Grenze International Journal of Engineering and Technology, June Issue

A Machine Learning-based Framework for Air Quality Analysis, Ozone Prediction, Procedural Visualization, and Carbon Footprint Assessment

Aarya Atul Thombre[1], Vaishnavi Rajendra Matsagar[2], Jay Ketan Sindhava[3], Sarthak Sadanand Nere[4], Rupali Ramdas Shevale[5] and Balasaheb S. Tarle[6]
[1-6]MVPS's Karmaveer Adv. Baburao Ganpatrao Thakare's College of Engineering, Nashik, Maharashtra, India
Email: aaryaathombre754@gmail.com, vaishno1702@gmail.com, jay.sindhava2504@gmail.com, sarthaknere09@gmail.com, rupalichavanke15@gmail.com, tarle.balasaheb@kbtcoe.org

***Abstract*— Air Pollution and ozone depletion have continued to be the leading environmental issues since the rise of industrialization and urbanization. The urgency to address this has reached a level where citizens have to replace checking the weather conditions with checking if the air around them is breathable. Although some platforms currently moni-tor Air Pollution or Ozone Depletion in real time, they do not bring together the monitoring, analysis and prediction of correlated environmental topics like Air Pollution, Ozone Depletion and Carbon Footprint. Moreover, the existing systems present these insights with overwhelming data that hinders their easy understanding by the common public. Our system presents a one-stop platform that monitors, analyzes, and predicts the air pollution and ozone depletion around the globe and presents the data in the form of graphs, charts and procedurally generated art. The outcomes are displayed on a responsive 3D Globe that redirects the user to the 2D map of the country, fol-lowed by region-wise dashboards for detailed analysis. Users can also calculate their personalized carbon footprint and analyze the impact of daily actions and make sustainable decisions. It aims to build a platform that can simulate Air Pollution and its ef-fects on Ozone Depletion over years using Procedurally Generated Art by collaborating Environment, Art, Data and Technology.**

***Keywords*— Air Pollution, Ozone Depletion, Carbon Footprint, Monitor, Analyze, Predict, Procedural Generated Art.**

I. INTRODUCTION

Rapid urbanization, industrial growth, and unsustainable lifestyle patterns have raised environmental degradation issues across the world. Two of these burning issues are Air Pollution and Ozone Depletion. These environmental issues are the events we face very frequently and significantly affect our daily lives. Major metropolitan cities are facing alarming number of particulate matter concentrations, greenhouse gas emissions, and environmental instability [4], [7]. Above these, changes in the ozone layer occur gradually over extended periods, making them difficult to observe and monitor in the short term [8], [9]. These environmental challenges are interconnected, dependent, dynamic and influenced by industrial activity and unsustainable decision making. Air pollution leads to a wide range of consequences like respiratory illnesses such as asthma, cardiovascular diseases and increased premature mortality [4].

Grenze ID: 01.GIJET.12.2.367_26

Ozone depletion increases exposure to the harmful ultraviolet (UV) radiation, leading to increased cases of skin cancer, cataracts and immune system suppression [9], [10]. Hence, environmental awareness has been of utmost importance with a major focus on topics such as Air Pollution and Ozone Depletion.
Environmental monitoring has entered a data-driven era with the rapid growth of smart cities, IoT-enabled air quality sensors, and satellite-based atmospheric observation systems [4], [13], [14]. Large volumes of environmental data are being generated at continuous time intervals, including particulate matter, nitrogen oxides, sulfur dioxide, carbon emissions, and ozone concentration levels [13], [14]. While this scientific data presents analysis, its access remains limited to technical dashboards and numerical reports that are difficult for non-experts to interpret.
Most platforms used for air pollution and ozone monitoring primarily present pollution levels through basic color indicators and isolated data views [14], [15]. They do not integrate monitoring, analysis, and predictive forecasting within a single framework [4]. Furthermore, the interrelationship between air pollution, the factors influencing it, ozone depletion, and its long-term recovery or decline is rarely addressed in a connected manner [8], [10].
With advancements in Artificial Intelligence, Machine Learning and Computer Graphics, complex environmental datasets can be processed, correlated, and transformed into meaningful predictive insights. Traditional statistical models such as linear regression that are frequently used for pollution forecasting often fail when it comes to nonlinear patterns and long-term dependencies in atmospheric data [3], [5]. Methods like Random Forest are better suited for mapping complex relationships in structured behavioral and environmental datasets, making them effective for personalized carbon footprint estimation [17]. For time-series prediction of air pollution and ozone concentration, recurrent neural networks (RNN), in particular, LSTM models have been widely explored for demonstrating seasonal variations and predictions [1], [5], [6].
Hence, the goal of the proposed system, Akashtattva, is to understand the scientific foundations of these environmental issues and make citizens, students, researchers and environmentalists have easy access to the monitoring, trend-analysis and prediction of Air Pollution and Ozone Depletion. Additionally, it includes a personalized carbon footprint calculator that actively engages users and encourages informed, sustainable decision-making in their daily activities [17].

