Journal of Intelligent Computing System (JICS), Vol.1, Issue. 1, pp. 1-15, (Jan 2026)
ISSN: 3139-3616 (online)

# Prediction and Analysis of Air Quality Index in South Asian countries using Deep Learning and SARIMAX

Souptik Kumar Chaki

Senior Analyst, Solutions Architect, Dell Technologies, Bengaluru
souptik2018@gmail.com

**Abstract:** Poor air quality can negatively affect the environment and human health. Air pollution contains unstable atoms that can attack healthy cells, damaging the body. Inhaling pollutant air affects the respiratory system, eyes, nose, throat, heart, and blood vessels. It deteriorates many breathing and lung illnesses, leading to cancer, or premature death. In South Asian countries, one of the main issues with suburban as well as urban areas is Air Quality Index (AQI). It is crucial to analyze and forecast current harmful emissions leading to poor air quality for sustainable urban development in the respective cities. Some of the gaps have been identified in existing methods, like the availability of the proper data, appropriate attributes, and other external factors like the population that can also be used to predict and forecast AQI. In this research, we introduced a hybrid Deep Learning (DL) model, including Bidirectional Long Short Term Memory (BD-LSTM), that uses SARIMAX for Time Series Forecasting to extract comprehensive features across spatiotemporal data exhibiting a 73.61% accuracy. Extensive model evaluations are done on data containing 10 different harmful pollutants that are responsible for creating a poor AQI along with the Population of the particular city which is taken as an added factor for AQI forecasting. Proposed model offers better accuracy of 6.7% than previous models.

***Keywords****—Machine Learning; Deep Learning; Air Quality Index; Time-Series Forecasting; Population.*

## 1. Introduction

The AQI is impacted by the increasing urbanization, population growth, along industrial expansion, which have resulted in several health risks, including congenital heart problems and lung diseases. Lung cancer, heart attacks, asthma, and other respiratory and cardiovascular conditions can result from prolonged exposure to poor air quality, according to World Health Organization (WHO). Approximately 4.2 million fatalities annually are attributed to lung and respiratory disorders, according to WHO estimates, and nearly 7 million premature deaths are resulted by poor air quality across that city [1]. According to the Yale and Columbia University researchers' study in June 2022 [2], India has been ranked 180, the lowest among all other countries, with a score of 18.9 on the 2022 Environmental Performance Index (EPI). In contrast, Bangladesh is ranked 177 with an EPI score of 23.1, and Pakistan is 176 with an EPI score of 24.5. The EPI is responsible for ranking 180 countries on performance factors such as climate change, air quality, etc. Aerosol particles containing delicate particulate matter having diameter of less than 2.5um are a significant concern in most South Asian cities. The harmful effects of poor AQI in urban and suburban areas can be understood by analyzing and forecasting the emission levels of various toxic gases. The monitoring stations in the city and the government agencies can take the necessary steps to control the emission level of these harmful gases [3]. Information regarding AQI and emission levels of toxic gases are maintained by different observational and monitoring stations in that particular city. The data being collected is highly dynamic and has various spatiotemporal relations; hence it is essential to analyze the information correctly so that the correct representation can be taken out of it and can also be used for future forecasting.

Criterion pollutants are considered most common pollutants that result in health issues. SO2, O3, and NO2 are the most common criterion pollutants. Health issues like airway inflammation in healthy individuals, worsened respiratory symptoms in asthmatics, trouble in meeting high oxygen demands during exercise, and critical respiratory situations particularly in children and the elderly are caused by short-term exposure to these pollutants. EU, EPA, and other national organizations have established acceptable air quality criteria. Criteria pollutant levels in the air can be indicated by AQI. Highest AQI noted for each criterion pollutant is the overall AQI. AQI levels also indicate the health concerns connected to exposure to a specific air quality. These health issues could develop over time or soon after being exposed to air pollution. The age and health status of the specific exposed individual may also affected due to these symptoms. Therefore, we must have a system in place to forecast the elevations in air pollution levels so that government agencies can take necessary steps to implement emergency response or on-demand pollution control measures to prevent additional increases [7]. As a result, the





*doi.org/10.67420/109319.1.1.1*

AQI would be easier to regulate to suit the general demands of the population. To protect the population from exposure to the pollutants once AQI forecast surpasses the allowed level, authorities can alert the public by looking into the forecast.

Due to complex and non-linear nature of data collected through several sources, which must be cleaned for future model prediction, air quality forecasting is an extremely difficult task. Analyzing diverse data collected from multiple sources is vital compared to ground-level data, which provides more direct information. The only challenge that is being faced in monitoring ground-level data is the uneven distribution of data, as the information received from urban areas is denser as compared to rural areas [4]. Machine learning-based prediction techniques are growing in popularity as a result of artificial intelligence (AI) and development of big data because knowledge of physical and chemical characteristics of air contaminants is not necessary to employ these kinds of models. The most widely used machine learning (ML) methods, which incorporate intricate non-linear correlations between air pollution concentrations and meteorological variables, include multiple linear regression (MLR), random forest (RF) [19], support vector regression (SVR) [20], and artificial neural networks (ANN) [21]. Numerous ANN frameworks, encompassing Neuro-Fuzzy Neural Network (NFNN) [22] along with Bayesian Neural Network [23], have been created to forecast air pollution in various research domains.
Multiple ML algorithms were used in an ensemble approach to provide a reliable as well as accurate assessment of pollution levels across Greater London region. In traditional methods, the pollution level was determined by various physical and chemical processes. Chemistry–Transport Models (CTM) were used to analyze the data and provide complete detailed information regarding the different monitoring sessions. However, it suffered from uncertainties due to the presence of inaccurate data. In the 21st century, much progress has been made using the IoT and various sensor technologies to collect accurate AQI data from multiple monitoring stations, like PM2.5, NOx, SO2, etc., [5]. Since many tractions were observed with the traditional forecasting models, DL models have proved to address these issues by providing more insights into them. With proper use of the DL models, the various inherent features of the data can be obtained and analyzed, which can further be used to forecast the behavior of the intrinsic attributes shortly. Along with DL in many cases, Multivariate Time Series like ARMA, ARIMA, and MLR regression analysis are also used to give a proper interpretation of future predictions of air quality in particular city.

Objective of this work is to understand techniques used to determine the AQI of a particular region using unsupervised ML algorithms like LSTM, BD -lSTM, and other supervised models like KNN and random forest. This method is analysed and gaps are identified win these techniques which leads to the proposed model which is BD-LSTM with SARIMAX that addresses the gaps identified in the previous techniques. Time series data has been taken along with other parameters as well like population of that particular area which was identified as a gap in the previous algorithms implemented. The proposed methodology has been tested and validated with new parameters and has been compared with other algorithms to provide better understanding of differences that are being encountered with other algorithms. In this paper, the concept of a DL-based model like LSTM, BD-LSTM, and EN-DC LSTM is implemented for predicting and forecasting AQI across cities belonging to South-Asian region.

**Organization of Paper**

This paper has been categorized into various sections, as specified. Section2 examines related work conducted by various researchers. Section3 discusses the various attributes considered for the dataset and will explain the proposed work. Section 4 discusses the proposed model formulation using BD–LSTM and SARIMAX time series and analyzes the various baseline DL-based models considered for predicting AQI. Finally, Section 5 concludes the proposed work, followed by future scope based on the results obtained.

## 2. Related Work

This section reviews the various research models implemented regarding air quality prediction.

