DEEP LEARNING AND CONVENTIONAL STATISTICAL MODELS FOR AIR QUALITY PREDICTION

[1]Dr. V.Gayatri, [2]Dr. R. Rajani

[1]Associate Professor, [2]Professor, [1,2]Department of Computer Science & Engineering, Geethanjali Institute of Science and Technology, Gangavaram, Andhra Pradesh, India

ABSTRACT

Rapid growth in urbanization and industrialization leads to an increase in air pollution and poor air quality. Because of its adverse effects on the natural environment and human health, it's been declared a "silent public health emergency". To deal with this global challenge, accurate prediction of air pollution is important for stakeholders to take required actions. In recent years, deep learning-based forecasting models show promise for more effective and efficient forecasting of air quality than other approaches. In this project, we made a comparative analysis of various deep learning-based single-step forecasting models such as long short term memory (LSTM), gated recurrent unit (GRU), and a statistical model to predict five air pollutants namely Nitrogen Dioxide (NO_2), Ozone (O_3), Sulphur Dioxide (SO_2), and Particulate Matter (PM2.5, and PM10). For empirical evaluation, we used a publicly available dataset collected inAir Quality Data in India (2015 – 2020)" dataset from Kaggle. It measures the concentration of air pollutants. The performance of forecasting models is evaluated based on three performance metrics: root mean square error (RMSE), mean absolute error (MAE) and R-squared (R2). The result shows that deep learning models consistently achieved the least RMSE compared to the statistical models with a value of 0.59. In addition, the deep learning model is also found to have the highest R2 score of 0.856.

INTRODUCTION

Over the past few years, air pollution has become a major global challenge. Air pollution has a direct impact not only on the environment but also on human health and well-being. It has been observed that air pollution leads to increase mortality and morbidity such as respiratory diseases, impaired cognitive function, cardiovascular diseases, and cancer. Each year, over 3 million deaths are recorded due to air pollution especially in low and middle income countries. In addition, the United Nations (UN) has defined sustainable development goals (SDG) such as 3, 7 and 11 where targets are set for 2030 to reduce deaths, illness and the adverse environmental effect in cities by improving air quality and other factors. Similarly, in the United Kingdom (UK), the government has set a target to reduce 35% of air pollution by 2040.

There are multiple factors involved in deteriorating air quality such as manufacturing, industrial emissions, transportation (in land, air and sea) emissions, dust, and coal consumption. Air pollution is the introduction of harmful materials and gases into the environment which is of great concern to humans and other living organisms. These harmful materials (solids, liquids or gases) are called pollutants. When these pollutants like PM2.5 (Particulate Matter) are produced in higher concentrations than usual, it reduces the quality of our environment and causes serious harmful health effects.

Education and rising public awareness relating to this issue requires interdisciplinary approaches with professionals and other stakeholders. Local councils are playing their part and have set up many air quality monitoring stations throughout the country to monitor the concentration of air pollutants. Data collected from such monitoring stations can be used in the prediction of pollutants. Prediction of air quality is important to control air pollution and to identify areas which require solutions to overcome air pollution and related impacts. However, how to model air quality accurately is a challenge on its own and depends on available data and

modelling approaches. The main contributions of this study include: We propose a combination of meteorological parameters such as temperature, wind speed and wind direction with lagged air quality feature which is based on the concentration value of the previous hour for the pollutant being predicted for the next hour. We also make use of the datetime index by splitting it into hour, day and month for additional features to improve prediction accuracy and reduction of error.We propose a comprehensive study to predict the five major air pollutants and a comparison of deep learning base models with statistical mRodel. We provide detail architectural and parameters information for both deep learning and statistical models, to predict each pollutant, which can be useful in various applications in the scope of smart cities or can provide benchmarking for more better and accurate models.

Urban air pollution is a pressing environmental and public health concern worldwide, with detrimental effects on human health, ecosystems, and the economy. The rapid urbanization and industrialization observed in recent decades have exacerbated air quality issues, leading to increased concentrations of pollutants such as particulate matter (PM), nitrogen dioxide (NO2), sulfur dioxide (SO2), and ozone (O3) in urban areas. These pollutants are primarily emitted by transportation, industrial activities, power generation, and residential heating, among other sources.

Given the complexity and dynamic nature of urban air pollution, accurate prediction of pollutant concentrations is crucial for effective air quality management, policy development, and public health protection. Traditional statistical models have been widely used for air pollutant prediction, leveraging statistical relationships between pollutant concentrations and various meteorological, geographical, and anthropogenic factors. However, with the advent of deep learning techniques and the availability of large-scale air quality data, there is growing interest in exploring the efficacy of deep learning models for air pollutant prediction.

LITERATURE SURVEY

Author: R. D. Brook

Title:" Cardiovascular effects of air pollution"

Project Conduct Year:2008

Description:

Air pollution is a heterogeneous mixture of gases, liquids and PM (particulate matter). In the modern urban world, PM is principally derived from fossil fuel combustion with individual constituents varying in size from a few nanometres to 10 μm in diameter. In addition to the ambient concentration, the pollution source and chemical composition may play roles in determining the biological toxicity and subsequent health effects. Nevertheless, studies from across the world have consistently shown that both short- and long-term exposures to PM are associated with a host of cardiovascular diseases, including myocardial ischaemia and infarctions, heart failure, arrhythmias, strokes and increased cardiovascular mortality. Evidence from cellular/toxicological experiments, controlled animal and human exposures and human panel studies have demonstrated several mechanisms by which particle exposure may both trigger acute events as well as prompt the chronic development of cardiovascular diseases

Author:M. Stafoggia and T. Bellander

Title:"Short-term effects of air pollutants on daily mortality in the Stockholm county—A spatiotemporal analysis"

Project Conduct Year:2020

Description:

