# AIR QUALITY

## MONITORING SYSTEM

Mentor : Dr. Mainak Thakur

B25MT01

1

RECAP

2

Challenges and Improved Approach

3

Dataset Description and Analysis

4

Modeling

5

Web Dashboard

6

Timeline

# Problem Statement

This project aims to build a spatiotemporal forecasting system that uses real-time sensor data and nearby station inputs to predict environmental conditions at a specific location using TinyML on low-power edge devices. A user-friendly interface will display forecasts through graphs and offer health recommendations based on predicted data.

# 1.RECAP

# RECAP

## Dataset

- Data from 6 Chennai stations (2023–2024)
- Format: CSV from CPCB & AQICN
- Attributes: Latitude, Longitude, Solar Radiation, Humidity, Wind Speed/Direction
- Contains NULL values

## Data Cleansing

- Outliers removed using Z-score & IQR
- Skewness corrected via Box-Cox & log transforms
- Missing values filled using KNN imputer

# RECAP

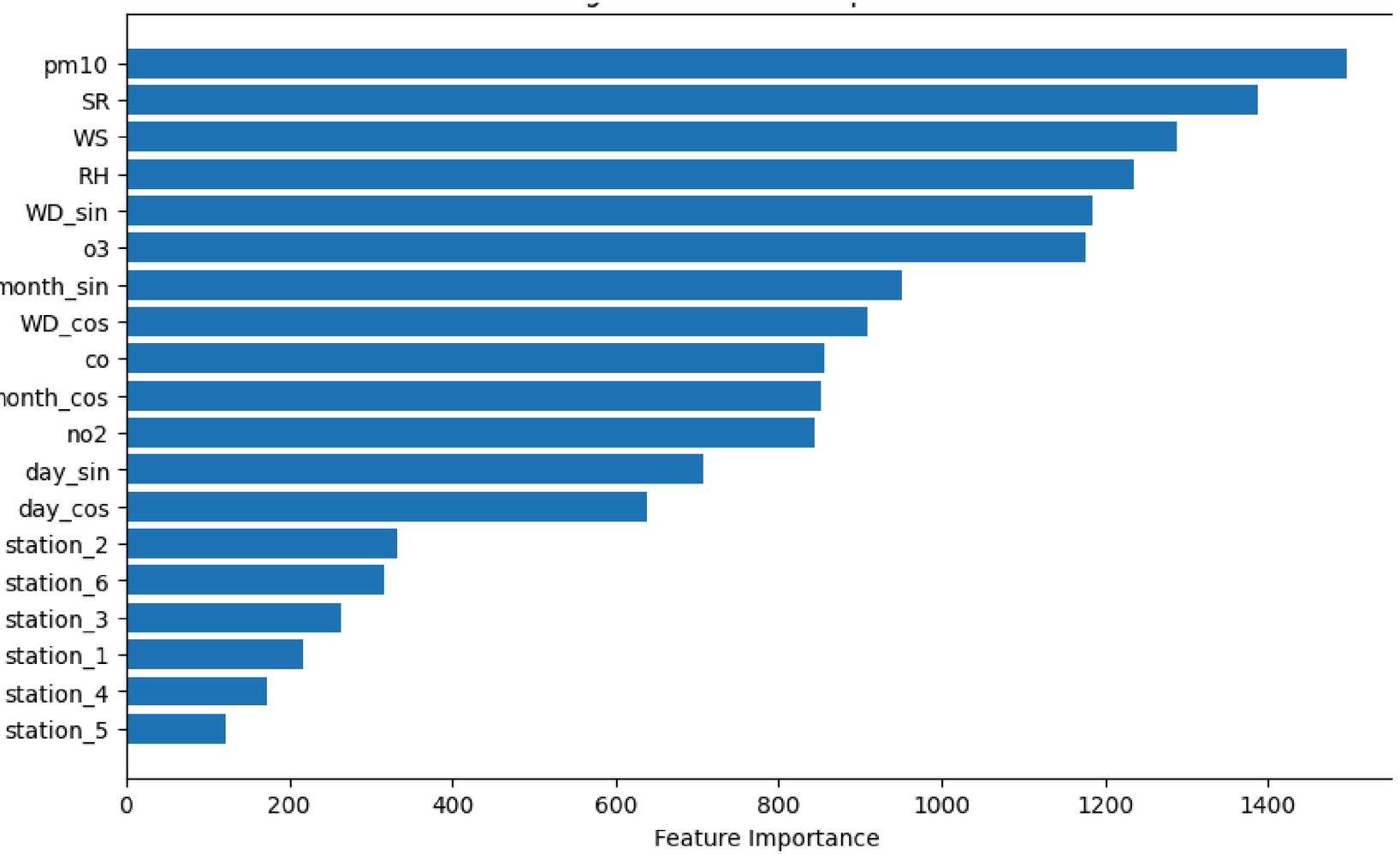
## Data Insights

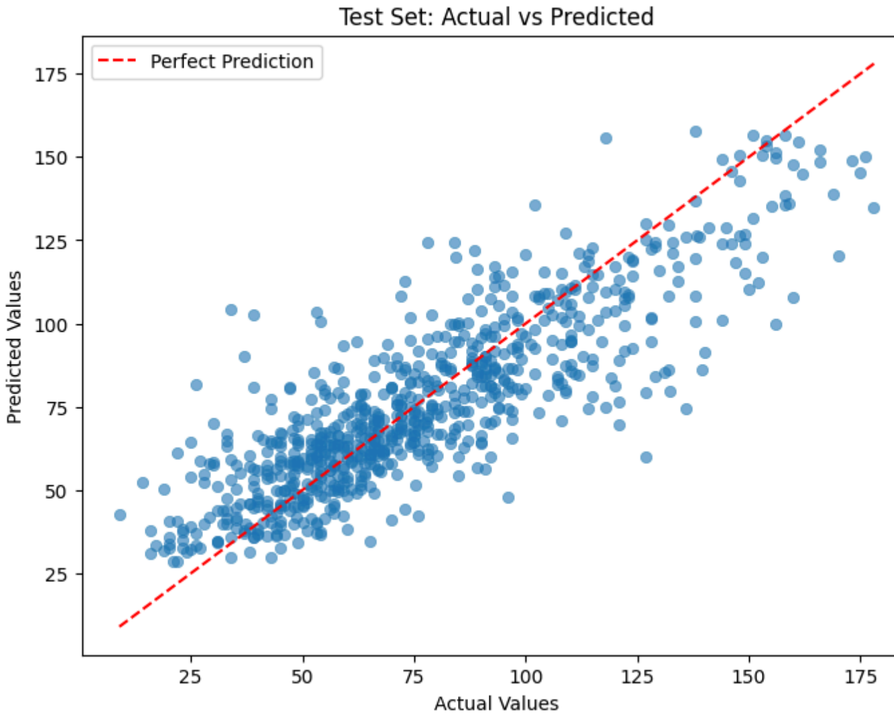
- PM2.5 spikes in winter due to low wind & inversion
- Higher PM2.5 at high temps → emissions, industries
- Skewed distribution: mostly 50–100 (Moderate AQI), with spikes

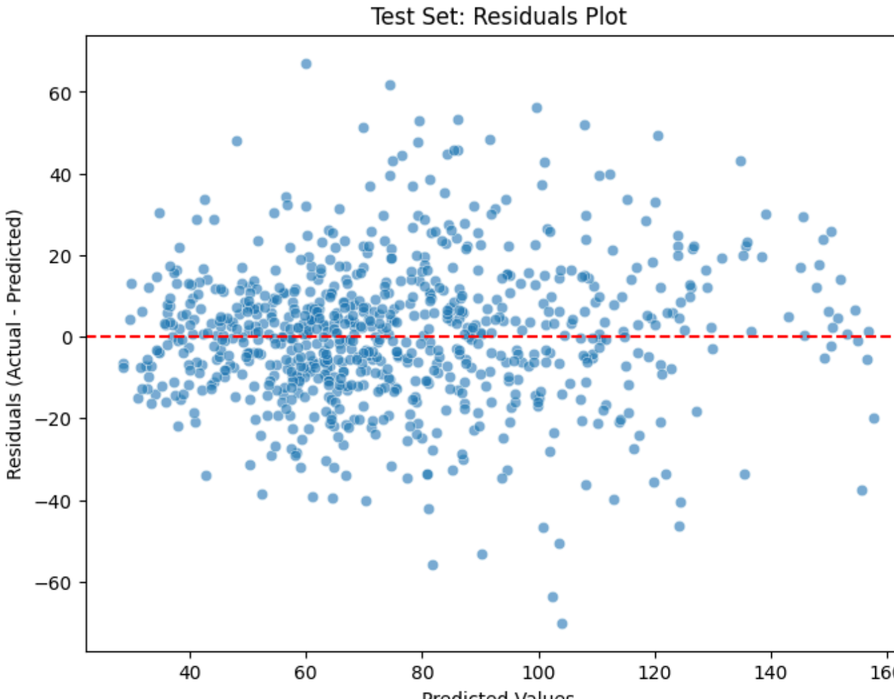
## Model Performance

- $R^2 = 0.72$, RMSE = 17.45, MAE = 13.05
- Some error spread at higher predictions
- Weather features are most important predictors