R. Janarthanan et al. (2021) suggested Support Vector Regression (SVR) as well as Long Short-Term Memory (LSTM) based DL model for classifying AQI values in Chennai. Authors refined the dataset through Exploratory Data Analysis (EDA) to substitute missing values as well as eliminate redundant data. Authors asserted that proposed model enhanced prediction accuracy compared to other existing models. X. Zhao et al. (2018) introduced a DL model called Recurrent Neural Network (RNN) for predicting Air Quality Classification (AQC) in 3 distinct industrial cities in US. Authors utilized sequential data regarding air quality issues, resulting in the suggested RNN model achieving superior accuracy in comparison to other supervised models encompassing Random Forest as well as SVM. Q.Liao *et al.,* [5] (2020) introduced different DL architectures like CNN, LSTM, and ANN, showcasing the importance of exploring the nonlinear relationships between different parameters required to analyze air pollution. They further examined the strength of DL networks on the various aspects of the diverse data, improving algorithm performance for atmospheric forecasting. S. Abirami et al. [6] (2021)

introduced a deep learning system called DL-Air, which comprises 3 fundamental components related to air quality forecasting; initial component is encoder designed to capture all spatial and different relationships within the data. Second proposed component was the STAA-LSTM model, an enhanced variant of LSTM designed to comprehend all temporal relationships and their correlations. Third crucial element is the Decoder, utilized to convert it into an actual forecast. The authors assessed proposed framework using actual air quality data from Delhi, where DL-Air algorithm demonstrated superior performance by decreasing the RMSE score by up to 30%. W.Mao *et al*., [7] (2021) proposed a DL-based model known as temporal sliding LSTM to predict air quality. This model was able to analyze and combine the optimum lag in time to understand the prediction of a bi-directional neural network architecture which included the hourly data of PM-2.5 concentrations across Jing-Jin-Ji region, which had most severe air quality issue in China. S.Du *et al.,* [8] (2019) proposed a novel DL-based model named one-dimensional CNN and BD-LSTM for forecasting air quality such as PM 2.5. The model learned various Spatial-temporal features of data, which helped to analyze the dependency between different attributes for the time–series prediction. Y. Lin *et al*., [9] (2018) presented an approach using a DL algorithm that is Geo-Context-Based Diffusion convolutional RNN to predict short-term PM-2.5 concentrations within particular area. Finally, the model described the relationships and similarities between different locations of air quality sensors. Y. Zhou *et al.,* [10] (2019) suggested deep multi-output LSTM model for predicting air quality. Proposed model had been evaluated by employing $PM_{2.5}$, $PM_{10}$, along with $NO_x$ at 5 air quality monitoring stations across Taipei City. Finally, Mean Squared Error (MSE) of 0.72 along with $G_{bench}$ of 0.81 was attained for testing stage at horizon t+4, respectively.

T. H Do *et al.,* [11] (2020) considered the concept of an IoT pipeline that addressed three things: acquisition, processing, as well as visualization of air pollution data, carried out across Antwerp City, Belgium. The authors deployed multiple IoT devices on different vehicles and reference stations used to measure various parameters like humidity, temperature, and air pollution. The proposed model was able to show correlation between air pollutants, which led to an enhanced estimation of AQI. The authors conducted different experiments to understand air pollutants' spatial and temporal qualities. They were compared with state-of-the-art approaches to demonstrate efficiency in their method. C.Y Lin *et al*., [12] (2021) authors used a Gated Recurrent Unit (GRU) architecture model to predict different predictive models such as GRUST14d, GRUAW13d, etc. The MLR-based GRU forecasting model was proposed using the regression technique to integrate used predictive models. The author concluded that the proposed model addressed different spatial and temporal situations and provided better forecasting accuracy. Z. Zhang *et al*.,[13] (2021) proposed a hybrid DL architecture named Variational Mode Decomposition BDLSTM (VMD-BD-LSTM), a combination of VMD and BD-LSTM. The proposed model was used to predict changes in emission levels of PM-2.5 in most populated cities across China. Authors concluded that proposed model is most stable predicting algorithm and provides better accuracy. A. Tiwari *et al*.,[14] (2021) proposed the latest versions of LSTM-based DL models like BD-LSTM as well as Encoder-Decoder LSTM (ED-LSTM) model to analyze and forecast AQI in Delhi. Proposed model had been evaluated by employing multivariate time series method for forecasting air quality in 10horizons. Finally, the authors claimed that multivariate BD-LSTM showed better accuracy than other models despite the effect of COVID-19 on AQI. N.Bruni *et al*.,[15] (2020) proposed a DNN model which captured $PM_{10}$ seasonal visibility and showed its impact on the air quality. According to this study, the region's long-term exposure to PM-2.5 is causing an increase in mortality rates; in 2005, there were 150,000 premature deaths (108,000–193,000), and in 2015, there were 204,000 (145,000–260,000).

S. Sannigrahi *et al.,*[16] utilized satellite as well as surface measurements to analyze and forecast air pollution levels among 20 cities. Authors conducted a time-series and spatial assessment of various harmful gases, such as NO2 and CO, observed from Google Earth Engine (GEE) as well as Tropospheric Monitoring Instrument (TROPOMI) applications. Finally, the authors concluded with how emissions can be reduced using various anthropogony methods, leading to various socio-economic benefits in those areas. Lee *et al.,*[17] (2021) estimated that The country's poor air quality index (AQI) is mostly caused by the fact that 77% as well as 9% of kilns are (illegally) located within 1km of schools as well as health facilities, respectively, and that at least 12% of Bangladesh's population (more than 18million people) reside within 1km of kiln. The author proposed an LSTM model which has been used to understand spatial status of $PM_{10}$. S. M Cabaneros et al., [20] have given a study where ANN models are implemented to forecast air pollution. The authors analyzed the various ANN methods implemented by researchers to forecast emission levels of harmful gases such as PM2.5 and PM10. In conclusion, authors found that most research to date has used Ad-Hoc, ensemble approaches along with Multi-Layer Perceptron (MLP) to anticipate AQI. M. Bruke et al. [18] (2021) investigated several ML techniques used currently for data construction, particularly in cases with extremely noisy and limited data. ML algorithms have been used to examine images from satellites in recent research techniques. In order to address spatial along temporal aspects of data, model took into consideration the sustainable development component, which emphasized about need for more satellite imagery with higher resolution. A gray model and rolling mechanism were used by Yves et al., [23] (2018) to improve CNN model. In 2016, Ravindra et al. [24] employed aggregated LSTM networks as well as compared outcomes employing SVR as well as gradient-boosted tree regression.

According to PM-2.5 concentrations during different time periods, Zaman et al. (2017) forecasted air quality using 3 ML techniques: multiple additive regression trees, deep feedforward neural networks, as well as a hybrid model based on LSTM networks. LSTM model was most effective in forecasting as well as controlling air pollution. In 2020, Bo and Zheng [22] investigated air pollution in 44 Chinese cities to assess if COVID-19 lockdowns had an impact. Human mobility mediated the decrease in the air quality index, as demonstrated by the scientists, who also found that lockdowns in 44 cities decreased human movement by 69.85%. In their 2021 investigation, Y. Zhou et al. [31] evaluated correlation between COVID-19 pandemic as well as the weather in Jakarta, Indonesia, taking into consideration changes in temperature, humidity levels, along precipitation totals. According to the investigators, there was a strong correlation between the COVID-19 pandemic and simply the average temperature. V. Sivakumar *et al.,* [32] (2022) investigated several models for the air pollutant prediction problem, including regression, multiple layer perceptrons, and physical and chemical models. The findings indicate that statistical models can compete with traditional physical as well as chemical models. K. Stebal *et al.,* [51] (2018) suggested a downscaling algorithm-based model to close the gap between Norwegian, Polish, and Romanian satellite data. The model's accuracy was 65%, according to the authors. N. F. K Zaman *et al.,* [52] (2019) suggested an ANN system to anticipate the PM level using a dataset of 16 Malaysian cities. K.Sidhu *et al.,* [54] (2024) proposed a model-based Dickery -Fuller Test which tested the stationary of the data before being fed to ML algorithms encompassing Random forest, CatBoost as well as SVM regressor. According to authors Random Forest performed better than other regression models. Waqas M *et al.,*[55] (2024) used LSTM and SARIMAX models to predict NO2 concentration in major U.S Cities. The models were trained on AQI data available from 2015- 2019. The paper highlighted the disparity between NO2 emissions during the pandemic time. Sadiku R *et al.,* [56] (2024) propose a study for dispatchable power generation between 2019 to 2023 The author used Sarimax model to train its data which predicted an accuracy of 69.5 %.

The authors reported an accuracy of 66% with their model. Similarly, A.R. Ganguly et al. (2021) [53] suggested the DeepAQ model to forecast AOD levels in Ghana, leveraging patterns identified in Los Angeles and New York, achieving an accuracy of 63%. Table1 summarizes accuracy of existing algorithms on basis of findings from literature review.

