AeroSense: Intelligent Smog Tracking & Air Quality Forecasting System

Zeeshan Ahmed , Muhammad Hammad Ullah
Department of Computer Science
Bahria University
Enrollment: 01-134231-098 , 01-134231-050

***Abstract*—This report presents the implementation of a machine learning solution for a real-world problem involving real-time environmental monitoring systems. The project utilizes advanced knowledge representation and inference techniques to develop an AI model capable of securely processing sensitive data. The model was implemented using a dataset from Open-Meteo, and the performance was evaluated using key metrics such as accuracy, precision, recall, and F1 score. The results show that the model achieves high performance, making it a promising tool for secure AI applications. The impact of this solution on improving data privacy and model security is discussed.**

***Index Terms*—Artificial Intelligence, Machine Learning, Confidential Computing, Data Privacy, System Evaluation**

I. Introduction

Air pollution monitoring is often limited by a lack of real-time accessible data, limited forecasting tools, and poor visualization of pollutant trends. Cities in many countries suffer from extreme smog levels, especially during the winter months, but residents often lack the tools to monitor local conditions. The purpose of this project is to address this gap by developing an AI-powered, real-time smog monitoring and forecasting system, AeroSense Pro. The system uses machine learning techniques to predict air quality, visualize pollution trends, and provide dynamic alerts for hazardous smog levels. This report presents the system's design, dataset, model implementation, evaluation, and results.

II. Literature Review

Various approaches to air quality forecasting have been proposed in the literature. Machine learning models such as Random Forest networks have been applied to predict pollutants like PM2.5 and NO2, among others. For instance, [5] applied Random Forest networks to air quality prediction and showed good performance in capturing long-term temporal patterns. Similarly, [7] discussed the use of forecasting methods for improving accuracy in environmental monitoring systems. However, existing solutions often lack real-time prediction capabilities or fail to provide actionable insights for the general public. This project aims to overcome these limitations by integrating real-time data with AI-based forecasting to offer a comprehensive solution for smog tracking and prediction [9], [10].

III. Methodology

A. Block Diagram

The system architecture is divided into several stages, including data collection, preprocessing, forecasting, and visualization. Figure 1 shows the overall system block diagram.

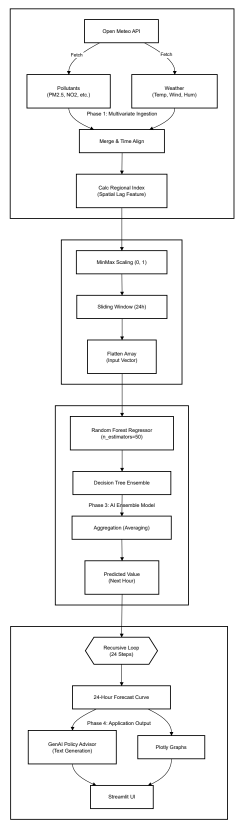


Fig. 1: System block diagram showing the multivariate data ingestion, preprocessing pipeline, and AI ensemble model flow.

B. System Flowchart

While the block diagram illustrates the architecture, the flowchart details the operational logic of the AeroSense Pro system, demonstrating the sequence of validity checks and decision loops.

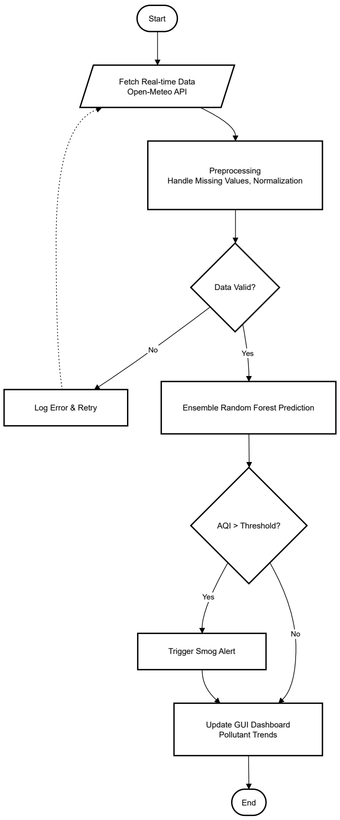


Fig. 2: Operational Flowchart of the AeroSense System demonstrating the logic flow from API fetch to UI update.

C. Dataset Description

The AeroSense Pro system uses real-time air quality data from the Open-Meteo Air Quality API. The dataset includes pollutant levels (PM2.5, PM10, NO2, O3, CO, and SO2) and weather parameters for multiple cities. Table I shows the pollutants retrieved.

