

A Bidirectional Long Short-Term Neural Network Model to Predict Air Pollutant Concentrations: A Case Study of Tehran, Iran

H. Ghayoumi Zadeh^{1,*}, A. Fayazi¹, Kh. Rezaee², O. Rahmani Seryasat³, H. Rabiee⁴

¹ Assistant Prof, Department of Electrical Engineering, Faculty of Engineering, Vali-E-Asr University of Rafsanjan, Rafsanjan, Iran.

² Assistant Prof, Department of Biomedical Engineering, Meybod University, Meybod, Iran.

³ Assistant Prof, Department of Electrical Engineering, Shams Higher Education Institute, Gorgan, Iran.

⁴ Mechatronics Faculty, Department of electrical Engineering, Karaj Branch, Islamic Azad University, Alborz, Iran.

ARTICLE INFO	ABSTRACT
<p>Article History: Received 13 January 2022 Received in revised form 15 February 2022 Accepted 5 June 2022 Available online 7 June 2022</p>	<p>Air pollution remains one of the most pressing environmental challenges in modern urban societies, driven by rapid industrialization, accelerated urban expansion, increasing vehicular traffic, and intensified anthropogenic activities. The presence of harmful substances in the atmosphere, including sulfur dioxide (SO₂), nitrogen dioxide (NO₂), ozone (O₃), particulate matter (PM_{2.5} and PM₁₀), and carbon monoxide (CO), poses significant health risks and environmental hazards. Monitoring and forecasting pollutant concentrations are critical for effective air quality management and policy-making. This study presents a novel hybrid model based on deep recurrent neural networks, particularly the Bidirectional Long Short-Term Memory (BiLSTM) architecture, for short-term air quality prediction. The model focuses on forecasting the Air Quality Index (AQI), a composite measure influenced by several pollutants with highly nonlinear and complex behavior. Daily average concentration data for O₃, PM_{2.5}, PM₁₀, NO₂, SO₂, and CO were collected from the Tarbiat Modares air quality monitoring station in Tehran's 6th district (latitude 35.71751, longitude 51.385909) over a 15-month period between March 2018 and June 2019. The proposed BiLSTM model demonstrated superior performance compared to traditional shallow learning techniques and Multi-Layer Perceptron (MLP) networks. The regression coefficients (R²) for O₃, PM_{2.5}, PM₁₀, NO₂, SO₂, and CO were 0.87, 0.62, 0.84, 0.67, 0.75, and 0.72, respectively. These results highlight the robustness and reliability of the BiLSTM-based approach for accurately predicting pollutant concentrations, thereby supporting timely decision-making in environmental health management.</p>
<p>Keywords: Air Pollution, Predicting, Deep Recurrent Neural Networks, Bidirectional Long Short-Term Model.</p>	

* Corresponding author: Vali-e-Asr University of Rafsanjan, Main Administration Building, 22 Bahman Square, Rafsanjan, Iran. Tel: (034) 31312398, Fax: (034) 31312202, Email: h.ghayoumizadeh@vru.ac.ir



1. INTRODUCTION

Among the most important human needs, the air is the most well-known vital need that if a disruption occurs in it, it will make life impossible [1]. Air pollution is one of the most important challenges of urban communities, and causes a lot of damage to human health. Concerns about air quality have increased due to increasing respiratory problems, especially in children, the elderly and people with respiratory diseases related to air pollution [2]. The pollution is a peripheral product of the following human activities: urban construction, the remaining waste resulting from activities such as the production of goods, transportation, energy and light heating living quarters, recreation and human labor [3]. Air pollution is one of the environmental issues that can cause a variety of environmental hazards. Environmental hazards including damage caused by air pollution attract a large amount of funds, human resource and manpower every year. The appropriate air pollution control is essential given that air quality is a general category [4]. One of the most important issues in the field of air pollution is the public awareness of the amount of pollution, because public awareness can inform managers and citizens about pollution, and the managers take control measures to reduce the air pollution. Moreover, the level of citizens' information and awareness about the amount of air pollution and the prediction of air pollution in accordance with current trends can attract their participation in reducing the pollution. Therefore, researchers trying to predict the level of air pollution [5]. Air pollution is one of the phenomena that changes in time and space; hence, the spatio-temporal predictions of ambient air pollution is of particular importance.

