SMART INDIA HACKATHON 2025

SMART INDIA HACKATHON 2025

TITLE PAGE

- Problem Statement ID – SIH25178
- Problem Statement Title- Short term forecast of gaseous air pollutants (ground-level O3 and NO2) using satellite and reanalysis data.
- Theme- Space Technology
- PS Category- Software
- Team ID- 117392
- Team Name- VyomCoders

Vyom Coders

Short term forecast of gaseous air pollutants (ground-level O3 and NO2) using satellite and reanalysis data.

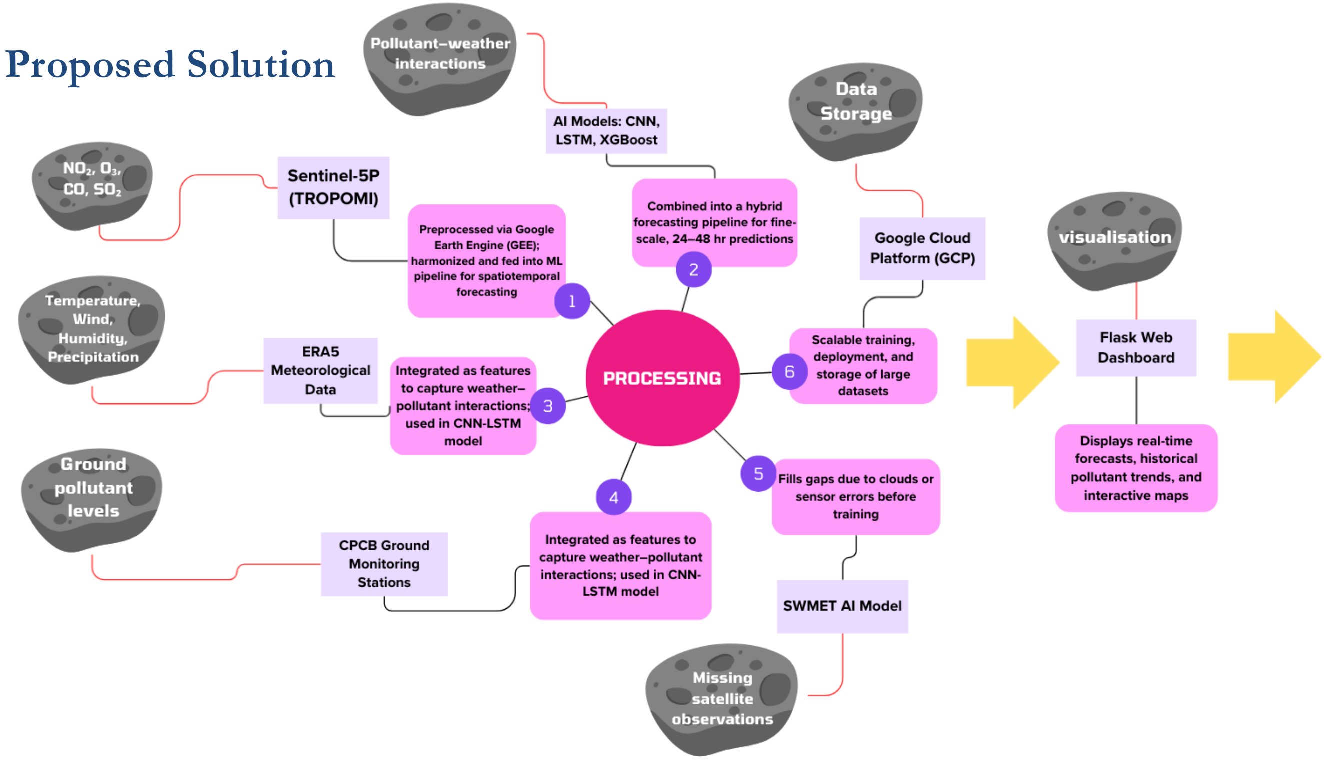
SMART INDIA HACKATHON 2025

An **advanced machine learning framework** for short-term forecasting of ground-level O_3 and NO_2 concentrations in Delhi. The solution leverages **satellite data and meteorological reanalysis** to build accurate 24-hour ahead predictions at hourly intervals. We developed an **advanced machine learning framework** for short-term forecasting of ground-level O_3 and NO_2 concentrations in Delhi. The solution leverages **satellite data and meteorological reanalysis** to build accurate 24-hour ahead predictions at hourly intervals.

Current systems Limitations:

- Satellite data cover large areas but miss local urban details and rapid changes.
- Available systems like MODIS AOD provides proper forecasts but lacks accuracy due to challenges like cloud and low sensitivity to surface pollution.
- Most models alone can't capture complex interactions between weather and multiple pollutants at fine scales.

