IJTE, Vol. 11, No. 2, Autumn 2023, pp. 1387-1400
Research Paper

Comparison of Regression and Deep Learning Approaches in Modeling Time Series to Predict Air Pollutant Concentration in City of Tehran

Sahar Koohfar[1], Fatemeh Bandarian[2], Amir Abbas Rassafi[3, *]

Received: 2021/11/11 *Accepted: 2023/08/15*

Abstract

The rapid growth of urbanization and the global population have resulted in climate change, air contamination, and various human health problems. Thus, estimating air pollution indices has become important to environmental science studies. With relevant data increasingly available, machine learning frameworks have been proposed as a particularly useful method to predict air pollution. Based on four years of Tehran's neighborhood air pollution data analysis, this paper proposes three machine learning approaches to predict NO_2 and CO concentration: Autoregressive Integrated Moving Average (ARIMA), Long Short-Term Memory Networks (LSTM), and Multiple Linear Regression (MLR). This paper compared the ability of the ARIMA, LSTM, and MLR machine learning methods to forecast the daily concentrations of NO_2 and CO at Punak air quality monitoring station, from 2017 to 2020. By applying four performance measurements, the ARIMA model displays the worst performance among the three models in all datasets with RMSE values of 47.39 and 1.29, and $R^2 = 0.012$ and 0.01 for NO_2 and CO respectively. The LSTM and MLR models achieve the best forecasting result with RMSE = 17.6 and 6.41, MAE = 10.59 and 4.33, R^2 = 0.458 and 0.46, and RRSE =1.06 and 1.10 for NO_2 forecasting and RMSE = 0.42 and 0.32, MAE = 0.24 and 0.25, $R^2 = 0.96$ and 0.98, and RRSE = 0.43 and 0.44 for CO forecasting.

Keywords: Air Pollutant, Machine Learning, Urban Area, NO_2, CO

* Corresponding author E-mail: rasafi@ikiu.ac.ir
1 Imam Khomeini International University, College of Engineering, Qazvin, Iran
2 Imam Khomeini International University, College of Engineering, Qazvin, Iran
3 Imam Khomeini International University, College of Engineering, Qazvin, Iran

1. Introduction

With the rapid growth of urbanization and industrialization, air pollution has become a significant environmental hazard and a global concern to human beings [Brunekreef & Holgate, 2002].

In recent decades, rapid economic growth and dramatic increases in the population and number of vehicles in Tehran, in particular, have resulted in high concentrations of suspended particles, like Carbon Monoxide (CO) and Nitrogen Dioxide (NO_2), that cause various environmental issues, especially air pollution [Mohammadi-Zadeh et al., 2017].

There are more than 8.6 million people living in Tehran, of whom 33.7% are employed, 20.9% are students, and 0.38 private cars are owned by each inhabitant. Due to its unique topography and climate, the establishment of thousands of industrial plants, the traffic of a variety of vehicles, and the daily consumption of more than 10.3 million liters of gasoline, 3.9 million liters of diesel, and 1.3 million cubic meters of CNG for transportation, Tehran is one of the most polluted cities in the world. The biggest contributor to Tehran's air pollution is indiscriminate energy use, and automobiles are responsible for more than 85% of the city's total vehicle emissions, with CO having the greatest percentage following with NO_2 [Torkian et al., 2012].

NO_2 is a major source of fine particulate pollution and causes multiple health problems, like lung irritation, lower resistance to respiratory infections, and increased sensitivity in people with chronic respiratory diseases, such as asthma [Juhos et al., 2008].

According to the EPA, short-term exposures (less than three hours) to the present NO_2 concentrations may alter lung function and airway responsiveness in people who already have respiratory illnesses, as well as increase respiratory infections in children between the ages of 5 and 12. Long-term exposure to NO_2 may raise one's susceptibility to respiratory infections and may change one's lungs, according to the EPA [Safriet & Brooks, 1989]

At ages 12 and 18, the presence of symptoms of anxiety, depression, conduct disorder, and attention deficit hyperactivity disorder was evaluated. At the age of 18, the individuals were interviewed to determine their psychiatric diagnosis. No links between exposure to pollution before the age of 12 and ongoing mental health issues were discovered. Even after adjusting for common risk variables, major depressive disorder probabilities at age 18 were significantly raised by age-12 pollution estimations [Roberts et al., 2019].

CO is a colorless, tasteless, and odorless gas that is produced due to the incomplete combustion of carbon-containing fuels, such as gasoline, wood, oil, or other fuel.

High concentrations of CO occur in areas with heavy traffic congestion, where as much as 95 percent of all CO emissions may come from automobile exhaust. Other sources include industrial processes, non-transportation fuel combustion, and natural sources such as wildfires [Wark et al., 1998].

High concentrations of CO may cause physiological and pathological changes, as well as death. However, lower levels of CO that accumulate in an open environment can still pose a health threat, especially to those who suffer from cardiovascular diseases, such as angina pectoris [Council, 2002].

Due to the direct harmful effects of air pollution on human health, air quality management authorities must develop proper and reliable methods to predict and evaluate concentrations of atmospheric pollutants and provide preventive and evasive actions and strategies.

The identification of reliable models for forecasting of harmful air pollutant concentrations is important especially for pollutant city such as Tehran. Using appropriate models for predicting air pollutant concentration can help to investigate the impacts of different undergoing policies that tries to reduce the air pollutant concentrations.

Although different studies have focused on forecasting particulate matters, to the best of authors' knowledge, predicting CO and NO_2 concentrations for Tehran city has not been addressed before.