II. Background and Related Work

Environmental monitoring has evolved with the availability of continuous observations from sensors, satellites, and open environmental databases. However, the collected information is often distributed across different platforms and presented as isolated measurements. Understanding air pollution, ozone behaviour, and carbon emissions together remains difficult because these factors are influenced by multiple changing conditions, including human activities and atmospheric processes.
A number of researchers have applied computational models to study air quality patterns. Earlier approaches mainly relied on statistical and machine learning techniques to estimate pollutant concentrations from previous records. Al-gorithms such as Random Forest, Support Vector Regression, and other regression-based methods have been ex-plored for identifying relationships between pollution levels and environmental variables [1; 3; 5]. More recently, sequence-based deep learning models have been adopted for forecasting tasks where previous observations influ-ence future values. LSTM networks have shown suitability for this type of data because environmental measurements usually contain gradual changes and repeating trends. Several studies have also examined practical issues such as incomplete sensor readings, irregular observations, and the effect of meteorological conditions on prediction perfor-mance [4; 7].
Ozone-related environmental studies follow a broader atmospheric perspective, where long-term chemical interac-tions and climatic factors are considered. Existing research has reported that ozone concentration changes are not immediate processes and can be affected by substances that remain active in the atmosphere for many years [8; 9]. While such studies provide important scientific understanding, the information is usually available through technical platforms that are not designed for everyday users.
Most current solutions focus on solving one environmental problem at a time. Prediction models, ozone monitoring systems, and carbon emission calculators generally work independently, creating a disconnect between environmen-tal data and personal impact. The proposed system combines these components by integrating prediction models, carbon footprint estimation, and procedural visualization. Instead of presenting environmental conditions only through numerical outputs, the framework converts complex data into interactive visual forms, allowing users to explore pollution changes, ozone behaviour, and sustainability-related impacts through a single platform.

III. Proposed System

A. Data Sources

The monitoring, analysis and prediction of Air Pollution and Ozone Depletion is data-heavy as real-time data is collected at frequent time-intervals. Hence, data has been collected from multiple sources to ensure data diversity and reliability. Data from the following sources has been used:

- AQI CN - For Air Pollution Data [15]
- Copernicus Atmosphere Monitoring Service (CAMS) - For Air Pollution and Ozone Depletion Data [13]
- Central Pollution Control Board (CPCB) - For Air Pollution Data [14]
- Kaggle Datasets: Ozone Level Detection Dataset and Individual Carbon Footprint Calculation Dataset [16; 17]

After studying the parameters that majorly impact these environmental problems, the data was collected and stored as shown in the database schema and parameter tables (Tables 1, 2, and 3).

TABLE I: AIR POLLUTION DATASET SCHEMA

Parameter	Description	Unit
Timestamp	Date and time of observation	–
Location	Monitoring station or geographic region	–
PM2.5	Fine particulate matter concentration	µg/m3
PM10	Particulate matter concentration	µg/m3
NO2	Nitrogen dioxide level	ppb
SO2	Sulfur dioxide level	ppb
CO	Carbon monoxide concentration	ppm
AQI – Air Quality Index	Air Quality Index value	–

TABLE II: OZONE DEPLETION DATASET SCHEMA

Parameter	Description	Unit
Timestamp	Date and time of ozone observation	–
Location	Geographic region of measurement	–
Ozone Concentration	Total column ozone level	DU
UV Index Temperature	Ultraviolet radiation intensity Atmospheric temperature	– C
Pressure	Atmospheric pressure	hPa
Wind Speed	Air circulation speed	m/s

