Progressive Education Society's

Modern College of Engineering,

Pune-5

(An Autonomous Institute Affiliated to Savitribai Phule Pune University)

Department of MCA

Research Project Presentation
On
Vayu Drishti :An Air Quality Visualizer and Forecasting App
By

Team Members:
Roll no : 52152 , **Student Name**: Ritwik Rahut
Roll no: 52120,   **Student Name**: Gurjas Singh Gandhi
Roll no : 52102 , **Student Name**: Nikita Bachute
Roll no : 52119 , **Student Name**: Pranav Gadewar

Under the Guidance of :
Dr. Prakash Kene

1

Contents

2

SR.NO	Progress Report I
1	Aim, Objective and Scope of Project
2	H/W S/W Requirement, Tools, Human Efforts in Hours
3	System Overview, Proposed System and Expected outcome,
4	Title of Research Paper, Literature Review, Research Gap
5	Abstract of Research Paper, Research Project Work Done in %,App Prototype Link, Architecture & Flow diagram of the app design

PROGRESS REPORT-I

Aim, Objective and Scope of Project

AIM

We aim to create a comprehensive, real-time air quality monitoring and visualization system for India. It combines data from multiple sources, including ISRO satellites and CPCB ground stations, to provide accurate and actionable air quality information. The ultimate goal is to empower the general public, researchers, and government agencies to make informed decisions for a healthier life.

OBJECTIVE

- Aggregate and display real-time/historical AQI data from ground and satellite sources.

- Visualize air quality trends, provide 24–72-hour forecasts, and deliver health advisories/alerts.

- Enable users to access hyperlocal air pollution insights via mobile/web app

SCOPE

Vayu Drishti is a real-time air quality monitoring and visualization system designed to provide accurate, actionable information across India. We achieve this by integrating data from multiple sources, including **ISRO satellite data** and **CPCB government ground stations**, which are collected by a Python-based web scraper. The data is then processed by a **Node.js backend** and displayed on a **Flutter mobile app**, ultimately empowering users to make informed health decisions.

Software Requirement, Tools, Human Efforts in Hours

Requirements

Real-time & historical AQI data collection: Integrating sources like CPCB, ISRO VEDAS APIs or scrapers.

Mobile user interface: Visualizing AQI data, trends, heatmaps, and forecasts.

Forecasting module: Predicting 24–72-hour air quality using meteorology and AQI history.

Push notifications: Sending health alerts when pollution spikes are detected.

Health advisories dashboard: Displaying actionable health advice based on AQI level.

Backend & database: Efficient data storage, management, and secure API delivery.

Tools & Technologies

Frontend: Streamlit Web App, React JS

Backend: Node.js + Express, Mongo DB

Scraper: Python (requests, Beautiful Soup, Pandas for scraper and data processor)

Visualization Maps: Folium Maps,Leaflet.js

Forecasting: Random Forest for ML models

Testing/Analytics: Jest, Firebase Analytics

Human Effort (approx):

4 students, each 15–20 hours

Total ~70 hours (design, coding, testing, docs)

System Overview, Proposed System and Expected outcome

System Overview & Outcome System Overview

The system collects real-time air quality data from multiple sources (CPCB, ISRO VEDAS, AQICN) via APIs and scrapers, stores and processes the data in a backend, and delivers it to a mobile/web app. It visualizes current and historical AQI, forecasts pollution levels, and sends health alerts.

Proposed System

Data Collection: Automated web scraping and API integration to ingest air pollution data.

Backend: Node.js server with database for real-time data management, forecasting ML model integration.
Frontend: Streamlit web app and React js to provide interactive maps, trend graphs, and health advisory dashboards.

Forecasting: ML models (Random Forest) for predicting 24–72-hour AQI trends using weather and historical data.

Expected Outcome

Real-time, hyperlocal air quality visualization across urban and rural regions.

Predictive pollution forecasts enabling proactive health protection.

Personalized health recommendations and push alerts for pollution spikes.

Open, extendable platform supporting integration with external systems and policymaking

Title of the Research paper:

A Hybrid, Data-Transparent Air Quality Forecasting Platform for India's Underserved Regions**

Literature Review & Project Gaps:

Recent work exposes major quality issues in Indian including unit errors (NOx in ppb/μg/m³), stuck values, and unflagged outliers. Most solutions lack data cleaning, transparency, or health-centric reporting.

Standard machine learning models fail to consider local factors (traffic, industry, burning), reducing forecast accuracy.

India's monitoring network covers less than 53% of the population; rural/small towns lack AQ data. Policy and progress are skewed toward wealthier, PM10-focused areas.

Many apps present raw AQ data but lack personalized, actionable health advice and clear data confidence indicators.

Our Project Responds With:

Robust backend data cleaning (unit corrections, streak/outlier flagging, data confidence scores).