Short-term exposure to air pollutants has been extensively related to daily mortality, however most of the evidence comes from studies conducted in major cities, and little is known on the extent of the spatial heterogeneity in the effects within areas including both urban and non-urban settings. We aimed to investigate the short-term association of air pollutants with daily cause-specific mortality in the Stockholm county, and to

test whether an association exists also outside the metropolitan area. We used a spatiotemporal random forest model to predict daily concentrations of fine and inhalable particulate matter ($PM_{2.5}$ and PM_{10}), nitrogen dioxide (NO_2) and ozone (O_3) at 1-km spatial resolution over Sweden for 2005-2016. We collected data on daily mortality for each small area for market statistics (SAMS) of the Stockholm county, to which we matched daily exposures to air pollutants and air temperature

Author:B. Paul and S. Louise

Title:"Air Quality: Policies Proposals and Concerns—House of Commons Library",

Project Conduct year:2022

Description:

Poor air quality is considered by the government to be "the largest environmental risk to public health in the UK". As well as human health, air pollution also has implications for the natural environment and for the economy. Due to the transboundary nature of air pollution, action to manage and improve air quality in the UK has been driven by both international agreements and EU legislation, as well as national and devolved legislation.

This briefing gives an overview of the current outdoor air quality legal framework, the changing governance and enforcement mechanisms following the UK's EU exit, forthcoming legislative changes and ongoing issues and concerns.

Information about road user charging schemes intended to reduce air pollution is set out in a separate Commons Library briefing, Clean Air Zones, Low Emission Zones and the London ULEZ.

Author:P. J. Landrigan

Title:"Air pollution and health"

Project Conduct Year:2017

Description:

Air pollution is a familiar environmental health hazard. We know what we're looking at when brown haze settles over a city, exhaust billows across a busy highway, or a plume rises from a smokestack. Some air pollution is not seen, but its pungent smell alerts you.

It is a major threat to global health and prosperity. Air pollution, in all forms, is responsible for more than 6.5 million deaths each year globally, a number that has increased over the past two decades.

EXISTING SYSTEM

Statistical Methods: These have been around since at least the 1970s, according to research papers on air quality forecasting.They use historical data on air pollutants, meteorology, and other factors to develop statistical models for predicting future concentrations.

Dispersion Modelling: This technique has been in use since the mid-20th century

It uses computer simulations to predict how pollutants will disperse in the atmosphere based on factors like wind speed, direction, and topography.

PROPOSEDSYSTEM

Deep learning models have transformed the field of air quality prediction by offering severalkey advantages over traditional methods. Here's how deep learning compares with the existing models

LSTM (LONG SHORT TERM MEMORY)

Recurrent neural networks (RNN) are networks with loops in them, enabling the information to persevere. When the gap between the related information and the place it is required is small, RNNs can learn to utilize the past information. Unfortunately, as the gap increases, RNNs become unfit to learn to associate the information.

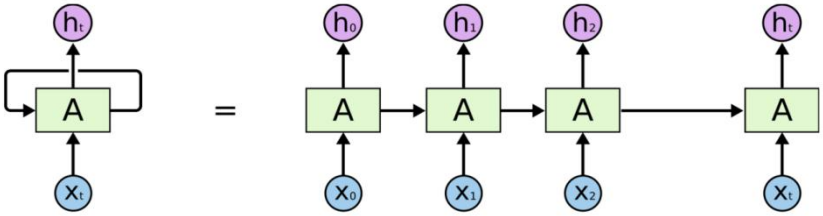

Figure 1 Recurrent Neural Network

LSTMs are an extraordinary sort of RNNs, equipped for adapting long term conditions. Recollecting information for long periods purposes their default behaviour. LSTMs also have a chain like structure, yet the repeating module has an alternate structure, not at all like RNNs. Rather than having a single neural network, there are four layers, cooperating in a unique manner. The way to LSTM is the cell state. The cell state is somewhat similar to a conveyor belt. It runs straight down the whole chain, with some minor linear connections. It is extremely simple for information to the cell state, carefully controlled by structures called gates.

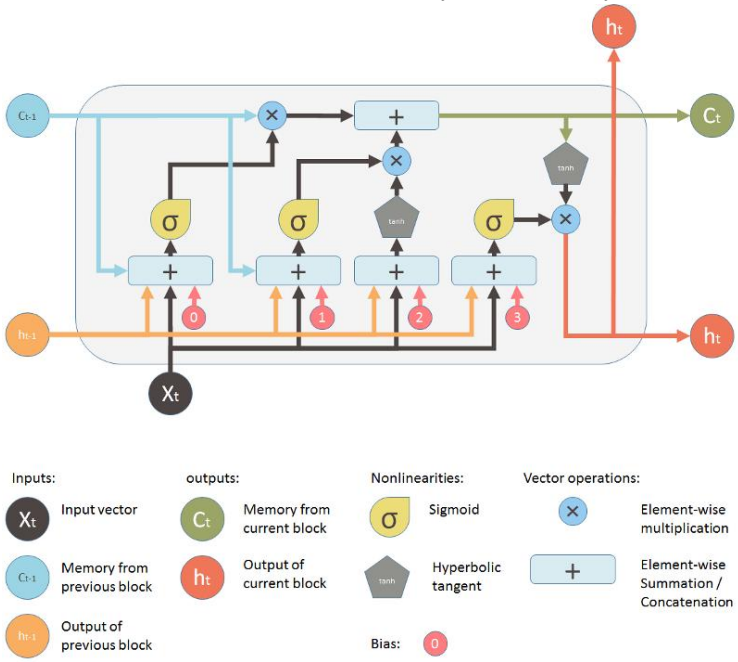

Figure.2. Basic LSTM Memory Block

LSTM networks are appropriate for classifying, processing and making predictions based on time series data, since there can be lags of obscure duration between important events in a time series. They were created to manage the exploding gradient and vanishing gradient problems that can be experienced when training traditional RNNs. The activation function of the LSTM gates is frequently the logistic function. The weight of these connections, which need to be learned during training, decide how the gates operate.