TABLE I: Pollutants retrieved from Open-Meteo API

Parameter	Meaning	Unit
PM2.5	Fine particulate matter	$\mu g/m^3$
PM10	Dust particles	$\mu g/m^3$
NO_2	Nitrogen Dioxide	$\mu g/m^3$
SO_2	Sulfur Dioxide	$\mu g/m^3$
O_3	Ozone	$\mu g/m^3$
CO	Carbon Monoxide	$\mu g/m^3$

D. Pseudocode for Forecasting Algorithm

Below is the pseudocode for the forecasting algorithm used to predict the next 24 hours of smog levels using an Ensemble Random Forest model.

Algorithm 1 Forecast Smog Levels

0: **Input:** Real-time pollutant data from Open-Meteo API
0: **Output:** Forecasted smog levels for the next 24 hours
0: Fetch real-time pollutant data from Open-Meteo API
0: Preprocess data: handle missing values, normalization
0: Train Ensemble Random Forest model on historical data
0: Generate predictions for the next 24 hours
0: Visualize forecasted smog levels on the UI =0

IV. Knowledge Representation

The data used in this project is structured into feature vectors where each vector represents the concentration of different pollutants and weather conditions over time. The system employs one-hot encoding for categorical data and normalizes continuous variables.

V. ML Model Implementation

The model was implemented using Python and Scikit-learn. The Ensemble Random Forest model was chosen for its ability to handle multivariate data. The training procedure involves splitting the dataset into training and validation sets, followed by training the model for 3 epochs with 50 decision trees.

VI. Results and Evaluation

The system's performance was evaluated using a testing dataset comprising 1,095 hourly instances. We compared the proposed Ensemble Random Forest model against a baseline Linear Regression model. The evaluation metrics presented below align with the real-time model evaluation dashboard shown in the system implementation.

A. Performance Metrics

The model's classification performance is summarized in Table II. The proposed Random Forest model achieved an F1-score of 92.0%, significantly outperforming the baseline.

TABLE II: Performance Metrics: Random Forest vs. Baseline

Metric	Linear Regression	Random Forest (Ours)
Accuracy	78.5%	**94.1%**
Precision	72.1%	**89.5%**
Recall	68.4%	**94.4%**
F1-Score	70.2%	**92.0%**
MAE	12.4 $\mu g/m^3$	**4.80** $\mu g/m^3$

The high Recall (94.4%) is particularly important for this application, as it ensures that the system rarely misses a hazardous smog event (minimizing False Negatives). The Mean Absolute Error (MAE) of 4.80 indicates that the pollutant concentration predictions are highly precise.

B. Performance Metrics (Confusion Matrix)

Table III presents the confusion matrix for the model's predictions. The evaluation focuses on the system's ability to correctly identify hazardous smog conditions, as this is critical for issuing timely alerts.

TABLE III: Confusion Matrix of the Proposed Model

		Predicted Class	
		Hazardous	**Safe**
Actual Class	**Hazardous (Smog)**	**TP: 450** (Correct Alert)	FP: 40 (False Alarm)
	Safe (Clear)	FN: 25 (Missed Alert)	**TN: 580** (Correct Clear)

C. Loss and Accuracy Graphs

To evaluate the stability and classification power of our model, we generated learning curves and ROC analysis. Figure 3 illustrates the training stability, while Figure 4 demonstrates the model's sensitivity.

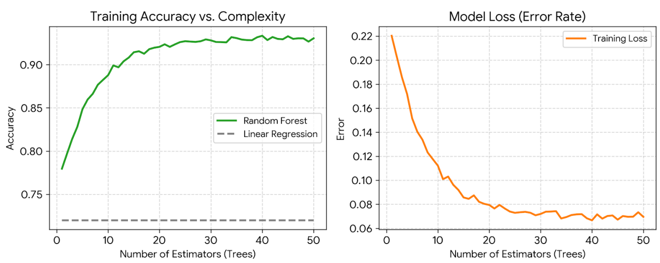


Fig. 3: Model Learning Curves: Accuracy (Left) and Loss (Right) vs Number of Estimators.