1.1. Related work

Two types of models can be used to model and temporal prediction of air pollution: deterministic and statistical models. Prediction of ozone concentration using partial differential equations to create a deterministic model is a complicated process that requires the use of a large number of chemical and physical variables, and very precise input data. Therefore, the development of these types of models is often difficult [6]. One such model is the Finardi et al., Which developed in 2008. They created a system for air pollution modeling using meteorological parameters and chemical transfer model [7]. In 1998, Pelidores et al. presented a dispersion model based on traffic, industrial centers, and wind rates to predict hourly air pollutant concentrations and compared it with univariate Box-Jenkins stochastic model. Their results showed that a model incorporating the stochastic model and the dispersion model can perform well [8].

To analyze the regression and prediction of hourly air pollution concentrations, Ping shi and Harrison created a model based on meteorological data and the concentration of, an ordinary least squares (OLS) model and a first-order autocorrelation (AR) model in 1997. Their findings showed that the AR model outperformed the other models [9]. Hubbard and Cobourn devised a multiple-linear regression model in 1998 that used just surface meteorological data from 1993 to 1996 to estimate daily domain-peak ground-level ozone concentration (O₃) [10]. Some research have been focused on modeling of air pollution using linear models, because conventional stochastic models are not capable of identifying and modeling the nonlinear relations of air pollution.

For example, in 2002, Dueñas et al. present an analysis of a year of ozone concentrations measured in the Málaga region. In this study, they used multivariate regression to predict ozone concentrations using meteorological parameters [11]. In recent years, neural networks has been used to model and predict air pollution that are able to identify the nonlinear relationships between the input and output data. Neural networks have a high ability to extract patterns from the data and solve the complex problems that are too complicated to be noticed by either humans or other computer techniques. While neural networks have different performance and properties, so that a kind of neural network may show a good performance problem, but another kind of neural network can solve the same problem. However, various neural networks have different performance and properties, so that a type of neural network may have good performance to solve a problem but another type of neural network to solve the problem poorly. Therefore, the use of neural networks and the comparison of their performance in predicting and modeling of air pollution seem to be necessary [12]. Among the researches that utilize the neural networks to predict air pollution can refer to Cai et al., in 2009, which modeling the air pollution using neural network and the parameters correlated with the concentrations of pollutants. Factors that influence pollutant concentrations are traffic-related, background concentration, meteorological and geographical. These parameters include traffic volumes, hours of day, weekday, pollutant concentrations in the past 1-3 hours, wind speeds and directions, temperature, solar radiation, rainfall,

humidity, distance to the street axis and street direction. They demonstrated that the models are able to produce accurate prediction of hourly concentrations of the pollutants respectively more than 10 h in advance [13]. In 2012, Cheng et al. used a type of neural network to improve regional air quality modeling; their model using meteorological parameters has a high potential for air quality modeling [14]. In 2007, Dutot et al. developed a model based on the Multi-Layer Perceptron (MLP) neural network for modeling and a real time forecasting of 24 hourly maximum ozone concentrations in the centre of France, in an urban atmosphere. Their model inputs are meteorological parameters and the output is maximum ozone concentrations. Their model had a good ability to predict the maximum ozone concentration in the next 24 hours. The neurozone model was used for real-time ozone prediction and updated annually in September with new data [15]. In 2008, Coman et al. utilized two static and dynamic static Multi-Layer Perceptron neural network models to predict ozone concentrations in the following day.

Their results indicate a rather good applicability of these models for a short-term prediction of ozone Coman, A., [6]. In 2008, al-Alawi et al. [16] used meteorological variables including wind speed, wind direction, solar radiation, relative humidity and temperature and concentration of pollutants in 5 stations to predict the concentration of pollutants in Kuwait. They created a combined model based on the artificial neural network (ANN) and the multiple regression combined with principal component analysis (PCR). Their results showed that combining the predictions from the PCR and ANN models can reduce the root mean square errors (RMSE) of ozone concentrations [16]. Several studies have also been conducted focusing on the spatial modeling of air pollution. A number of studies have emphasized the use of geostatistical techniques for interpolation of data. In 2005, Pummakarnchana et al. developed an air pollution monitoring system for increased public awareness and enhanced public participation in Bangkok. They have been used the local deterministic and geostatistical interpolation methods for spatial prediction [17]. In 2009, Shad et al. proposed a combined model consisting of genetic algorithm (GAs) and fuzzy linear membership kriging to predict air pollution in Tehran.