**Table 1.** Existing algorithm accuracy

| Name of Algorithm | Accuracy in Percentage |
|---|---|
| LSTM | 52.4% |
| EN-DC LSTM | 51.66% |
| BD-LSTM | 58.61% |
| VMD-BD-LSTM | 63.2 |

As per the analysis mentioned above, the researchers have considered specific countries like China and Malaysia and some specific cities in India like Chennai and Delhi. Most of the investigation was done on PM2.5 as well as PM10 composition. In present investigation, we propose DL models like LSTM, BD–LSTM, and time series forecasting using SARIMAX on an extended version of the dataset. The six countries have been considered across South–Asia as India, Nepal, Pakistan, Bangladesh, Bhutan, and Sri Lanka, in which increased urban population is a significant issue. An extensive increase in the population leads to poor AQI within cities of South Asian countries.

## 3. Material and Methods

As per analysis by previous researchers emissions have increased because of the increasing population which has, in turn, led to an increase in both primary and secondary pollutants which affects the ozone level concentrations, having said that it shows that any increase in population or population density can put some changes in the ambient air quality. The way the population is growing in South Asia, the urban spread of industrial emissions is also increasing where a large portion of the airshed is being occupied with industrial and chemical pollutants. With rise in poor air quality critical health outcomes can be seen in public offspring as they will be exposed to harmful pollutants in air which can be compounded by the greater vulnerability as and when the population increases. This section shows the methods in which the models have been implemented along with the framework. Here the goal is to analyze and predict the changes in the AQI for the coming years. The time is taken from 2016-2020, where the data has been maintained by various ground-based monitoring stations, which play a crucial role in determining early warning signals for poor AQI and defining the necessary steps to control air pollution. Regional and synoptic meteorology is essential in determining air pollution concentrations in different meteorological conditions. Factors like Lower wind and higher wind speed can be one of the reasons for increasing the traffic pollutants in the air. In addition, other factors like humidity levels are a statistical factor in increasing industrial pollutants like CO2, Nitrogen (NO), SOX, etc. Due to mixed precipitation in various areas, high humidity is observed along with wet deposition, which lowers air pollutant concentration across respective areas. In higher concentrations of PM2.5, visibility can be reduced due to different particle compositions. In some cases, air pollutants like ozone (O3) can be reduced by the presence of other contaminants in cloud cover as they help absorb as well as scatter solar radiation.

### 3.1. Dataset

The dataset considered for the model formulation has been collected from two economic data repositories such as Kaggle [29] and Central Pollution Control Board (CPCB) [30]. Kaggle dataset consists of 26 cities in India. These belong to two categories: high population growth with poor AQI (such as Delhi, Bangalore, Ahmedabad, etc.) and less population with better AQI (such as Shillong, Coimbatore, etc.). CPCB dataset contains the emission level of harmful gases present in the cities of South Asian nations like Nepal, Pakistan, Bangladesh, Bhutan, as well as Sri Lanka, along with their AQI. The dataset comprises 12 pollution metrics encompassing PM2.5, PM10, NO, NO2 NOx, CO, SO2, O3, Benzene, Toluene, as well as Xylene, which performs critical functions in calculating AQI of respective cities. In addition, the city's population is considered a factor contributing to AQI in the city. The dataset contains 26947 entries within a timeframe of 2015-2020. Further, dataset has been divided into 2 sets where 70% original dataset is considered for training while remaining 30% for testing proposed model.

Table 2 shows range of values for 14 parameters considered in the dataset with the time frame of 1$^{st}$ January 2016 and 10$^{th}$ December 2020 from different monitoring stations.

**Table 2**. Details about the Features Considered

| Pollutant Name | Min | Max | Mean | Std |
|---|---|---|---|---|
| PM10 | 0.03 | 1000.0 | 110.424 | 74.74 |
| NO | 0.02 | 390.68 | 16.2 | 21.88 |
| NO2 | 0.01 | 362.21 | 28.263 | 23.804 |
| Xylene | 0.0 | 170.37 | 2.15 | 4.17 |
| CO | 0.0 | 175.81 | 2.264 | 6.84 |
| Benzene | 0.0 | 455.03 | 3.06 | 14.884 |
| NOx | 0.0 | 467.63 | 30.139 | 28.399 |
| O3 | 0.01 | 257.73 | 34.239 | 20.740 |
| SO2 | 0.01 | 187.02 | 13.77 | 17.00 |
| NH3 | 0.01 | 352.89 | 21.424 | 21.891 |
| Toluene | 0.0 | 454.85 | 7.73 | 17.835 |
| PM2.5 | 0.04 | 949.99 | 71.0344 | 65.0644 |
| Population | 529 | 13205697 | 3738839 | 3967419 |

### 3.2. Data Preprocessing

To estimate different particulate matter, test representation skills are necessary. Satellite data is eliminated by averaging pixel values that fall within specific radius across all significant measuring stations. Radius selection may be altered according to requirements for different levels of sensitivity that may be utilized to determine minimum radius by taking advantage of correlation between daily particulate matter as well as AQI while keeping non-missing data for full specified time-period. AOD averages out just small proportion of all pixels inside designated radius for data that is present.

### 3.3. Methods

In this section, the various unsupervised algorithms are implemented and compared to check spatial features for various particulate matters that are being used to determine AQI.
The methods are listed below in sub-sections.

#### 3.3.1. SARIMAX time series.

SARIMAX is ARIMA model's extended version that has been used for forecasting
dependent values based on the time–series data. ARIMA models are generally made of two important components which are moving-average term and Autoregressive Term. In ARIMA the values are seen as a weighted sum of previous values whereas in SARIMAX the same technique of weighted sum is applied but not with all the residuals. The SARIMAX algorithms have been used on the datasets because of the seasonal data with both seasonality and external factors. The formula for SARIMAX is shown in Equation 1

$$Yt=\phi p\cdot Yt-p+\theta q\cdot \epsilon t-q+\Phi P\cdot Yt-sP+\Theta Q\cdot \epsilon t-sQ+Xt\cdot \beta+\epsilon t \qquad Eq.\ (1)$$

The seasonality component captures recurring patterns in data over a specific period. In this study, the seasonality component considered is the population of a city on an annual basis. To compute seasonality, the Seasonal Moving Average (SMA) technique is employed. SMA is determined by calculating the average of values over a given time period, as illustrated in Equation 2 below

$$SMA(t) = (Y_{\{t-k+1\}} + \ldots + Y_t)/k \qquad \text{Eq. (2)}$$

Here $Y_t$ represents observed time series data at time t.
k represents number of periods in the moving average.

### 3.3.2. Bi-directional LSTM

An expansion of conventional LSTM, bidirectional LSTM can enhance the ordinal classification problem model's performance. Two LSTMs are trained via bidirectional LSTMs rather than just one of the input sequences.
The first is feedback as it is, the second is the opposite of feedback. This can provide more information to the ne twork and lead to a faster and more thorough investigation of the issue.
BD-LSTM is hybrid of bidirectional RNNs as well as LSTM models that has access to long-range context in both input channels. Input sequences with preset starting and ending points are the target of BD-LSTM. They analyze each element's past as well as future states in a sequence, with one LSTM processing the data from start to finish along with others from finish to start. Their combined outcomes predict appropriate labels, if available at each time step. If the labels are available at each time step, their combined outputs forecast them. In contrast to a unidirectional technique, this method retains information from past as well as future at every time step by combining information from past as well as future LSTM model states. In sequence processing tasks including continuous voice recognition, phoneme classification, speech synthesis, as well as sentence classification, BD-LSTM has been widely used.

## 4. Proposed Methodology

Proposed Algorithm is implemented on dataset, which is used to forecast AQI in Southeast Asia. The First proposed algorithm implements DL-based models like LSTM, BD-LSTM, and SARIMAX time series to determine the AQI forecast for Southeast Asia. The Air pollution database which contains data from satellite data, monitoring stations, and metrological data goes through scaling and feature extraction techniques after which it is trained on DL-based models (such as LSTM and BD-LSTM) to identify the dataset's Spatio-Temporal features. Subsequently, the processed data serves as input to the SARIMAX time series algorithm to forecast AQI and emission levels of other harmful gases for the upcoming year. Further, two hidden layers are considered, with each layer having around 50 cells for the LSTM models. The input layer dimension is taken as (5,11), corresponding to $(N, N_f)$, along with ten output layers considered for the experiment.

