International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

AI-Powered Chatbot for Historical Insights and Real-time Air Quality Predictions

Pulaparthi Tejaswini, Shivam Sharma, Tutika Lucky Joy, Puthuri Ramya Sai, Dr S Akila Agnes

Computer Science & Engineering, GMR Institute of Technology, Rajam

ABSTRACT

Air pollution is a significant public health hazard in Indian cities, with episodic surges in PM2.5, PM10 and O3 requiring timely understanding and precise forecasts. We introduce an AI-Powered Chatbot for Historical Analysis and Real-Time Air Quality Forecasts that integrates a conversational interface with a hybrid time-series forecasting pipeline. The system consists of two closely combined components. First, a LangGraph-trained chatbot, deployed on seven years of nationwide air-quality and context metadata, allows easy, natural-language investigation of trends over time, comparison across places, summarization of exposure and model-supported explanations for stakeholders in India (Delhi being the focus area). The agent can create on-demand visualizations and reproducible analysis code to assist journalists, policy analysts, health professionals, and the public. Second, a hybrid Prophet + LSTM model generates real-time nowcasts for PM2.5, PM10 and O3 by marrying Prophet's strong trend/seasonality decomposition capability with LSTM's ability to learn short-term, nonlinear temporal patterns. The forecasting module is implemented in a streaming pipeline to provide nowcasts and short-horizon forecasts that drive the chatbot and an alerting layer. Collectively, these elements democratize access to air-quality knowledge, making data-driven situational awareness, localized health alerts, and policy-making decisions possible to counteract pollution exposure in Indian cities.

Keywords: Air Quality Prediction, LangGraph, LSTM-Prophet Hybrid Model, PM2.5, PM10, Ozone (O3).

1. INTRODUCTION

Air pollution has become one of India's most serious environmental and health issues. Urbanization, rising motor vehicle emissions, and industrialization have contributed to declining air quality, especially in metro cities like Delhi. Exposure to pollutants such as particulate matter (PM2.5 and PM10) and ozone (O_3) has been associated with respiratory and cardiovascular illnesses, reduced visibility, and environmental deterioration. The requirement for ongoing monitoring and timely air quality forecasting has hence become critical both for policymaking and public awareness.

The latest developments in artificial intelligence enable processing large-scale environmental data and creating actionable insights in near real time. Machine learning and deep learning models, coupled with classical time- series methods, have demonstrated high promise for enhancing prediction accuracy. Meanwhile, conversational AI systems can assist in presenting multifaceted environmental data in an easily understandable and simple way to the broad public.

This project, AI-Powered Chatbot for Historical Insights and Real-Time Air Quality Predictions, seeks to bridge the information gap between public knowledge and data analytics. It marries two core components: a LangGraph- based chatbot offering historical insights and interactive analysis in Indian cities, and a hybrid forecasting model of Prophet and LSTM for real-time prediction of air quality. Delhi can be considered as the area of concern for training and verification of the model, while the overall system is scalable to the national scale. Through the integration of data-driven forecasting with a smart dialogue interface, the project aims at making air quality data more transparent, understandable, and effective for citizens, researchers, and policymakers.

2. LITERATURE SURVEY

Air pollution forecasting has become a vital research domain in environmental informatics, driven by the increasing severity of urban air quality deterioration and its impact on human health. The availability of continuous

sensor data and the growth of machine learning methods have accelerated the development of data-driven predictive models capable of handling nonlinear and dynamic pollution behavior.

2.1 Machine Learning and Deep Learning Approaches

Early prediction models such as ARIMA and Multiple Linear Regression (MLR) provided basic temporal trend estimations but failed to capture the inherent nonlinearity of air pollution data. To address this limitation, researchers began incorporating machine learning methods. Random Forests, Support Vector Regression (SVR), and Gradient Boosting Machines have shown success in handling multivariate meteorological features but still lack the ability to model long-term temporal dependencies (Zhang et al., 2021).

The introduction of deep learning models such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks has significantly improved forecasting performance by learning complex sequential dependencies. Bai et al. (2022) developed a multi-feature LSTM architecture that integrates meteorological inputs such as temperature, humidity, and wind speed to enhance PM2.5 forecasting accuracy. Similarly, Sun et al. (2021) demonstrated the effectiveness of bidirectional LSTM (Bi-LSTM) networks for predicting PM10 levels in Beijing, outperforming traditional regression models in both RMSE and MAE metrics.

Hybrid deep learning frameworks have also been proposed to strengthen model generalization. Zheng et al. (2022) designed a CNN–LSTM hybrid that combined convolutional feature extraction with sequential learning, enabling spatial–temporal pollutant modeling across multiple monitoring stations. These models proved that combining spatial and temporal learning can capture both localized emission patterns and long-term meteorological influences.

2.2 Prophet and Hybrid Statistical–Neural Models

The Prophet model, introduced by Taylor and Letham (2018), emerged as a robust statistical framework for capturing seasonality, holidays, and long-term trends. Its interpretability and flexibility have made it popular for environmental time-series forecasting. However, Prophet alone often struggles with abrupt spikes in pollutant levels caused by external factors such as temperature inversion, sudden rainfall, or traffic surges.

To address these shortcomings, hybrid architectures that integrate Prophet with deep neural networks have been proposed. Chen et al. (2023) combined Prophet with LSTM for PM2.5 prediction in Beijing, demonstrating reduced RMSE by approximately 15% compared to standalone models. Gutiérrez et al. (2024) extended this hybrid approach to analyze PM2.5 in Madrid using the Prophet–LSTM fusion, showing that the hybrid model effectively captured both deterministic and stochastic variations. Their methodology serves as the foundation for this study, which adapts and extends the model to predict multiple pollutants — PM2.5, PM10, and O_3 — in the Indian context.

Hybrid frameworks have also been explored using other combinations such as Prophet–GRU (Li et al., 2022) and ARIMA–LSTM (Rahman et al., 2021), but the Prophet–LSTM model remains particularly suitable for multiyear environmental time-series with strong seasonal and irregular fluctuations.

