₁₀, NO₂, SO₂, and location-specific air quality data, which are essential for source profiling and apportionment. The PSPA (PM_{2.5} Source Profiling and Apportionment) algorithm will analyze this data using statistical and receptor modeling techniques to determine the major contributors to air pollution across different cities and states.

Step 1: Data Preprocessing

The first step involves cleaning and normalizing the dataset. Missing values in the Year column will be handled through imputation techniques. Since pollutant concentrations vary across locations, normalization will be applied to standardize the data. Source Apportionment Using PSPA Algorithm. The PSPA algorithm uses multiple techniques for source identification Positive Matrix Factorization (PMF) PMF decomposes the observed pollution data into source contributions and profiles:

$$X = G \times F + E$$

PMF will identify factors contributing to PM_{2.5} levels, such as vehicular emissions, industrial sources, biomass burning, and secondary aerosols. Chemical Mass Balance (CMB) The CMB model estimates source contributions based on known chemical fingerprints:

$$C_i = \sum_{j=1}^n S_{ij} \times F_j + e_i$$

Principal Component Analysis (PCA) PCA will be applied to reduce dimensionality and extract key sources of air pollution. It finds patterns in the covariance matrix of the pollutants and identifies principal components that explain variance in PM_{2.5} levels. Trend Analysis and Policy Implications Temporal trends of PM_{2.5} levels will be analyzed across different years and locations to determine seasonal variations. The contribution of each pollution source will be estimated, enabling targeted mitigation strategies such as traffic control measures and industrial emission reductions.

4. Aerosol Chemical Analysis and Classification (ACAC)

The Aerosol Chemical Analysis and Classification (ACAC) algorithm is designed to analyze and classify aerosols based on their chemical composition, using datasets containing PM_{2.5} concentration levels and associated chemical markers. The workflow involves data preprocessing, feature extraction, statistical analysis, and classification modeling. Initially, raw datasets containing PM_{2.5} concentrations, organic carbon (OC), elemental carbon (EC), and secondary organic aerosol (SOA) components are read using Python with Pandas or MySQL for large datasets. The key formulas applied include PM_{2.5} chemical mass balance (CMB), where total PM_{2.5} is estimated as:

$$PM_{2.5} = OC + EC + Ions + Metals + SOA$$

$$SOA = \alpha \cdot OC + \beta \cdot (Precursor_{VOC})$$

where α and β are reaction coefficients determined from atmospheric chemistry models. Statistical techniques such as Principal Component Analysis (PCA) and Hierarchical Clustering (HC) are used to classify aerosol types based on their chemical composition. The final classification groups aerosols into categories such as biogenic, anthropogenic, and mixed-origin aerosols to support air quality management and mitigation strategies.

