Air Pollution and LST Prediction and its Impact on Ecosystem

Allen Joshua, Mohammad Zaid, Lavena Babu

Department of Information Technology, The Bombay Salesian Society's

Don Bosco Institute of Technology , Mumbai-400070

2023 - 2024

Abstract

Mumbai, the commercial hub of India, is a bustling metropolis where the urban landscape undergoes continual transformation. However, this rapid urbanization has led to environmental degradation, with recent research revealing a persistent issue of pollutant levels exceeding advised thresholds. This phenomenon poses a significant threat to human health in the region. The research utilizes meteorological data collected from four weather stations (Powai, Sion, Colaba, and Borivali) spanning the years 2000–2023 to analyze pollutant concentrations, including PM2.5, PM10, and air temperature. The study reveals an alarming trend of exponential growth in contaminant levels, highlighting the urgent need for intervention. Deep learning techniques such as RNN, LSTM, GRU and BiLSTM are employed for future prediction, with the results demonstrating the superiority of BiLSTM algorithms over other models. The findings suggest that certain areas, particularly Sion, are at risk of becoming pollution hotspots, exacerbating health hazards for residents.

By leveraging insights from this study, Mumbai metropolitan planners and authorities can make informed decisions to mitigate pollution and promote sustainable growth, thereby safeguarding the well-being of the city's inhabitants.

Background Information

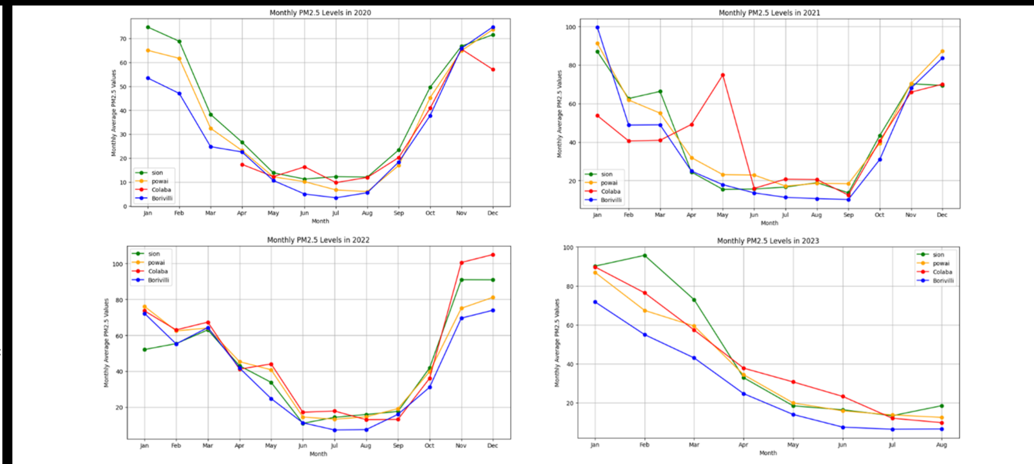
Tian et al. found that the random forest methodology was the most accurate in forecasting PM2.5 levels in China's Pearl River Delta. Using meteorological parameters including BP, TEMP, and RHR, they estimated the ground-level concentrations of PM2.5 using six machine learning techniques: RF, BPNN, SVM, ENN, GBRT and GAM.

GIS and Python can be utilized to provide precise and timely air pollution predictions, according to Zhalehdoost and Taleai's review. Individuals can then take action to protect themselves from air pollution exposure by using these estimates to guide public health measures. Moreover, certain challenges identified in the assessment need addressing to enhance the accuracy of air pollutant's forecasts and broaden availability of air pollution prediction maps.

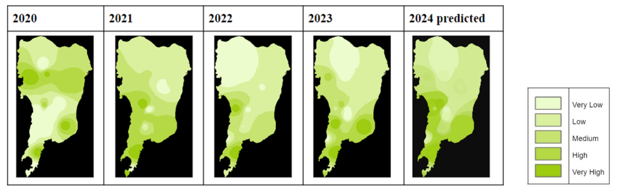
E. M. Jovanovska et al.'s research concentrates on diverse methods for assessing and forecasting pollution in urban areas. In addition to discussing the limits that various models encounter, the paper emphasizes the significance of data quality and relevance for precise forecasts. The research intends to contribute to the progress of precise and trustworthy urban air pollution predictions by highlighting the critical role of data quality and identifying the difficulties associated with various models.

In Zhan's study, a number of models were compared, including Long Short-Term Memory (LSTM), Back Propagation Neural Network (BPNN), Gated Recurrent Unit (GRU), Convolutional Neural Networks (CNN) and Bi-directional Long Short-Term Memory (BiLSTM). Based on their findings, BiLSTM performs best. Additionally, according to this study, the air quality index is greater in winter than it is in summer.

Results



Error\Model	RNN	LSTM	GRU	BiLSTM
MAE	0.0329	0.0612	0.0350	0.0295
MSE	0.0018	0.0041	0.0018	0.0012
RMSE	0.0428	0.0645	0.0425	0.0349



Introduction

In recent decades, Mumbai has experienced a significant surge in its population, transforming into a bustling metropolis housing over 26 million people. As the city expands, it grapples with a pressing environmental dilemma – air pollution.