TABLE III: CARBON FOOTPRINT CALCULATOR INPUT PARAMETERS

Parameter	Description	Unit
Name	User name	–
Place	User location	–
Car Travel	Distance travelled by car per month	km/month
Electricity Consumption	Monthly electricity usage	kWh/month
AC Usage	Daily air conditioner usage	hours/day
Trees Planted	Number of trees planted	count
Vehicle Type	Type of vehicle used	–
Diet Type	Dietary pattern of the user	–
Estimated Carbon Emissions	Calculated carbon footprint	kg CO2/year

B. Prediction Model

To predict the future variations of Air Pollution and Ozone Depletion, traditionally, algorithms like decision trees and Random Forest were used. However, these algorithms often struggle to capture sequential relationships. Researches shows that LSTM - Long Short-Term Memory gives the best results with the prediction of environmental data. LSTM is a specialized type of RNN i.e. Recurrent Neural Network that works with time-series and sequential data that aligns well with environmental data. Hence, we use LSTM networks to predict the weekly and monthly statistics. In the proposed system, LSTM networks are trained on historic, sequential time-series data. These sequential inputs allow the model to identify the underlying patterns and relationships between environmental variables, to make accurate prediction of future air pollution levels and ozone concentration trends.

The operations of the LSTM cell can be represented as follows:

Forget Gate: Determines the information to be discarded

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (1)$$

Input Gate: Determines the new information to be stored in the cell state.

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (2)$$

Output Gate: Determines the part of the cell state to be passed as the output.

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (3)$$

Hidden State (Final Output):Calculated using the updated cell state and the output gate.

$$h_t = o_t \cdot \tanh(C_t) \quad (4)$$

where x_t represents the input vector at time step t, h_{t-1} represents the hidden state from the previous time step, and C_t represents the memory cell state. The function σ denotes the sigmoid activation function and tanh denotes the hyperbolic tangent activation function.

C. Carbon Footprint Model

The increasing concentration of greenhouse gases in the atmosphere has significantly contributed to global climate change. Carbon dioxide emissions generated through daily human activities such as transportation, electricity consumption, and diet habits play a major role in the overall carbon footprint of individuals [18]. Monitoring and calculating these emissions is important for promoting healthy practices and environmental awareness.

To assess and overcome this challenge, a Carbon Footprint Calculator was developed as part of Akashtattva. This calculator calculates the monthly carbon emissions of an individual based on several lifestyle parameters such as transportation usage, electricity consumption, air conditioning usage, diet habits, and tree plantation activities. The total carbon footprint is calculated by aggregating emissions from these activity sources as shown in Eq. 5 [18]. The calculator uses a machine learning Regression model to predict carbon emissions more accurately and stores user data in a PostgreSQL database for further analysis.

Carbon Emission Calculation Strategy:

The total carbon footprint is calculated using:

$$C_{total} = C_{vehicles} + C_{electricity} + C_{AC} + C_{diet} \quad (5)$$

Where:

- $C_{vehicles}$ represents emissions from transportation, calculated as $C_{vehicle}$ = Distance × Emission Factor$_{vehicle}$ (Emission factors obtained from US Environmental Protection Agency (EPA) guidelines) [18].
- $C_{electricity}$ represents emissions from electricity usage, calculated as $C_{electricity}$ = Electricity Consumption ×
- Emission Factor$_{electricity}$ (Emission factor based on Indian power grid emission intensity ≈ 0.82 kg CO_2/kWh) [18].
- C_{AC} represents emissions from air conditioning usage, calculated as C_{AC} = AC Hours × Power Rating × Emission Factor$_{electricity}$ (Emission factor derived from electricity grid emission intensity) [18].
- C_{diet} represents emissions from diet habits. Typical values are Vegetarian diet ≈ 1.5–2 kg CO_2/day and Non-vegetarian diet ≈ 3–5 kg CO_2/day (Values derived from studies on food production emissions) [18].
- Tree carbon offset is calculated as Carbon Offset = Trees Planted × Sequestration Rate (Sequestration rate ≈ 180 kg CO_2/tree/year in tropical conditions) [18].
- Net carbon footprint is calculated as Net Carbon = Gross Carbon − Tree Offset.