2. Literature Review

In recent years, machine learning methods have emerged as a prominent and useful technique for predicting air quality problems, particularly pollutant concentrations in urban areas [Srivastava et al., 2019]. For modeling and predicting air pollution indices specifically, advanced statistical models based on machine learning are largely used, including ARIMA (Autoregressive Integrated Moving Average), LSTM (Long Short-Term Memory Networks), and MLR (Multi-Linear Regression). For example, Kumar applied the ARIMA modeling approach to forecasting the maximum daily surface O_3 concentrations at Brunei Darussalam. The researchers showed that the ARIMA model performed well when forecasting one day ahead for the daily maximum O_3 concentrations, with the normalized mean square error as 0.02 and mean absolute percentage error (MAPE) as 13.14% [Kumar et al., 2004]. Similarly, Robeson and Steyn studied forecasting techniques for daily maximum ozone concentrations. The results showed that the ARIMA model had nearly the same predictive capabilities as the persistence model. The mixed deterministic/stochastic model performed the worst Several previous studies have proposed using LSTM for more accurate time-series predictions [Connor et al., 1994]. For example, Zhao applied LSTM and ANN (artificial neural network) approaches to model local variations of PM10 contamination. The results showed that the LSTM approach performed better than ANN [Zhao et al., 2019]. Athira investigated the effectiveness of LSTM, RNN (Recurrent Neural Network), and Gated Recurrent Unit (GRU) models for forecasting Particulate Matter 10 (PM10). Their study demonstrated that all three models performed comparatively well [Athira et al., 2018].

Xayasouk Then extended these models to examine the ability of LSTM and DAE (Deep Auto Encoder) to predict fine PM concentrations and compared the conclusions in terms of Root Mean Square Error (RMSE). The models monitored hourly air quality data from 25 positions in Seoul, South Korea. Fine PM concentrations were forecasted at an optimal learning rate of 0.01 for 100 epochs with batch sizes of 32 for the LSTM model and 64 for the DAE model, which is optimal. The results showed that the suggested algorithm could forecast fine PM concentrations with a suitable accuracy using both models, though the LSTM performed slightly better [Xayasouk et al., 2020].

Abdullah Proposed using MLR models to predict the PM10 concentrations in Peninsular Malaysia during different monsoon seasons with meteorological factors as predictors. The researchers reported the Root Square Error (R^2) for the NEM, Inter Monsoon 1, SWM (South West Monsoon), and Inter Monsoon seasons as 0.68, 0.58, 0.57, and 0.63, respectively. Based on these results, MLR models are appropriate for forecasting PM10 concentrations at the local level for each monsoon [Abdullah et al., 2017]. Dragomir evaluated the efficiency of a multiple regression model and ANN for predicting NO2 concentrations in Braila, Romania. The results of this statistical analysis indicated a relatively similar accuracy between the two approaches [Dragomir et al., 2015]. Dey forecasted criteria pollutants in Durgapur, India using MLR and Principal Component Regression (PCR). The results revealed that the MLR outperformed the PCR and that the PCR could not enhance the performance of the MLR [Dey et al., 2009].

In the present work, three machine learning algorithms namely, ARIMA, LSTM, and MLR are proposed to predict CO and NO_2 concentrations. Daily concentration of CO and NO_2 data from Punk air quality station were analyzed. Models are applied and the performance of these models was evaluated in terms of their performance indicators.

3. Methodology

In the present work, three machine learning algorithms namely, ARIMA, LSTM, and MLR are proposed to predict CO and NO_2 concentrations. Daily concentration of CO and NO_2 data from Punk air quality station were analyzed. Models are applied and the performance of these models was evaluated in terms of their performance indicators.
To forecast the concentrations of NO_2 and CO in Tehran, this paper used three time series models: ARIMA, LSTM, and MLR. Their formulas are explained in the following sections.

3.1. ARIMA Model

ARIMA is one of the most prominent models for univariate forecasting and was successfully developed by Box and Jenkins [Jenkins et al., 2011] ARIMA model assumes that the datasets are stationary, with a constant mean and variance over time. The ARIMA order (p, d, q) is represented in the following formula:

$$y_t^* = \Delta^d y_t \tag{1}$$

$$y_t^* = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i y_{t-i}^* + \sum_{j}^{q} \theta_j y_{t-j}^*, \tag{2}$$

Where y_t^* is differencing time series, y_t and ε_t are the actual value and random error (or random shock) at time t, **Δ** is the differencing operator θ_j and φ_i are the autoregressive parameters, and c is the constant. Parameters p, d, and q are non-negative integers, p is known as the order of the model, d is the number of Mathematical expressions of a LSTM network with a single forget gate is:

$$f_t = \sigma_g(W_f x_t + U_t h_{t-1} + b_f) \tag{3}$$

$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \tag{4}$$

$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \tag{5}$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \sigma_g(W_c x_t + U_c h_{t-1} + b_c) \tag{6}$$

$$h_t = o_t \circ \sigma_h(c_h), \tag{7}$$

where c_t is the output of the cell; f_t , i_t, and o_t are the forget gate, input gate, and output gate, respectively; x_t represents the input factor; h_t and h_{t-1} are the state vector at time t and t-1, respectively; W is the weight parameter; σ is the sigmoid function; and ○ represents the function (e.g., tanh or sigmoid) of the LSTM algorithm. Differencing passes, and q is the number of the moving average.

3.2. LSTM Model

LSTM is a kind of recurrent neural network that can learn long-term dependencies and learns over many time steps. It was first introduced in 1997 by Hochreiter [Hochreiter & Schmidhuber, 1997] and was improved in 2000 by [Gers et al., 1999]. LSTMs contain three gates: an input gate, an output gate, and a forget gate. The input gate controls the amount of new value entering the cell, the forget gate controls the extent to which the value stays in the cell, and the output gate controls the amount of cell value that is used to calculate the output activation of the LSTM units. LSTM networks are usually depicted in the following way:

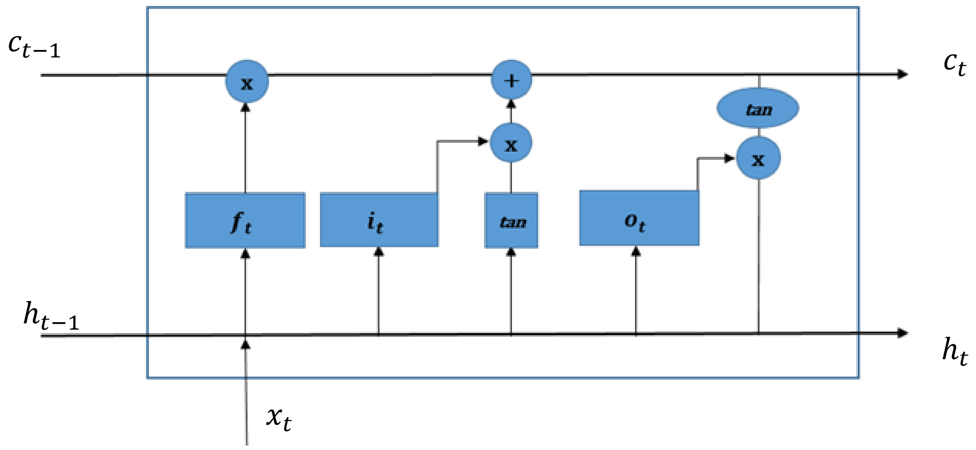

Figure 1. shematic description of LSTM

3.3. MLR Model

Linear regression is an appropriate statistical tool for investigating time series and quantitative data correlations, extrapolating trends, and estimating parameters. The MLR model examines a dependent variable (X) and an independent variable (Y) and estimates the relationship between the variables by drawing a regression line that best fits the data. This model is outlined as follow:

$$Y = a + b_1(X_1) + b_2(X_2) + \cdots + b_k(X_k) + e \tag{8}$$

Where Y is the dependent variable (target value); $b_1, b_2, \dots, b_k$ are the regression coefficients that are associated with X_1, X_2, …, X_k, respectively; a is the constant term, and e is the random error that represents the difference between the observed and fitted linear relationship [Frank R. Giordano & Horton, 2000].

The most common method to calculate b_k and e is least-square estimation, that minimize the sum of the squared residual. The equation is outlined as follow:

$$\sum_{t=1}^{T} e^2 = \sum_{t=1}^{T} (Y - a - b_1(X_1) - b_2(X_2) - \cdots - b_k(X_k))^2, \tag{9}$$

4. Performance Measurement

To estimate the forecasting accuracy of each model and compare the performances, four common error evaluation metrics were used: the root mean square error (RMSE) [Zhang, 2007] between the measured and predicted values; the mean absolute error (MAE) [Hamzaçebi, 2008] between the average of the forecasted and the actual values; coefficient of determination (R^2); and the root relative squared error (RRSE) , which takes the total squared error and normalizes it by dividing by the total squared error of the simple predictor. This process reduces the error to the same dimensions as the quantity being predicted. The RMSE, MAE, R^2, [Figueiredo Filho et al., 2011] and RRSE [Willmott & Matsuura, 2005] formulas are defined as follows:

$$RMSE = \sqrt{\sum_{i=1}^{N} (P_m - P_r)^2/N}, \tag{10}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |P_m - P_r| \tag{11}$$

$$R^2 = \frac{[\sum_i (P_m - \overline{P_m})(P_r - \overline{P_r})]^2}{\sum_i (P_m - \overline{P_m})^2 \sum_i (P_r - \overline{P_r})^2}, \tag{12}$$

$$RRSE = \sqrt{\frac{\sum_{i=1}^{N} (P_r - T_i)^2}{\sum_{i=1}^{N} (T_i - \overline{T})^2}}, \tag{13}$$

Where N is the number of measured values, P_m and P_r are measured and predicted values, and $\overline{P_m}$ and $\overline{P_r}$ are the mean value of P_m and P_r. T_i is the targeted value and the formula of $\overline{T}$ is as follow:

$$\overline{T} = \frac{1}{N}\sum_{i=1}^{N} T_i \tag{14}$$

4.1. Materials and Methods

To estimate the daily concentrations of CO and NO2, a four-step procedure was followed: (1) data collection, (2) data preprocessing, (3) machine learning, and (4) evaluation as shown in Figure 2.

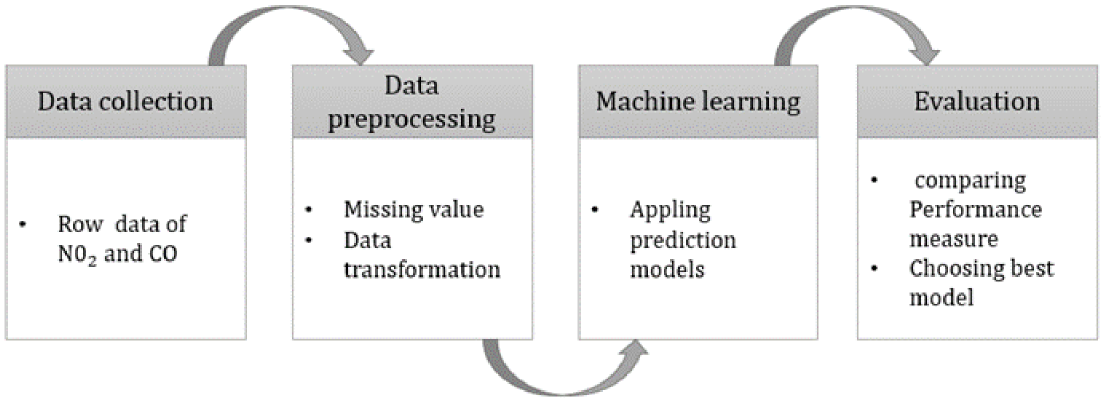

Figure 2. Workflow for predicting CO and NO_2 concentrations

5. Data Collection

5.1. Study Area

Tehran, which is the capital of Iran and is inhabited by more than 8.5 million people, has one of the most polluted air in the world. It contains 24 air quality monitoring stations, as shown in Figure 3.

In this study, the data from air quality monitoring station in Punak was used to forecast the air pollution index. This station is located in the northwest of Tehran, with a longitude of 35.75° N and 51.33° E, where the population is about 860,000 people, according to the latest statistics. The rapid growth of residential, commercial, and industrial areas in Punak has contributed to the poor air quality.