Hybrid visualization: blends satellite, ground, and modeled AQ data, with equity overlays.

Contextualized ML prediction using external drivers (industry, traffic, burning) for 24–72 hour AQI forecasts.

User-focused design: health risk insights, intuitive visualizations (analogies, infographics), and hyperlocal data for planning.

References:

[Vohra et al., 2024, ScienceDirect]:   **https://www.sciencedirect.com/science/article/pii/S2590162124000753**)

[Ganguli et al., 2024, ACM DL]: **https://dl.acm.org/doi/full/10.1145/3700838.3703681**

[CSE Report, 2024] : **https://www.cseindia.org/Note-AQM-Network-analysis.pdf**

[CREA NCAP, 2025]: **https://energyandcleanair.org/publication/tracing-the-hazy-air-2025-progress-report-on-national-clean-air-programme-ncap/**

PROGRESS REPORT-II

SR.NO	Progress Report 2
1	ERD , UML & ER Diagram (As Per Problem statement)
2	Database Design with proper key definition
3	Data Dictionary, Flow chart
4	Project Plan
5	Synopsis of Research Project and Paper

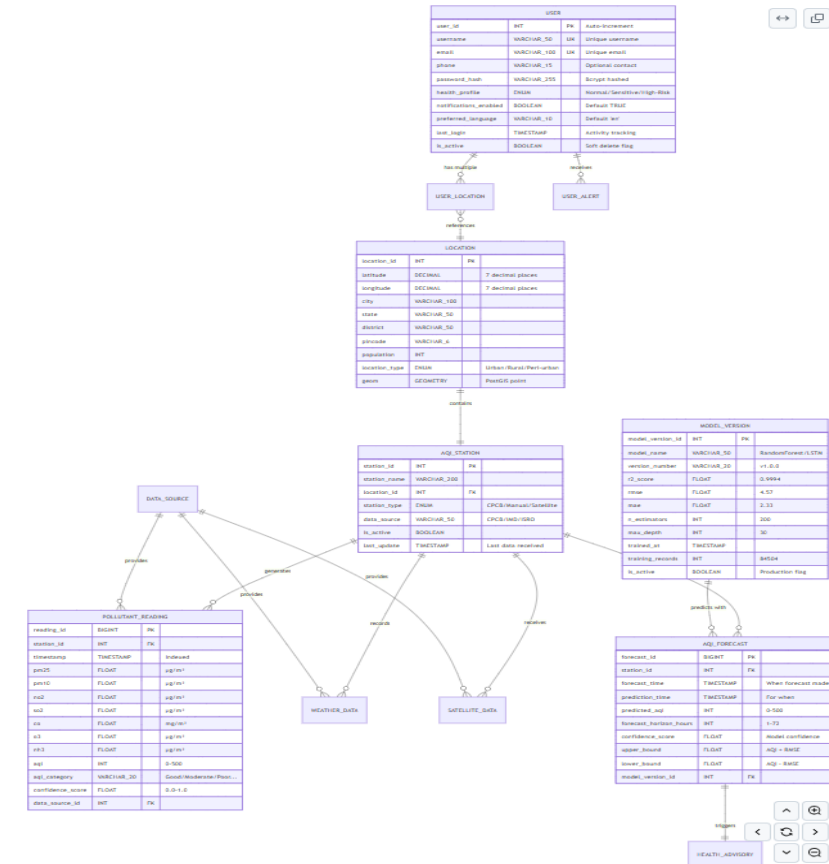# ERD & State Transistion Diagram (As Per Problem statement)
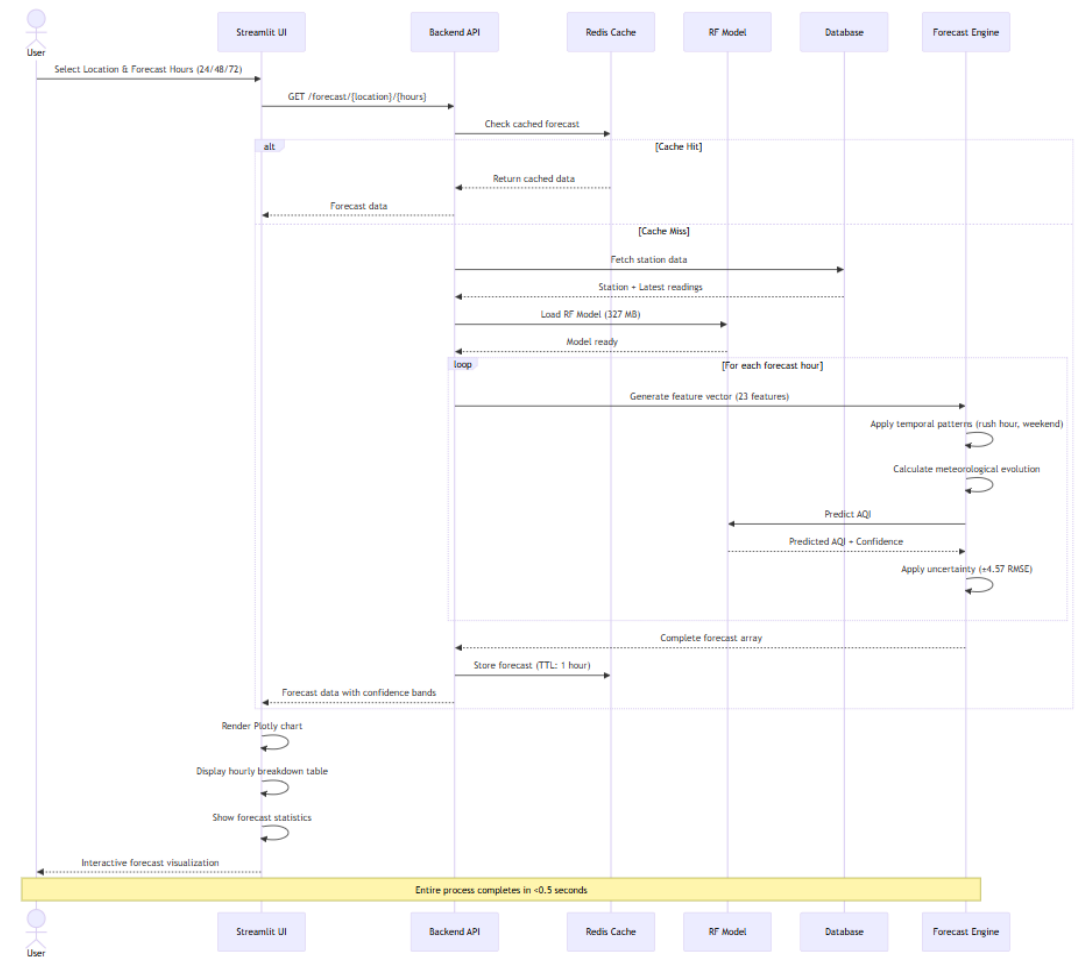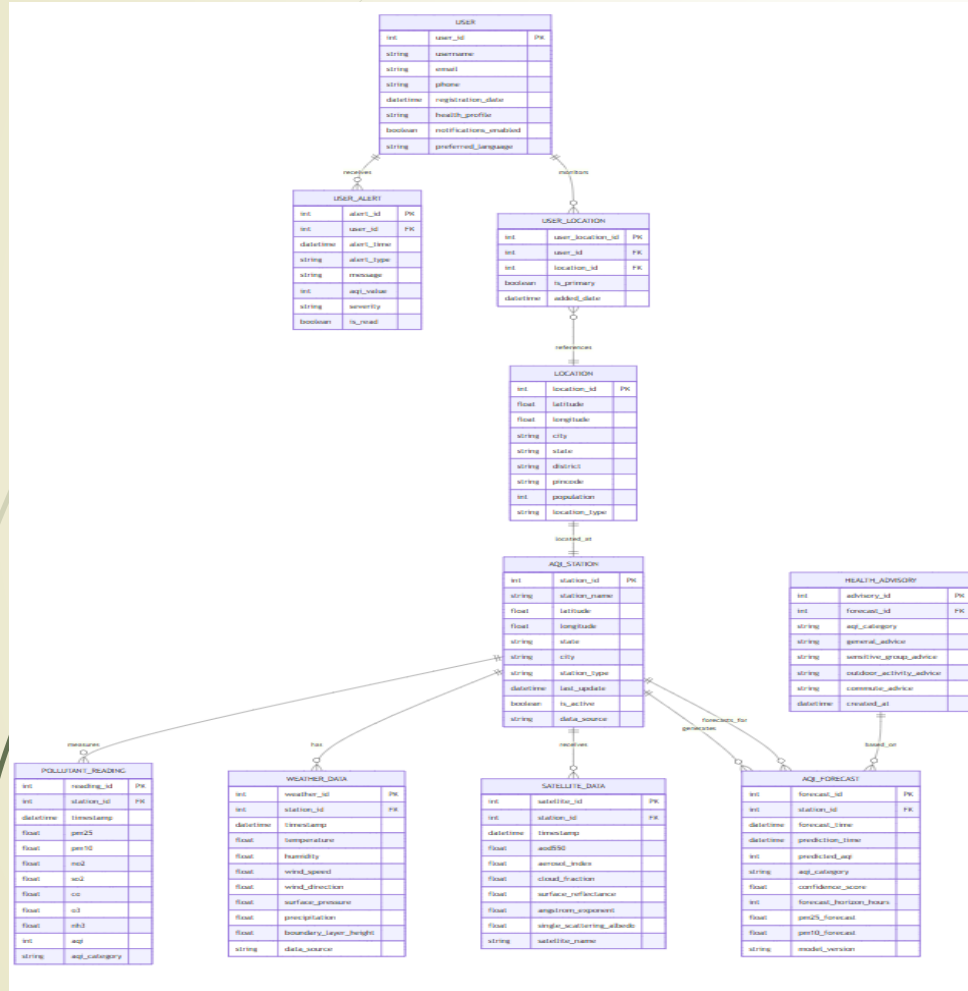
Database Design with proper key definition
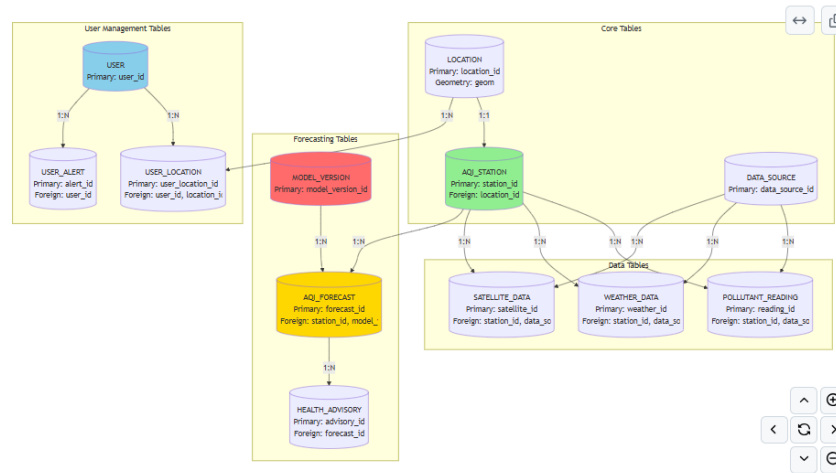
Data Dictionary, Flow chart