D. Procedural Visualization of Environmental Data

Data Visualization plays an important role for complex and large data like the Air Pollution and Ozone Depletion. Hence, to simplify and dynamically provide comparative data visualization procedurally generated art is used to show air pollution and ozone depletion around the globe. Procedurally Generated art creates visualizations dynam-ically and is widely used in coding randomized games. Here, it is used to create visualizations dynamically using the current real-time air pollution or ozone data and provide a birds-eye view to the user, as illustrated in Figure. 1 and Figure. 2. Further detailed visual analytics are provided through country-level and city-level dash-boards, as shown in Figure. 3 and Figure. 4. Analysis Additional analytical interfaces such as weekly air pollu-tion analysis and ozone prediction are presented in Figure. 5 and Figure. 6.

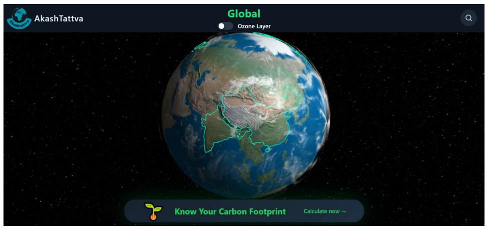


Figure. 1: Procedural Art-Air Pollution

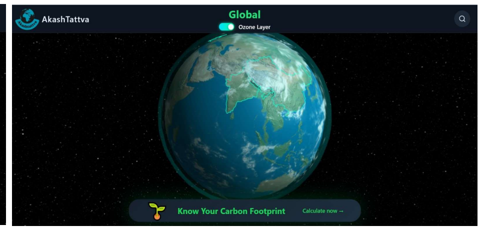


Figure. 2: Procedural Art - Ozone

Figure. 3: Country-level Analysis

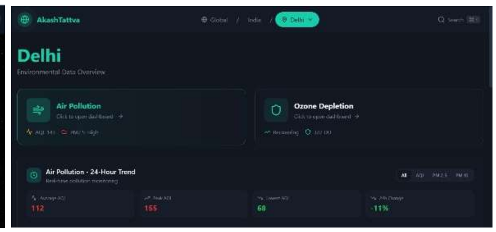


Figure. 4: City-wise Analysis

Figure. 5: Weekly air pollution analysis

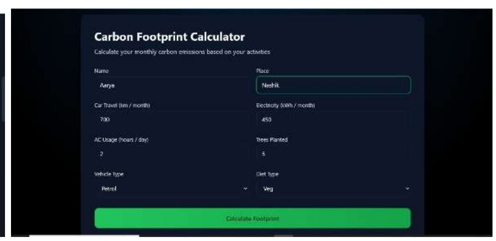


Figure. 6: Carbon- Foot-print Calculator

The homepage welcomes the user with a 3D interactive globe that can be panned, zoomed in and out or can search a location manually, as illustrated in Figure. 1. A toggle bar is provided to visualize the air pollution or ozone depletion around the globe. When the toggle bar is off, a comparative visualization of air pollution of different places around the world where density of cloud-like texture represents the pollution at that place, as depicted in Figure. 1. When the toggle bar is on, the visualization of ozone layer, ozone holes present and shown around the various loca-tion on the earth, as shown in Figure. 2. On clicking on a particular country, the map of that country with AQI of different cities marked with colours appears on the screen (Figure. 3). Further, on selecting a city a dual dashboard for air pollution and ozone is displayed (Figure. 4). The air pollution card had weekly analysis and prediction in form of graphs and marked with colours (Figure. 5). The Ozone card had a slide bar where the user can see the fu-ture prediction of ozone depletion or recovery followed by historic trend graph (Figure. 6). Lastly, the carbon footprint is an interactive module that takes input from user and gives personalized carbon footprints and suggestions for sustainable choices.

E. System Architecture

The proposed platform follows a modular approach with 3 main modules: Frontend, Backend and Data Module communicate with each other while delivering services to the end user. The architecture diagram shown in Figure. 7 depicts the working and engagement of the modules in the proposed work.

FRONTEND MODULE: The Frontend Module is responsible for and is a striking feature of the Application where the WebApp welcomes you with a 3D Interactive Globe with Air condition mapped region-wise.