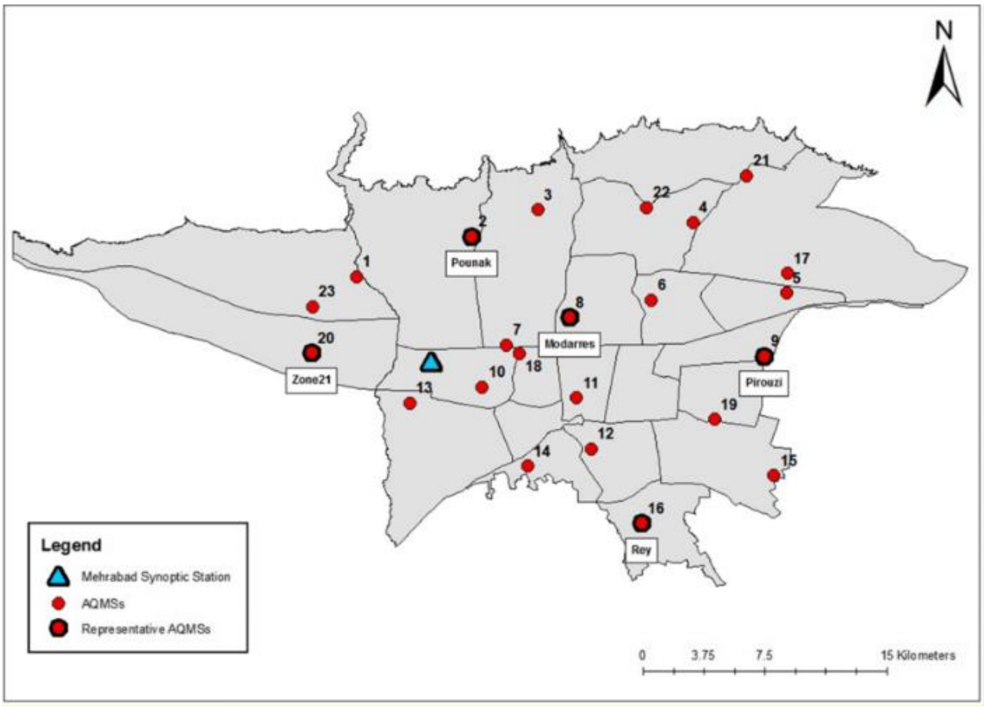

Figure 3. Tehran air pollution monitoring stations and Location of Punak, Tehran, Iran [Zohdirad et al., 2022]

5.2 Data Source

The dataset used for this experiment included the daily concentrations of NO_2 and CO measured at the Punak air quality monitoring station, as reported by the official website from the Air Quality Control Company in Tehran Province. The Concentration unit of NO_2 is in micrograms per cubic meter (µg/m3), and for CO is in milligrams per cubic meter (mg/m3). Table below illustrates the index parameter of NO_2 and CO.

Table 1. Air Quality Index [EPA, 2001]

Air Quality Category (Range)	NO_2 (µg/m3)	CO (mg/m3)
Good	0-40	0-1.0
Satisfactory	41-80	1.1-2.0
Moderate	81-180	2.1-10
Poor	181-280	10-17
Very poor	281-400	17-34
Sever	400 +	34 +

The data was recorded from March 23, 2017 to September 6, 2020 and contained 1,996 datasets from Punak air quality monitoring station in District 4 of Tehran (http://airnow.tehran.ir/home/DataArchive.aspx) Row data's descriptive statistics and trends are shown in table 1, Figure 4 and 5 respectively.

Table 2. Statistical summary of NO2 and CO before treatment

Variable	Observation	Minimum	Maximum	Mean	Std. deviation
NO_2 (µg/m3)	1996	5.0	134.0	72.33	27.66
CO (mg/m3)	1996	7	90	26.91	12.08