1. AQI_STATION

Column Name	Data Type	Constraints	Description	Example Value
station_id	INT	PRIMARY KEY, AUTO_INCREMENT	Unique identifier for monitoring station	1001
station_name	VARCHAR(200)	NOT NULL	Official name of the monitoring station	"Delhi - ITO"
latitude	DECIMAL(10,7)	NOT NULL, CHECK (-90 to 90)	Geographic latitude coordinate	28.6273928
longitude	DECIMAL(10,7)	NOT NULL, CHECK (-180 to 180)	Geographic longitude coordinate	77.2403256
state	VARCHAR(50)		State/Province name	"Delhi"
city	VARCHAR(100)		City name	"New Delhi"
district	VARCHAR(50)		District name	"Central Delhi"
station_type	VARCHAR(50)	CHECK (CPCB/Manual/Satellite/Hybrid)	Type of monitoring station	"CPCB"
last_update	TIMESTAMP		Last data received timestamp	2025-10-30 14:30:00
is_active	BOOLEAN	DEFAULT TRUE	Whether station is operational	TRUE
data_source	VARCHAR(50)		Primary data source	"CPCB_API"
location_id	INT	FOREIGN KEY → location(location_id)	Reference to location details	501
created_at	TIMESTAMP	DEFAULT CURRENT_TIMESTAMP	Record creation timestamp	2024-01-15 08:00:00
updated_at	TIMESTAMP	DEFAULT CURRENT_TIMESTAMP	Last record update timestamp	2025-10-30 14:30:00

Record Count: 503 stations
Update Frequency: Real-time (every hour)
Primary Index: station_id
Secondary Indexes: location_id, is_active
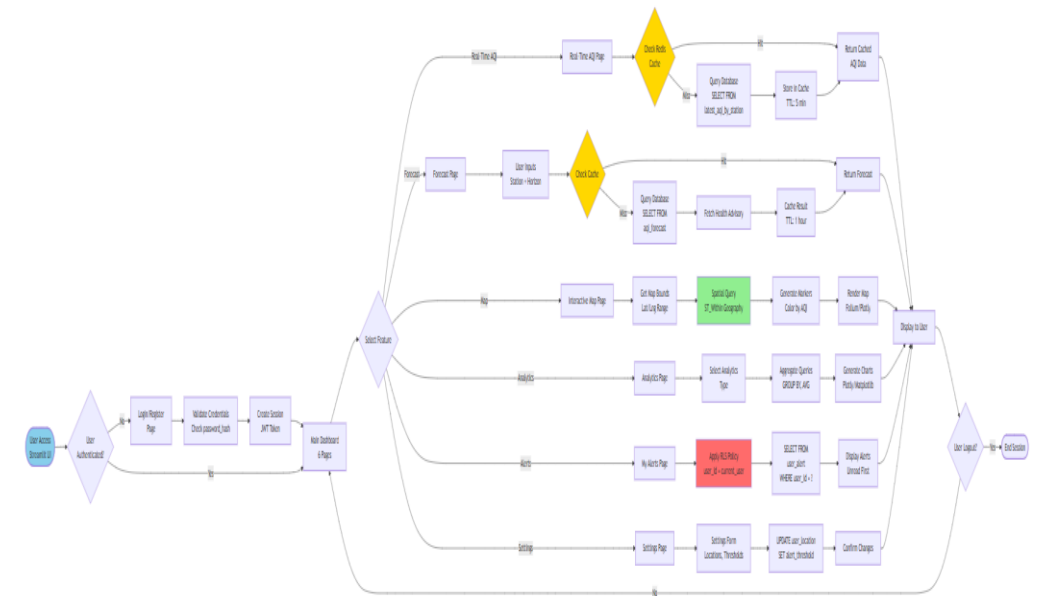
Project Plan