# RECAP


pm10
SR
WS
RH
WD_sin
o3
month_sin
WD_cos
co
month_cos
no2
day_sin
day_cos
station_2
station_6
station_3
station_1
station_4
station_5
0
200
400
600
800
1000
1200
1400
Feature Importance


Test Set: Actual vs Predicted
Perfect Prediction
Predicted Values
Actual Values
25
50
75
100
125
150
175


Test Set: Residuals Plot
Residuals (Actual - Predicted)
60
40
20
0
−20
−40
−60
40
60
80
100
120
140

# 2.Challenges and Improved Approach

# Where can we Improve?

GNN (Initial Choice):

- Captures spatial dependencies between 5 Chennai stations
- Modeled stations as graph nodes connected by proximity

Challenges with GNN:

- High computational cost – unsuitable for edge deployment
- Inference delay – graph message passing caused time lag
- Data sparsity – limited historical data from some stations reduced model performance and generalizability

# Why TinyML + LSTM Was a Better Fit?

- TinyML enables real-time predictions on low-power edge devices—no need for cloud processing
- LSTM captures temporal patterns in pollution data, ideal for time series forecasting
- Two-layered LSTM balances accuracy and model size, perfect for TinyML deployment
- Achieved faster inference with minimal latency, enabling real-time PM2.5 forecasting

# 3. Data Description and Analysis

# Data Collection

Ground station data of the Chennai state containing 5 locations data for the year 2024 for every 15 mintues.