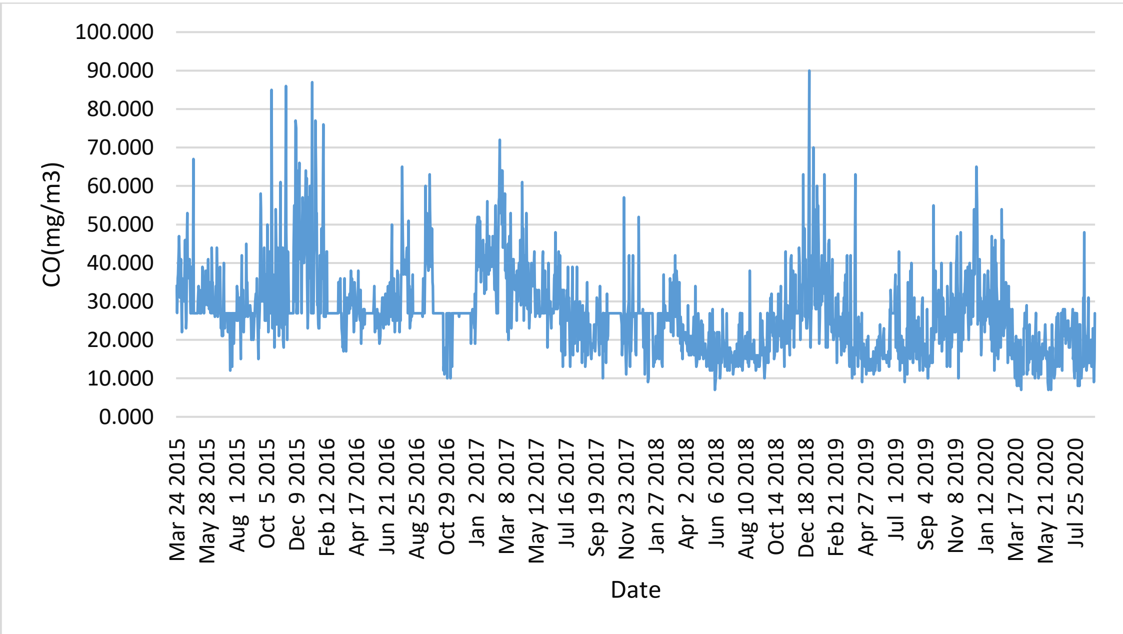

Figure 4. Daily CO concentration

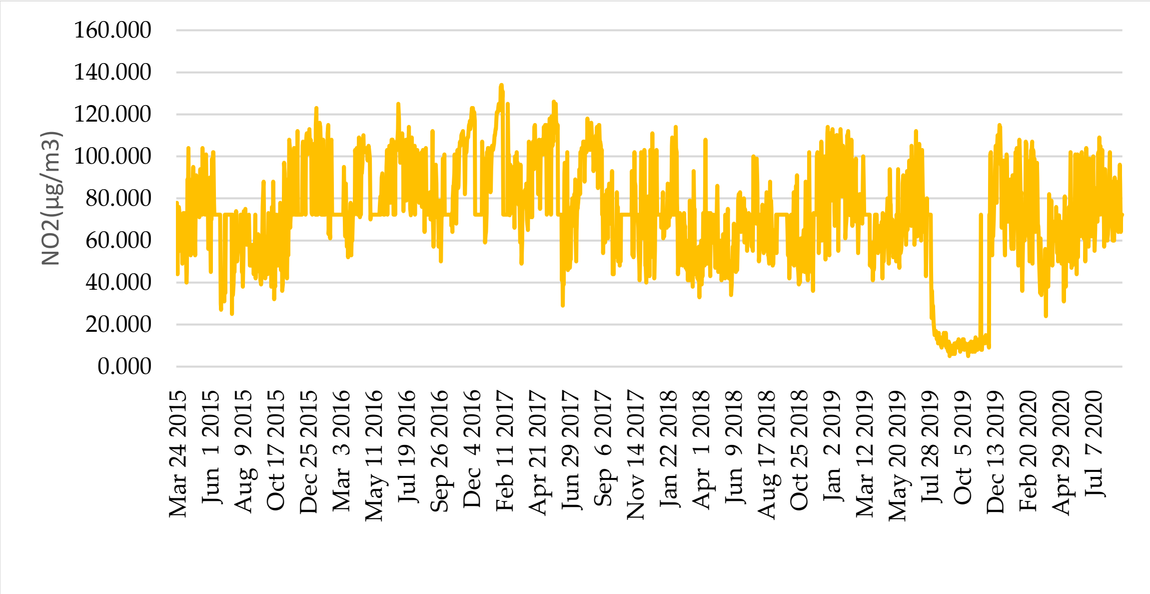

Figure 5. Daily NO2 concentration

5.3. Data Preprocessing

To eliminate the effect of the missing values on the model training, the k-Nearest Neighbors (KNN) imputer was applied to the row dataset. The KNN process uses the closest k-neighbors in a multi-dimensional space and the mean value of the k-neighbors found in the dataset, to estimate the missing data points. Table 3 shows the summary statistic of the dataset after treatment:

Table 3. Statistical summary of NO2 and CO after treatment

Variable	Observation	Minimum	Maximum	Mean	Std. deviation
NO_2 (µg/m3)	1996	5.0	134.0	72.33	25.15
CO (mg/m3)	1996	7	90	26.91	10.89

6. Results

In this study, concentration data of CO and NO_2 has been obtained from the open data website and was preprocessing to check for the missing values and eliminate their effect on the research. The clean dataset was split to training and testing dataset. The percentage of training and testing dataset are 70% and 30% of the total dataset. Machine learning algorithms are applied to the dataset. The prediction results and the accuracy of the models, using performance measurement are shown in this section. Using mentioned algorithms, One-month ahead concentration for both NO_2 and CO has been forecasted.

6.1. NO2 Prediction

Table 4 shows the measured and predicted values of NO_2 using the ARIMA, LSTM, and MLR models. Among the three models, the LSTM and MLR closely resembled the original one and had the highest performance accuracy, with RRSE values of 1.06 and 1.10, and R^2 values of 0.458 and 0.46 respectively. On the other hand, the ARIMA model performed the worst. The RMSE of ARIMA was 47.39, and its RRSE was 6.08. Both values were higher than those for LSTM and MLR models. Fig. 6 presents the plot of the observed versus predicted concentrations of NO_2 by the LSTM and MLR models.

Table 4. Statistical analysis of the measured and predicted values of NO_2

Measurement parameter	ARIMA	LSTM	MLR
RMSE	47.39	17.6	6.41
MAE	43.6	10.59	4.33
RRSE	6.08	1.06	1.10
R^2	0.038	0.458	0.46