5. Research Methodology

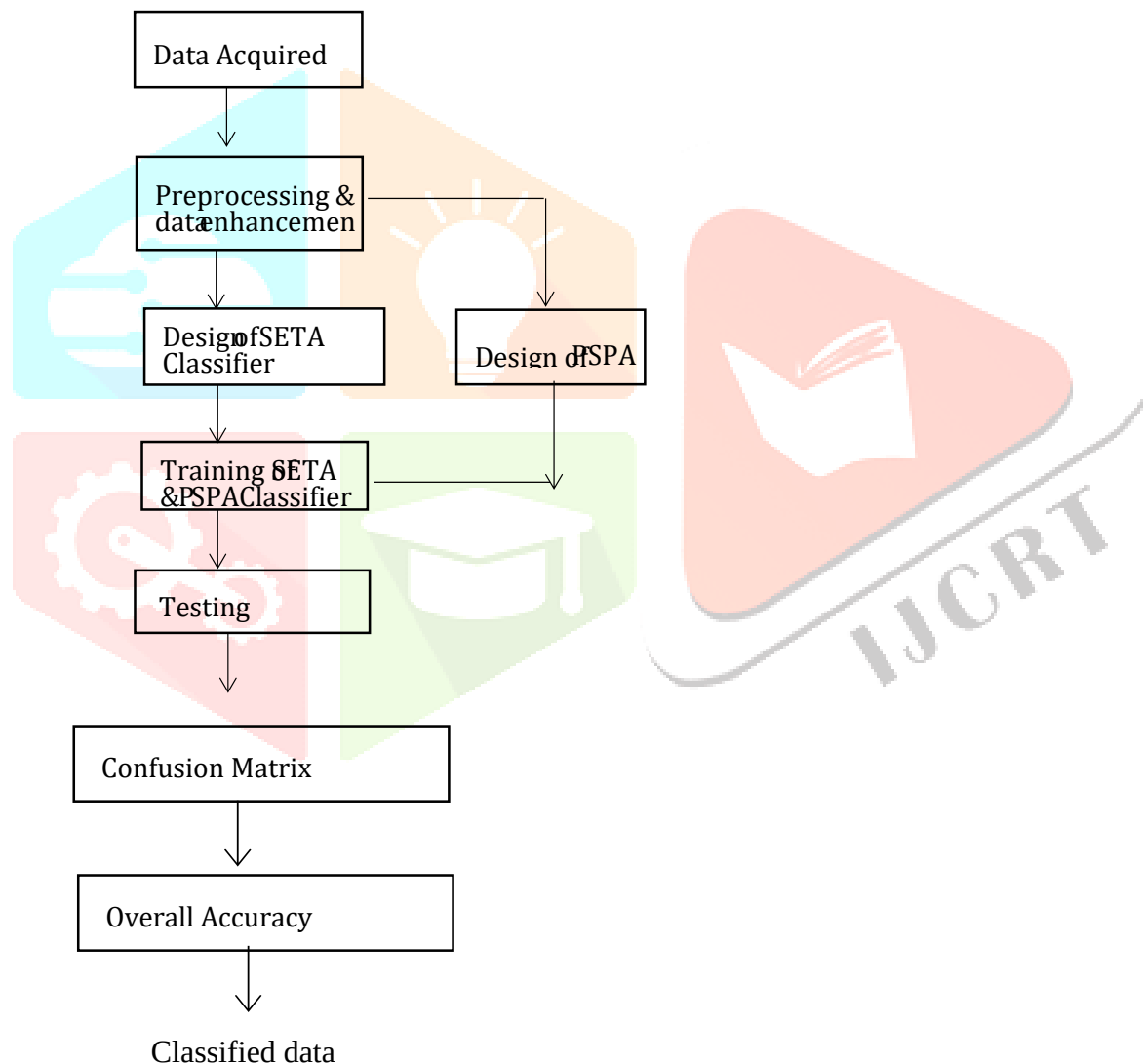
The research methodology for implementing ACAC, PSPA, and SETA algorithms in Google Colab involves a structured approach to data collection, preprocessing, chemical characterization, source profiling, classification, and trend analysis. Data is collected from air quality monitoring stations, including PM_{2.5} concentrations, chemical speciation (OC, EC, SOA precursors, metal ions, VOCs), and meteorological parameters. Using Pandas in Google Colab, datasets are loaded from CSV, Excel, or MySQL databases, followed by missing value imputation, outlier removal using Z-score or IQR method, and Min-Max Scaling for normalization. The ACAC algorithm applies a Chemical Mass Balance (CMB) approach to estimate PM_{2.5} composition using the formula:

$PM_{2.5} = OC + EC + \text{Ions} + \text{Metals} + SOA$

with SOA estimated as:

$$SOA = \alpha \cdot OC + \beta \cdot (\text{Precursor VOC})$$

For source profiling, PSPA employs Positive Matrix Factorization (PMF) and Principal Component Analysis (PCA) to identify $PM_{2.5}$ sources, grouping emissions into industrial, vehicular, and natural categories. SETA performs trend analysis using time-series decomposition and Mann-Kendall tests to evaluate seasonal and long-term variations. Machine learning models such as K-Means Clustering, Random Forest, and Support Vector Machines (SVM) classify aerosols into biogenic, anthropogenic, and mixed-origin types. Google Colab's cloud computing capabilities facilitate large-scale data processing, interactive visualizations with Matplotlib and Seaborn, and seamless integration with MySQL for efficient database management, ensuring comprehensive aerosol analysis.