Contains NULL Values

Data Source

- CPCB WEBSITE
- AQICN
- Format : CSV

**Main Attributes include**

- PM 2.5
- NO2
- SO2
- Relative Humidity
- Wind Speed

# Data Preprocessing

1

## Selected Key Features:

- Included date, $NO_2$, Relative Humidity (RH), PM2.5, and wind speed
- Focused on features directly influencing air quality

2

## Date-Time Feature Engineering:

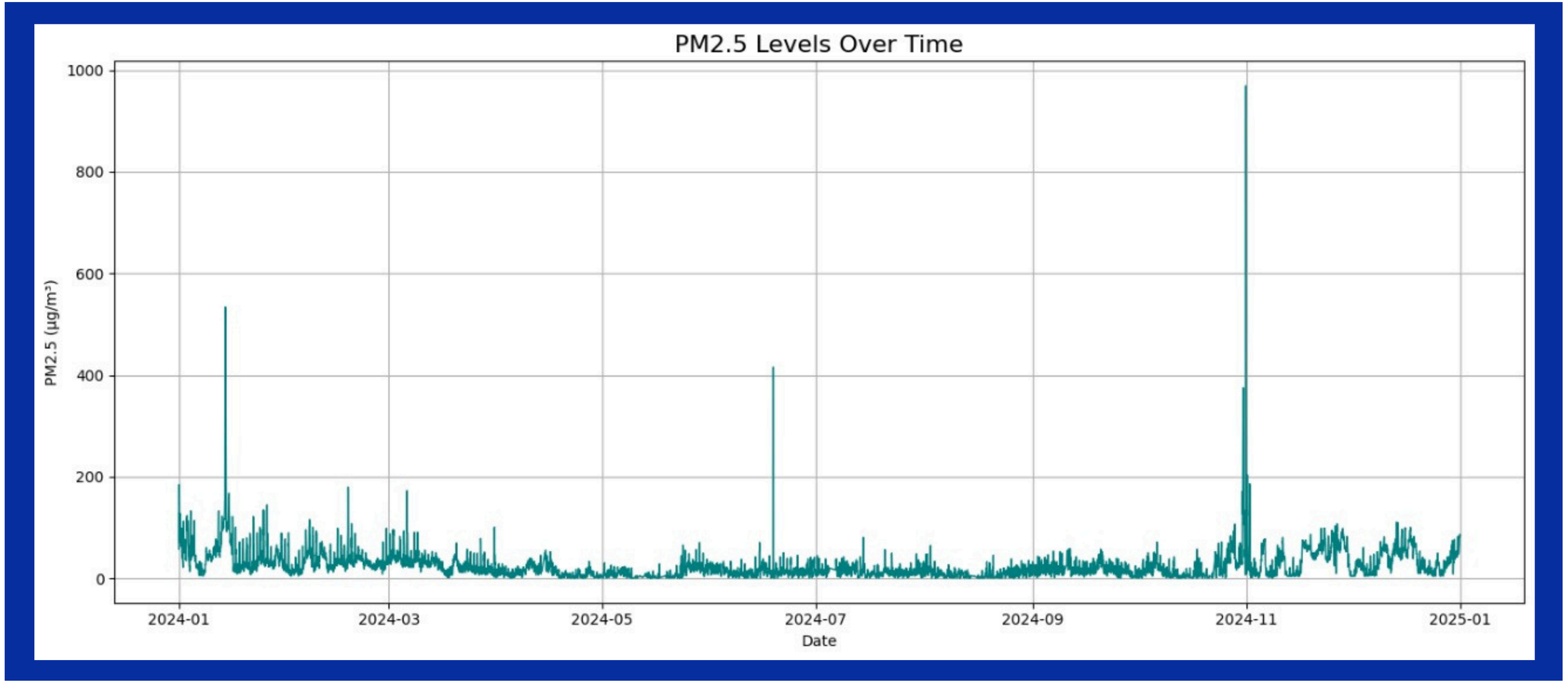
- Split date into month, day, hour, and quadrant
- Quadrant divides each hour into four 15-minute intervals for finer temporal resolution