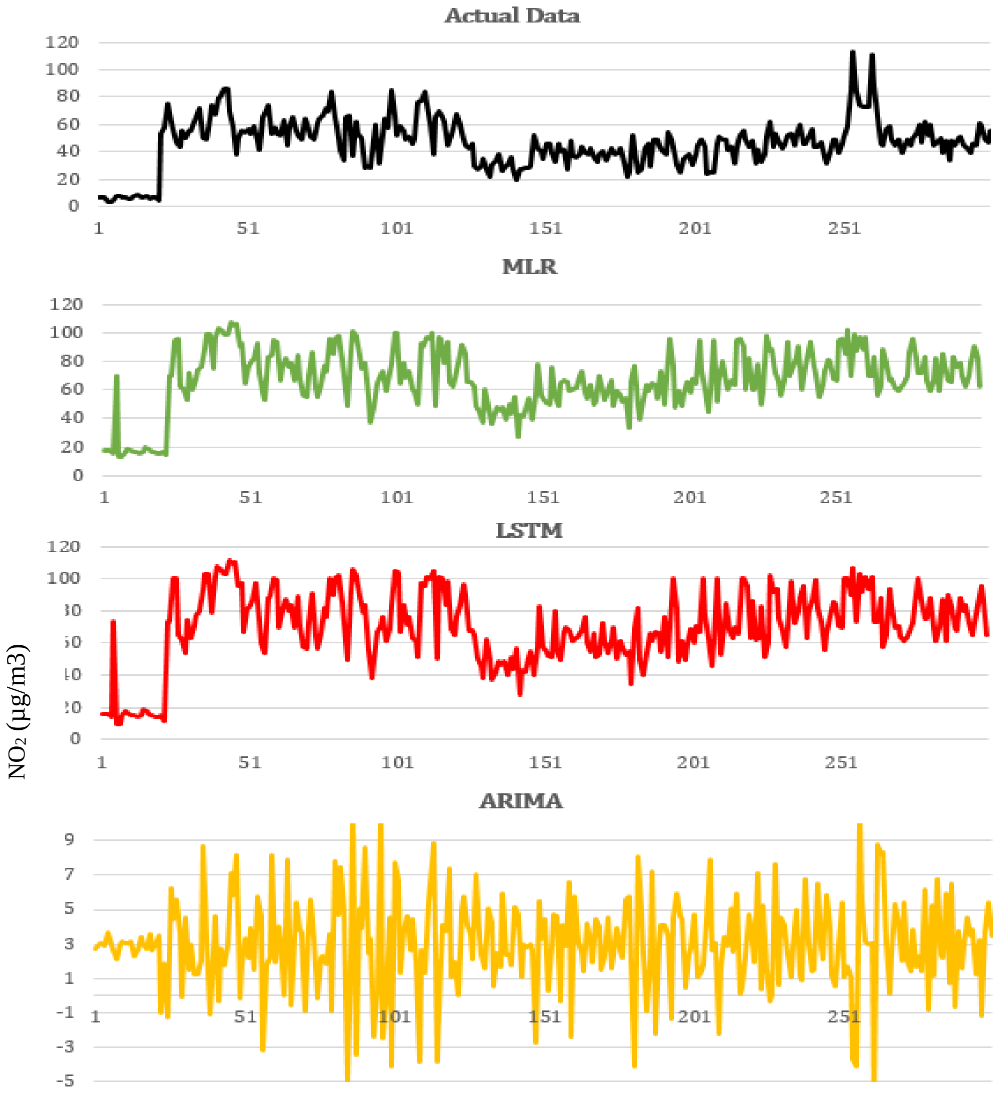

Figure. 6. Actual and predicted concentrations of NO_2

6.2. CO Prediction

Table 5 shows the statistical evaluation of three models for CO concentration. For the CO Dataset, the MLR performed slightly better than LSTM and was more effective than the ARIMA model, with an RMSE of 0.32, RRSE of 0.38, R^2 of 0.981, and MAE of 0.25. In contrast, the actual and forecasted values for the ARIMA model were significantly different, with R^2 value of 0.01. Figure 7 presents a comparison of the measured and predicted values of CO obtained from the three approaches. Also, Figure 8 represents the scattered plots between measured and modeled concentrations of NO_2 and CO

Table 5. Statistical analysis of the measured and predicted values of CO

Measurement parameter	ARIMA	LSTM	MLR
RMSE	1.29	0.42	0.32
MAE	1.67	0.25	0.24
RRSE	3.95	0.43	0.38
R^2	0.01	0.968	0.981