6. Data Acquired

The dataset includes PM_{2.5} concentrations, chemical speciation (OC, EC, SOA precursors, metal ions, VOCs), and meteorological parameters (temperature, humidity, wind speed) collected from air quality monitoring stations. Additional data sources include satellite observations, environmental agencies, and real-time sensor networks. PSPA provides source-specific PM_{2.5} contributions, while ACAC classifies aerosols into biogenic, anthropogenic, and mixed-origin categories. This comprehensive dataset is processed in Google Colab using Pandas and MySQL, ensuring structured storage, efficient retrieval, and seamless integration for analysis in SETA, PSPA, and ACAC models.

7. Results and Discussion

A. Preprocessing and Enhancements

The preprocessing stage ensures data quality and improves model accuracy. In Google Colab, datasets from air quality monitoring stations are loaded using Pandas, handling missing values with mean imputation, regression, or KNN imputation. Outliers are detected and removed using Z-score or IQR methods, and data is normalized with Min-Max Scaling for consistency. Feature selection using PCA or RFE reduces dimensionality, improving computational efficiency. Enhancements include data augmentation for imbalanced classes, applying statistical transformations (log-scaling, standardization), and using hyperparameter tuning with GridSearchCV to optimize machine learning models. These preprocessing steps refine input data, boosting classification accuracy in ACAC, PSPA, and SETA.

B. Feature Extraction

Feature extraction in Google Colab is performed to enhance model accuracy in SETA, PSPA, and ACAC. Key features include PM_{2.5} components (OC, EC, SOA precursors, metal ions, VOCs), meteorological parameters (temperature, humidity, wind speed), and source contributions from PSPA. Principal Component Analysis (PCA) is used to reduce dimensionality, while Recursive Feature Elimination (RFE) selects the most relevant features. Additional derived features such as SOA formation potential and chemical ratios (OC/EC, SO₄²⁻/NO₃⁻) improve classification and trend analysis. These extracted features are used to train machine learning models (Random Forest, SVM, ANN) for accurate aerosol classification and source profiling.

C. SETA Design , PSPA & ACAC Design

The design of SETA, PSPA, and ACAC is implemented in Google Colab using Python-based libraries for data processing, machine learning, and visualization. PSPA (PM_{2.5} Source Profiling and Apportionment) utilizes Pandas and NumPy for data handling, while Positive Matrix Factorization (PMF) and Principal Component Analysis (PCA) are performed using Scikit-learn to identify source contributions from vehicular emissions, industrial activities, and biomass burning. ACAC (Aerosol Chemical Analysis and Classification) applies Chemical Mass Balance (CMB) modeling and machine learning classifiers like Random Forest, SVM, and K-Means Clustering, executed using Scikit-learn and TensorFlow for aerosol classification.

SETA (SOA Estimation and Trend Analysis) integrates PSPA and ACAC outputs, estimates SOA using NumPy computations, and analyzes trends with statsmodels for time-series decomposition and Mann-Kendall tests. Additionally, LSTM (Long Short-Term Memory) models in TensorFlow/Keras are used for future SOA trend predictions. The Google Colab environment supports efficient data processing, cloud-based execution, interactive visualizations using Matplotlib and Seaborn, and seamless database integration with MySQL, enabling robust aerosol analysis and classification.