A RNN utilizing LSTM can be trained in a supervised fashion, on a set of training sequences, using an optimization algorithm, gradient descent, joined with back propagation through time to calculate the gradients needed during the optimization process, in order to change weights.

Gates are an approach to alternatively let data through. They are made out of a sigmoid neural net layer and a point wise multiplication operation.

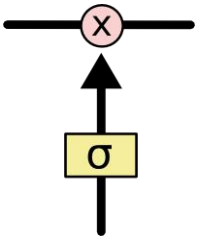

Figure 3. Block diagram of a gate

The sigmoid layer yields numbers somewhere in the range of 0 and 1, depicting the amount of every component ought to be let through. A value of 0 signifies “let nothing through”, while a value of 1 signifies “let everything through”. An LSTM has three of these gates, to secure and control the cell state.

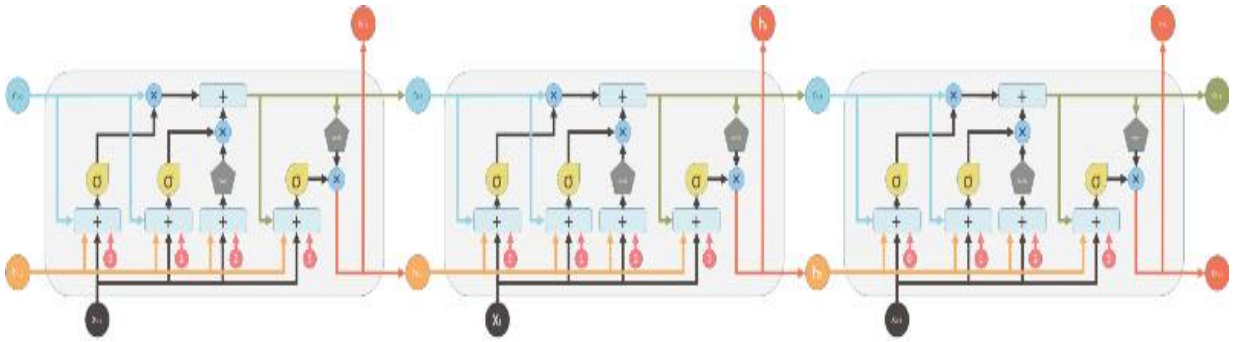

Figure 4 LSTM network with memory blocks

The initial phase in our LSTM is to choose what data we are going to discard from the cell state. This choice is made by the sigmoid layer, called the “forget gate” layer. It looks at h_{t-1} and $x_t$and yields a number somewhere in the range of 0 and 1 for every cell state C_{t-1}. 1 signifies “totally keep this” and 0 implies “totally dispose of this”.

DATASET

The “Air Quality Data in India (2015 – 2020)” dataset from Kaggle contains air quality data and the Air Quality Index (AQI) at both hourly and daily levels from various stations across multiple cities in India1. This dataset is used to analyze the air quality in India over a span of five years, from 2015 to 20201.

The purpose of this dataset is to help researchers, data scientists, and decision-makers understand the air quality situation in India, identify trends, and make informed decisions to prevent diseases caused by air pollution.

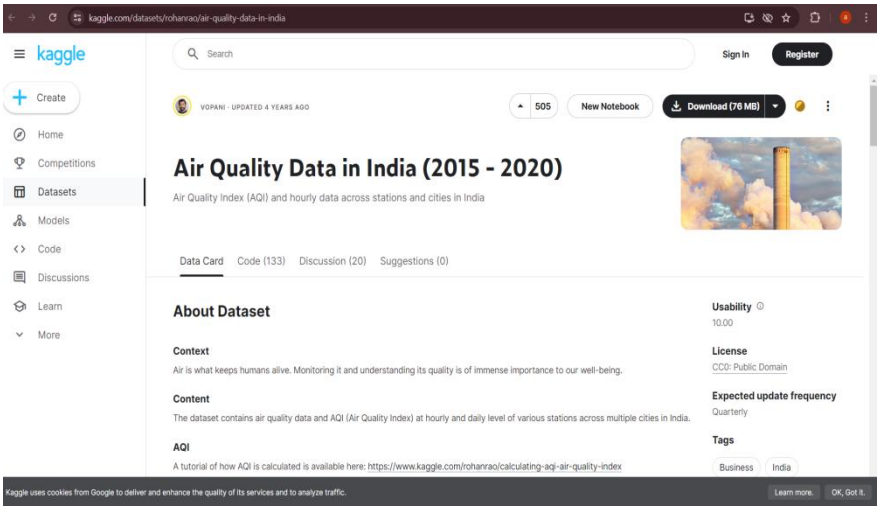

Fig 5 Air Quality Data in India (2015 – 2020)

TABLE 1 Statistics of Meteorological

	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3
count	24172.000000	17764.000000	24463.000000	24459.000000	22993.000000	18314.000000	24405.000000	24245.000000	24043.000000
mean	67.476613	118.454435	17.622421	28.978391	32.289012	23.848366	2.345267	14.362933	34.912885
std	63.075398	89.487976	22.421138	24.627054	30.712855	25.875981	7.075208	17.428693	21.724525
min	0.040000	0.030000	0.030000	0.010000	0.000000	0.010000	0.000000	0.010000	0.010000
25%	29.000000	56.777500	5.660000	11.940000	13.110000	8.960000	0.590000	5.730000	19.250000
50%	48.785000	96.180000	9.910000	22.100000	23.680000	16.310000	0.930000	9.220000	31.250000
75%	80.925000	150.182500	20.030000	38.240000	40.170000	30.360000	1.480000	15.140000	46.080000
max	914.940000	917.080000	390.680000	362.210000	378.240000	352.890000	175.810000	186.080000	257.730000

Data

Data Flow Diagram:

A data-flow diagram is a way of representing a flow of data through a process or a system. The DFD also provides information about the outputs and inputs of each entity and the process itself. A data-flow diagram has no control flow — there are no decision rules and noloops.