3

## Missing Value Treatment:

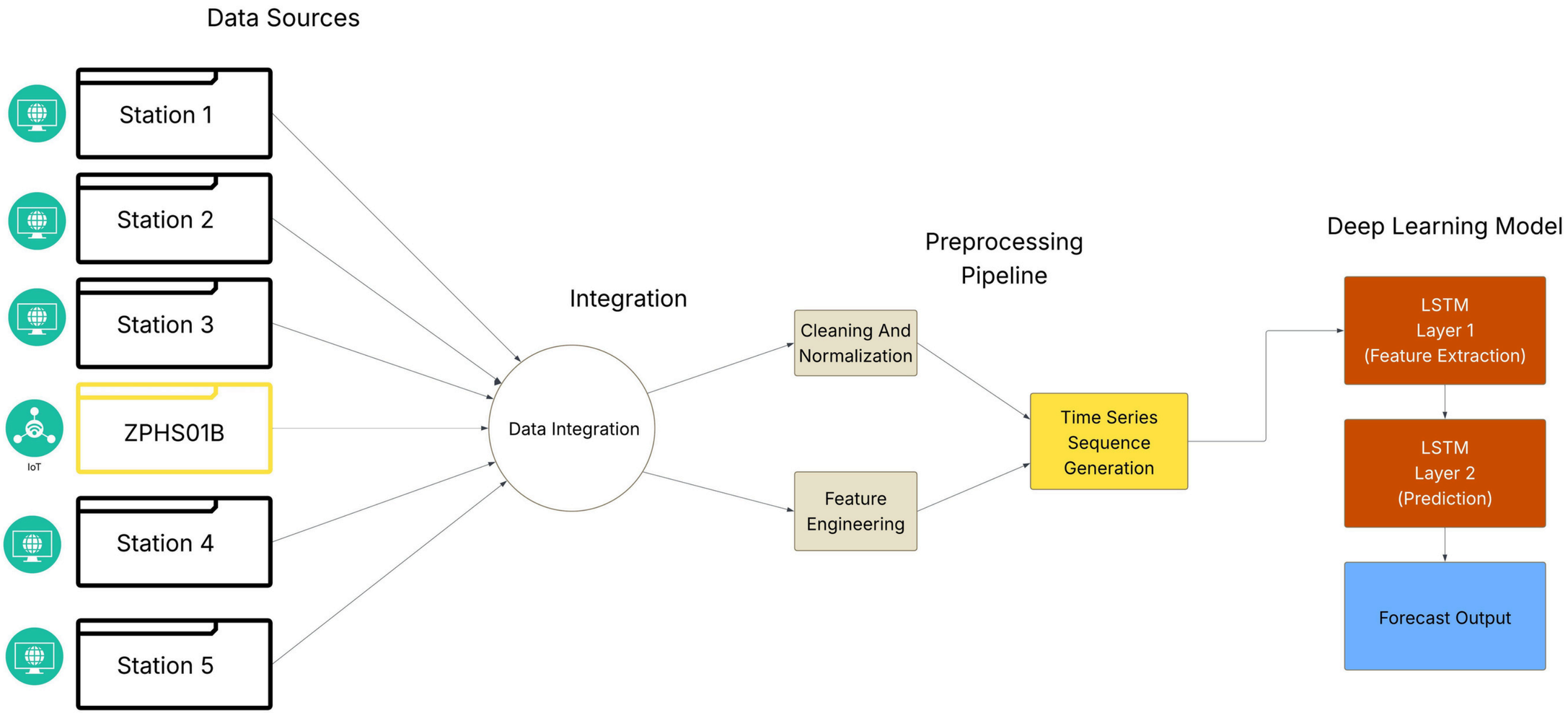
- Applied forward fill to propagate last valid value
- Followed with backward fill to fill any remaining gaps
- Ensured smoother and more continuous time series data