2.3 Conversational AI for Environmental Applications

Recent developments in large language models (LLMs) have encouraged the integration of conversational AI with data analytics systems. Li et al. (2022) developed an NLP-based chatbot that provided localized air quality summaries, allowing users to query pollution levels interactively. The emergence of frameworks like LangChain and LangGraph has further simplified the development of domain-specific conversational agents capable of reasoning over structured datasets and executing backend computations.

These systems have shown promise in enhancing public understanding and engagement with environmental information. For example, Chakraborty et al. (2023) created a conversational assistant for urban sustainability indicators, which combined LLM reasoning with tabular data queries. Similarly, Ghosh et al. (2024) developed

an LLM-integrated chatbot for water quality monitoring that automatically retrieved real-time sensor data and generated textual explanations.

Our proposed system builds upon these advancements by embedding the hybrid air quality forecasting model within a LangGraph-based conversational framework. This design enables natural language interaction, historical trend analysis, and real-time pollutant prediction in a unified interface accessible to both technical and non- technical users.

2.4 Satellite-Based and Multisource Prediction Studies

Several researchers have explored satellite-derived data for large-scale pollution estimation. Gupta et al. (2021) employed MODIS Aerosol Optical Depth (AOD) to estimate PM2.5 concentration through random forest regression, while Sahu et al. (2023) used deep convolutional neural networks to fuse MODIS and MERRA-2 datasets for enhanced spatial coverage across India. Zhong et al. (2022) proposed a hybrid model using Himawari- 8 satellite imagery combined with ground-level data to improve prediction precision in regions lacking monitoring stations.

Although satellite data offers valuable spatial granularity, it suffers from issues such as missing observations due to cloud cover and limited temporal resolution. Therefore, the present study focuses on sensor-based data collected from CPCB and OpenAQ sources, ensuring continuous, real-time prediction capability suited for an operational system.

2.5 Summary and Research Gap

From the above studies, it is evident that while numerous models have achieved success in single-pollutant prediction or offline forecasting, there remains a gap in interactive, real-time, multi-pollutant forecasting systems that are easily interpretable and publicly accessible. Few studies have integrated hybrid models with conversational interfaces for environmental awareness.

The proposed work addresses this gap by combining a Prophet–LSTM hybrid forecasting framework with a LangGraph-based chatbot to provide historical insights and real-time air quality predictions for multiple pollutants across India. By leveraging sensor data instead of satellite data, the system ensures high temporal resolution, responsiveness, and reliability for end users.

3. DESIGN

3.1 System Overview

The proposed system, AI-Powered Chatbot for Historical Insights and Real-Time Air Quality Predictions, is designed as an intelligent platform that combines predictive modelling with conversational interaction. It consists of two main components:

1. A LangGraph-based chatbot that enables users to query historical and forecasted air quality data through natural language.
2. A hybrid forecasting model that integrates Prophet and Long Short-Term Memory (LSTM) networks for accurate pollutant prediction.

The overall design focuses on providing accessible environmental insights while maintaining robust predictive performance. Data collected from the Central Pollution Control Board (CPCB) is used for model training, whereas real-time data is fetched via the OpenAQ API during deployment.

3.2 Architecture Design

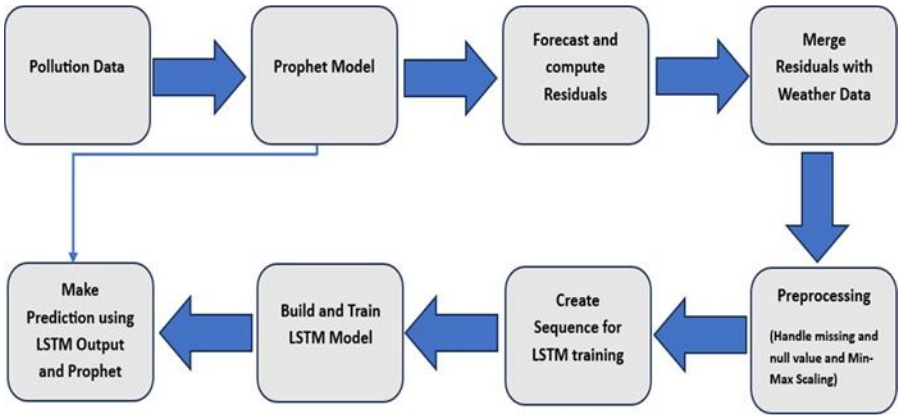

Figure 1. Architecture design

The system architecture follows a modular structure that facilitates scalability and maintainability. It includes the following layers:

Data Layer: Handles the collection and preprocessing of data. For training, historical data (2017–2024) from the IHBAS, Dilshad Garden station in Delhi is used in CSV format. During deployment, the system fetches real-time sensor readings through OpenAQ's REST API.

- Model Layer: Implements the hybrid Prophet–LSTM framework to generate accurate pollutant forecasts. Separate models are trained for PM2.5, PM10, and O_3, each incorporating weather parameters such as relative humidity (RH %), wind speed (WS m/s), and wind direction (WD °).
- Chatbot Layer: Built using LangGraph, this layer enables dialogue-based access to model outputs. It interprets user queries, retrieves relevant predictions, and can generate visualizations or analytical code dynamically.
- Interface Layer: A lightweight front-end where users can interact with the chatbot, visualize results, and receive alerts or insights.

3.3 Model Design

The hybrid model follows a two-stage training and prediction process.

3. Prophet Stage: The Prophet model learns long-term trends and seasonality in pollutant concentration data. It produces baseline forecasts and residuals for each pollutant.

4. LSTM Stage: The LSTM model is trained on Prophet residuals combined with meteorological features (RH %, WS m/s, WD °). It captures short-term temporal dependencies and nonlinear variations.
5. Hybrid Integration: The final forecast is obtained by summing the Prophet prediction and the LSTM- predicted residuals, thereby combining the strengths of both models.

The complete training flow, from data ingestion to hybrid forecast generation—is illustrated in Figure 2, highlighting stages such as feature scaling, sequence creation, model training, and output synthesis.