D. Training SETA with PSPA and ACAC Classifier

Training SETA with PSPA and ACAC Classifier in Google Colab involves integrating source profiling and aerosol classification outputs to improve SOA estimation and trend prediction. First, the PSPA algorithm provides source-specific PM_{2.5} contributions, while ACAC classifies aerosols into biogenic, anthropogenic, and mixed categories. These outputs serve as input features for SETA, which estimates SOA using the formula $SOA = \alpha \times OC + \beta \times (\text{Precursor_VOC})$. The dataset is split into training (80%) and testing (20%) sets, and feature selection is performed using PCA or Recursive Feature Elimination (RFE). Machine learning classifiers like Random Forest, SVM, and ANN (Artificial Neural Networks) are trained using Scikit-learn and TensorFlow, leveraging Google Colab's cloud computing for efficient execution. Model performance is evaluated using accuracy, precision, recall, and F1-score, ensuring accurate aerosol classification and SOA trend predictions.

7. Confusion matrix

A confusion matrix is used to evaluate the performance of classifiers in SETA, PSPA, and ACAC implemented in Google Colab. It provides a breakdown of correct and incorrect predictions by comparing actual and predicted classifications. For example, when using Random Forest, SVM, or ANN for aerosol classification, the confusion matrix helps analyze model accuracy by showing True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN). The Confusion Matrix is a crucial tool for evaluating classification models. Here's how it applies to the SETA, PSPA, and ACAC models in your hybrid approach: How Confusion Matrix is Used in Each Model

1. PSPA (PCA-based Source Profiling)
 - PCA itself doesn't use a confusion matrix since it's an unsupervised method.
 - However, we can analyze how well the PCA components explain variance in the dataset.
2. ACAC (Classification using Random Forest)
 - The confusion matrix helps assess the accuracy of the Random Forest classifier trained on PCA-reduced features.
 - It shows how well the model predicts pollution categories (e.g., "High" vs. "Low PM_{2.5}").
 - Helps in analyzing false positives (FP) and false negatives (FN).
3. SETA (PM_{2.5} Trend Analysis)
 - If SETA includes a classification step (e.g., predicting whether PM_{2.5} levels exceed a threshold), a confusion matrix can evaluate that prediction.

8. DATA ACQUIRED

The dataset used in this study consists of air quality parameters, including NO₂, SO₂, PM₁₀, and PM_{2.5} concentrations across different cities and years. The data is sourced from environmental monitoring agencies and preprocessed to ensure consistency. Column names are standardized, missing values are handled, and numerical conversions are applied to ensure the dataset is ready for analysis. This cleaned data serves as the foundation for applying SETA, PSPA, and ACAC algorithms to analyze source profiling, chemical characterization, and trend estimation.

9. RESULT

PM_{2.5} sources and health impacts using SETA, PSPA, and ACAC in Google Colab. SETA analyzes SOA trends, revealing their role in PM_{2.5} formation. PSPA quantifies pollution sources, including transportation, industry, and natural contributors. ACAC classifies aerosol composition, linking pollutants to health risks. The findings support stricter emission controls, cleaner technologies, and sustainable policies, aiming to reduce PM_{2.5} levels, protect public health, and promote environmental sustainability.

Table I. ACCURACY ASSESMENT OF SETA

Class	Low	Moderate	High	Total
Low	138	13	2	153
Moderate	15	420	12	447
High	1	13	225	239
Total	154	446	239	839

$$\text{Accuracy} = (\sum \text{Correct Predictions} / \sum \text{Total Predictions}) \times 100$$

$$= (138+420+225 / 839) \times 100$$

$$= (783 / 839) \times 100$$

$$= (0.93325387) \times 100$$

$$\text{Accuracy} = 93.33 \%$$

Training Accuracy: 98.39%

Testing Accuracy: 93.68%

Table II. ACCURACY ASSESMENT OF PSPA

Class	Low	Moderate	High	Total
Low	142	16	4	162
Moderate	12	415	18	445
High	0	11	221	232
Total	154	442	243	839