# Data Visualization



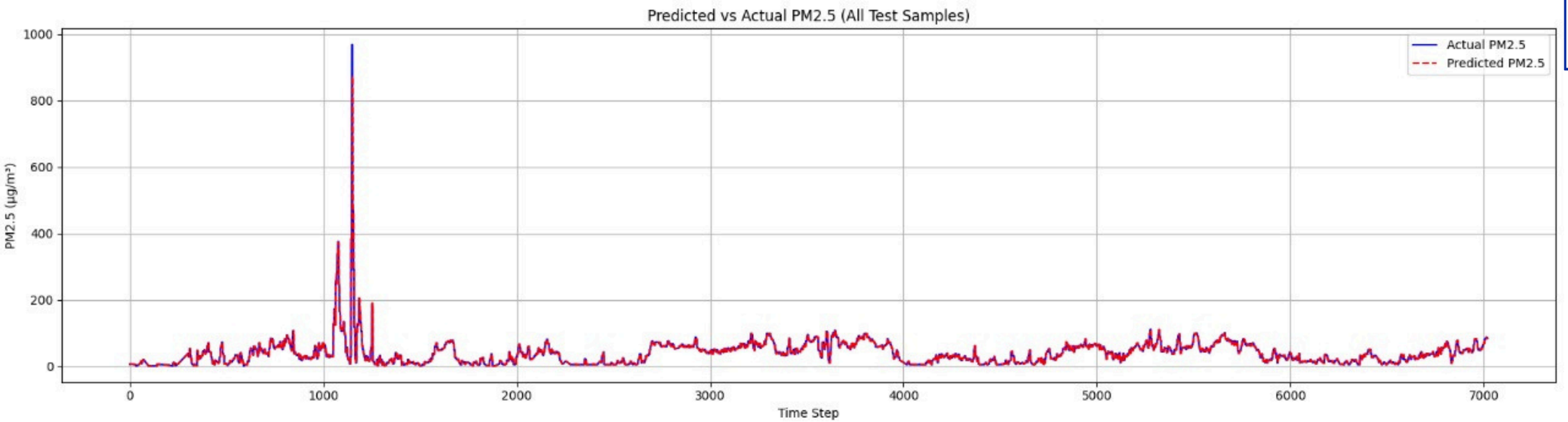
# 4.Modeling

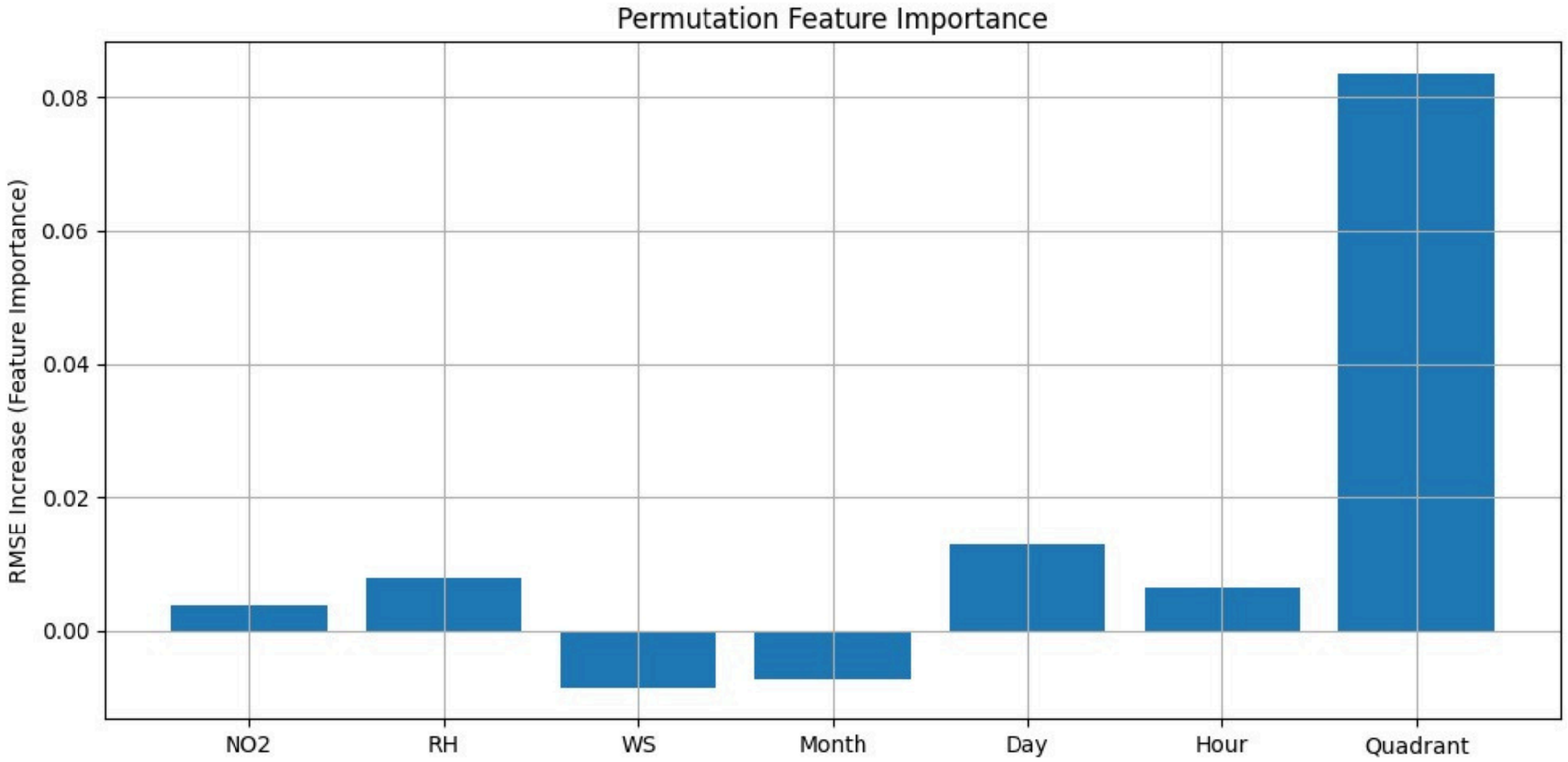
# Model Architecture



# Temporal Forecasting

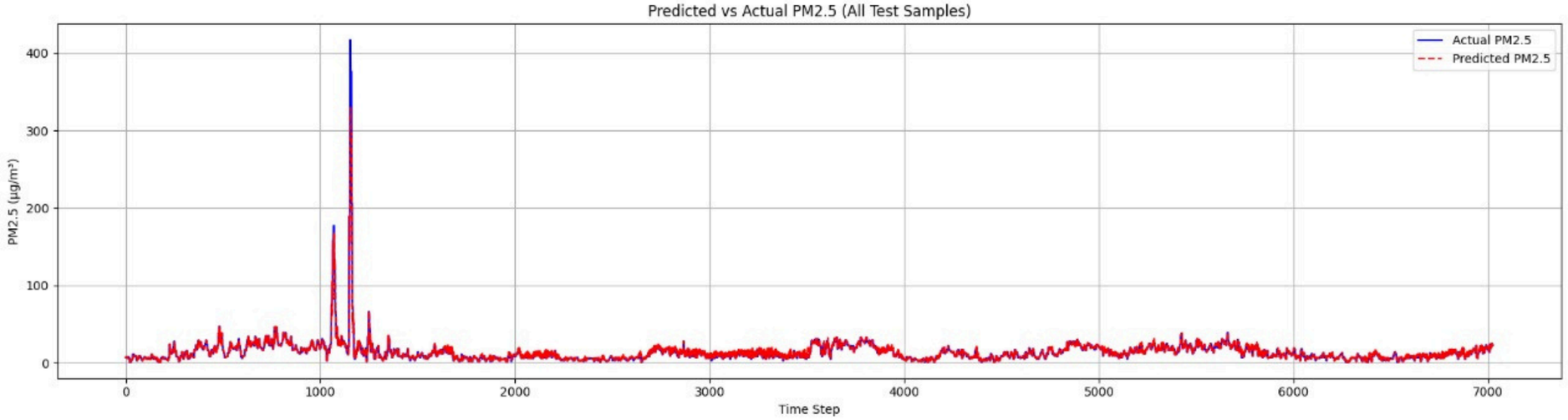
- We performed single-step temporal forecasting using past 32 time steps of features like PM2.5, $NO_2$, RH, wind speed and date-time
- A two-layer LSTM model was trained on the scaled data to predict future PM2.5 concentrations at single station.
- The model achieved strong performance with an MAE of 2.12, RMSE of 12.01, and an $R^2$ score of 0.91.
- These results indicate the model's effectiveness in capturing temporal patterns for accurate air quality forecasting, but failing to capture the effect of metrological factors on PM2.5