Proposed Solution



OUTPUT

- Fine-scale (city-level) O_3 / NO_2 24–48 hr forecasts
- High-resolution spatiotemporal pollution maps
- Real-time air quality monitoring
- Historical pollutant trends visualization
- Alerts for unsafe pollutant levels
- Scalable national-level forecasting

Vyom Coders

TECHNICAL APPROACH

SMART INDIA HACKATHON 2025

1. Data Sources:

Sentinel-5P (TROPOMI): Satellite-based measurements of NO_2, O_3, CO, and SO_2. Daily high-resolution data (~3.5 × 5.5 km) for forecasting air quality.

ERA5 Meteorological Data: Temperature, wind, humidity, and precipitation to capture weather effects on pollutants.

Ground Stations (CPCB, optional): Real observed data to validate predictions.

2. Platforms:

Google Earth Engine (GEE): Access, preprocess, and analyze satellite data.

Google Cloud Platform (GCP): Cloud storage and scalable compute for training and deployment.

3. AI Models:

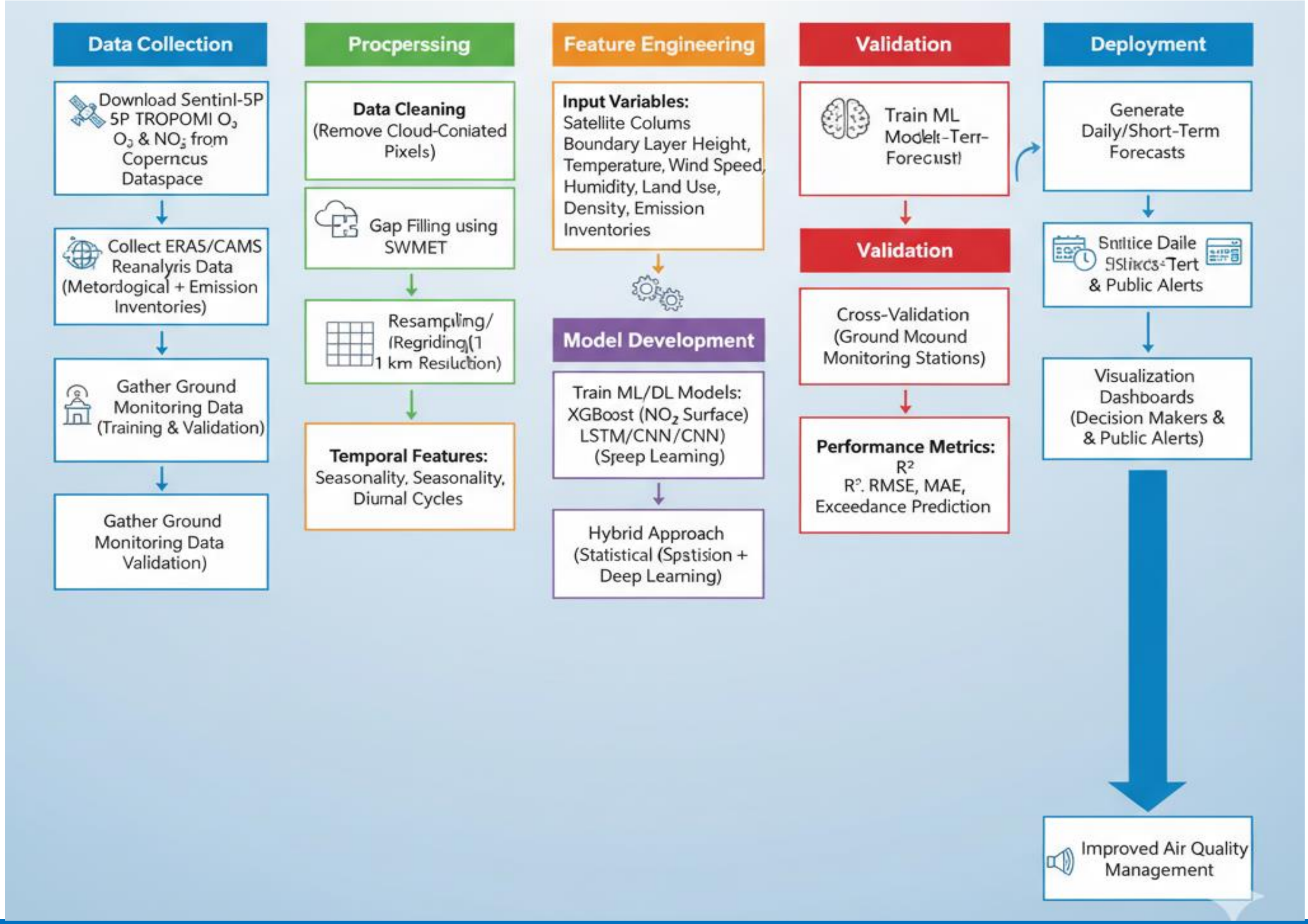
XGBoost: Regression model for baseline prediction and feature importance.

LSTM: Captures temporal patterns in pollutant levels.

CNN: Extracts spatial features from gridded data.