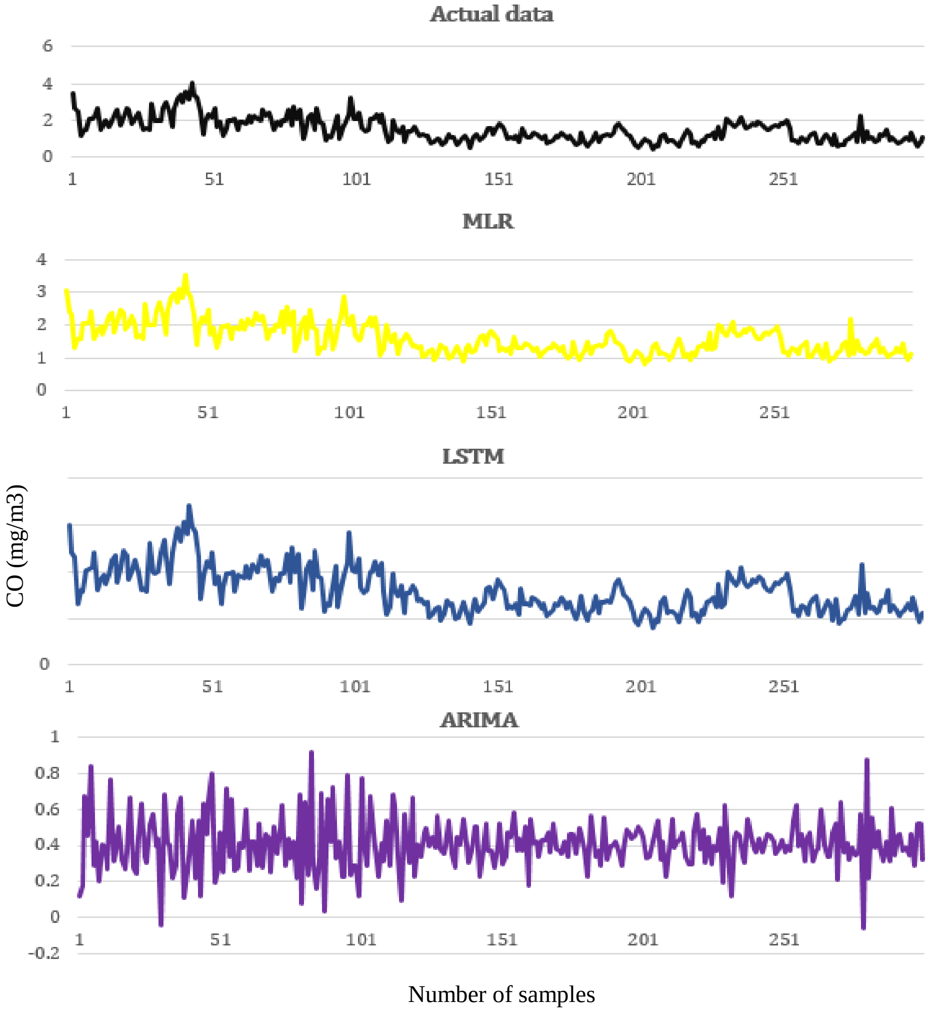

Figure 7. Actual and predicted concentrations of CO by three approaches

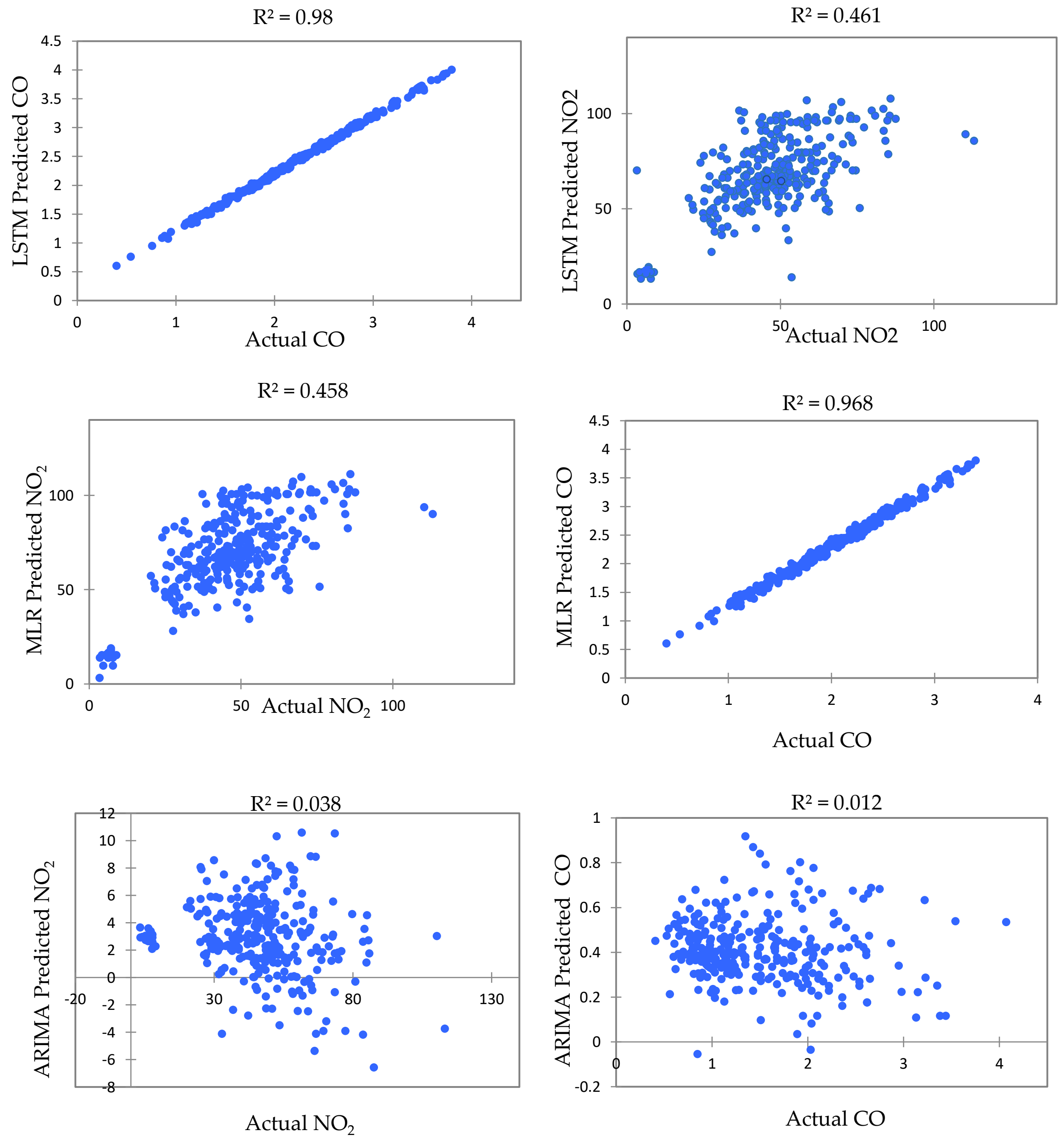

Figure 8. Scatterplots of observed and predicted NO_2 and CO derived from the three algorithms

7. Disscution and Conclution

Pollution index forecasting has become significantly important in populated, industrial cities due to its direct impact on human health. The overall goal of this study was to compare widely applied machine learning algorithms models for the forecasting of NO_2 and CO concentration levels. This information can be used to develop comprehensive strategies and policies for reducing the concentrations of air pollutants such as NO_2 and CO.

To evaluate the predictive performance of the proposed models, the models were used to generate one-month ahead forecasts of NO_2 and CO concentration based on air quality data collected in Punak, Tehran.

In this work, pollution data was used from the air quality monitoring station in Punak, Tehran from 2017 to 2020, and then missing data were removed for use in machine learning algorithms. The ability of the ARIMA, LSTM, and MLR machine learning methods to forecast the daily concentrations of NO_2 and CO were compared. Performance of the three models was evaluated using four performance measures comparing the predicted values to the observed values: RMSE, MAE, RRSE, and R^2. Based on the results gained from the models, it can be concluded that both the LSTM and MLR models performed better than the ARIMA model when forecasting NO_2 and CO concentrations.

The LSTM and MLR closely resembled the original and had the highest performance accuracy for NO_2 Datasets, with RRSE values of 1.06 and 1.10, and R^2 values of 0.458 and 0.46, respectively. The ARIMA model, on the other hand, performed the worst. ARIMA's RMSE was 47.39, and its RRSE was 6.08. With an RMSE of 0.32, RRSE of 0.38, R^2 of 0.981, and MAE of 0.25 for the CO Dataset, the MLR performed slightly better than the LSTM and was more effective than the ARIMA model. The actual and forecasted values for the ARIMA model, on the other hand, were significantly different, with an R^2 value of 0.01. This finding is consistent with those of previous studies, which have shown that nonlinear models (such as neural networks) tend to outperform ARIMA models.

One limitation of this study lies in the fact that air quality concentrations were the only variables used for model development. Using other variables such as temperature and humidity is recommended in future studies.

This study focused on a dataset for the Punak air quality monitoring station. Due to the importance of air quality, the researchers recommend validating the performance of these models using the datasets from several additional air quality monitoring stations under different conditions.

Furthermore, the researchers recommend using deterministic models for air pollution index forecasting, as they can be combined with statistical models to obtain more accurate results. More comprehensive data will lead to more accurate predictions. Thus, using larger datasets and more complicated models will likely enhance the accuracy of the predictions for this important issue.

8. Conflict of Interest

The authors declare no conflict of interest.

9. References

– Abdullah, S., Ismail, M., & Fong, S. Y. (2017). Multiple Linear Regression (MLR) models for long term Pm 10 concentration forecasting during different monsoon seasons. Journal of Sustainability Science and Management, 12(1), 60–69.