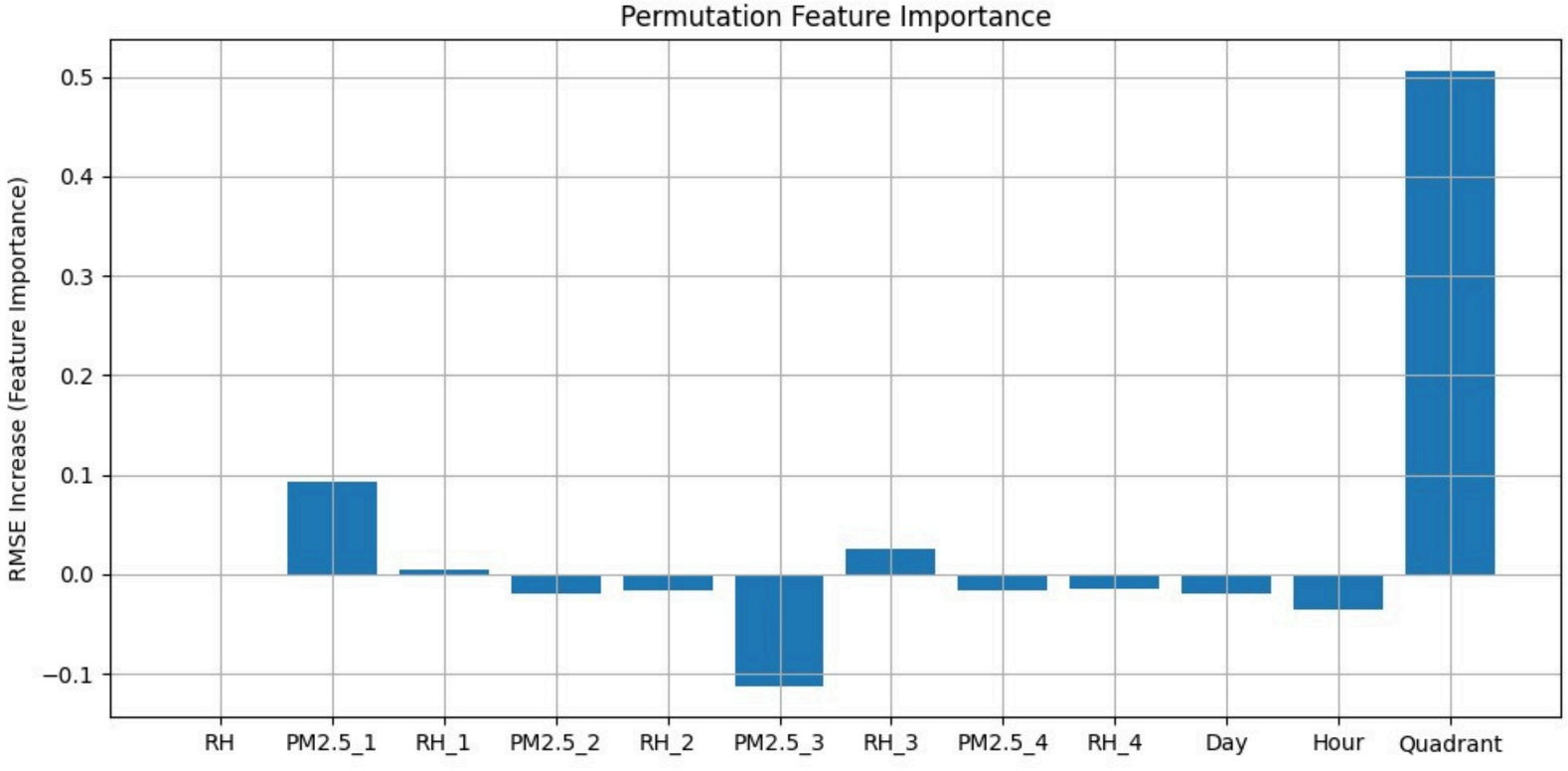

Predicted vs Actual PM2.5 (All Test Samples)
Actual PM2.5
Predicted PM2.5
PM2.5 (µg/m³)
Time Step
0
200
400
600
800
1000
0
1000
2000
3000
4000
5000
6000
7000


Permutation Feature Importance
RMSE Increase (Feature Importance)
0.00
0.02
0.04
0.06
0.08
NO2
RH
WS
Month
Day
Hour
Quadrant

# Spatial Temporal

- We implemented spatio-temporal forecasting using PM2.5 and RH data from four surrounding stations, along with time and location features.
- The feature set included local and neighboring station data (PM2.5, RH), as well as day, hour and quadrant.
- The two-layer LSTM model achieved good performance with MAE of 1.23, RMSE of 5.64, and $R^2$ of 0.90.
- However, the model struggled to fully capture spatial dependencies, as feature importance analysis showed low contribution from surrounding station values.


Predicted vs Actual PM2.5 (All Test Samples)
Actual PM2.5
Predicted PM2.5
PM2.5 (µg/m³)
Time Step
0
100
200
300
400
0
1000
2000
3000
4000
5000
6000
7000


Permutation Feature Importance
RMSE Increase (Feature Importance)
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
RH
PM2.5_1
RH_1
PM2.5_2
RH_2
PM2.5_3
RH_3
PM2.5_4
RH_4
Day
Hour
Quadrant