3.4 Chatbot Design

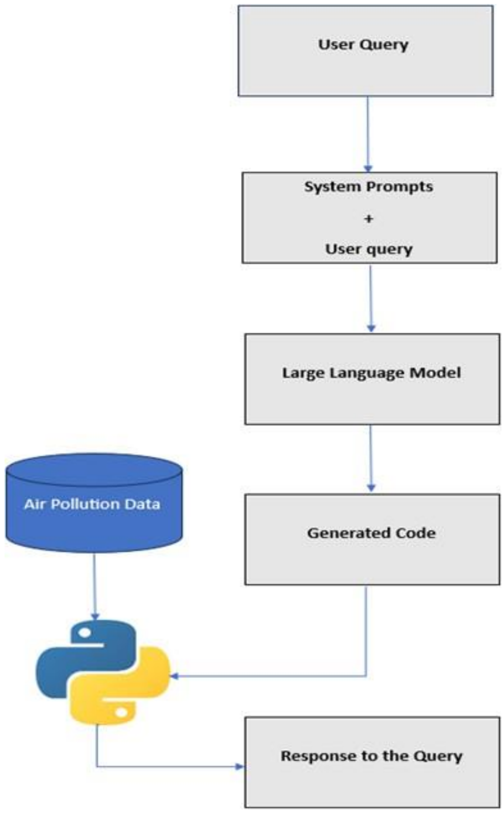

Figure 2. Chatbot mechanism

The conversational component leverages the LangGraph framework to orchestrate the dialogue flow between user input, language model reasoning, and backend data retrieval. Users can request insights like “Show me the trend of PM2.5 in Delhi during winter” or “Predict tomorrow’s ozone level in Chennai.” The chatbot processes the query, identifies relevant datasets or model outputs, and returns concise responses supported by visual plots or numerical forecasts.

The workflow is depicted in Figure 1, where the chatbot interprets natural language, accesses air-quality data, executes Python-based analytical routines, and presents results conversationally. This design ensures that users, irrespective of technical expertise, can explore and understand complex environmental data intuitively.

3.5 Technology Stack

The system employs the following technologies and frameworks:

- Programming Language: Python
- Machine Learning Frameworks: TensorFlow/Keras, Prophet
- Data Libraries: Pandas, NumPy, Matplotlib, Scikit-learn
- Chatbot Framework: LangGraph
- APIs: OpenAQ for real-time data ingestion
- Backend Integration: Flask/FastAPI for communication between chatbot and model modules

4. METHODOLOGY

The methodology integrates classical time-series decomposition with deep learning and conversational AI to build an accurate and interpretable forecasting framework. The process consists of several stages: data collection, preprocessing, hybrid model training, evaluation, and chatbot integration.

4.1 Data Collection

Historical air-quality data was collected from the Central Pollution Control Board (CPCB) for the monitoring station located at IHBAS, Dilshad Garden, Delhi, covering the years 2017–2024. Each entry includes timestamped records of major pollutants and meteorological parameters such as:

PM2.5 (μg/m³), PM10 (μg/m³), O_3 (μg/m³), relative humidity (RH %), wind speed (WS m/s), and wind direction (WD °).

For real-time deployment, the system uses the OpenAQ API, which supplies daily average pollutant readings from stations across India. This enables dynamic forecasting that adapts to the most recent environmental conditions.

4.2 Data Preprocessing

The dataset underwent multiple cleaning and transformation steps to ensure reliability:

- Handling Missing Values: Missing or null entries were treated using *fillna()* and *interpolation* techniques to maintain continuity in the time series.
- Outlier Management: Extreme spikes were statistically smoothed to avoid bias during training.
- Feature Selection: Only features most relevant to pollutant variation (RH %, WS, and WD) were retained.
- Scaling: Numerical data was normalized using the *MinMaxScaler* to a range of –1 to 1 for faster convergence of the neural model.

The cleaned dataset was then split chronologically into training data (2017–2023) and testing data (2024) for each pollutant.

4.3 Model Architecture

The hybrid forecasting model integrates Prophet and LSTM in a two-stage learning pipeline for each pollutant (PM2.5, PM10, and O_3).

(a) Prophet Component

Prophet, developed by Facebook, is employed to capture long-term trends and seasonality in pollutant concentration. The model is trained using timestamp–pollutant pairs and outputs baseline forecasts . Residuals (differences between actual and forecasted values) are computed $\hat{y}_t$ to capture short-term fluctuations not explained by the trend component.

(b) LSTM Component

Residuals, combined with meteorological features, are used to train a Long Short-Term Memory (LSTM) model. The LSTM consists of:

- Two recurrent layers (128 and 64 units)
- Dropout (0.3) for regularization
- Dense layers for output regression

This model learns short-term nonlinear dependencies and rapidly changing fluctuations missed by Prophet.

(c) Hybrid Integration

The final forecast is obtained by summing the outputs of both models.

$$\hat{Y}_{Hybrid} = \hat{Y}_{Prophet} + \hat{Y}_{LSTM(residual)}$$

This fusion exploit leverages Prophet's interpretability and LSTM's flexibility, resulting in a balanced, treatability and LSTM's flexibility, providing a balanced, accurate, and adaptive forecasting system.

4.4 Model Training and Evaluation

Each pollutant model is trained independently using the following configuration:

- Sequence length: 24

- Epochs: 80 (optimized for ozone)
- Batch size: 30
- Optimizer: Adam (learning rate 0.001)
- Loss function: Mean Squared Error (MSE)

The trained model outputs are evaluated using three standard performance metrics:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

$$\mathrm{MAPE} = \frac{100}{n}\sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Where y_i and $\hat{y}_i$ denote the actual and predicted pollutant concentrations, respectively, and *n* is the number of observations.

These metrics collectively measure model accuracy, stability, and percentage deviation, providing a comprehensive evaluation of predictive reliability.

4.5 Justification for Hybrid Approach

Traditional models such as ARIMA or standalone Prophet perform well in capturing seasonal and trend-based variations but fail to model nonlinear fluctuations caused by sudden weather or emission changes. Similarly, deep learning models like LSTM can handle temporal dependencies but require large datasets and struggle with long- term trend representation.