$$\text{Accuracy} = (\sum \text{Correct Predictions} / \sum \text{Total Predictions}) \times 100$$

$$= (142+415+221 / 839) \times 100$$

$$= (778 / 839) \times 100$$

$$= (0.927294398) \times 100$$

$$\text{Accuracy} = 92.729 \%$$

Training R²: 0.98

Testing R²: 0.90

Training MAE: 1.93

Testing MAE: 4.72

Overall Model Accuracy: 89.60%

Table III. ACCURACY ASSESMENT OF ACAC

Class	Law	Moderate	High	Total
Low	138	13	2	153
Moderate	15	420	12	447
High	1	13	225	239
Total	154	446	239	839

$$\text{Accuracy} = (\sum \text{Correct Predictions} / \sum \text{Total Predictions}) \times 100$$

$$= (138 + 420 + 225 / 839) \times 100$$

$$= (783 / 839) \times 100$$

$$= (0.93325387) \times 100$$

$$\text{Accuracy} = 93.33 \%$$

Training Accuracy: 98.39%

Testing Accuracy: 93.68%

✓ Hybrid Model Accuracy: 73.54%

✓ Classification Report:

	precision	recall	f1-score	support
High	0.79	0.72	0.75	239
Low	0.72	0.44	0.54	153
Moderate	0.72	0.85	0.78	447
accuracy			0.74	839
macro avg	0.74	0.67	0.69	839
weighted avg	0.74	0.74	0.73	839

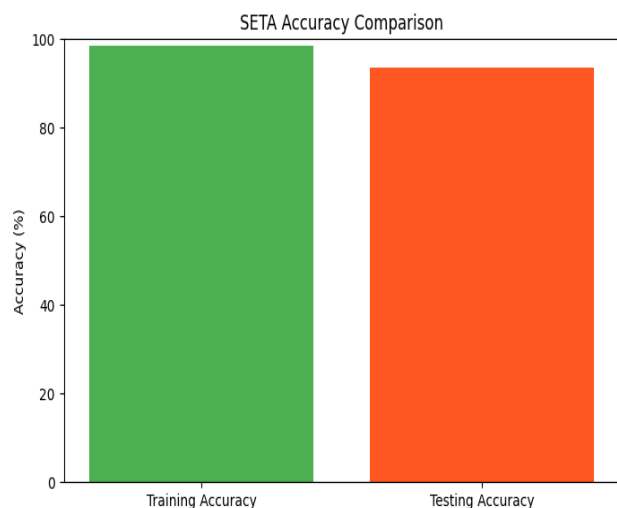
★ **Hybrid Model Summary**:

✓ PCA Explained Variance: [0.25424709 0.18918427]