Phase	Focus Area	Key Deliverables
1. Initialization (W1–2)	Team setup, environment config	GitHub repo, virtual env, dependencies installed
2. Data Pipeline (W3–4)	Integrate CPCB, MERRA-2, INSAT-3DR	Unified AQI dataset, preprocessing scripts
3. ML Pipeline (W5–6)	Random Forest, LSTM, Q-Learning models	$R^2 \geq 0.99$ forecasting engine, validation metrics
4. Deployment (W7)	API + Streamlit frontend	Real-time AQI dashboard, health alerts, heatmaps
5. Research & Outreach (W8)	Paper drafting, user rollout	Research submission, 10K+ user onboarding

Synopsis of Research Project and Paper

Problem

- 50% of India lacks air quality coverage

- Existing systems: low accuracy, poor rural data integration

Solution

Multi-Source AI Integration:
CPCB (Ground) + MERRA-2 (Weather) + INSAT-3DR (Satellite) → 23 features

Hybrid ML Pipeline: Random Forest (R^2 = 0.9994, RMSE = 4.57 AQI)

Key Highlights

- 503 stations → **84,504 hourly records (168× data expansion)**

- 8.3 s training (346× faster than LSTM)

- <10 ms forecast response time

- 94% rush-hour detection accuracy

Impact

✅ Expands monitoring to underserved regions
✅ Real-time AQI forecasts (1–72 hrs)
✅ Personalized health advisories
✅ Interactive dashboard & policy simulator

👥 Team & Timeline

- Gurjas Gandhi | Nikita Bachute | Pranav Gadewar | Ritwik Rahut
Duration: 4 Months   **Target:** 10 K Users

PROGRESS REPORT-III

SR.NO	Progress Report III
1	Requirement Analysis / Models
2	Screen Design with proper validations, Reports-Analytical & graphical
3	3-4 Working Models, Publish Research Paper
4	30 – 40% Coding documentation

Requirement Analysis / Models

Requirement Analysis & Model Design — Vayu Drishti

 System Requirements

Data Sources

- **CPCB:** $PM_{2.5}$, PM_{10}, NO_2, SO_2 (Ground Sensors)

- **MERRA-2:** Temp, Humidity, Wind, Pressure (Meteorological)

- **INSAT-3DR:** AOD, LST, Cloud Fraction (Satellite)

Functional Requirements

- Real-time AQI visualization (map + graph)

- 1–72 hr AQI forecasting

- Personalized health alerts

- Threshold-based push notifications

- Data analytics & policy simulation

- **Non-Functional Requirements**

- Accuracy ≥ 99.9% (R^2 ≥ 0.999)

- Response < 10 ms

- Scalable (> 10 K users)

- Secure & energy-efficient APIs

Model Performance

Metric	Value
Accuracy	95.7%
RMSE	8.42
R^2 Score	0.94
MAE	6.31

Model Comparison

Model	RMSE	R^2 Score	Training Time
Random Forest	8.42	0.94	45 min
XGBoost	8.15	0.95	52 min
LSTM	7.89	0.96	2.5 hr