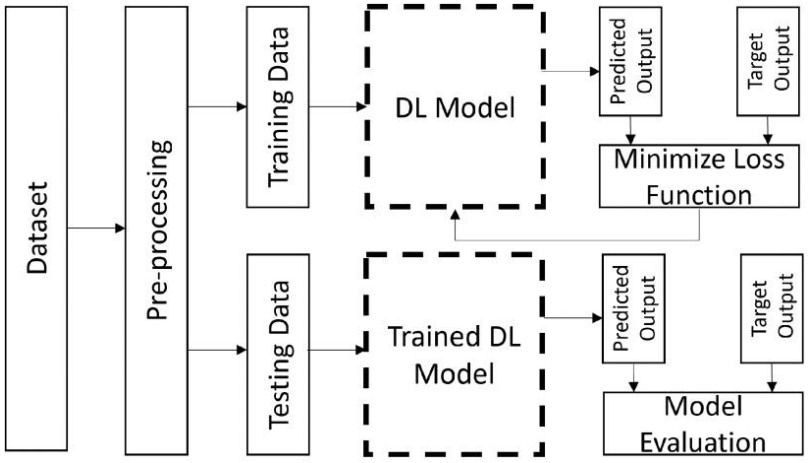

Figure 6 Data flow diagram

MODELTRAINING AND TESTING

This Section provides details about the data preparation, model training, hyperparameters optimization, and testing of the single-step forecasting models.

Data Pre-Processing

Generally, any dataset may contain outliers, and invalid values, and data may need to be normalized as per forecasting model requirements. Outliers are extreme or odd values that are unlike other dataset values and their presence may affect the overall distribution of data. However, outliers must be removed to improve the forecasting model's performance. Likewise, the dataset may have a missing or periodic sequence of values known as invalid values that needs to be removed or replaced by some estimated values before modeling. Finally, normalization is performed on the dataset by re-scaling data to fall in the predefined range. Normalization can help the forecasting models to perform better on the data with a smaller scale and improve the convergence speed of the models. The dataset used in this paper is pre-processed for outliers using interquartile range method (IQR) and invalid values by removing them. In pre-processing to replace the missing values, we are grouping data into month, day and hour. Then, the missing values are replaced by taking an average of the available concentration values on same month, day and hour in all years of the dataset. This approach allows for a greater spread of values for the missing data.

The meteorological data is considered an addition input features for the forecasting models. Other than the features available in the dataset, we have also created lagged feature by taking the concentration value of the previous hour of the pollutant being predicted and the datetime index is split into hour, day, and month to create additional features. Therefore, for the prediction of each pollutant, we have considered a combination of features such as previous hour concentration, meteorological (temperature, wind speed and wind direction) and temporal (day, month, hour) information for next hour prediction. The workflow for pre-processing of the dataset is shown in Fig. 1. The input features are normalised using Min-Max normalization as defined by , where xmin and xmax are the minimum and maximum values of data x . Fig. 2 shows a sample of NO_2 before and after pre- processing data.

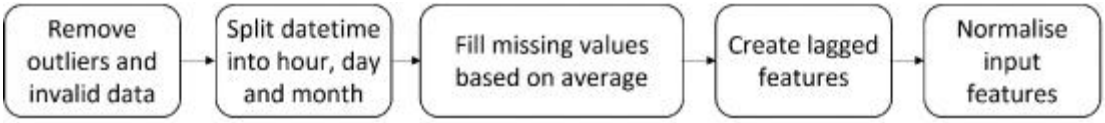

FIGURE 7 Pre-processing of dataset.

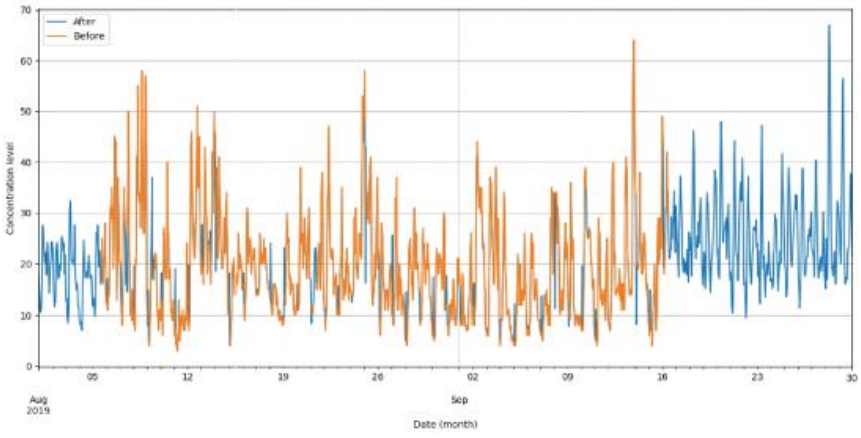

FIGURE.8. NO_2 data sample (over 2 months) representing addition of missing values