4. Tools

METHODOLOGY AND PROCESS OF AIR QUALITY FORECASTING IMPLEMENTATION



Vyom Coders

FEASIBILITY AND VIABILITY

SMART INDIA HACKATHON 2025

What makes this solution out of the Box:

- Seamless Multimodal Fusion: Integrates TROPOMI satellite data, ECMWF forecasts, and ground stations for unparalleled spatial-temporal coverage.
- Advanced Deep Learning: Employs CNN-LSTM with attention mechanisms to capture nonlinear pollutant–meteorology interactions at urban scales.
- Real-Time Operational Design: Cloud-native architecture and optimized inference enable sub-hourly, city-wide forecasts for immediate public health and policy action.
- Human resource: Along with the smart tech developers, on site industry experts have been involved in the development and research of this project.

Government backing through National Clean Air Programme and Digital India initiatives.

Societal relevance: These systems provide actionable forecasts that help protect public health, support planning, and guide farm and urban policies.

Proven operational models in India (WRF-Chem, SAFAR, AQ-EWS) demonstrate technical capability and institutional expertise at IITM Pune.

Technically feasible: Proven models (like WRF-Chem, AQ-EWS, SAFAR) already operate in Indian cities; expertise exists at institutes like IITM Pune.

Feasibility

Scalable to multiple Indian cities—requires investment in 3,000–5,000 additional monitoring stations and regional HPC centers.

Scalable: Can be expanded to other cities , but needs investment in monitoring networks and high-performance computing for model runs.

Vyom Coders

IMPACT AND BENEFITS

SMART INDIA HACKATHON 2025

Our approach stands out because:

Current available system
1.SAFAR
2.AQ-EWS
3.AeroVision

SAFAR: 10km regional resolution, limited to 4 cities
AQ-EWS: 400m only for Delhi-NCR, no national coverage
AeroVision: 25km grid resolution

Our Proposed Solution Advantage:

- TROPOMI satellite data at 3.5×7km resolution combined with ERA5 at 31km provides consistent high-resolution national coverage
- Unified framework scalable across all Indian megacities simultaneously
- Hourly forecasting at sub-urban scale vs. current daily/3-day systems

Benefits to the users

AI/ML Integrity:

1.CNN-LSTM hybrid architecture capturing complex spatiotemporal relationships vs. traditional regression.
2. Attention mechanisms for automatic feature importance weighting.
3.Multi-task learning for simultaneous O_3/NO_2 prediction with shared meteorological drivers

Data Integration Deficiencies

1.TROPOMI provides daily global coverage with superior sensitivity to NO_2/O_3 compared to MODIS
2.Real-time meteorological fusion using ECMWF operational forecasts (1-hour latency) vs. ERA5's 5-day delay
3.Ground-satellite-model triangulation compensating for individual data source limitations

Accuracy and Performance Issues:

1.Literature demonstrates R^2 >0.84 for satellite-based CNN-LSTM models
2.RMSE improvements of 20-40% over traditional CTM approaches
3.Multi-pollutant validation with European CAMS achieving >0.9 correlation after post-processing

Vyom Coders

RESEARCH AND REFERENCES

SMART INDIA HACKATHON 2025

- **Wang, Y., et al. (2021).** *Short-term forecasting of air pollutants: A comprehensive review of data sources, methods, and applications. Science of the Total Environment, 775,* 145944. Elsevier.
  - **Why used:** This paper provides a detailed review of different approaches and data sources for short-term air pollutant forecasting, which supports the methodology section of the presentation. ScienceDirect Link
- **Giani, P., Anav, A., De Marco, A., & Sicard, P. (2022).** *Air pollutant forecasting: Short-term prediction of O_3 and NO_2 using satellite and reanalysis data. Frontiers in Environmental Science.*
  - **Why used:** This research focuses specifically on forecasting ozone (O_3) and nitrogen dioxide (NO_2) using satellite and reanalysis datasets, which directly aligns with the pollutants studied in this project. PMC Link
- **Zhang, Y., et al. (2020).** *Deep Learning for Air Quality Forecasts: A Review. Current Pollution Reports, 6,* 399–415.
  - **Why used:** This review highlights how deep learning techniques (CNN, RNN, LSTM) are applied for air quality prediction, supporting the technical framework and AI-based methodology in the PPT. ResearchGate Link
- **European Centre for Medium-Range Weather Forecasts (ECMWF). (2025).** *ERA5: Fifth Generation of ECMWF Atmospheric Reanalyses of the Global Climate*
  - **Why used:** ERA5 reanalysis dataset provides global-scale historical weather and atmospheric data, which is essential as an input dataset for pollutant forecasting models. ERA5 Reference

Comparative Insight:

Unlike traditional models (**ARIMA**, **CTMs** like **SAFAR**, or ensemble systems such as **CAMS**), our solution directly learns from **satellite** and **reanalysis data**—eliminating dependence on emission inventories or dense sensor networks. This enables faster, scalable, and region-specific air quality forecasting with minimal computational cost.