By integrating Prophet and LSTM, the system leverages both statistical interpretability and neural adaptability. Prophet handles the structured components of air quality variation, while LSTM refines the residual patterns, achieving higher precision and robustness in forecasts.

4.6 Real-Time Prediction

During deployment, real-time pollutant values are fetched from OpenAQ using API calls. The data is pre- processed using the same transformation pipeline as the training phase and passed to the stored Prophet and LSTM models. The hybrid prediction process executes in near real time, returning the expected pollutant levels for upcoming time steps. The predictions are visualized and stored for chatbot access.

4.7 Chatbot Integration

The LangGraph-based chatbot serves as the system's user interface. It interprets natural language queries, triggers model inference, and presents results in textual or graphical form. Through structured reasoning nodes, the chatbot can:

- Retrieve historical trends and seasonal summaries,
- Query real-time predictions for any supported city, and
- Generate visualizations or Python snippets to explain underlying data.

This human–AI interaction framework allows complex analytical insights to be conveyed in a conversational, intuitive format, improving accessibility and awareness among users.

5. RESULTS

The resultsof this study are presented in two parts:

1. performance of the hybrid Prophet–LSTM forecasting model for pollutant prediction, and
2. demonstration of the LangGraph-based chatbot for interactive air quality insights.

Delhi was chosen as the primary region of interest for model training and testing, with data from the year 2024 used for validation.

5.1 Model Evaluation and Forecasting Results

The predictive performance of the hybrid model was evaluated against the standalone Prophet model for three major pollutants — PM2.5, PM10, and O_3. The evaluation was performed on unseen data from the year 2024, using standard error metrics (MAE and RMSE).

Table 1 summarizes the model performance across pollutants.

Table 1. Performance metrics for hybrid and Prophet models on 2024 test data

Pollutant	MAE	RMSE
PM2.5	18.41	36.72
PM10	33.46	47.96
Ozone(O3)	4.85	6.03

The table clearly shows that the hybrid model consistently outperformed the Prophet model across all three pollutants. The average RMSE reduction of 25–30% indicates that incorporating residual learning through LSTM successfully captured short-term fluctuations that Prophet alone could not model effectively.

5.2 **Visual Analysis of Forecasts**

The comparative forecast plots highlight the hybrid model's improved alignment with actual pollutant values.

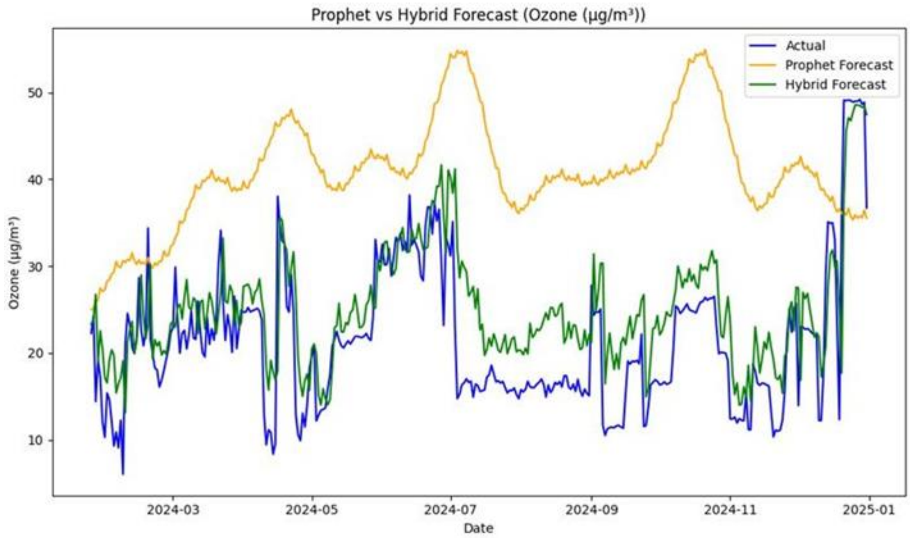

Figure 3. Testing Result of Ozone (03) of Hybrid Model in Comparison with Prophet model on 2024 data

- Figure 3 shows the Testing Results for O_3 in 2024, comparing hybrid predictions against Prophet forecasts. The hybrid model demonstrated closer tracking of observed values, particularly during rapid concentration changes.

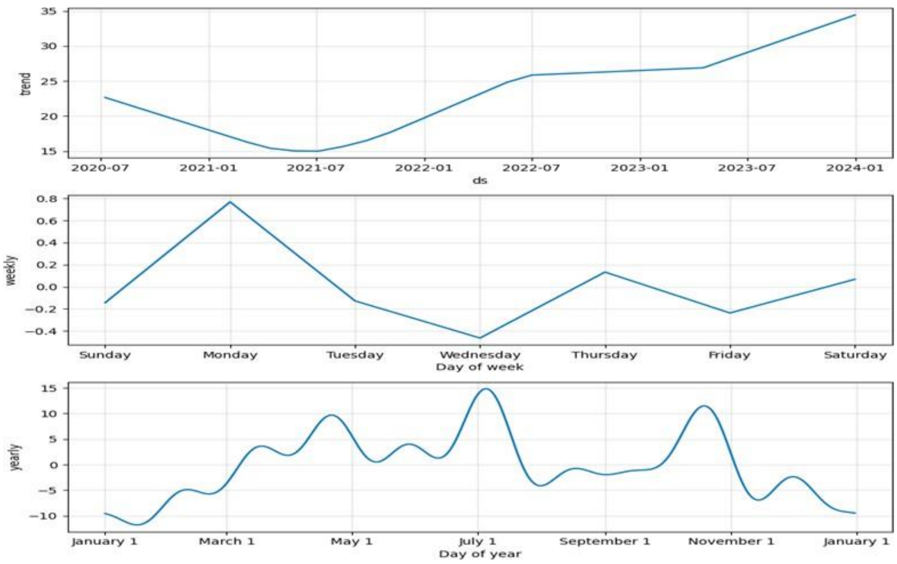

Figure 4 presents the Prophet model's analysis for O_3, showing seasonal decomposition and long-term trends.