Model Parameters and Tuning

The training of the ARIMA model and associated parameters are optimised and tested to forecast air pollutants. We trained the model for predicting all five air pollutants in the dataset. However, for analysis purpose, only NO_2 is explained as a use case. To find optimum values of model parameters such as p, d and q, investigation of ACF and partial autocorrelation (PACF) is required, based on the data and its associated differences. As ARIMA works on the stationary data, we observe ACF of actual data to find if the dataset is stationary or non-stationary. Fig. 7.3 shows that the ACF of NO_2 has all positive values and gradually dropping which indicates that the data is non-stationary. By comparing ACF of 1st order differencingwith the 2nd order differencing , we can observe that the lag at 1 of 2nd order differencing is negative, which indicates that the data may get over differenced. Whereas, this is not the case in the ACF of 1st order differencing and is sufficient to make the data stationary. Thus, we found that the value of d is 1 which defines order of differencing (d).

xnorm=x−xminxmax−xmin(13)

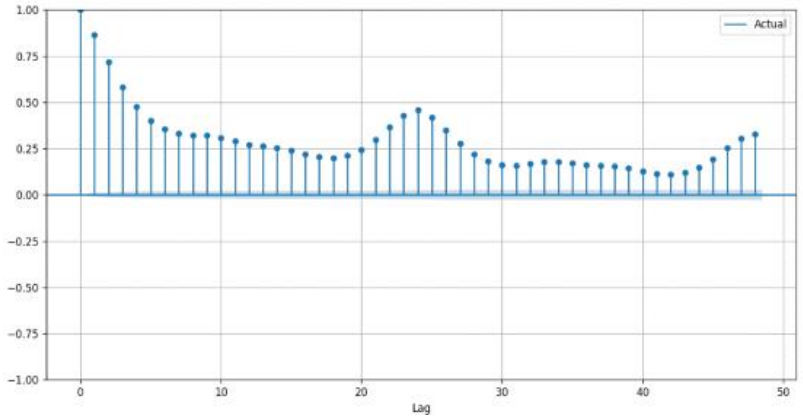

FIGURE 9. ACF of NO_2 data.

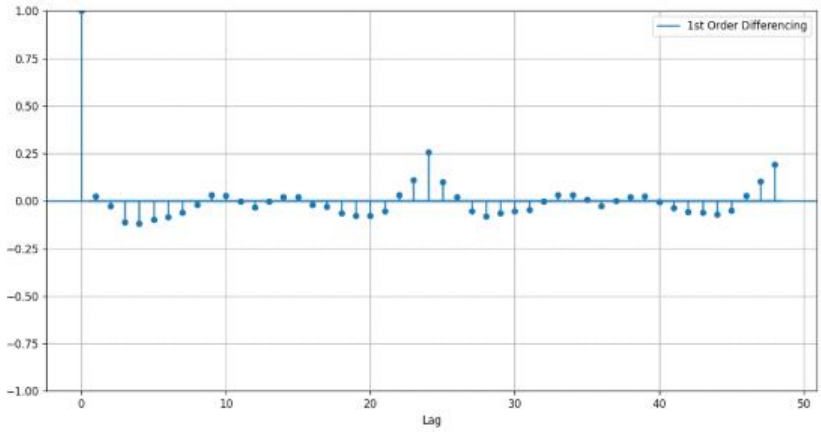

FIGURE 10. ACF of 1st order difference of NO_2 data

Similarly, the order of MA terms can be found based on over differenced data by looking at ACF cuts off point. In our case, 2nd order differencing indicates over differenced due to the lag at 1 being negative and lag at 1 also shows the cut off point, which leads to q equals to 1.

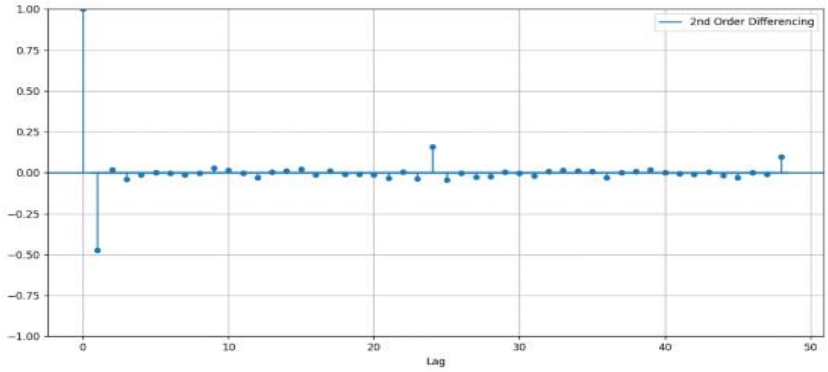

FIGURE 11. ACF of 2nd order difference of NO_2 data.

The order of AR terms is found based on the first cut off point, where PACF at lag 1 is positive. Fig. 6 shows the PACF of NO_2 data. In our case as shown in Fig. 10, lag at 1 of PACF of 1st order differencing is positive and shows the cuts off point, which leads to p equals to 1. Hence, we found (1, 1, 1) as an optimum set of parameters (p, d, q) for NO_2. The detail of the optimum set of parameters for the rest of the pollutants is given Table. 2. In summary, the optimum parameters such as d and q are both found to be 1 for all the pollutants. However, only the value of p is either 0 or 1 for the respective pollutant.