- Interactive 3D Globe: The Interactive 3D Globe is made using Three.js and the base code is driven by tns-3d-globe which has an MIT License and is open source. Users can navigate various regions on the globe, zoom in and out, and visualize comparative pollution across regions.
- Air Pollution: Upon deep diving into a particular region, the user can visualize the air condition of that region with analysis of measures like Air Quality Index (AQI), PM2.5, PM10, and other pollutants.

- ➢ Ozone Depletion: Upon deep diving into the ozone depletion of a particular region, the UI displays a visualization of the current ozone layer with a slider. Moving the slider across future years provides predictions of the effects on the ozone layer backed by an LSTM machine learning model.
- ➢ Know Your Carbon Footprint: This is a standalone module that allows users to check the carbon footprint of activities they perform that impact the environment. It encourages users to reduce their carbon footprint by adopting activities that nurture nature.

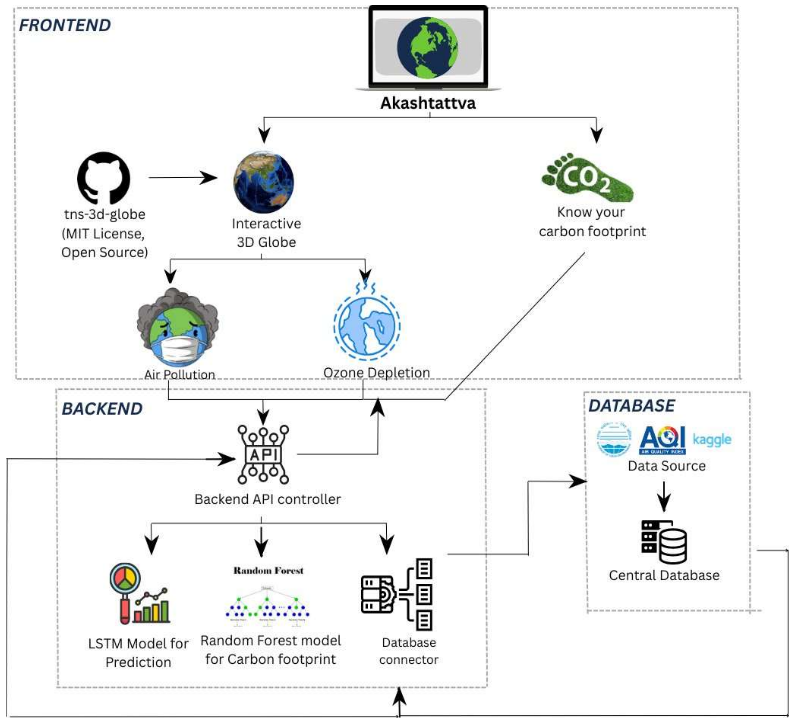


Figure. 7: Architecture Diagram

BACKEND MODULE: This module forms the heart of the system as it interacts with the Frontend Module as well as the Data Module. It is responsible for the processing and Logic of the entire system.

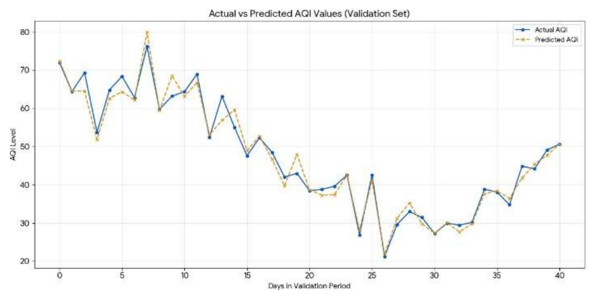
- ➢ Backend API controller: The Frontend Module interacts with the Backend Module through the Backend API. The Backend API receives the requests initiated at the frontend by the user. The tech stack Fast API is being used as the Backend API due to its nature to handle real-time and large data.
- ➢ LSTM Machine Learning Model: The LSTM Machine Learning Model predicts the Ozone Depletion based on past and current trends.
- ➢ Database Connector: The database connector is called when we need to fetch data from the Data Module. SQLAlchemy is used as the Database connector. It directly interacts with the Central Database to fetch the necessary data.

DATABASE MODULE: The database module is responsible for fetching data through API, storing it in a system-atic format in central database and managing orchestration of the workflow.