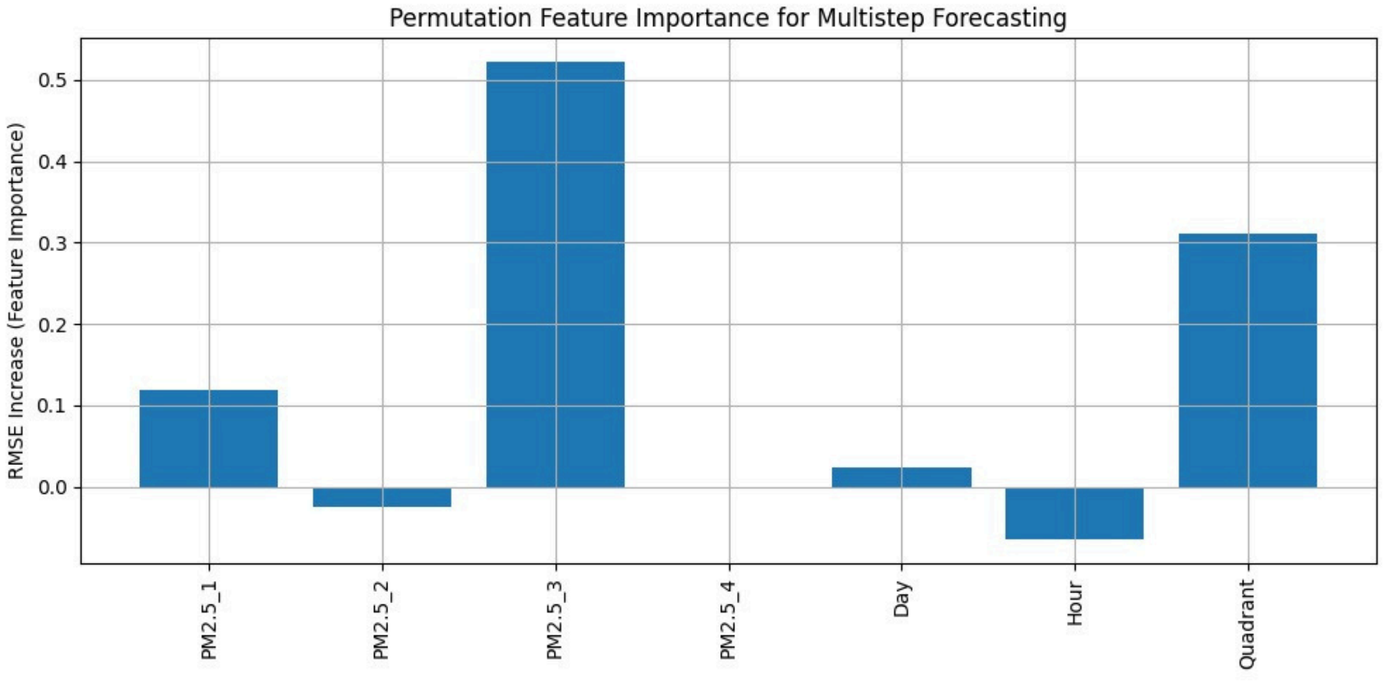
# Multi Step Spatio-Temporal Forecasting

- To enhance spatial representation, a multi-step forecasting model was developed using PM2.5 data from the target station and four neighboring stations, along with temporal and spatial features (day, hour, quadrant).

- The model was trained with 12 input time steps to predict PM2.5 levels over the next 4 time steps, addressing the time-lag in pollutant transport.

- While the overall accuracy slightly declined (MAE: 1.29, RMSE: 6.33, $R^2$: 0.87) compared to the single-step model, the multi-step approach provided improved temporal context.

- Notably, feature importance analysis revealed increased relevance of surrounding station data, indicating better spatial awareness and pollutant dispersion modeling.



# 5. Web Dashboard

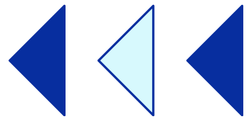
# Air Quality Visualization Website

## Why We Need It

- Makes model predictions accessible to users in a visual and intuitive way
- Helps in tracking air quality trends over time
- Supports informed decision-making for public health and safety
- Enables real-time monitoring without technical expertise
- Useful for policymakers, researchers, and citizens

## Real-time visualization of:

- Weather details
- AQI values
- Integrated using OpenWeather API



Home Cities Countries Resources My Location

Search for a city...

Sign In Sign Up

# Real-time Air Quality Index (AQI)

Get current air quality data for your location

Using Your Location View Map Refresh Data

## New Delhi

India

**88** AQI

Poor

| PM2.5 | PM10 | Temp |
|---|---|---|
| 37.6 µg/m³ | 70.3 µg/m³ | 34°C |

## Weather

Varadaiahpalem, IN

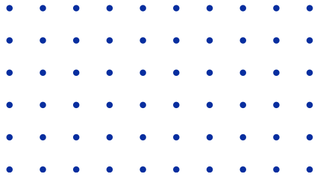
27 °C

Clouds

| Humidity | Wind |
|---|---|
| 78% | 7.58 m/s |
| **AQI** | **Updated** |
| 104 | 21:43:05 |

## AQI Scale

| | |
|---|---|
| Good | 0-50 |
| Moderate | 51-100 |
| Poor | 101-150 |
| Very Poor | 151-200 |
| Severe | 201+ |



Home Cities Countries Resources My Location Indian Cities

Search for a city...