TABLE 1 Summary of Deep Learning Model Parameters

Parameters	Value
No. of layer	1, 1, 5
No. of cells in each layer	5, 5, 30
Dropout layer	0, 0.1, 0.2
Dense layer	1, Relu
Optimiser	Adam
Tuning algorithm	Hyperband

TABLE 2 Optimised Hyperparameters of Single-Step Forecasting Models for All the Pollutants

Models	Parameters	NO_2	O_3	SO_2	PM2.5	PM10
ARIMA	(p, d, q)	(1,1,1)	(1,1,1)	(0,1,1)	(0,1,1)	(0,1,1)
LSTM	Layers	4	3	2	3	3
	Cells	25,10,15,20	20,10,10	30,5	15,15,25	25,20,15
	Dropout rate	0.0	0.0	0.1	0.1	0.0
GRU	Layers	3	3	2	1	4
	Cells	25,30,15	15,10,10	25,25	20	15,20,15,5
	Dropout rate	0.1	0.0	0.0	0.1	0.1

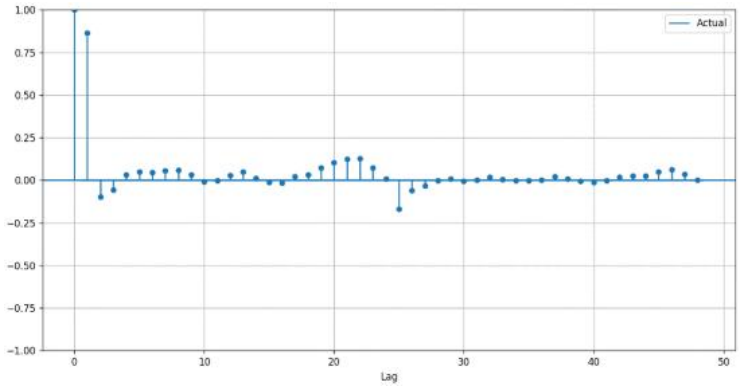

FIGURE 12 PACF of NO_2 data

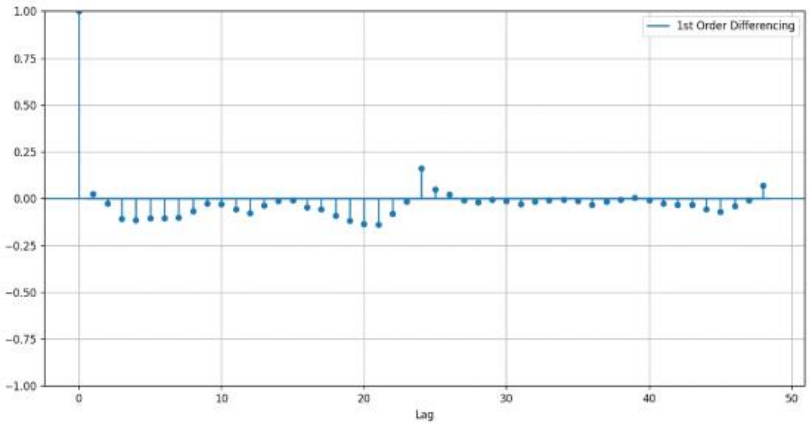

FIGURE 13 PACF of 1st order difference of NO_2 data

The dataset is split into training, validation, and testing sets with ratio of 70%, 20% and 10%, respectively. In each split, the indices kept higher than previous set, which will avoid the shuffling (i.e., inappropriate in time-series). Fig. 8 shows the workflow of deep learning training and testing with key components. Fig. 9 shows the bounds and step size of parameters for model hyper-parameters optimisation in terms of cell size per layer, total number of layers and dropout rate for both LSTM and GRU models. In DL models, input layer pass features to model where we have considered a maximum of 5 layers and each layer can have minimum of 5 and maximum of 30 cells. In DL models, we are optimising number of layers in the range from 1 to 5 with a step size of 1. In case of number of cells, we are considering minimum of 5 to maximum of 30 cells by increasing with a step size of 5.

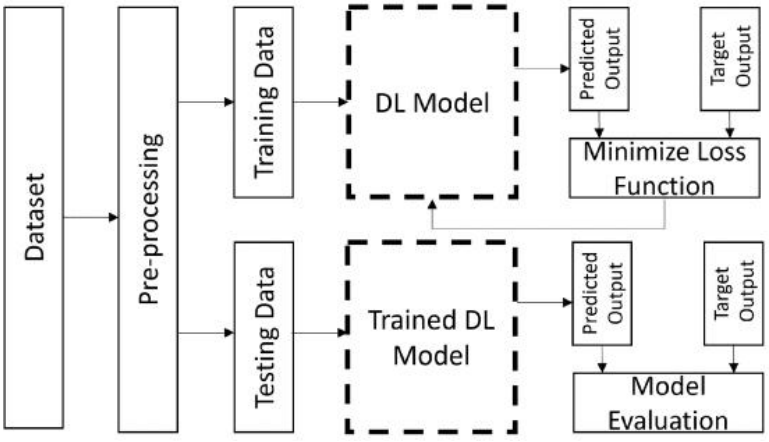

Figure 14 Workflow of deep learning model training and testing.