- ➢ Data Source: The data source fetches historic as well as real time data through its API. Data Sources including AQICN [15], CPCB [14], Copernicus [13] and Kaggle datasets [16; 17] have been used.
- ➢ Central Database: The Central Database stores data in the required format, all the data needed by the application is stored in Central Database, here PostgreSQL, and connected to Database connector to maintain flow of information from Backend to Fronted.

IV. Experimental Results and Discussion

The experimental results demonstrate that the model training for air pollution using LSTM, ozone depletion predic-tion using LSTM and Carbon footprint calculator using Random Forest algorithm. The performances of the models were evaluated using the standard evaluation metrics.



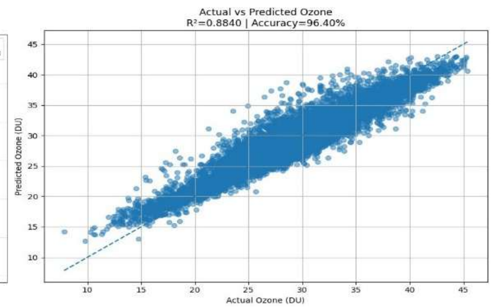


Figure. 8: Actual v/s Predicted air pollution Figure. 9: Ozone Prediction with accuracy (R2 = 0.8840, Accuracy = 96.40%)

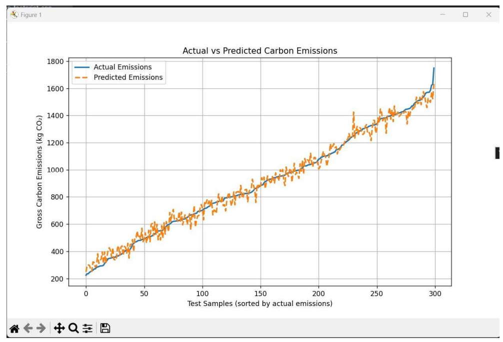


Figure 10. Actual v/s Predicted Carbon Footprint

The air pollution prediction model shows close alignment between the actual and predicted AQI values as illustrated in Fig. 8. The ozone depletion prediction model achieved an R2 score of 0.8840 with an overall prediction accuracy of 96.40% as shown in Fig. 9. The Carbon footprint calculator developed using Random Forest also demonstrates effective prediction performance as shown in Fig. 10. The air pollution prediction model shows close alignment be-tween the actual and predicted AQI values as illustrated in Fig. 8. The ozone depletion prediction model achieved an R2 score of 0.8840 with an overall prediction accuracy of 96.40% as shown in Fig. 9. The Carbon footprint calculator developed using Random Forest also demonstrates effective prediction performance as shown in Fig. 10. These find-ings demonstrate that the proposed framework can effectively support environmental monitoring, analysis, and predictive assessment.

V. Conclusion

This research presents Akashtattva, a platform developed to study environmental changes by bringing together mul-tiple aspects that are usually handled separately. The system collects and processes air quality and atmospheric in-formation, applies prediction models to identify future trends, and provides users with a way to understand how their daily activities contribute to carbon emissions. The implemented framework uses historical environmental observa-tions to estimate possible changes in air pollution and ozone concentration. The prediction component helps in ana-lysing patterns over time rather than only displaying current conditions. Along with this, the carbon footprint module allows individuals to calculate their estimated emissions based on factors such as transportation, electricity usage, and lifestyle choices. This creates a connection between large-scale environmental changes and personal actions. A major focus of the proposed platform is the way environmental information is presented. Instead of limiting users to numerical values and conventional charts, the system uses procedural visualization to convert environmental meas-urements into interactive representations. This makes complex atmospheric data easier to explore and supports better understanding of pollution distribution and ozone behaviour. The developed architecture also provides a structured approach for storing and managing environmental records, allowing future expansion with additional datasets and monitoring sources. The combination of data analysis, predictive modelling, visualization, and carbon assessment demonstrates the possibility of creating more engaging environmental awareness tools. Future improvements can include incorporating more regional data sources, refining prediction performance with larger datasets, and adding advanced visual simulations for long-term environmental scenarios.