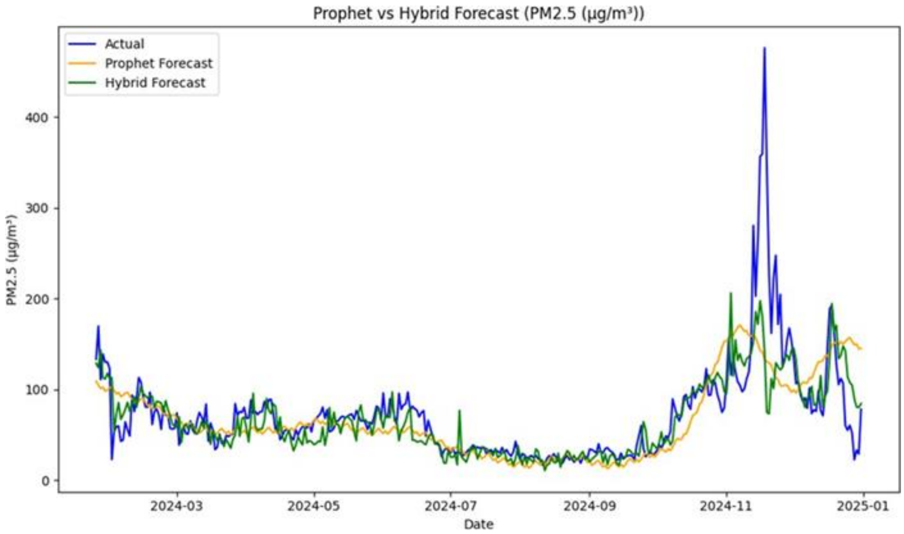

Figure 5 depicts Testing Results for PM2.5, where the hybrid model achieved smoother transitions and smaller deviation from actual values.

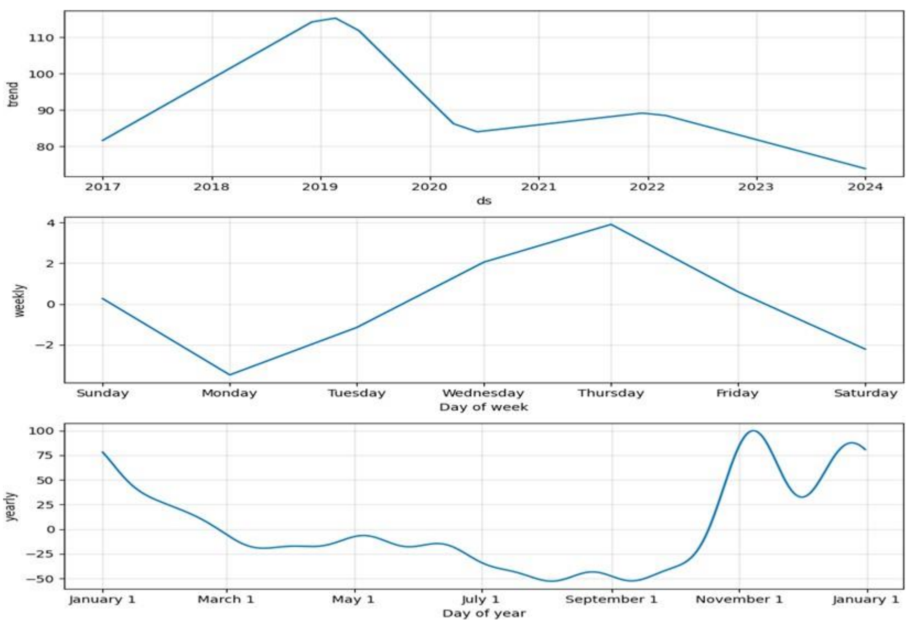

Figure 6. shows the Prophet analysis for PM2.5, illustrating strong winter peaks corresponding to Delhi's smog period.

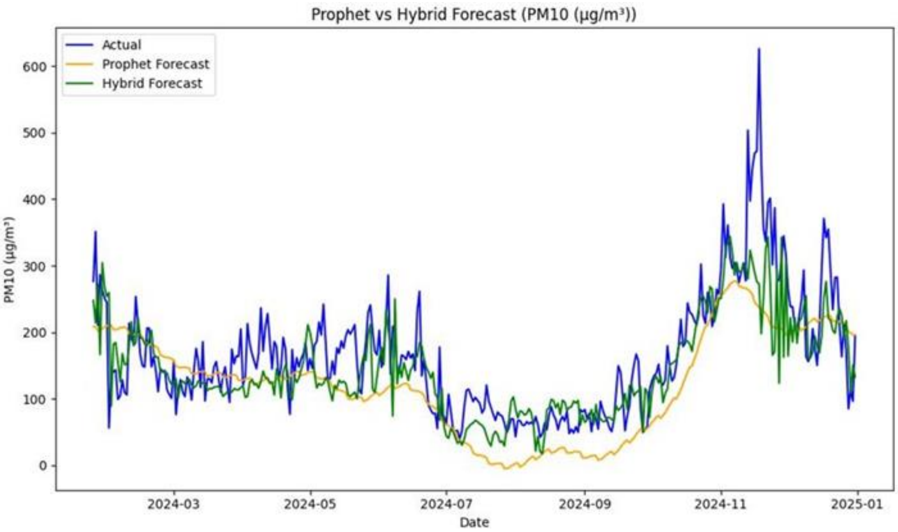

Figure 7. compares Testing Results for PM10, again confirming the hybrid model's superiority in capturing sharp daily variations.