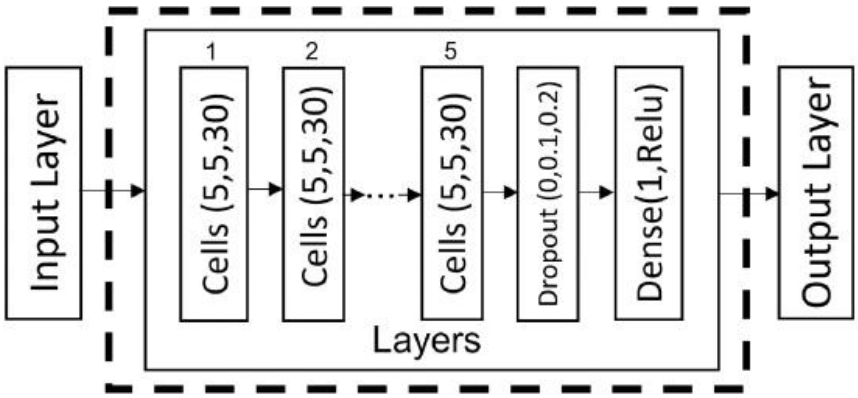

FIGURE 15 Architecture of deep learning model

In the dropout layer, which randomly drops out number of cells to handle over fitting, Dropout rate is also considered as an optimisation parameter. At last, a fully connected dense layer with Relu activation function is used. We are using Adam optimiser during training and for the tuning of hyperparameters, the Hyperband algorithm [31] is used which tunes the number of units in a layer, the amount of layers used and the learning rate of the model. Here, the Hyperband algorithm optimise the hyperparameters by minimising the validation loss during model training to reduce the training time. The summary of the parameters with architectural details of deep learning models is given in Table. 3. After hyperparameter optimisation, the LSTM model used the least number of layers that is 2 for SO_2 and the maximum of 4 layers for NO_2. Whereas, the GRU model used the minimum number of layers that is 1 for PM2.5 and a maximum of 4 layers for PM10. Moreover, LSTM and GRU models used an optimised number of cells in a range of [35 70] and [20 70] for considered pollutants, respectively. In addition, both DL models used optimised dropout rates with the value of 0 or 0.1 for the respective pollutant. The detail of the optimum set of hyperparameters for all the pollutants is given in Table. 4 and these parameters can be further used to evaluate the complexity and computational effort.

Performance Metrics

Performance metrics are measurements used to evaluate the effectiveness, efficiency, and success of a system, process, or activity. They are crucial for assessing performance against goals and objectives and for making informed decisions to improve performance. Performance metrics can vary widely depending on the context, industry, and specific goals, but some common types include:

- Key Performance Indicators (KPIs): These are specific metrics that measure progress toward organizational goals. They are typically quantifiable and directly linked to performance objectives
- Quality Metrics: These metrics assess the quality of products, services, or processes. They may include defect rates, error rates, customer satisfaction scores, and adherence to standards.
- Financial Metrics: These metrics evaluate the financial performance of an organization, such as revenue, profit margins, return on investment (ROI), and cost per unit.
- Operational Metrics: These metrics focus on the efficiency and effectiveness of operational processes. Examples include cycle time, throughput, capacity utilization, and resource allocation.
- Customer Metrics: These metrics measure aspects of the customer experience, such as customer satisfaction, retention rate, net promoter score (NPS), and customer lifetime value (CLV).

RESULTS AND DISCUSSION

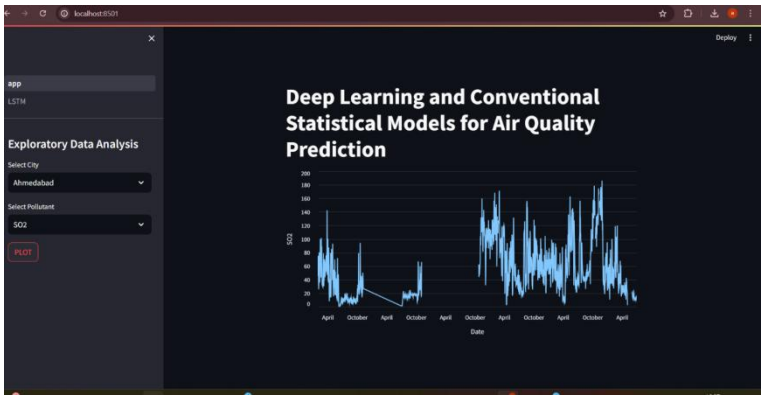

Fig 16. Selection of city and pollutant

We've developed a Streamlit application tailored for analyzing air quality pollution datasets, focusing on specific cities and eight distinct pollutants. The screenshot provided exemplifies the functionality of the application. It showcases the interface designed to facilitate exploration and examination of air quality data. Each city's air quality and its corresponding pollutant levels are comprehensively examined within the application. The tool is specifically structured to allow users to delve into detailed analyses of pollution trends and variations across different locations. The screenshot serves as a preview of the user-friendly interface designed to facilitate efficient data exploration. With this application, users can gain insights into the correlation between urban areas and pollutant concentrations. The interface provides a seamless experience for accessing and interpreting complex air quality data sets. Through interactive visualizations and data filters, users can navigate and analyze pollution data with ease. Overall, the application offers a robust platform for understanding and addressing air quality concerns in various cities.

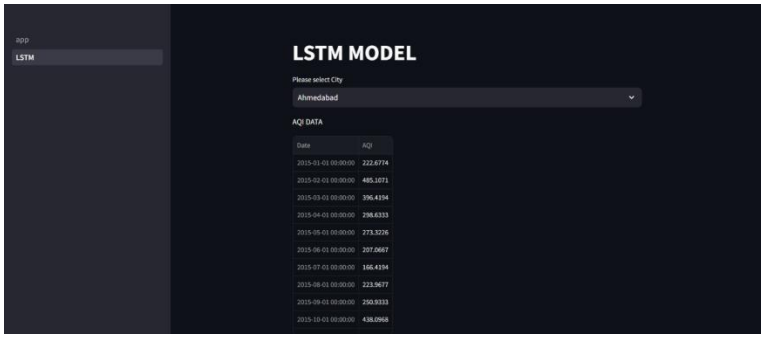

Fig 17 Air quality index of the data with time

The subsequent screenshot features an LSTM (Long Short-Term Memory) model designed to predict specific features such as SO2, NO2, CO, or AQI (Air Quality Index). Within the screenshot, a graphical representation illustrates the performance of the LSTM model. This graph displays distinct colors to represent various data segments: blue denotes training data, green represents testing data, and red indicates future predictions. Through this visualization, users can easily discern the model's accuracy across different phases of data processing. The LSTM model's predictive capabilities are showcased through its ability to generate future projections of air quality metrics. The clear differentiation of colors aids in understanding the model's performance throughout its training, testing, and prediction stages. Users can readily assess the model's efficacy in capturing patterns and trends within the air quality data. The graphical depiction enables intuitive interpretation of the model's predictive accuracy over time. With this visual representation, users can evaluate the reliability of the LSTM model in forecasting air quality metrics. Overall, the screenshot provides valuable insights into the LSTM model's predictive performance and its potential application in forecasting air pollution levels.