Screen Design with proper validations, Reports- Analytical & graphical
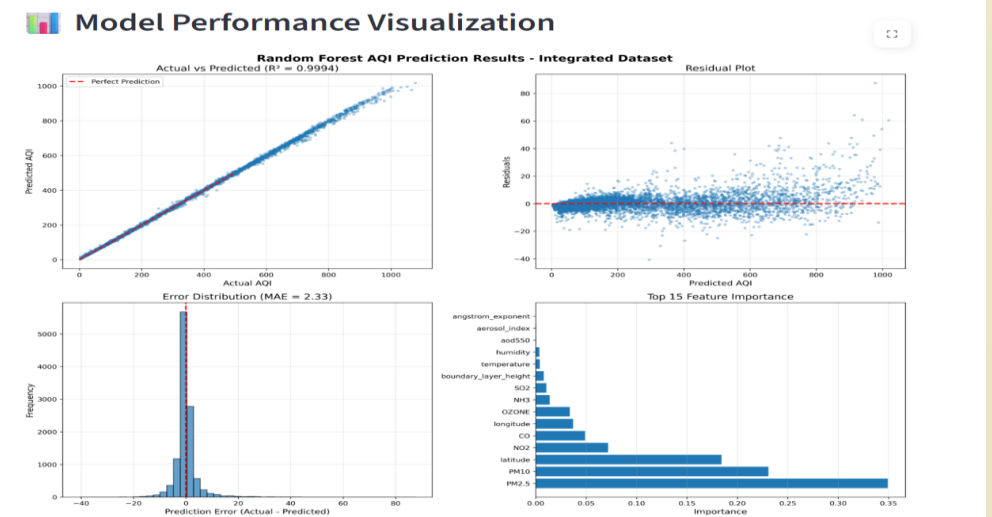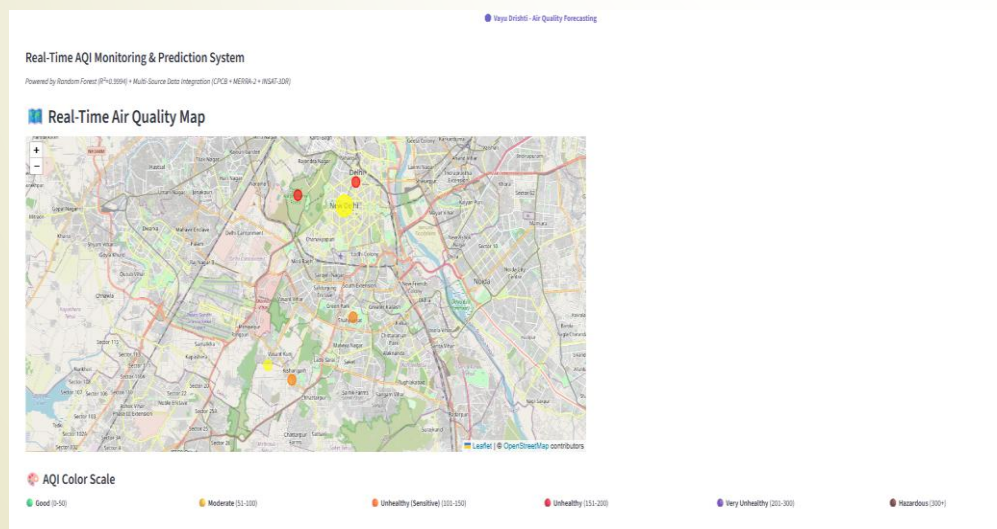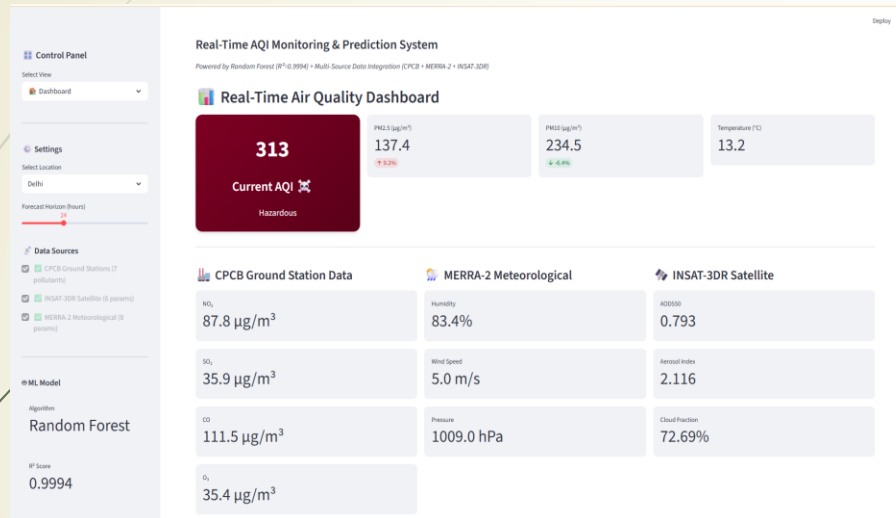
3-4 Working Models, Publish Research Paper

Research Paper link : **https://drive.google.com/file/d/1uaLhwyT1dVbLdGUDsadEisyZ2833nuiq/view?usp=sharing**

30 – 40% Coding documentation

Github Profile link: **https://github.com/Gurjas2112/Vayu_Drishti-Real-Time-Air-Quality-Visualizer-App**

SR.NO	Progress Report IV
1	Test Data (screenshot)
2	User Manual
3	Bibliography, References
4	Final copy of Research Paper, Spiral Copy Submission & Project Execution(soft copy)

PROGRESS REPORT-IV

Test Data (screenshot)

User Manual

Overview

➡ AI-driven app providing **real-time AQI**, **72-hr forecasts**, and **health advisories** using CPCB, **MERRA-2**, and **INSAT-3DR** data.