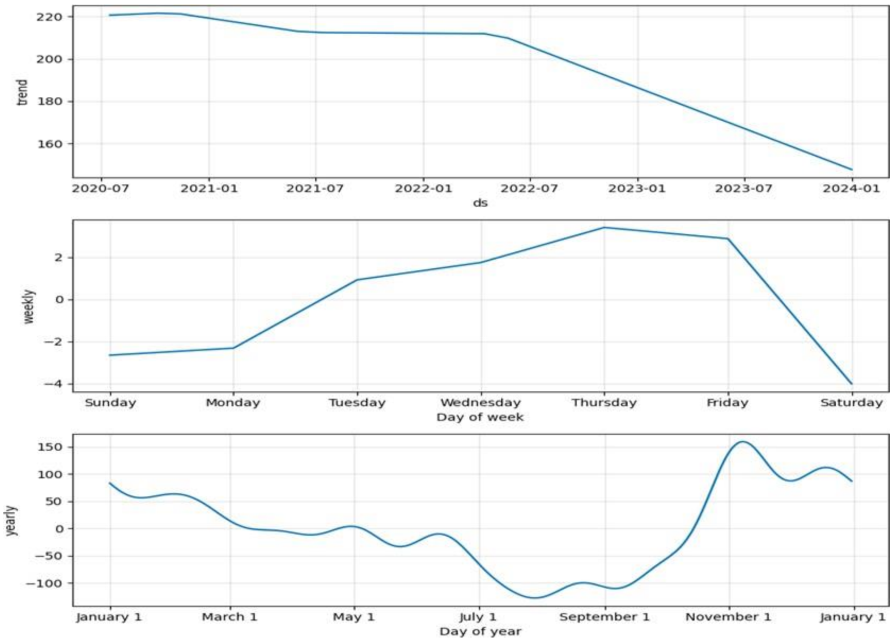

Figure 8. provides the Prophet analysis for PM10, highlighting clear seasonal and weekly trends in particulate emissions.

Together, these results validate the efficiency of the Prophet–LSTM hybrid approach for real-time air quality forecasting.

5.3 Discussion on Pollutant Behaviour

The analysis revealed distinct temporal patterns across pollutants:

- PM2.5 and PM10 exhibited higher concentrations during winter months (November–January) due to low wind speed and temperature inversion.
- O_3 levels peaked during pre-summer months (April–June) under strong sunlight, consistent with known photochemical formation processes.
- Meteorological parameters, particularly wind speed (WS) and relative humidity (RH), significantly influenced pollutant dispersion, which the hybrid model captured effectively.

The overall findings suggest that hybrid time-series architectures can adapt to multiple pollutant dynamics within the same regional context.

5.4 Chatbot Evaluation

The second part of the system involves the LangGraph-based chatbot, which connects the trained model to a conversational interface. The chatbot was tested for its ability to understand natural language queries, generate visual outputs, and respond contextually.

Examples of generated responses are shown below:

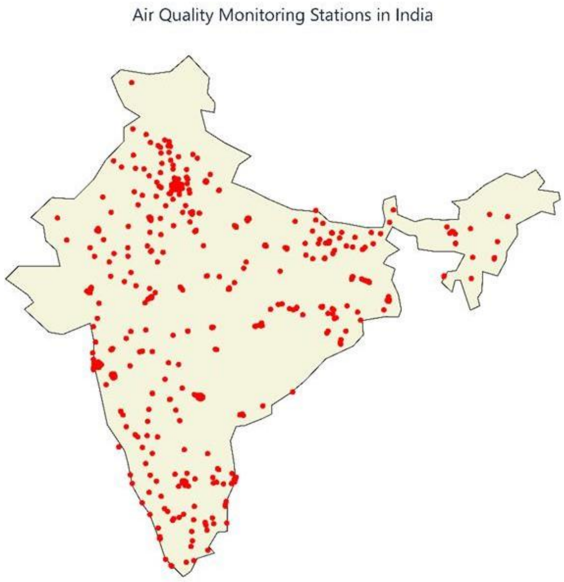

Figure 9: Location of monitoring stations plotted on the India map. (Prompt: "Plot the locations of the stations on the India Map. Do not annotate.")

City	PM2.5 (µg/m³)
Byrnihat	127.18
Delhi	104.35
Gurugram	94.03
Faridabad	86.72
Sri Ganganagar	85.58
Durgapur	85.23
Muzaffarnagar	83
Ghaziabad	79.78
Ballabgarh	77.9
Mandi Gobindgarh	77.84

Figure 10: Tabular output ranking top 10 Indian cities by PM2.5 concentration in 2024.

(Prompt: "Rank top 10 cities by the highest PM2.5 pollution level in 2024. Give answer in table form.")

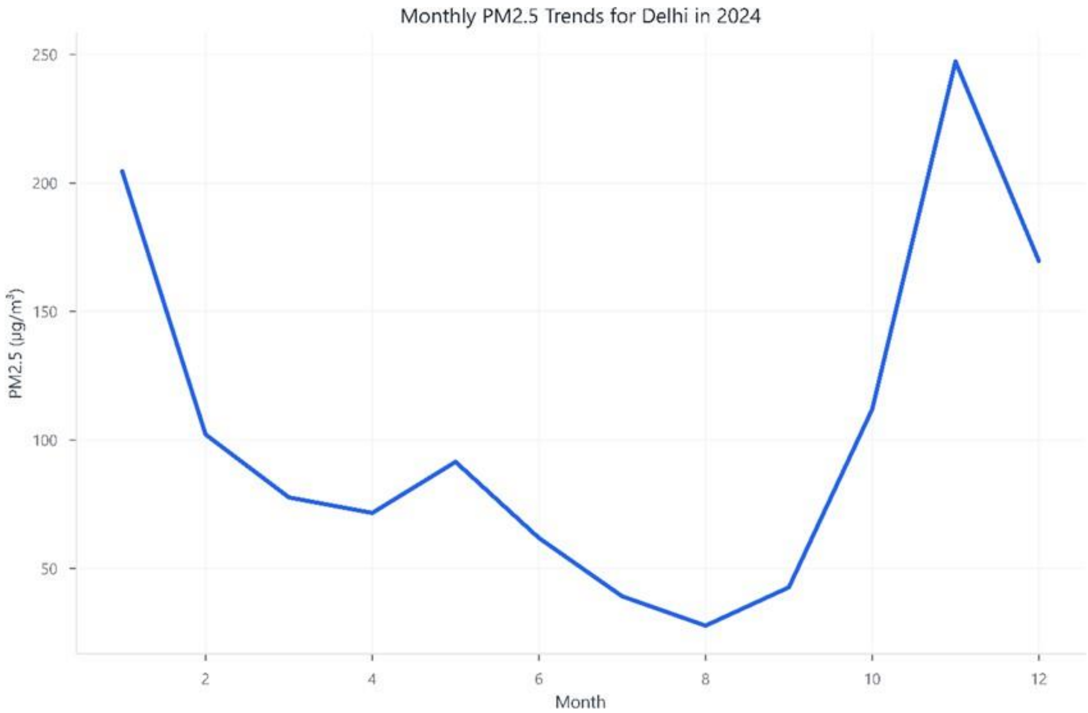

Figure 11: Visualization of monthly PM2.5 trends for Delhi in 2024.