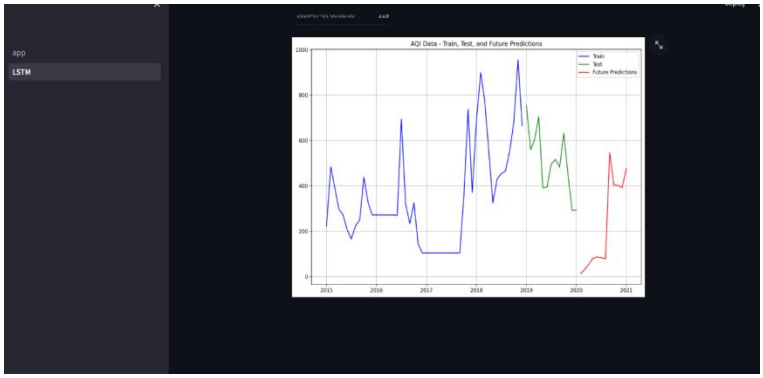

Fig 18 Train, test and prediction of the data

Conclusion

Air pollution is a global health challenge and its accurate prediction is vital to reduce health risks and environmental concerns. This work aims at single-step air pollution prediction for most of the pollutants (e.g. NO_2, O_3, SO_2, PM2.5, PM10) using various forecasting approaches based on DL and statistical models. The performance of the forecasting models is tested using evaluation metrics such as RMSE, MAE and R2 . At the broader level, among all the forecasting models and pollutants, LSTM achieved the lowest RMSE and MAE of 0.591 and 0.396 respectively, in predicting SO_2 time series data, whereas the highest RMSE and MAE is found to be 9.354 and 6.065 respectively, for NO_2 by ARIMA model. In terms of R2 , among all forecasting models, both DL models performed similar in achieving the highest score of around 86% while predicting O_3. On the other end, GRU model is the one found to be having least predictive accuracy of around 55% for SO_2. Overall findings through results revealed that among all considered forecasting models, DL models outperform

statistical models consistently in achieving the least error in terms of RMSE and MAE for all the pollutants and attained better predictive accuracy in terms of R2 for most of the pollutants. While ARIMA model could only perform better in predicting two pollutants (i.e. SO_2 and PM10) in terms of R2 score only, however with the highest error value of RMSE and MAE for all of the pollutants. In future, we aim to target multi-step prediction and improve the performance of the DL models using new feature engineering approaches and relating optimization of hyper-parameters of the models

FUTURESCOPE

As air pollution continues to be a pressing global issue, the project holds immense potential for further expansion and enhancement. One avenue for future development lies in the integration of real-time data streams from various sources such as sensors, satellites, and weather stations. By incorporating live data feeds, the application can provide up-to-the-minute insights into air quality dynamics, enabling more timely and proactive interventions. Additionally, leveraging advancements in machine learning and AI techniques can enhance the accuracy and predictive capabilities of the models employed within the application. This could involve implementing more sophisticated algorithms, such as ensemble methods or deep learning architectures, to better capture complex patterns in air pollution data.

Furthermore, there's an opportunity to scale the project to encompass a broader geographical scope, encompassing additional cities and regions worldwide. By expanding the coverage area, the application can offer insights into air quality trends on a global scale, facilitating cross-regional comparisons and enabling policymakers to implement targeted interventions where they are most needed. Collaborations with governmental agencies, environmental organizations, and research institutions could further enrich the project by providing access to proprietary data sources and expertise. Additionally, incorporating user feedback mechanisms and community engagement features can foster a sense of ownership among stakeholders and empower individuals to contribute to air quality monitoring efforts. Overall, the project has the potential to evolve into a comprehensive, data-driven platform for addressing air pollution challenges and promoting environmental sustainability on a global scale.

References

1. G. Box, G. Jenkins, G. Reinsel, and G. Ljung, Time Series Analysis: Forecasting and Control (Wiley Series in Probability and Statistics). Hoboken, NJ, USA: Wiley, 2015.
2. M. Stafoggia and T. Bellander, "Short-term effects of air pollutants on daily mortality in the Stockholm county, Sep. 2020.
3. A. Bekkar, B. Hssina, S. Douzi, and K. Douzi, ‘‘Air quality forecasting using decision trees algorithms,’’ Mar. 2022,
4. B. Wang, W. Kong, H. Guan, and N. N. Xiong, ‘‘Air quality forecasting based on gated recurrent long short term memory model in Internet of Things,’’2019.
5. B. Paul and S. Louise, Air Quality: Policies Proposals and Concerns—House of Commons Library, Jan. 2022,
6. N. Zaini, L. W. Ean, A. N. Ahmed, M. A. Malek, and M. F. Chow, ‘‘PM2.5 forecasting for an urban area based on deep learning and decomposition method,’’ Oct. 2022
7. K. Abutalip, A. Al-Lahham and A. El Saddik, " Digital twin of atmospheric environment: Sensory data fusion for high-resolution PM 2.5 estimation and action policies recommendation ", 2023.
8. J. Ma, Y. Ding, J. C. P. Cheng, F. Jiang, Y. Tan, V. J. L. Gan, et al., "Identification of high impact factors of air quality on a national scale using big data and machine learning techniques", J. Cleaner Prod., , Jan. 2020.
9. I. Manisalidis, E. Stavropoulou, A. Stavropoulos and E. Bezirtzoglou, "Environmental and health impacts of air pollution: A review", . 14, 2020.

10. S. Ameer, M. A. Shah, A. Khan, H. Song, C. Maple, S. U. Islam, et al., "Comparative analysis of machine learning techniques for predicting air quality in smart cities", , 2019.