In this experiment, ADAM optimizers having a batch size of 25 each are executed with 200 epochs to determine the activation function. We used the Relu activation function on hidden layers to support LSTM model. Relu functions are used to map the derivative of the positive ones to 1, and on the other hand, the negative ones are mapped to 0. Eq (3) and Eq (4) show the formula used for Relu activation functions.

$$F(x) = \max(a,0) \text{ if } x >= 0 \qquad \textit{Eq. (3)}$$

$$F(x) = \alpha\max(a,0) \text{ if } x < 0 \qquad \textit{Eq. (4)}$$

In contrast to RELU functions, there is an exponential RELU function, which is used to overcome shortcomings like zero gradient issues and bias shift issues of normal RELU functions. Eq. (5) and Eq. (6) show the formula for Exponential RELU functions.

$$F(x) = \max(a,0) \text{ if } x > 0 \qquad \textit{Eq. (5)}$$

$$F(x) = \alpha\,(e^x - 1) \text{ if } x <= 0 \qquad \textit{Eq. (6)}$$

In proposed model, dataset has been divided into training, testing, as well as validation sets for tuning data using hyperparameter methods to obtain a better prediction. The proposed model of the research work has been demonstrated in Figure1.

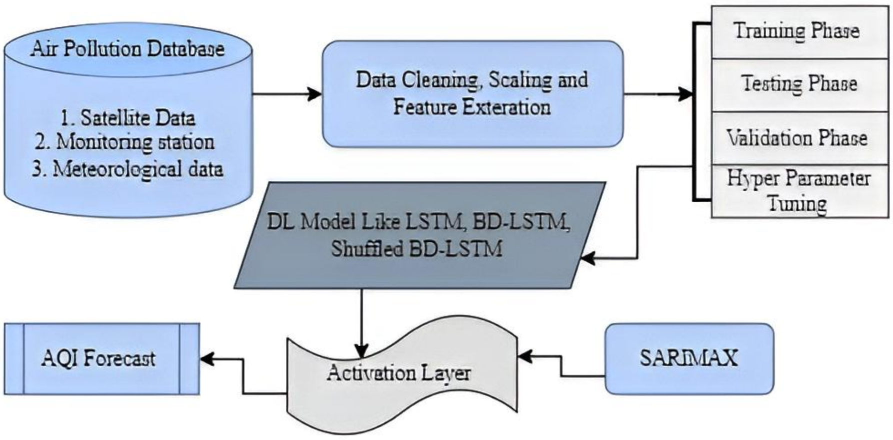


**Figure 1**. Proposed Model

### 4.1. Architecture and Working

This section describes DL-based proposed algorithms for analyzing and forecasting AQI by considering the emission levels of various industrial and chemical pollutants, along with the city's population.

**Algorithm 1**. AQI estimation using Shuffled BD-LSTM and SARIMAX time series SARIMAX time series

**Input**: -

Emission Levels of 12 dangerous pollutants include PM2.5, PM10, NO, NO2 NOx, CO, SO2, O3, Benzene, Toluene, as well as Xylene, along with city's population.

**Output:** AQI Score w.r.t to the emission levels in that particular city

**Data**: Air Pollution Data

Step 1: - Initialize preprocessing of the Spatio-Temporal dataset

Step 2: -Use Standard Scalar to transform the variables into a standardized format.

Scaled_dataset = S(D) *Eq. (7)*

D = Non-Scalar dataset

S = Scalar Function

Step 3: - Perform Feature Extraction Process to extract the necessary attributes required for further model training.

Step 4: - DL algorithm approach (LSTM, BD–LSTM, Shuffled BD–LSTM) along with SARIMAX time series.

$y_{outj}$ (t) = denotes output gate with activation time t.

$y^{inj}$ (t) = denotes input gate with activation time t

$net_{outj}$ = denotes network of hidden cells before the output gate, along with the summation of weight attributes on the output layers

$net_{inj}$ = Represents the network of hidden cells after the input gate, along with a summation of weight attributes on the input layers.

$S_i$ = SARIMAX time layer.

For i=1 to n:

Where n represents number of subsets of data fed into the model.

$y_{outj}(t)n = f_{outj}(net_{outj}(t)n); \; y^{inj}(t) + S_i = f_{inj}(net_{inj}(t));$ *Eq. (8)*

where

$net_{outj} = \sum u \; W_{outj} \; u \; y^u(t-1)$ *Eq. (9)*

and

$net_{inj} = \sum u \; W_{inj} \; u \; y^u(t-1)$ *Eq. (10)*

$net_{cj} = \sum u \; W_{cj} \; u \; y^u(t-1)$ *Eq. (11)*

Here $W_{outj}$ and $W_{inj}$ represent the weights used between different layers for making a connection between the neurons. $net_{outj}$ and $net_{inj}$ determine the net output and net input for different layers. $net_{cj}$ and $W_{cj}$ determine the averaged output and weights for different layers.

Step 5: - Evaluate performance of various DL models by employing performance metrics like RMSE and MSE scores.

DL-based models like LSTM, BD-LSTM, and Shuffled LSTM are used to analyze the attributes' spatiotemporal features, which helps forecast AQI. As per the proposed algorithm, LSTM, BD-LSTM consists of 4 layers: input gate, output gate, as well as forget gate, an activation function like tanh, sigmoid, etc. In the proposed algorithm first dataset is cleaned by finding the missing values and replacing them with standard values like mean, median, or mode. The dataset is then put into a scalar function to scale all the attributes to one unit type using Eq. (6). Once the data is scaled, the necessary attributes required for forecasting AQI are extracted. The updated dataset is fed into the LSTM model where the input gates $y_{inj}$ (t) and the output gates $y_{outj}$ (t) with the activation time t. Eq. (8) defines the output obtained by calculating the results obtained from all hidden layers. Eq. (9), Eq. (10), and Eq. (11) determine the weights calculated between the hidden layers. The weights assigned to connections between the layers are averaged out to get a more scaled output.

## 5. Experiments and Results

This section compares performance of DL-based algorithms like LSTM, BD-LSTM, and other unsupervised learning techniques. Their performance is compared and analyzed to understand the temporal behavior of the harmful gases for the near future prediction using SARIMAX time series for one year ahead forecast.

### 5.1. Experiment Setup

Experiments have been visualized to offer analysis and knowledge of the quantities of several harmful gases that contributed to a low AQI index between 2015 and 2020. For comparing performance of LSTM network models by employing Adam optimizer with various regression algorithms, such as Random Forest, SVM, and MLR.Create training data with different strategies, especially while doing regression techniques to understand the correlation between different attributes and their impact AQI bucket. The experiment also provides a comparison between univariate and multivariate approaches of different models. After which it provides one one-year year-ahead forecast using SARIMAX time series.

The technical Process that has been adopted for applying LSTM models is as follows. A good amount of different hidden neurons also along with supporting hyperparameters are considered for the experiment. For LSTM models 2 hidden layers are considered with each layer having around 50 cells. The input layer dimension is taken as (5,11) which corresponds to (N, Nf) along with 10 output layers that are considered for the experiment.

### 5.2. Performance Metrics

Evaluation and performance measures are crucial for determining the accuracy of models trained on training data and their outcomes in testing data. These metrics are mostly used to optimize the performance and fine-tune it to get a better result. The evaluation and Performance metrics considered in this experiment are as follows:-

#### 5.2.1. MSE (Mean Squared Error)

MSE is basic statistic that estimates squared difference between actual as well as anticipated values. Squared difference is conducted in this case basically to avoid any cancellation caused due to negative terms although if there are any outliers present in the dataset it penalizes them as it is very robust to the outliers and it affects the score heavily if they are present. Equation 11 shows the formula used for Mean Squared error.

$$MSE = (1/n) * \Sigma \text{ (actual – forecast)}^2 \qquad \textit{Eq. (12)}$$

Here n denotes number of observations.

#### 5.2.2 RMSE (Root Mean Squared Error)

RMSE is just an extended version of MSE where it puts a simple square root on MSE. The advantages observe red over here is that the loss encountered while training the dataset can be interpreted properly. Equation 12 shows the formula for RMSE.

$$RMSE = \sqrt{[\ \Sigma(Pi - Oi)^2 / \ n\ ]} \qquad \textit{Eq. (13)}$$

Here n denotes number of observations where Pi as well as Oi denote the actual as well as predicted values respectively.

### 5.3. Results

This section talks about the results obtained from various versions of LSTM models along with SARIMAX time series.