– Athira, V., Geetha, P., Vinayakumar, R., & Soman, K. P. (2018). DeepAirNet: Applying Recurrent Networks for Air Quality Prediction. Procedia Computer Science, 132, 1394–1403. https://doi.org/10.1016/j.procs.2018.05.068

– Brunekreef, B., & Holgate, S. T. (2002). Air pollution and health. Lancet, 360(9341), 1233–1242. https://doi.org/10.1016/S0140-6736(02)11274-8

– Connor, J. T., Martin, R. D., & Atlas, L. E. (1994). Recurrent neural networks and robust time series prediction. IEEE Transactions on Neural Networks, 5(2), 240–254. https://doi.org/10.1109/72.279188

– Council. (2002). The Ongoing Challenge of Managing Carbon Monoxide Pollution in Fairbanks, Alaska. In The Ongoing Challenge of Managing Carbon Monoxide Pollution in Fairbanks, Alaska. National Academies Press. https://doi.org/10.17226/10378

– Dey, S., Sibanda, P., Gupta, S., & Chakraborty, A. (2009). Analyzing and predicting the criteria pollutants over a tropical urban area by using statistical models. 2.

– Dragomir, C. M., Voiculescu, M., Constantin, D. E., & Georgescu, L. P. (2015). Prediction of the NO2 concentration data in an urban area using multiple regression and neuronal networks. AIP Conference Proceedings, 1694(2). https://doi.org/10.1063/1.4937255

– EPA. (2001). Latest findings on national air quality: 2000 status and trends. EPA Publications, 454 K-01–002, 2–26.

– Figueiredo Filho, D. B., Silva Júnior, J. A., & Rocha, E. C. (2011). What is R2 all about? Leviathan (São Paulo), 3, 60. https://doi.org/10.11606/issn.2237-4485.lev.2011.132282

– Frank R. Giordano, W. P. F. and, & Horton, S. B. (2000). A course in mathematical modeling. In Richard Stratton (Vol. 37, Issue 05). Richard Stratton.

– Gers, F. A., Schmidhuber, J., & Cummins, F. (1999). Learning to forget: Continual prediction with LSTM. IEE Conference Publication, 2(470), 850–855. https://doi.org/10.1049/cp:19991218

– Hamzaçebi, C. (2008). Improving artificial neural networks' performance in seasonal time series forecasting. Information Sciences, 178(23), 4550–4559. https://doi.org/10.1016/j.ins.2008.07.024

– Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

– Jenkins, B. O. X., Approach, A., Predicting, T. O., Fdi, N. E. T., & In, I. (2011). Box-Jenkins ARIMA approach to predicting net FDI inflows in Zimbabwe. 87737.

– Juhos, I., Makra, L., & Tóth, B. (2008). Forecasting of traffic origin NO and NO2 concentrations by Support Vector Machines and neural networks using Principal Component Analysis. Simulation Modelling Practice and Theory, 16(9), 1488–1502. https://doi.org/10.1016/j.simpat.2008.08.006

– Kumar, K., Yadav, A. K., Singh, M. P., Hassan, H., & Jain, V. K. (2004). Forecasting daily maximum surface ozone concentrations in brunei darussalam—an ARIMA modeling approach. Journal of the Air and Waste Management Association, 54(7), 809–814. https://doi.org/10.1080/10473289.2004.10470949

– Mohammadi-Zadeh, M. J., Karbassi, A., Bidhendi, G. N., Abbaspour, M., & Padash, A. (2017). An Analysis of Air Pollutants' Emission Coefficient in the Transport Sector of Tehran. Open Journal of Ecology, 07(05), 309–323. https://doi.org/10.4236/oje.2017.75022

– Roberts, S., Arseneault, L., Barratt, B., Beevers, S., Danese, A., Odgers, C. L., Moffitt, T. E., Reuben, A., Kelly, F. J., & Fisher, H. L. (2019). Exploration of NO 2 and PM 2.5 air pollution and mental health problems using high-resolution data in London-based children from a UK longitudinal cohort study. Psychiatry Research, 272(2), 8–17. https://doi.org/10.1016/j.psychres.2018.12.050

– Safriet, D. W., & Brooks, G. (1989). Estimating air toxics emissions from coal and oil combustion sources.

– Srivastava, C., Singh, S., & Singh, A. P. (2019). Estimation of air pollution in Delhi using machine learning techniques. 2018 International Conference on Computing, Power and Communication Technologies, GUCON 2018, 304–309. https://doi.org/10.1109/GUCON.2018.8675022

– Torkian, A., Bayat, R., Najafi, M. A., Arhami, M., & Askariyeh, M. H. (2012). Source Apportionment of Tehran ' s Air Pollution by Emissions Inventory. International Emission Inventory Conference, August, 41. www.epa.gov/ttnchie1/conference/ei20/finalprogram.pdf

– Wark, K., Warner, C. F., & Davis, W. T. (1998). Air Pollution: Its Origin and Control (3rd Edition) (3rd Editio).

– Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research, 30(1), 79–82. https://doi.org/10.3354/cr030079

– Xayasouk, T., Lee, H. M., & Lee, G. (2020). Air pollution prediction using long short-term memory (LSTM) and deep autoencoder (DAE) models. Sustainability (Switzerland), 12(6). https://doi.org/10.3390/su12062570

– Zhang, G. P. (2007). A neural network ensemble method with jittered training data for time series forecasting. Information Sciences, 177(23), 5329–5346. https://doi.org/10.1016/j.ins.2007.06.015

– Zhao, J., Deng, F., Cai, Y., & Chen, J. (2019). Long short-term memory - Fully connected (LSTM-FC) neural network for PM2.5 concentration prediction. Chemosphere, 220, 486–492. https://doi.org/10.1016/j.chemosphere.2018.12.128

– Zohdirad, H., Montazeri Namin, M., Ashrafi, K., Aksoyoglu, S., & Prévôt, A. S. H. (2022). Temporal variations, regional contribution, and cluster analyses of ozone and NOx in a middle eastern megacity during summertime over 2017–2019. Environmental Science and Pollution Research, 29(11), 16233–16249. https://doi.org/10.1007/s11356-021-14923-1.