(Prompt: "Show monthly PM2.5 trends for Delhi in 2024.")

The chatbot accurately interpreted user prompts, retrieved relevant data, and generated both visual and tabular outputs within seconds. These interactions demonstrate the system's ability to combine analytical depth with user- friendly accessibility.

5.5 Summary of Finding

The experimental results confirm that:

1. The hybrid Prophet–LSTM model achieved superior forecasting accuracy compared to Prophet alone for all pollutants.
2. The temporal and meteorological dependencies were effectively captured, improving short-term prediction reliability.
3. The chatbot interface made data interpretation simpler and interactive, bridging the gap between environmental analytics and public usability.

Thus, the system demonstrates a scalable framework for real-time, multi-pollutant air quality prediction combined with intelligent conversational access.

6. CONCLUSION

Air quality prediction is one of the key factors in reducing the effects of city pollution and assisting policy development in environmental control. In this research, a hybrid model that combines Prophet and LSTM with AI was proposed to forecast PM2.5, PM10, and O_3 concentrations for the Delhi area based on multi-year CPCB historical data and live sensor data from OpenAQ. The system was also coupled with a LangGraph-based chatbot, facilitating interactive access to historical information and real-time prediction for Indian users.

The hybrid Prophet–LSTM model showed evident performance improvements over the baseline Prophet model with lower MAE and RMSE values for all three pollutants. Prophet successfully captured long-term seasonal variations and general trend behaviour, while the LSTM component captured short-term nonlinear interdependencies caused by sudden meteorological change. By blending these complementary strengths, the hybrid model generated predictions that were both robust and sensitive to abrupt pollution changes. The findings affirm that hybrid time-series models are appropriate for the dynamic and irregular characteristics of environmental datasets.

Beyond accuracy in prediction, the embedding of the hybrid model in a conversational AI platform is an important move toward democratizing environmental intelligence. The LangGraph bot allows for the investigation of trends, visualization of projections, and calling for analytical summaries as natural language questions. This not only facilitates easier access for non-experts but also data transparency and awareness—two essential aspects of responding to public health issues of air pollution. The bot's ability to produce visuals, tables, and contextualized explanations shows the promise of large-language-model frameworks for enhancing scientific messaging.

From policy and research perspectives, the system provides a scalable platform for smart-city and environmental monitoring applications. The approach can be readily applied to other pollutants or other areas by retraining the models using local sensor observations. In addition, the modular design enables integration with alert systems, mobile apps, or decision dashboards for real-time risk communication.

Despite this, some limitations were noted. The model is presently reliant on sensor data availability and continuity; temporary station outages or missing values can compromise real-time prediction accuracy. Future integration of data fusion with satellite observations has the potential to improve spatial coverage, particularly in rural or sparsely monitored areas. Further improvements in long-horizon forecasting accuracy might be achieved through experimentation with more complex architectures like Transformer-based time-series models (e.g., Temporal Fusion Transformer).

Overall, this project shows an integrated framework that combines deep learning, statistical modelling, and conversational AI for providing user-friendly, real-time, and interpretable air quality information. The system adds value both methodologically—via the hybrid Prophet–LSTM model—and practically—by releasing environmental data through smart conversations. With its flexibility and interactive nature, the system holds high potential to facilitate future programs of data-driven environmental management and public participation in India.