### 5.3.1. AQI Forecasting using DL-based Models along with SARIMAX time series.

As per the correlation matrix shown in Figure 2. the darker bricks indicate a strong relationship between the parameters. On the other hand, the lighter bricks show a weak relationship between parameters. Figure 2 shows that the additional factor population does not correlate with all the harmful gases in the dataset, with an increase in the population of PM10, NO2, NH3, O3, and Toluene being directly impacted, which impacts the AQI at the atmospheric level. On the other hand, the variants of NO, like NOx, NO2, etc., directly correlate with each other, which comes from industrial waste. Figure 3 shows the trend where AQI is getting affected in the Southeast Asia region with the increase in the population. Because of this, the emission levels of other industrial and chemical pollutants are rising. There have been constant high spikes from 2016- 2019. However, after 2020, the spike dropped down because of the global lockdown, which had a huge influence on improving the air quality in those particular places where it was particularly poor or classified as hazardous.

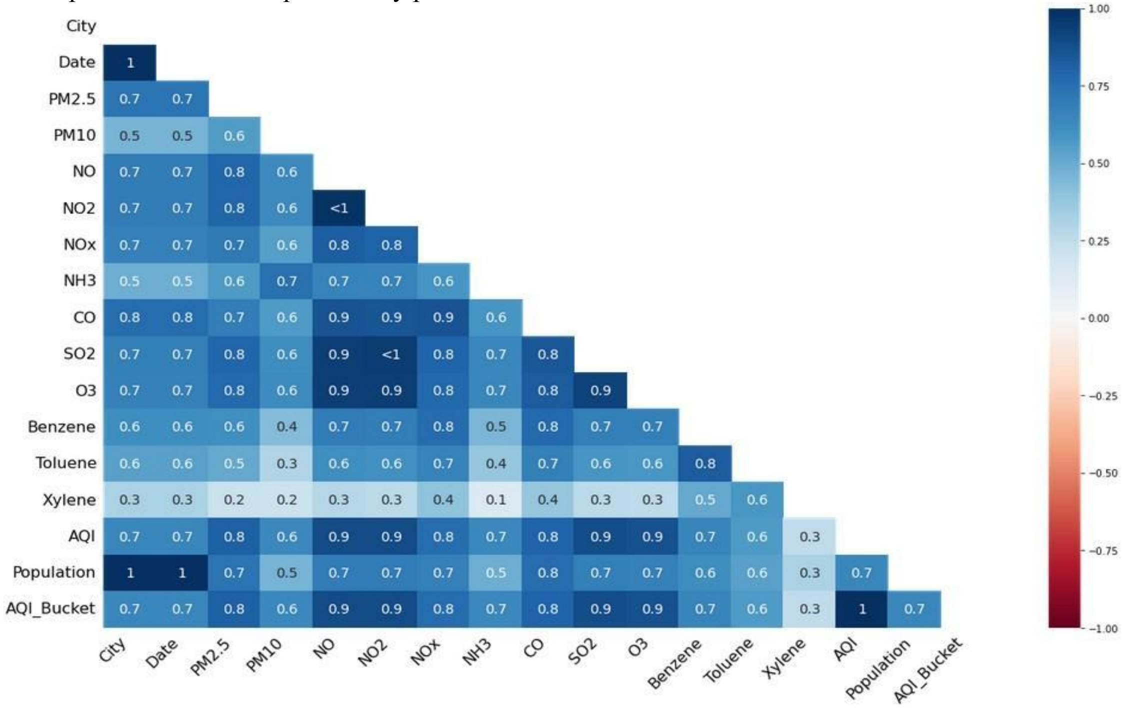


**Figure 2**. Correlation matrix of the Dataset

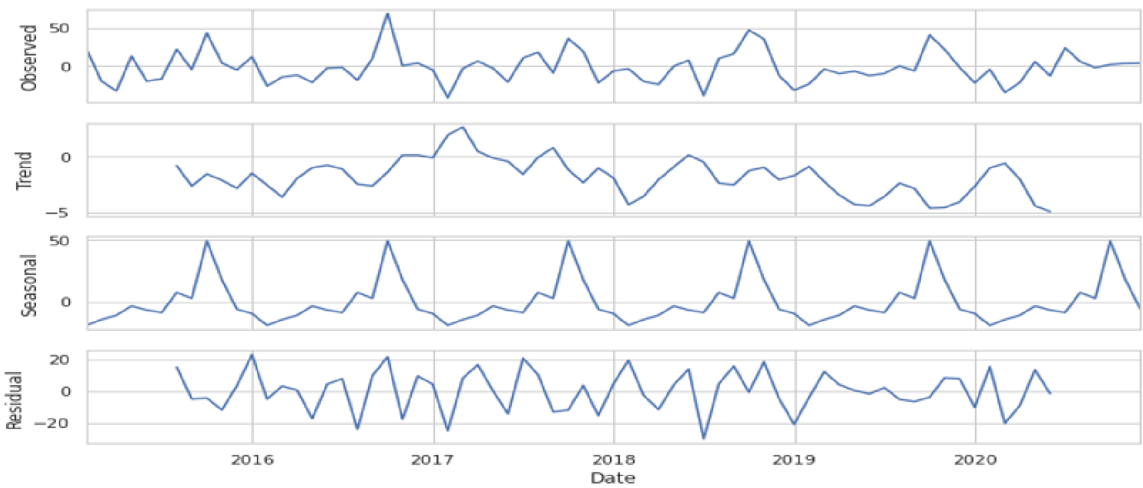


**Figure 3**. Trend In The AQI Index from 2015-2020

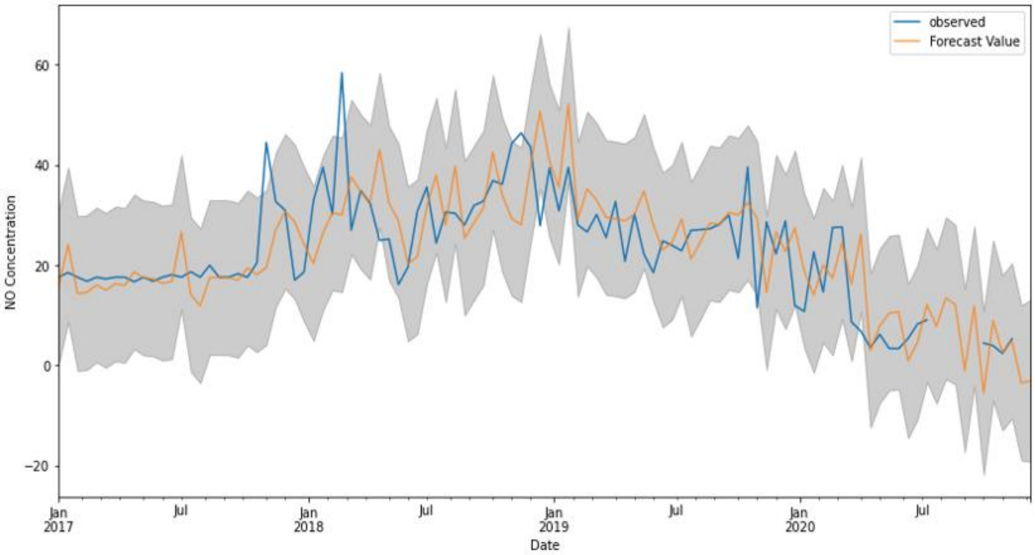


**Figure 4**. The concentration of 'NO' from 2017-2020

In Figure 4, we can see the concentration of NO variants from 2017-2020. It also depicts that the concentration had gone up from 2018 to 2019, which was the pre-lockdown period. NO is generally termed an industrial pollutant released by various industries as waste. The concentration of NO had come down after 2020 because non –functional industrial groups were functioning correctly before Jan 2020. Analysis of used DL algorithms and SARIMAX time series proposed model has been analyzed using MSE and Root RMSE metrics as shown in Table 3. MSE and RMSE are calculated using the following equations.

Here, n denotes number of observations, whereas Pi as well as Oi represent the actual as well as anticipated values, respectively. Finally, it is concluded that Shuffled BD –LSTM and the SARIMAX time series algorithm perform better than all implemented LSTM models.