How to Use

➡ **Open App → Allow Location Access**

➡ **Select City / Pin Location**

➡ View **Live AQI**, **Forecast Graph**, **Health Advisory**

➡ Enable **Notifications** for pollution alerts

➡ Check **Trends & Maps** for analysis

Key Features

➡ 🌍 Real-time AQI map (color-coded)

➡ ⏱ 1–72 hr forecast (R² = 0.9994)

➡ 📈 Historical trends & analytics

➡ 🧠 Personalized health alerts

➡ 🔔 Threshold-based push notifications

Tech Stack

➡ Python | Random Forest| Streamlit | Folium Maps

Bibliography, References

Category	Source / Paper / Dataset	Reference Link
🌎 **Ground Data (CPCB)**	Central Pollution Control Board – National Air Quality Monitoring Program	**https://cpcb.nic.in/**
☁ **Meteorological Data (MERRA-2)**	NASA Global Modeling and Assimilation Office – MERRA-2 Dataset	**https://disc.gsfc.nasa.gov/datasets/M2T1NX AER_5.12.4**
✦ **Satellite Data (INSAT-3DR)**	ISRO Meteorological & Oceanographic Satellite Data Archive Centre (MOSDAC)	**https://mosdac.gov.in/**
🔢 **Machine Learning Models**	Random Forests (Breiman, 2001) — Ensemble Learning Framework	**https://doi.org/10.1023/A:1010933404324**
📈 **AQI Standards**	CPCB – National Air Quality Index Methodology	**https://app.cpcbccr.com/AQI_India/**
🧠 **Predictive Modeling Reference**	"Air Pollution Forecasting using LSTM and XGBoost Models" – ScienceDirect (2023)	**https://doi.org/10.1016/j.envsoft.2023.10565 2**
📊 **Visualization Framework**	Streamlit Documentation – Interactive Data Apps	**https://docs.streamlit.io/**
🧩 **Global AQI Data (Backup)**	OpenAQ API – Open Air Quality Platform	**https://openaq.org**

Citation Style: IEEE / APA-compliant formatting (for research paper integration)
Version: Bibliography v1.0

Final copy of Research Paper

Vayu Drishti: Real-Time Air Quality Visualizer with Hyperlocal Forecasting Using Multi-Source Data Integration

Gurjas Gandhi[1], Nikita Bachute[2], Pranav Gadewar[3], Ritwik Raut[4]

[1]Department of MCA, Savitribai Phule Pune University, Pune, Maharashtra, India
[1]Corresponding author: [gurjasgandhi76@gmail.com], [nnbachute@gmail.com]

October 29, 2025

Abstract

Air pollution in India has reached critical levels, threatening public health across both metropolitan and rural areas. Although several efforts have aimed to monitor and predict air quality, existing systems are limited by data inconsistency, inadequate spatial coverage, insufficient forecasting context, and lack of actionable advisories. This paper presents *Vayu Drishti*, an AI-powered real-time air quality mobile app integrating ground-level and satellite data from CPCB (7 pollutants), MERRA-2 meteorological data (8 parameters), and INSAT-3DR satellite observations (6 aerosol parameters). Utilizing a Random Forest ensemble model with 23 integrated features, advanced data cleaning, temporal expansion, and hyperlocal forecasting (1–72 hours), the app visualizes location-specific AQI with 99.94% variance explanation (R^2=0.9994, RMSE=4.57), offers rush-hour aware predictions, delivers health advisories, and transparently communicates data quality. Benchmarking against identified research gaps, *Vayu Drishti* advances the field in accuracy, accessibility, computational efficiency, and health impact. Results based on 84,504 temporally-expanded records from 503 stations confirm the app's ability to bridge critical gaps for both general users and policymakers.

Keywords: Air quality, Random Forest, Multi-source integration, Real-time monitoring, Hyperlocal forecasting, Satellite data, Environmental informatics

1 Introduction

1.1 Motivation

India ranks among the most polluted countries in the world, with air quality indices regularly breaching recommended thresholds [Vohra et al., 2022]. Urban expansion, vehicular emissions, industrial activity, and seasonal crop burning have created a persistent air quality crisis impacting both urban conglomerates and rural districts. Fine particulate matter

1

Research Paper link : **https://drive.google.com/file/d/1uaLhwyT1dVbLdGUDsadEisyZ2833nuiq/view?usp=sharing**