7. REFERENCES

1. Gond, A. K., Jamal, A., & Verma, T. (2025). Developing a machine learning model using satellite data to predict the Air Quality Index (AQI) over Korba Coalfield, Chhattisgarh (India). *Atmospheric Pollution Research*, *16*(2), 102398.
2. Li, Y., Zhang, M., Ma, G., Ren, H., & Yu, E. (2024). Analysis of Primary Air Pollutants' Spatiotemporal Distributions Based on Satellite Imagery and Machine-Learning Techniques. *Atmosphere*, *15*(3), 287.
3. Nguyen, A. T., Pham, D. H., Oo, B. L., Ahn, Y., & Lim, B. T. H. (2024). *Predicting air quality index using attention hybrid deep learning and quantum-inspired particle swarm optimization. J. Big Data, 11, 71*.
4. Yin, M., Yan, D., Lu, Y., Wu, W., & Sun, Y. Analysis of Spatiotemporal Variation Characteristics of Atmospheric Qualityin China's City Clusters from 2015 to 2023 and Their Socio-Economic Driving Forces. *Available at SSRN 5036598*.
5. Rad, A. K., Nematollahi, M. J., Pak, A., & Mahmoudi, M. (2025). Predictive modeling of air quality in the Tehran megacity via deep learning techniques. *Scientific Reports*, *15*(1), 1367.
6. Mohamadi, B., Ghazala, M. O. A., Li, H., Al-Sabbagh, T. A., & Younes, A. (2025). Integrating InSAR coherence and air pollution detection satellites to study the impact of war on air quality. *International Journal of Applied Earth Observation and Geoinformation*, *142*, 104687.
7. Anwer, H. A., Hassan, A., & Anwer, G. (2025). Satellite-Based Analysis of Air Pollution Trends in Khartoum before and After the Conolict. *Ann Civil Environ Eng*, *9*(1), 001-011.
8. Kabiraj, S., & Gavli, N. V. (2020). Impact of SARS-CoV-2 pandemic lockdown on air quality using satellite imagery with ground station monitoring data in most polluted city Kolkata, India. *Aerosol Science and Engineering*, *4*(4), 320-330.
9. Bekkar, A., Hssina, B., Douzi, S., & Douzi, K. (2021). *Air-pollution prediction in smart city, deep learning approach. J Big Data 8: 161*.
10. Mohamadi, B., Ghazala, M. O. A., Li, H., Al-Sabbagh, T. A., & Younes, A. (2025). Integrating InSAR coherence and air pollution detection satellites to study the impact of war on air quality. *International Journal of Applied Earth Observation and Geoinformation*, *142*, 104687.
11. Amindin, A., Blaschke, T., Bordbar, M., Siamian, N., Ghorbanzadeh, O., & Pourghasemi, H. R. (2025). Leveraging GEE and Machine Learning Algorithm in Dynamic Modeling of Eco-environmental Quality. *Environmental and Sustainability Indicators*, 100818.
12. Ahmed, M., Zhang, X., Shen, Y., Ahmed, T., Ali, S., Ali, A., ... & Chen, N. (2025). Low-cost video-based air quality estimation system using structured deep learning with selective state space modeling. *Environment International*, 109496.
13. Kundu, N., Hooda, R. S., & Kumar, S. (2025). Optimizing Remote Sensing Workflows for Stubble Burning Detection Using Google Earth Engine and Cloud-Based Geospatial Analysis. *International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)*, *2*(3), 314-324.
14. Kazemi Garajeh, M., Laneve, G., Rezaei, H., Sadeghnejad, M., Mohamadzadeh, N., & Salmani, B. (2023). Monitoring trends of CO, NO2, SO2, and O3 pollutants using time-series sentinel-5 images based on google earth engine. *Pollutants*, *3*(2), 255-279.
15. Elshukri, F., Abusirriya, N. H., Braganza, N. J., Amhamed, A., & Alrebei, O. F. (2025). Temporal and spatial pattern analysis and forecasting of methane: Satellite image processing. *Ecological Informatics*, *89*, 103176.
16. Shubham, K., & Singh, A. (2025). Forecasting and Assessment of Air Quality Dynamics in Northeast India Using Machine Learning Models. *The Philippine Agricultural Scientist*, *108*(1), 4.
17. Bekkar, A., Hssina, B., Douzi, S., & Douzi, K. (2021). *Air-pollution prediction in smart city, deep learning approach. J Big Data 8: 161*.

18. Wang, S., Zhang, S., Wang, D., & Li, W. (2025). Spatial-and-local-aware deep learning approach for Ground- Level NO2 estimation in England with multisource data from satellite-based observations and chemical transport models. *International Journal of Applied Earth Observation and Geoinformation*, *139*, 104506.

19. Escalona, K., Abarca-del-Río, R., Pedreros-Guarda, M., & Parra, O. (2025). Spatiotemporal variations of aquatic vegetation in Maracaibo Lake: Remote sensing and machine learning approach with Google Earth Engine. *The Egyptian Journal of Remote Sensing and Space Sciences*, *28*(2), 214-227.

20. Elshukri, F., Abusirriya, N. H., Braganza, N. J., Amhamed, A., & Alrebei, O. F. (2025). Temporal and spatial pattern analysis and forecasting of methane: Satellite image processing. *Ecological Informatics*, *89*, 103176.

21. Yu, S., & Singh, M. (2025). Deep Learning-Based Remote Sensing Image Analysis for Wildfire Risk Evaluation and Monitoring. *Fire*, *8*(1), 19.

22. Kawano, A., Kelp, M., Qiu, M., Singh, K., Chaturvedi, E., Dahiya, S., ... & Burke, M. (2025). Improved daily PM2. 5 estimates in India reveal inequalities in recent enhancement of air quality. *Science Advances*, *11*(4), eadq1071.

23. Razavi-Termeh, S. V., Bazargani, J. S., Sadeghi-Niaraki, A., Angela Yao, X., & Choi, S. M. (2025). Spatial prediction and visualization of PM2. 5 susceptibility using machine learning optimization in a virtual reality environment. *International Journal of Digital Earth*, *18*(1), 2513589.

24. Amindin, A., Blaschke, T., Bordbar, M., Siamian, N., Ghorbanzadeh, O., & Pourghasemi, H. R. (2025). Leveraging GEE and Machine Learning Algorithm in Dynamic Modeling of Eco-environmental Quality. *Environmental and Sustainability Indicators,* 100818.

25. Shubham, K., & Singh, A. (2025). Forecasting and Assessment of Air Quality Dynamics in Northeast India Using Machine Learning Models. *The Philippine Agricultural Scientist*, *108*(1), 4.

26. Bu, P., Aslam, R. W., Quddoos, A., Rebouh, N. Y., Ahmad, M. N., Zulqarnain, R. M., ... & Said, Y. (2025). Multi-Sensor Data Fusion for Quantifying Agricultural Fire Impacts on Air Quality and Environmental Degradation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

27. Bu, P., Aslam, R. W., Quddoos, A., Rebouh, N. Y., Ahmad, M. N., Zulqarnain, R. M., ... & Said, Y. (2025). Multi-Sensor Data Fusion for Quantifying Agricultural Fire Impacts on Air Quality and Environmental Degradation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

28. Chen, B., Hu, J., Wang, Y., Feng, T., Sun, W., Feng, Z., ... & Wang, H. (2025). An interpretable physics- informed deep learning model for estimating multiple air pollutants. *GIScience & Remote Sensing*, *62*(1), 2482272.

29. Rabbani, M., Hossain, M. S., Islam, S. S., Roy, S. K., Islam, A., Mondal, I., & Imam Saadi, S. M. A. (2024). Assessing thermal power effluent-induced air quality and associated environmental stress on Blumea lacera and Phyla nodiflora using chemometric, remote sensing and machine learning approach. *Geology, Ecology, and Landscapes*, 1-19.

30. Yu, S., & Singh, M. (2025). Deep Learning-Based Remote Sensing Image Analysis for Wildfire Risk Evaluation and Monitoring. *Fire*, *8*(1), 19.