**Table 3.** LSTM network analysis

| Algorithm | Train RMSE score | Test RMSE score | Predicted accuracy |
|---|---|---|---|
| LSTM + SARIMAX | 2142.92± 31.5 | 1056.68± 37.45 | 0.492 |
| ED-LSTM + SARIMAX | 2101.77± 34.16 | 1135.46± 24.12 | 0.534 |
| BD-LSTM + SARIMAX | 1827.36± 11.34 | 963.26± 34.04 | 0.5161 |
| BD-LSTM(Shuffled)+ SARIMAX | 1650.48± 39.21 | 1204.85± 33.29 | 0.7361 |

**5.4. Analysis.**

As per the models implemented in sections 4.2 and 4.3, Shuffled BD-LSTM and SARIMAX time series provide better accuracy of 73.61% among the various DL-based models implemented. The DL-based models address the Spatiotemporal features of the sparse dataset, which were necessary to forecast AQI using the SARIMAX time series. In contrast, baseline regression models face the challenges of understanding the time series data needed to predict AQI and emission levels of other gases in that area.

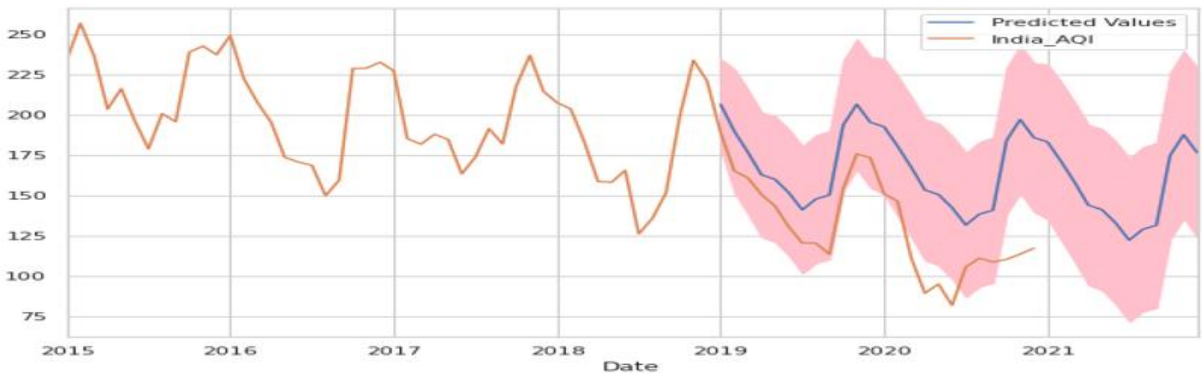


**Figure 5.** Prediction of AQI values for the year 2021

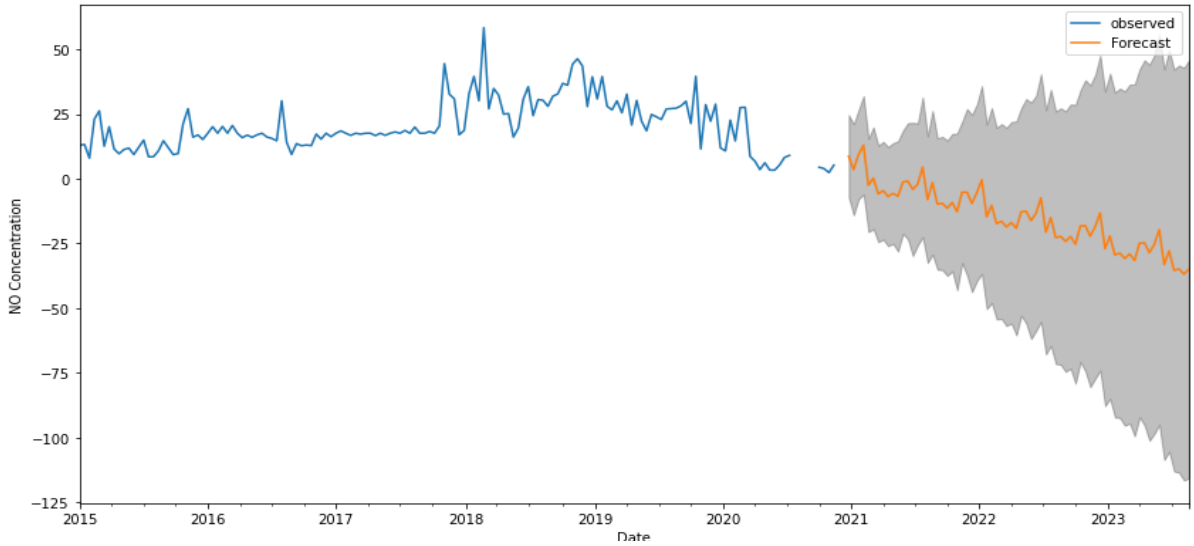


**Figure 6.** NO Concentration prediction till the year 2023

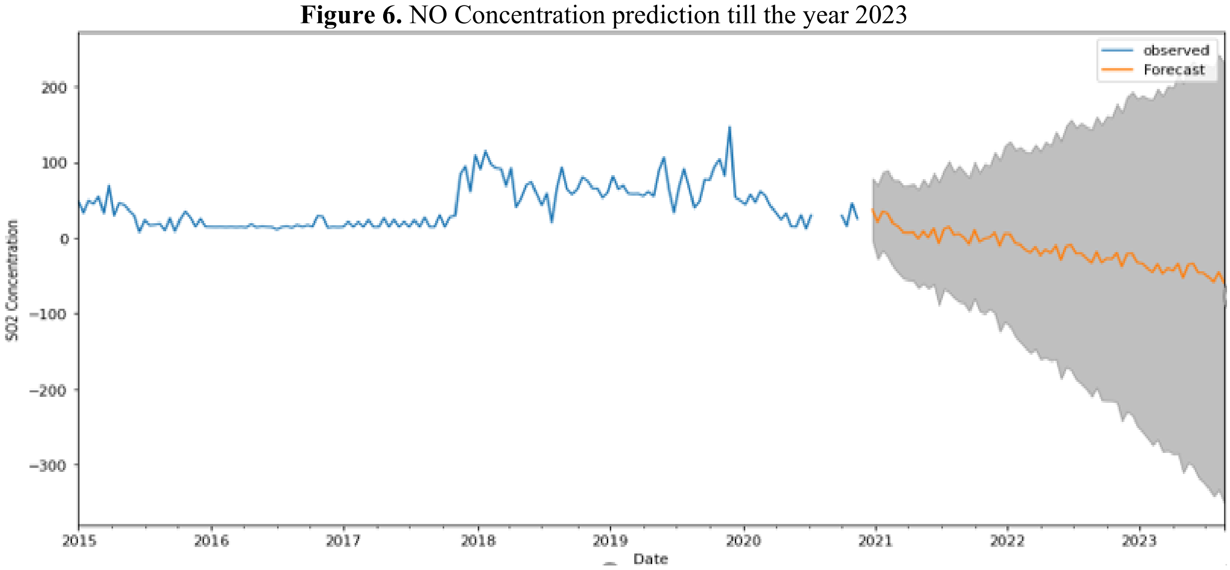


**Figure 7.** SO2 concentration predicted till the year 2023

Figure 5 shows how the AQI will grow in 2021 based on the analysis done in 2016-2020. It also shows that the AQI in the South Asian region will be affected most as the emission levels and the various harmful concentrations grow. Figure 6 and Figure 7 shows the concentration forecasting of harmful chemicals like NO and SO2, which are prime contributors to having a poor AQI in the South –Asian region. These figures show that the concentration levels of NO and SO2 will not come down until 2023 unless necessary steps are taken to control the emission levels. The AQI and the corresponding risk to public health increase in line with air pollution levels. People with respiratory or cardiovascular conditions, the elderly, and children are usually the first to be impacted by poor air quality. Government agencies typically advise individuals to limit their outdoor physical activity or even stay indoors when the AQI is high. Public awareness of the harmful health effects of local air pollution is the primary objective of the AQI. Various Government agencies are making more effort in creating infrastructure for solar, wind energy, and electric vehicles. According to the data, it is evident that air pollution will rise as the population grows, but this can be mitigated by changing the energy sources used in the future. In the future, we may speculate about the different socioeconomic strategies that will improve our ability to predict AQI. By 2040, renewable energy is predicted to account for almost 43% of global electricity demand, compared to 27% in 2014. Increasing population density in many developing countries will increase pollution, but this will change w ith utility mode.

China as an example shows that increasing population reduces pollution through clean energy and public transp ortA study on the impact of weather and air pollution on the spread of COVID19 in 219 cities in China found th at air quality indicators were positively associated with newly confirmed conditions and that the increase in case s was related to weather. Such studies are needed in Asia, for estimating air quality indicators which will be useful for prevention measures. More green and renewable energy must be promoted in these areas to prevent any more damage caused to the environment.