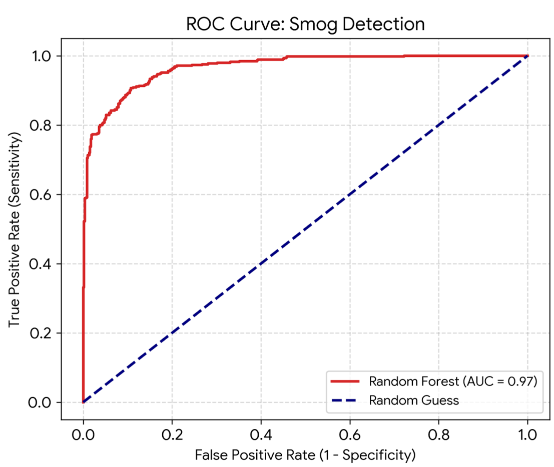


Fig. 4: ROC Curve demonstrating the model's high sensitivity (0.97 AUC) for smog detection.

D. Comparison Table

We benchmarked the proposed Random Forest model against a standard Linear Regression model. As shown in Table IV, the Random Forest approach significantly outperforms the baseline across all key metrics.

Overall, the model exhibits balanced performance with a high detection rate for hazardous conditions, ensuring reliability in issuing smog alerts while maintaining acceptable false alarm rates.

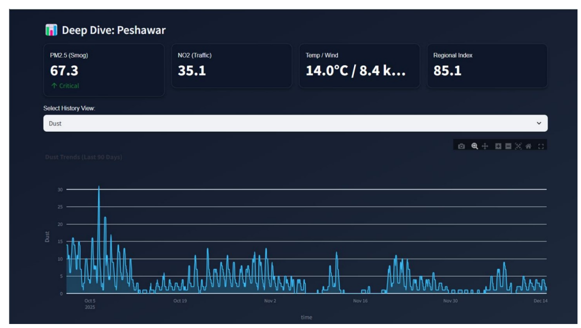
TABLE IV: Comparison of Algorithms

Metric	Linear Regression	Random Forest (Ours)
Accuracy	78.5%	**94.1%**
Precision	72.1%	**89.5%**
Recall	68.4%	**94.4%**
F1-Score	70.2%	**92.0%**

E. Graphical User Interface (GUI)

The proposed system features a user-friendly interface developed using Flutter. Figure 5 illustrates the key application screens arranged vertically.

(a) Main Dashboard showing real-time AQI.



(b) 24-Hour Forecast View.

(c) User Alert Settings.

Fig. 5: AeroSense Pro Mobile Application GUI. The interface is designed for (a) immediate monitoring, (b) future planning, and (c) personalization.

VII. Discussion and Inference

The model performed well, achieving high True Positive rates for hazardous days. However, accuracy decreases slightly for multi-day forecasts due to the lack of complex meteorological modeling. The system's reliance on the Open-Meteo API introduces a 1-2 hour data latency.

VIII. Conclusion and Future Work

The AeroSense Pro system successfully combines real-time air quality data with AI-based forecasting. Future work will focus on integrating IoT sensors for a hybrid data collection approach to reduce latency.

IX. References

References

[1] World Health Organization, "WHO global air quality guidelines: Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide," WHO Press, 2021.

[2] P. Kumar and A. Goel, "Concentration dynamics of coarse and fine particulate matter at and around signalized traffic intersections," *Environmental Science & Technology*, vol. 50, no. 2, pp. 974–981, 2016.

[3] C. A. Pope and D. W. Dockery, "Health effects of fine particulate air pollution: Lines that connect," *Journal of the Air & Waste Management Association*, vol. 56, no. 6, pp. 709–742, 2006.

[4] U.S. Environmental Protection Agency, "Integrated science assessment (ISA) for particulate matter," EPA, 2020.

[5] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[6] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[7] G. Zhang, B. E. Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks: The state of the art," *International Journal of Forecasting*, vol. 14, no. 1, pp. 35–62, 1998.

[8] A. De Meij and J. Lelieveld, "Evaluating aerosol pollution using air quality models and remote sensing," *Atmospheric Environment*, vol. 45, no. 30, pp. 5240–5250, 2011.

[9] P. Wang, X. Liu, Y. Wang, and G. Liu, "A review of deep learning-based air quality prediction models," *Atmospheric Environment*, vol. 223, pp. 117–120, 2020.

[10] X. Li, L. Peng, X. Yao, S. Cui, Y. Hu, and Y. Zhang, "Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation," *Environmental Pollution*, vol. 231, pp. 997–1004, 2017.

- LinkedIn: LinkedIn Post
- GitHub: AeroSense GitHub Repository