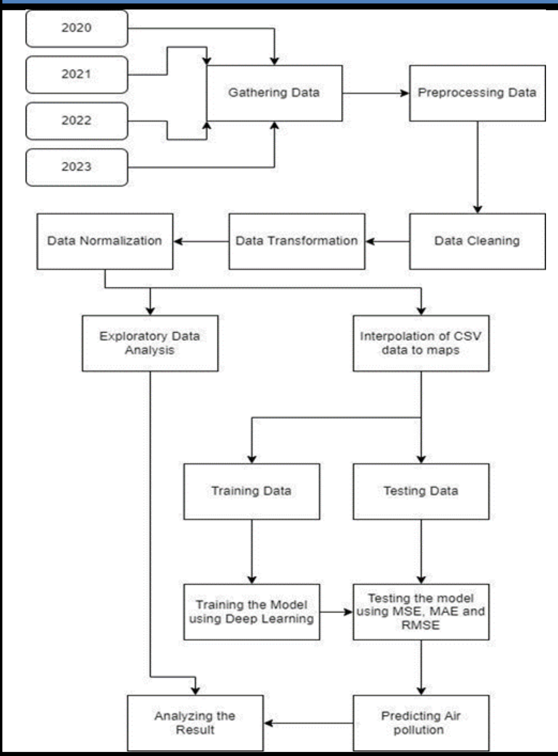
The escalating levels of pollutants in the atmosphere, compounded by rising temperatures, present a formidable crisis. Beyond mere city's appearance, this poses a grave threat to the well-being of its inhabitants.

Mumbai's air pollution predominantly stems from human activities, industrial operations, construction activities, and vehicular emissions. The deteriorating air quality, characterized by elevated levels of fine particulate matter (PM), poses significant health hazards. These tiny particles have the potential to infiltrate deep into the respiratory system and even enter the bloodstream, precipitating respiratory and cardiovascular complications.

The escalating prevalence of pollution is alarming, contributing to a surge in respiratory ailments such as chronic obstructive pulmonary disease (COPD) and asthma, particularly affecting vulnerable demographics such as children and the elderly. Prolonged exposure to airborne pollutants has been correlated with cardiovascular issues, adverse pregnancy outcomes, and premature mortality. Understanding fluctuations in pollutant levels and their health implications is paramount.

This initiative aims to develop a predictive model for air pollution and temperature by harnessing historical data and meteorological variables. Aligned with a global health agenda, the project integrates geographic information systems (GIS), data analytics, and public health interventions to combat pollution-related diseases and empower communities.

Flow diagram and Algorithm



1. Data Preparation: Divide data into input-output pairs where input is a sequence of data points and output is the subsequent point.
2. Model Initialization: Set up LSTM, RNN, or GRU model with desired layers, units, and hyperparameters. Randomly initialize weights.
3. Forward Propagation: Feed input sequences, predict output sequences by passing through recurrent layers and update hidden states.
4. Loss Calculation: Compute loss between predicted and target sequences using MSE.
5. Backpropagation: Conduct backpropagation through time (BPTT) to compute gradients and update weights.
6. Gradient Optimization: Use optimization algorithm like Adam to adjust model weights, minimize loss and enhance performance.
7. Repeat: Iterate steps 3 to 6 for multiple epochs for refining model's performance.
8. Validation: Assess model's performance on separate validation dataset, monitoring metrics and prevent overfitting.
9. Testing: Once model achieves satisfactory performance on validation set, deploy it to make predictions on unseen test data.

References

[1] H. Tian, Y. Zhao, M. Luo, Qingqing, Y. Han, Z. Zeng, "Estimating PM2.5 from multisource data: A comparison of different machine learning models in the Pearl River Delta of China", Urban Climate, vol 35, Jan 2021

[2] S. Siami-Namini, N. Tavakoli and A.S. Namin, "The Performance of LSTM and BiLSTM in Forecasting Time Series", IEEE International Conference on Big Data, 2019, pp.

[3] C. Zhan, S. Li, J. Li, Y. Guo, Q. Wen and W. Wen, "Prediction of Air Quality in Major Cities of China by Deep Learning," 16th International Conference on Computational Intelligence and Security (CIS), Guangxi, China, pp. 68-72, 2020

[4] Bu-Yo Kim, Yun-Kyu Lim.. "Short-term prediction of particulate matter (PM10 and PM2.5) in Seoul, South Korea using tree-based machine learning algorithms", National Institute of Meteorological Sciences, https://doi.org/10.1016/j.apr.2022.101547.

[5] H.S. Kim, I. Park "Development of daily PM10 and PM2.5 prediction system using a deep long short-term memory neural network model",Gwangju Institute of Science and Technology, Korea,
https://doi.org/10.5194/acp-2019-268

[6] P. Muthukumar et al, "Satellite Image Atmospheric Air Pollution Prediction through Meteorological Graph Convolutional Network with Deep Convolutional LSTM", https://doi:10.1109/CSCI51800.2020.00094.

[7] Doreswamy et al, "Forecasting Air pollution particulate matter (PM2.5) using machine learning regression models", Mangalore University