## 6. Conclusion and Future Scope

Rapid economic development as well as population growth have resulted in poor air quality at an alarming rate. The poor air quality in the environment can be unsafe for human lives as it could lead to severe health hazards like asthma and heart attacks. For controlling the city's emission levels, proper analysis and forecasting of AQI are essential for maintaining sustainable development. DL algorithms have shown great potential in providing detailed analyses while working with a diverse dataset. This paper applies various DL-based algorithms like LSTM, ED-LSTM, and BD-LSTM to the dataset to analyze and forecast AQI. Among the LSTM models, Stuffed BD –LSTM and the SARIMAX time series algorithm performed better with an accuracy of 73.61%. As per the analyses and the results obtained from various models, PM2.5, PM 10, SO2, and NOx emission levels with population play a significant role in creating a poor AQI in the city. The proposed models provide an adequate solution to monitor and forecast the AQI level in the city. However, it is further suggested to consider new parameters like forest fires, earthquakes, and all sorts of natural calamities, which can provide us with more inputs and insights about controlling harmful emissions in Southeast Asian countries. Additional learning approaches, such as multi-task as well as transfer learning, may be combined with Bayesian inference to produce better models that include existing DL models for COVID-19 infections within South Asian countries. In the future, Bayesian DL algorithms can be employed to provide robust uncertainty quantification in predictions, expanding the Bayesian neural networks employed for time series prediction. Furthermore, we foresee applying the proposed paradigm in South-Asian countries with declining air quality. The various authorities can create a web-based architecture to give effective weekly and monthly planning, which may include how traffic is controlled during different seasons. The technology can potentially be enhanced to predict air quality concerning global forest fires.

**Declarations:**

***Ethics Approval:*** No Human subject or animals are involved in the research.

***Consent to Participate:*** All authors have mutually consented to participate.

***Consent to Publish:*** All the authors have consented the Journal to publish this paper.

***Data Availability Statement:*** Authors declare that all the data being used in the design and production cum layout of the manuscript is declared in the manuscript.

***Funding Statement***: Nil

**Conflicts of Interest:** None

## REFERENCE

[1] "7 million premature deaths annually linked to air pollution," August 2020, [Online; accessed 15-July-2022]. [Online]. Available: https://www.who.int/news/item/25-03-2014-7-million-premature-deaths-annually-linked-to-air-pollution.

[2] "25-most-polluted-cities-world-2022-ranking," April 2022, [Online; accessed 13-July-2022]. [Online]. Available: https://smartairfilters.com/en/blog/25-most-polluted-cities-world-2022-rankings/.

[3] Janarthanan, R., P. Partheeban, K. Somasundaram, and P. Navin Elamparithi. "A deep learning approach for prediction of air quality index in a metropolitan city." *Sustainable Cities and Society* 67 (2021): 102720.

[4] Zhao, Xiaosong, Rui Zhang, Jheng-Long Wu, and Pei-Chann Chang. "A Deep Recurrent Neural Network for Air Quality Classification." *J. Inf. Hiding Multim. Signal Process.* 9, no. 2 (2018): 346-354.

[5] Liao, Qi, Mingming Zhu, Lin Wu, Xiaole Pan, Xiao Tang, and Zifa Wang. "Deep learning for air quality forecasts: a review." *Current Pollution Reports* 6, no. 4 (2020): 399-409.

[6] Abirami, S., and P. Chitra. "Regional air quality forecasting using spatiotemporal deep learning." *Journal of Cleaner Production* 283 (2021): 125341.

[7] Mao, Wenjing, Weilin Wang, Limin Jiao, Suli Zhao, and Anbao Liu. "Modeling air quality prediction using a deep learning approach: Method optimization and evaluation." *Sustainable Cities and Society* 65 (2021): 102567.

[8] Du, Shengdong, Tianrui Li, Yan Yang, and Shi-Jinn Horng. "Deep air quality forecasting using hybrid deep learning framework." *IEEE Transactions on Knowledge and Data Engineering* 33, no. 6 (2019): 2412-2424.

[9] Lin, Yijun, Nikhit Mago, Yu Gao, Yaguang Li, Yao-Yi Chiang, Cyrus Shahabi, and José Luis Ambite. "Exploiting spatiotemporal patterns for accurate air quality forecasting using deep learning." In *Proceedings of the 26th ACM SIGSPATIAL international conference on advances in geographic information systems*, pp. 359-368. 2018.

[10] Zhou, Yanlai, Fi-John Chang, Li-Chiu Chang, I-Feng Kao, and Yi-Shin Wang. "Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts." *Journal of cleaner production* 209 (2019): 134-145.

[11] Do, Tien Huu, Evaggelia Tsiligianni, Xuening Qin, Jelle Hofman, Valerio Panzica La Manna, Wilfried Philips, and Nikos Deligiannis. "Graph-deep-learning-based inference of fine-grained air quality from mobile IoT sensors." *IEEE Internet of Things Journal* 7, no. 9 (2020): 8943-8955.

[12] Lin, Chi-Yeh, Yue-Shan Chang, and Satheesh Abimannan. "Ensemble multifeatured deep learning models for air quality forecasting." *Atmospheric Pollution Research* 12, no. 5 (2021): 101045.

[13] Zhang, Zhendong, Yongkang Zeng, and Ke Yan. "A hybrid deep learning technology for PM2. 5 air quality forecasting." *Environmental Science and Pollution Research* 28, no. 29 (2021): 39409-39422.

[14] Tiwari, Animesh, Rishabh Gupta, and Rohitash Chandra. "Delhi air quality prediction using LSTM deep learning models with a focus on COVID-19 lockdown." *arXiv preprint arXiv:2102.10551* (2021).

[15] Zani, N. Bruni, G. Lonati, M. I. Mead, M. T. Latif, and P. Crippa. "Long-term satellite-based estimates of air quality and premature mortality in Equatorial Asia through deep neural networks." *Environmental Research Letters* 15, no. 10 (2020): 104088.

[16] Sannigrahi, Srikanta, Prashant Kumar, Anna Molter, Qi Zhang, Bidroha Basu, Arunima Sarkar Basu, and Francesco Pilla. "Examining the status of improved air quality in world cities due to COVID-19 led temporary reduction in anthropogenic emissions." *Environmental research* 196 (2021): 110927.

[17] Lee, Jihyeon, Nina R. Brooks, Fahim Tajwar, Marshall Burke, Stefano Ermon, David B. Lobell, Debashish Biswas, and Stephen P. Luby. "Scalable deep learning to identify brick kilns and aid regulatory capacity." *Proceedings of the National Academy of Sciences* 118, no. 17 (2021): e2018863118.

[18] Burke, Marshall, Anne Driscoll, David B. Lobell, and Stefano Ermon. "Using satellite imagery to understand and promote sustainable development." *Science* 371, no. 6535 (2021): eabe8628.

[19] Helbich, Marco, Yao Yao, Ye Liu, Jinbao Zhang, Penghua Liu, and Ruoyu Wang. "Using deep learning to examine street view green and blue spaces and their associations with geriatric depression in Beijing, China." *Environment international* 126 (2019): 107-117.

[20] Cabaneros, Sheen Mclean, John Kaiser Calautit, and Ben Richard Hughes. "A review of artificial neural network models for ambient air pollution prediction." *Environmental Modelling & Software* 119 (2019): 285-304.

[21] Ghimire, Sujan, Ravinesh C. Deo, Nawin Raj, and Jianchun Mi. "Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms." *Applied Energy* 253 (2019): 113541.

[22] Zheng, Bo, Guannan Geng, Philippe Ciais, Steven J. Davis, Randall V. Martin, Jun Meng, Nana Wu et al. "Satellite-based estimates of decline and rebound in China's CO2 emissions during COVID-19 pandemic." *Science advances* 6, no. 49 (2020): eabd4998.

[23] Rybarczyk, Yves, and Rasa Zalakeviciute. "Machine learning approaches for outdoor air quality modelling: A systematic review." *Applied Sciences* 8, no. 12 (2018): 2570.

[24] Ravindra, Khaiwal, Maninder Kaur Sidhu, Suman Mor, Siby John, and Saumyadipta Pyne. "Air pollution in India: bridging the gap between science and policy." *Journal of Hazardous, Toxic, and Radioactive Waste* 20, no. 4 (2016): A4015003.