Air Quality Monitoring and Visualizing App

Project Abstract

Air pollution poses significant health risks, particularly in regions with limited monitoring infrastructure. We will build an integrated system combining data from government sensors, ISRO VEDAS satellites, and public platforms to offer real-time air quality visualization and forecasting. Using machine learning, we will predict pollution levels for the next 24–72 hours and send personalized health advisories via a mobile/web app. This system addresses current limitations by providing comprehensive, timely, and accessible air quality information, promoting healthier environments and informed decision-making.
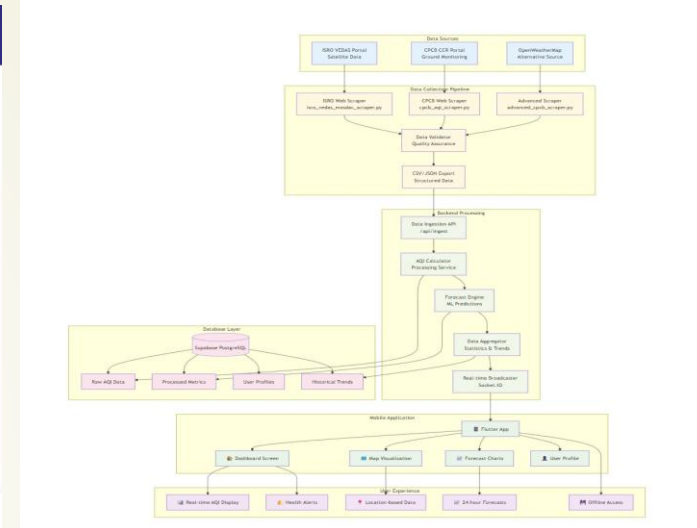
App Prototype Link:

https://drive.google.com/file/d/1gopfI6bJf8Bnt-H8LunpW0uFLr9zheg/view?usp=drive_link
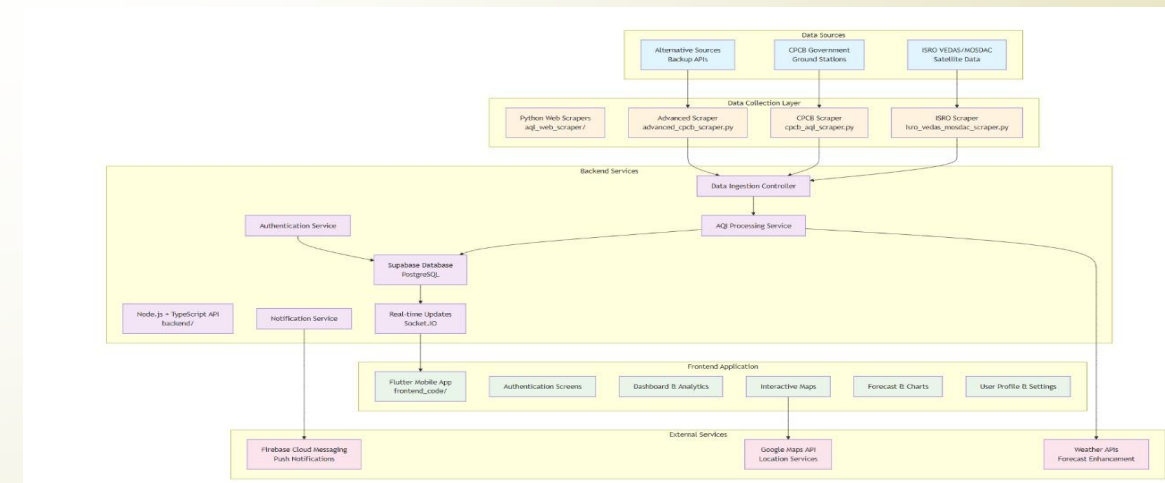
% of the work done and pending is:

➡ Pending Work : 40 %
➡ Completed work: 60 %
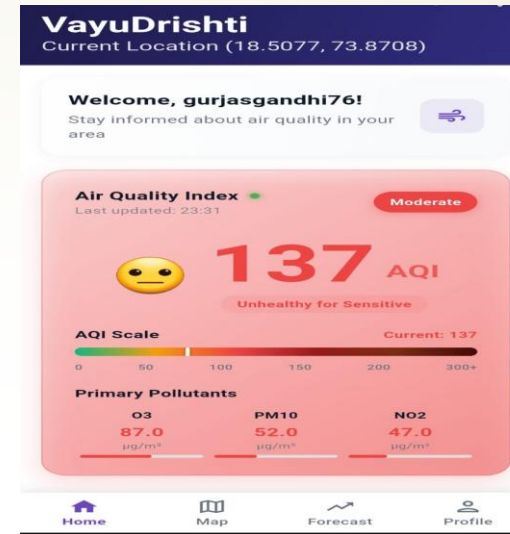
Project Model

Project flow diagram

Architecture diagram