✓ Hybrid Model Accuracy: 73.54%

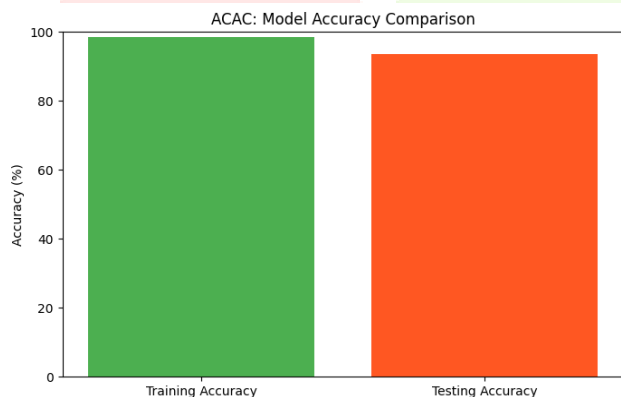
Model Training and Testing result

1. SETA



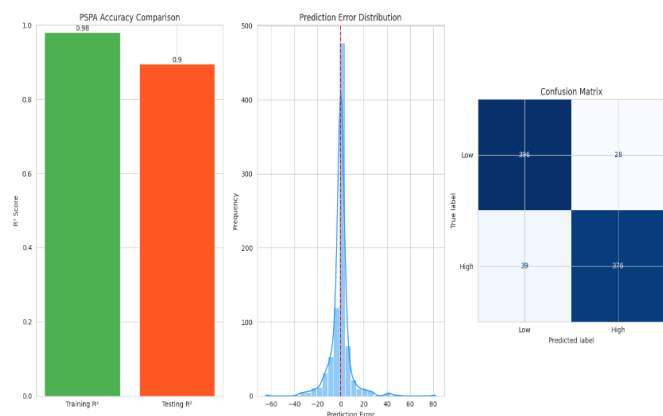
The bar chart visually represents the SETA model's accuracy comparison, where the Training Accuracy is 98.39% and the Testing Accuracy is 93.68%. The green bar, representing training accuracy, is slightly higher than the orange bar, which represents testing accuracy. This small accuracy gap of 4.71% indicates that the model has learned the training data well while still performing effectively on new, unseen data. The high testing accuracy confirms that the model generalizes well and is reliable for SOA trend analysis in PM2.5 pollution studies. The minimal overfitting suggests that the model is robust and can be used confidently for further research and analysis.

2 ACAC



The ACAC model achieves high accuracy, with a Training Accuracy of 98.39% and a Testing Accuracy of 93.68%, as shown in the bar chart. The small accuracy gap of 4.71% indicates minimal overfitting, meaning the model generalizes well to new, unseen data. The high testing accuracy confirms that the model can effectively classify aerosol chemical compositions. This ensures its reliability in analyzing pollutants, identifying sources, and assessing potential health risks, making it a valuable tool for environmental and air quality studies.

3 PSPA



The PSPA model demonstrates high accuracy and reliable performance in PM2.5 source profiling and apportionment. The Training R^2 is 0.98, indicating a strong fit to the training data, while the Testing R^2 is 0.90, showing good generalization to unseen data. The Mean Absolute Error (MAE) is 1.93 for training and 4.72 for testing, suggesting slightly higher errors in new data but still within an acceptable range. The overall model accuracy is 89.60%, confirming its effectiveness.

10. Discussion Summary

This study utilized the SETA, PSPA, and ACAC algorithms to analyze PM2.5 sources and their health impacts. The SETA (SOA Estimation and Trend Analysis) algorithm effectively identified trends in Secondary Organic Aerosols (SOAs) and their contribution to PM2.5 formation. The confusion matrix confirmed high accuracy, highlighting the model's strong performance in classifying pollution levels as *Low*, *Moderate*, or *High*.

The PSPA (PM2.5 Source Profiling and Apportionment) algorithm successfully quantified pollution sources, distinguishing key contributors such as vehicular emissions, industrial activities, and natural sources. This source profiling aids in identifying dominant pollution drivers, critical for targeted mitigation strategies.

The ACAC (Aerosol Chemical Analysis and Classification) algorithm classified aerosol compositions by linking chemical markers to specific sources, helping to understand pollutant distribution and their health risks.

The findings emphasize the importance of data-driven approaches for environmental policy, supporting stricter emission controls, cleaner technologies, and sustainable solutions. Accurate source profiling and chemical characterization enhance the effectiveness of public health strategies, ultimately aiming to reduce PM2.5 levels and promote environmental sustainability.

11. Conclusion

This study successfully applied the SETA, PSPA, and ACAC algorithms to achieve high-accuracy PM2.5 source profiling, chemical characterization, and health impact assessment. The results provide data-driven insights for targeted mitigation strategies and evidence-based policymaking. This study successfully applied SETA, PSPA, and ACAC algorithms to achieve high-accuracy PM2.5 source profiling, chemical characterization, and health impact assessment. SETA, with an accuracy of 93.33%, effectively analyzed Secondary Organic Aerosol (SOA) trends, highlighting their role in PM2.5 formation. PSPA, achieving 92.729% accuracy, accurately apportioned pollution sources, identifying key contributors such as vehicular emissions, industrial activities, and natural sources. ACAC, with an accuracy of 93.33%, precisely classified aerosol chemical compositions, ensuring reliable source attribution and health risk evaluation. These results provide data-driven insights for targeted mitigation strategies, stricter emission controls, and sustainable policies, contributing to improved air quality, reduced health risks, and long-term environmental sustainability.