[25] Zaman, Nurul Amalin Fatihah Kamarul, Kasturi Devi Kanniah, and Dimitris G. Kaskaoutis. "Estimating particulate matter using satellite based aerosol optical depth and meteorological variables in Malaysia." *Atmospheric Research* 193 (2017): 142-162.

[26] Ranjan, Avinash Kumar, Aditya Kumar Patra, and A. K. Gorai. "A review on estimation of particulate matter from satellite-based aerosol optical depth: Data, methods, and challenges." *Asia-Pacific Journal of Atmospheric Sciences* 57, no. 3 (2021): 679-699.

[27] Stebel, Kerstin, Iwona S. Stachlewska, Anca Nemuc, Jan Horálek, Philipp Schneider, Nicolae Ajtai, Andrei Diamandi et al. "Samira-satellite based monitoring initiative for regional air quality." *Remote sensing* 13, no. 11 (2021): 2219.

[28] Yadav, Nishant, Meytar Sorek-Hamer, Michael Von Pohle, Ata Akbari Asanjan, Adwait Sahasrabhojanee, Esra Suel, Raphael Arku et al. "Deep Transfer Learning on Satellite Imagery Improves Air Quality Estimates in Developing Nations." *arXiv preprint arXiv:2202.08890* (2022).

[29] "Air Quality Data in India from 2015-2020," July 2020, [Online; accessed 03-April-2022]. [Online]. Available: https://www.kaggle.com/datasets/rohanrao/air-quality-data-in-india.

[30] "Air Quality Data ," June 2020, [Online; accessed 05-April-2022]. [Online]. Available: https://cpcb.nic.in/

[31] Y. Zhou, S. De, G. Ewa, C. Perera, and K. Moessner, "Data-Driven air quality characterization for urban environments: a case study," *IEEE Access*, vol. 6, Article ID 77996, 2018.

[32] C. Srivastava, S. Singh, and A. P. Singh, "Estimation of air pollution in Delhi using machine learning techniques," in *Proceedings of the 2018 International Conference on Computing, Power and Communication Technologies (GUCON)*, pp. 304–309, Greater Noida, India, September 2018.

[33] R. Raturi and J. R. Prasad, "Recognition of future air quality index using artificial neural network," *International Research Journal of Engineering and Technology (IRJET)*, vol. 5, pp. 2395–0056, 2018.

[34] U. Mahalingam, K. Elangovan, H. Dobhal, C. Valliappa, S. Shrestha, and G. Kedam, "A machine learning model for air quality prediction for smart cities," in *Proceedings of the 2019 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*, pp. 452–457, Chennai, India, March 2019.

[35] V. Sivakumar, G. R. Kanagachidambaresan, V. Dhilip kumar, M. Arif, C. Jackson, and G. Arulkumaran, "Energy-efficient markov-based lifetime enhancement approach for underwater acoustic sensor network," *Journal of Sensors*, vol. 2022, Article ID 3578002, 10 pages, 2022.

[36] J. Sethi and M. Mittal, "Ambient air quality estimation using supervised learning techniques," *ICST Transactions on Scalable Information Systems*, vol. 6, Article ID 159628, 2019.

[37] P. Hajek and V. Olej, "Predicting common air quality index - the case of Czech microregions," *Aerosol and Air Quality Research*, vol. 15, no. 2, pp. 544–555, 2015.

[38] S. Ameer, M. A. Shah, A. Khan et al., "Comparative analysis of machine learning techniques for predicting air quality in smart cities," *IEEE Access*, vol. 7, Article ID 128325, 2019.

[39] A. R. Behesht Abad, S. Mousavi, N. Mohamadian et al., "Hybrid machine learning algorithms to predict condensate viscosity in the near wellbore regions of gas condensate reservoirs," *Journal of Natural Gas Science and Engineering*, vol. 95, Article ID 104210, 2021.

[40] M. Rajabi, S. Beheshtian, S. Davoodi et al., "Novel hybrid machine learning optimizer algorithms to prediction of fracture density by petrophysical data," *Journal of Petroleum Exploration and Production Technology*, vol. 11, no. 12, pp. 4375–4397, 2021.

[41] A. R. Behesht Abad, P. S. Tehrani, M. Naveshki et al., "Predicting oil flow rate through orifice plate with robust machine learning algorithms," *Flow Measurement and Instrumentation*, vol. 81, Article ID 102047, 2021.

[42] O. Hasbeh, M. Ahmadi Alvar, K. Y. Aghdam, H. Ghorbani, N. Mohamadian, and J. Moghadasi, "Hybrid computing models to predict oil formation volume factor using multilayer perceptron algorithm," *Journal of Petroleum and Mining Engineering*, vol. 23, no. 1, pp. 17–30, 2021.

[43] F. Jafarizadeh, M. Rajabi, S. Tabasi et al., "Data driven models to predict pore pressure using drilling and petrophysical data," *Energy Reports*, vol. 8, pp. 6551–6562, 2022.

[44] G. Zhang, S. Davoodi, S. S. Band, H. Ghorbani, A. Mosavi, and M Moslehpour, "A robust approach to pore pressure prediction applying petrophysical log data aided by machine learning techniques," *Energy Reports*, vol. 8, pp. 2233–2247, 2022.

[45] S. Tabasi, P. Soltani Tehrani, M. Rajabi et al., "Optimized machine learning models for natural fractures prediction using conventional well logs," *Fuel*, vol. 326, Article ID 124952, 2022.

[46] M. Rajabi, O. Hazbeh, S. Davoodi et al., "Predicting shear wave velocity from conventional well logs with deep and hybrid machine learning algorithms," *Journal of Petroleum Exploration and Production Technology*, 2022.

[47] S. Beheshtian, M. Rajabi, S. Davoodi et al., "Robust computational approach to determine the safe mud weight window using well-log data from a large gas reservoir," *Marine and Petroleum Geology*, vol. 142, Article ID 105772, 2022.

[48] Z. K. Masoud, S. Davoodi, H. Ghorbani et al., "Band, Permeability prediction of heterogeneous carbonate gas condensate reservoirs applying group method of data handling," *Marine and Petroleum Geology*, vol. 139, 2022.

[49] N. Mohamadian, H. Ghorbani, D. A. Wood, and M. A. Khoshmardan, "A hybrid nanocomposite of poly(styrene-methyl methacrylate- acrylic acid)/clay as a novel rheology-improvement additive for drilling fluids," *Journal of Polymer Research*, vol. 26, no. 2, p. 33, 2019.

[50] N. Mohamadian, H. Ghorbani, D. A. Wood, and H. K. Hormozi, "Rheological and filtration characteristics of drilling fluids enhanced by nanoparticles with selected additives: an experimental study," *Advances in Geo-Energy Research*, vol. 2, no. 3, pp. 228–236, 2018.

[51] K. Stebal et al., "SAMIRA: Satellite Based Monitoring Initiative for Regional Air quality," RemoteSens., vol.13, no.11, 2021, doi: 10.3390/rs13112219.

[52] A. R. Ganguly, N. Yadav, M. V. Pohle, "Deep learning Transfer on Satellite Imagery Improves Air Quality Estimates In Developing Nations," arXivp:2202.08890, 2022, doi: 10.48550/arXiv.2202.08890 .

[53] N.F.K Zaman, K. Devi , K. Dimitris and G. Kaskaoutis , “Estimating Particulate Matter using satellite based aerosol optical depth and meteorological variables in Malaysia”,
[54] Atmospheric Research ,vol. 193, 2019,pp.142-162,doi: 10.1016/j.atmosres.2017.04.019.
[55] [54]. Sidhu, Kamaljeet Kaur, Habeeb Balogun, and Kazeem Oluwakemi Oseni. "Predictive Modelling of Air Quality Index (AQI) Across Diverse Cities and States of India using Machine Learning: Investigating the Influence of Punjab's Stubble Burning on AQI Variability." *arXiv preprint arXiv:2404.08702* (2024).
[56] Waqas, Muhammad, Majid Nazeer, Man Sing Wong, Wu Shaolin, Li Hon, and Joon Heo. "Impact of urban spatial factors on NO2 concentration based on different socio-economic restriction scenarios in US cities." *Atmospheric Environment* 316 (2024): 120191.
[57] Folorunso, O., Sadiku, R., & Hamam, Y. (2024). Dispatchable generation analysis and prediction by using machine learning: A case study of South Africa. *e-Prime-Advances in Electrical Engineering, Electronics and Energy*, *9*, 100701.