Sign In Sign Up

# India Air Quality Index (AQI)

Real-time air pollution monitoring data for cities in India

## New Delhi

India

**114** AQI

Poor

| PM2.5 | PM10 | Temp |
|---|---|---|
| 42.0 µg/m³ | 126.0 µg/m³ | 29°C |

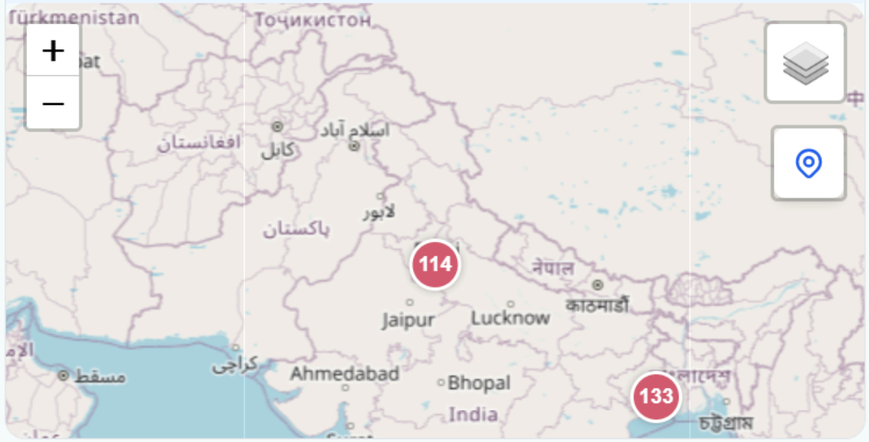
## Major Air Pollutants

View Details

| PM2.5 | PM10 | $O_3$ |
|---|---|---|
| Particulate Matter < 2.5 µm | Particulate Matter < 10 µm | Ozone |
| 42 µg/m³ | 126 µg/m³ | 54 µg/m³ |

Map View

Enlarge




## India's Metro Cities

| New Delhi | 114 |
|---|---|
| Mumbai | 89 |

# Your Location Weather

This page automatically detects your location and displays the current weather and air quality information.

## Varadaiahpalem

Andhra Pradesh IN

Lat: 13.5533, Lon: 80.0262

**82**

AQI - Moderate

**32** °C

Clouds

| | |
|---|---|
| Humidity<br>46% | Wind<br>3.24 m/s |
| PM2.5<br>N/A µg/m³ | PM10<br>42.0 µg/m³ |

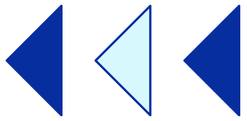
Last updated: 08:57:20

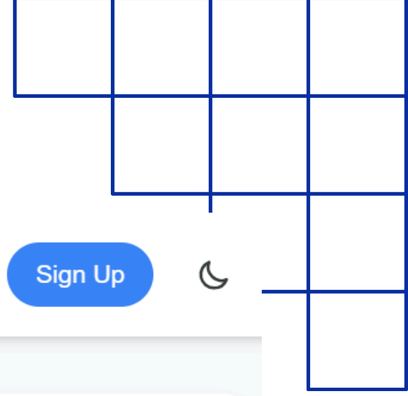
## About This Feature

This component uses your browser's geolocation API to determine your current location. It then fetches the following information:

- Your current coordinates (latitude and longitude)
- Your city name and country (using reverse geocoding)
- Current weather conditions (temperature, humidity, wind)
- Air quality index (AQI) and pollutant levels

**Privacy Note:** Your location data is only used to fetch weather and air quality information. It is not stored or shared with any third parties.





Home Cities Countries Resources My Location Indian Cities

Search for a city...

Sign In Sign Up

# Air Quality Index (AQI) in Indian Cities

Real-time air quality data for cities across India

Refresh Data Back to Home

Search cities...

All Cities Major Cities Tier-2 & Tier-3 Smaller Cities

## 36 Cities Found

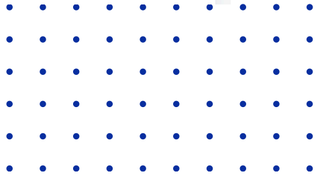
Last updated: 01:09:26

**Delhi**
India
127 AQI
Poor
PM2.5 127.0 µg/m³
PM10 52.0 µg/m³
Temp 29°C

**Mumbai**
India
96 AQI
Moderate
PM2.5 96.0 µg/m³
PM10 57.0 µg/m³
Temp 28°C

**Chennai**
India
82 AQI
Moderate
PM2.5 82.0 µg/m³
PM10 46.0 µg/m³
Temp 34°C

**Bengaluru**
India
65 AQI
Moderate
PM2.5 65.0 µg/m³
PM10 51.0 µg/m³
Temp 23°C



# 6.Timeline

# Work Done Till Now

- Gone through the research papers, proposing various methodologies.
- Collected the ground station and sensor data.
- Cleaned the data.
- Data Analysis of important features.
- Built various TinyML models to forecast PM2.5 values
- Finalized the plan of action and improvements to do

# Future Work

- Forecast various pollutants using TinyML model.
- Displaying the forecast and pollution data to the user through web application.
- Sharing users important tips based on health recommendations.

# Timeline

**1**

**2**

Progress as mentioned above.

**3**

Complete Website

**4**

Personalised Health Recommendations

# References


- **TinyML with Meta-Learning on Microcontrollers for Air Pollution Prediction**, I. N. K. Wardana(1), S. A. Fahmy(2), J. W. Gardner(3), IEEE Transactions on Instrumentation and Measurement, 2023, DOI: 10.1109/TIM.2023.3293177.
- **Optimizing TinyML: The Impact of Reduced Data Acquisition Rates for Time Series Classification on Microcontrollers**, R. Samanta(1), B. Saha(2), S. K. Ghosh(3), R. B. Roy(4), Proceedings of the 2023 International Joint Conference on Neural Networks (IJCNN), DOI: 10.1109/IJCNN54540.2023.10191595.
- **Canva – For slide and UI design** (https://www.canva.com)
- **OpenWeather API – For fetching real-time weather data** (https://openweathermap.org/api)
- **WAQI (World Air Quality Index) API – For real-time AQI data** (https://aqicn.org/api)
- **CPCB (Central Pollution Control Board) – Official Indian AQI monitoring data** (https://cpcb.nic.in)

# Our Team

Jyothish Varma - S20220010127

Nandhitha Shaganti - S20220010198

Hemanth Chakravarthy - S20220010095

# Thank You