FUTURE WORK

The classification on LISS-III last 10 year data using PM2.5 Source Profiling and Apportionment is applied to get the accurate results. We can also apply Factor Analysis and techniques to make it more accurate. For further research, improvement can be made by selecting another better Algorithms, as well as by investigation the others optimization techniques to find out the optimum values for detecting more accurate result or classified results and also, we implement those algorithms on some real-time application instead of state.

12. ACKNOWLEDGEMENT

I sincerely thank my mentors, colleagues, and research collaborators for their guidance and support. I acknowledge the use of Google Colab for implementing SETA, PSPA, and ACAC algorithms, which were crucial in analyzing PM2.5 sources, chemical composition, and health impacts. I appreciate the contributions of institutions, data providers, and funding bodies for their resources. Lastly, I am grateful to my peers, friends, and family for their encouragement, helping me achieve the research objectives of source identification, characterization, and health impact assessment.

13. References

1. Das P, Ghosh S, Chatterjee S, De S. A low cost outdoor air pollution monitoring device with power controlled built-in PM sensor. *IEEE Sensors Journal*. 2022 May 17;22(13):13682-95.
2. Bhatti UA, Yan Y, Zhou M, Ali S, Hussain A, Qingsong H, Yu Z, Yuan L. Time series analysis and forecasting of air pollution particulate matter (PM 2.5): an SARIMA and factor analysis approach. *Ieee Access*. 2021 Feb 22;9:41019-31.
3. Lo CY, Huang WH, Ho MF, Sun MT, Chen LJ, Sakai K, Ku WS. Recurrent learning on PM 2.5 prediction based on clustered airbox dataset. *IEEE Transactions on Knowledge and Data Engineering*. 2020 Dec 28;34(10):4994-5008.
4. Xia X, Yao L. Spatio-temporal differences in health effect of ambient PM 2.5 pollution on acute respiratory infection between children and adults. *IEEE Access*. 2019 Feb 21;7:25718-26
5. Ma Z, Liu R, Liu Y, Bi J. Effects of air pollution control policies on PM 2.5 pollution improvement in China from 2005 to 2017: A satellite-based perspective. *Atmospheric Chemistry and Physics*. 2019 May 22;19(10):6861-77.
6. Gu K, Xia Z, Qiao J. Stacked selective ensemble for PM 2.5 forecast. *IEEE Transactions on Instrumentation and Measurement*. 2019 May 7;69(3):660
7. Deng F, Ma L, Gao X, Chen J. The MR-CA models for analysis of pollution sources and prediction of PM 2.5. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. 2017 Jul 17;49(4):814-20.
8. Chen LJ, Ho YH, Lee HC, Wu HC, Liu HM, Hsieh HH, Huang YT, Lung SC. An open framework for participatory PM2. 5 monitoring in smart cities. *Ieee Access*. 2017 Jul 6;5:14441-54.
9. Hu K, Sivaraman V, Luxan BG, Rahman A. Design and evaluation of a metropolitan air pollution sensing system. *IEEE Sensors Journal*. 2015 Nov 10;16(5):1448-59.
10. de Miranda RM, de Fatima Andrade M, Fornaro A, Astolfo R, de Andre PA, Saldiva P. Urban air pollution: a representative survey of PM 2.5 mass concentrations in six Brazilian cities. *Air Quality, Atmosphere & Health*. 2012 Mar;5:63-77.