References

[1] K. Chatterjee et al., “Future Air Quality Prediction Using LSTM Based on Hyper-Heuristic Multi-Chain Model,” IEEE Access, vol. 12, pp. 123678–123700, 2024.

[2] S. Zhang et al., “Revealing Influence of Meteorological Conditions on Air Quality Prediction Using Explainable Deep Learning,” Environmental Modelling and Software, vol. 180, pp. 105–118, 2024.

[3] A. Patel, S. Roy, and M. Jain, “A Comparative Analysis Study for Air Quality Prediction in Smart Cities Using Regression Techniques,” International Journal of Smart City Applications, vol. 5, no. 2, pp. 100–112, 2023.

[4] H. Kumar and T. Singh, “A Complete Air Pollution Monitoring and Prediction Framework,” IEEE Access, 2023.

[5] D. S. Gupta and R. Verma, “Air Quality Prediction Using LSTM and Hybrid Regression Techniques,” Environmental Informatics, vol. 19, no. 3, pp. 55–71, 2024.

[6] H. Chen, M. Guan and H. Li, ”Air Quality Prediction Based on Integrated Dual LSTM Model,” IEEE Access, vol. 9, pp. 93285–93297, 2021. doi: 10.1109/ACCESS.2021.3093136.

[7] A. Abalo-Garc´ıa et al., “MPD: A Meteorological and Pollution Dataset: A Comprehensive Study of Machine and Deep Learning Methods for Air Pollution Forecasting,” IEEE Access, vol. 13, pp. 41282–41310, 2025.

[8] R. Rigby and A. Smith, Atmospheric Burden of Ozone Depleting Substances and Forecasting Ozone Layer Recovery. WMO/UNEP, 2023.

[9] A. Tumwebaze, D. Twinomujuni, E. Baluku and R. Komakech, ”Ozone Layer Depletion and Its Impacts on the Environment and Public Health: A Review,” African Journal of Environment and Natural Science Research, vol. 8, pp. 83–94, 2025. doi: 10.52589/AJENSR-9P6IXPAP.

[10] A. E. M. van den Oever, S. Puricelli, D. Costa, N. Thonemann, M. Lavigne Philippot and M. Messagie, ”Revisiting the Challenges of Ozone Depletion in Life Cycle Assessment,” Cleaner Environmental Systems, vol. 13, Article 100196, 2024. doi: 10.1016/j.cesys.2024.100196.

[11] M. -K. Kim, T. -A. Yoo and J. -B. Chung, ”Toward Sustainable AI: A Scoping Review of Carbon Footprint and Environmental Impacts Across Training and Inference Stages,” IEEE Access, vol. 14, pp. 18881–18896, 2026. doi: 10.1109/ACCESS.2026.xxxxx.

[12] R. C. E. Silva et al., “Procedural Generation of 3D Maps With Snappable Meshes,” IEEE Access, vol. 10, pp. 43093–43115, 2022.

[13] Copernicus Atmosphere Monitoring Service (CAMS), ”Atmosphere Data Store (ADS),” European Centre for Medium-Range Weather Forecasts (ECMWF). Available: https://atmosphere.copernicus.eu/data. Accessed: Mar. 13, 2026.

[14] Central Pollution Control Board (CPCB), ”National Air Quality Index (AQI) Data Portal,” Ministry of Environment, Forest and Climate Change, Government of India. Available: https://airquality.cpcb.gov.in/AQI India/. Accessed: Mar. 13, 2026.

[15] World Air Quality Index Project, ”Air Quality Programmatic API,” Available: https://aqicn.org/api/. Accessed: Mar. 13, 2026.

[16] Prashant111, ”Ozone Level Detection Dataset,” Kaggle Dataset. Available: https://www.kaggle.com/datasets/prashant111/ozone-level-detection. Accessed: Mar. 13, 2026.

[17] M. Duman, ”Individual Carbon Footprint Calculation Dataset,” Kaggle Dataset. Available: https://www.kaggle.com/datasets/dumanmesut/individual-carbon-footprint-calculation. Accessed: Mar. 13, 2026. IPCC, ”2006 IPCC Guidelines for National Greenhouse Gas Inventories,” Intergovernmental Panel on Climate Change. Available: https://www.ipcc-nggip.iges.or.jp/public/2006gl/. Accessed: Mar. 13, 2025.