The results showed that the proposed model has a good accuracy over the linear model [18]. Meanwhile, interpolation techniques are not able to properly model the concentration of air pollutants. Because it only considers the concentration of the pollutant; and the parameters affecting air pollution do not participate in the model. The information needed to predict air pollution in a system is time variant. In order to use the time series, comprehensive and complete data seems to be necessary. By analyzing, studying and modeling this time series can be used to predict the concentration of pollutants. Given the fact that the behavior of a time series about the concentration variables of pollutants may be random and have chaotic behaviors, the use of complex and powerful neural networks seems to be necessary for better predictions. For this purpose, in this paper attempts to use BI-LSTM deep neural networks to obtain better results over the other neural network that has been ever used to predict the concentration of pollutants.

2. MATERIALS AND METHODS

2.1. Dataset

Tehran is the biggest and most populous city in Iran, which due to overcrowding and traffic congestion is affected by severe air pollution. In recent decades, because of excessive population growth, extreme increase in vehicle traffic and the concentration and accumulation of the industries, Tehran has been encountered with serious environmental crises such as air pollution [19].

Air pollution in Tehran sometimes exceeds the standard level, which means that they have numerous effects on the health of Tehran citizens. In some cases it causes a complete closure of the city. Prediction of air pollution for public awareness can have an important role in air pollution control. Air pollution has two spatial and temporal aspects. The analysis and prediction of air pollution just from the spatial aspect and regardless of the temporal aspect is not appropriate. Although the temporal aspect of air pollution is more important than its spatial aspect, because of temporal variations of urban air pollution are more tangible. However, spatial prediction also helps identify the areas that the pollution is in worse condition and then try to improve conditions. A time series is a sequence of numerical data points in successive order about specific phenomena that are taken by a variable over time. In this study, among various air pollutants, the AQIs (O₃, PM_{2.5}, PM₁₀, NO₂, SO₂, CO) prediction has been considered. The data on O₃, PM_{2.5}, PM₁₀, NO₂, SO₂ and CO concentrations is average daily report of the 6th district of Tehran, Tarbiat Modares station, which was collected between March 2018 and June 2019. The station is located at 35.71751 latitude and 51.385909 longitudes. The time series of the concentration of pollutants are shown in Figure 1.

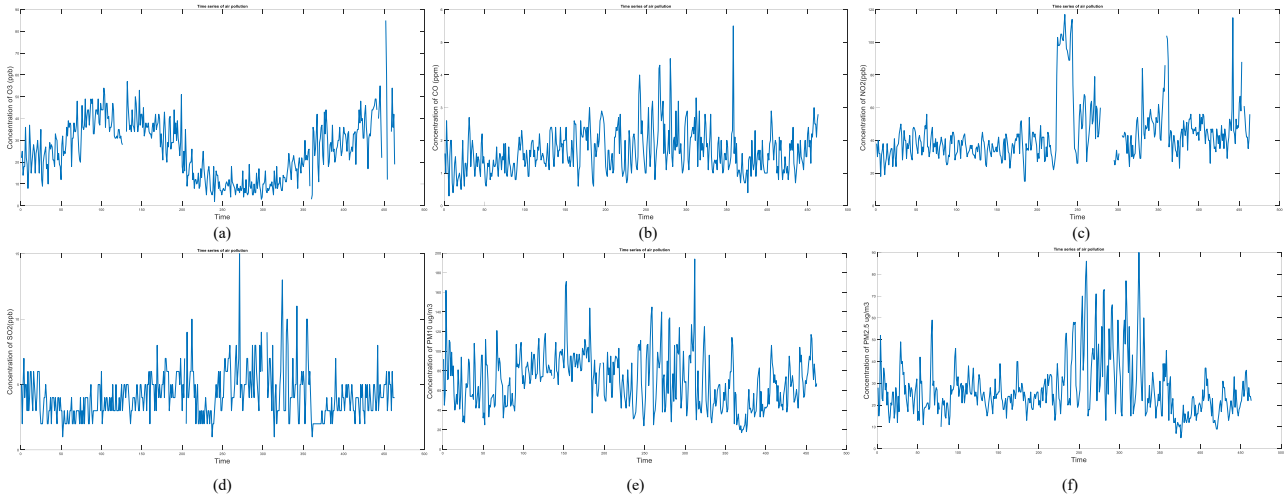


Fig. 1. Time series of concentration of pollutants O₃, PM_{2.5}, PM₁₀, NO₂, SO₂ and CO

Specifications of the concentration of pollutants in the related database are shown in Table 1.

Table 1. Statistical specification of the concentration of pollutants used in the database

Data Set	var	SD _x	X _{min}	X _{max}	X _{mean}	Unit
O3	177.12	13.3	2	85	25.93	PPb
CO	0.4489	0.67	0.3	5.5	1.6691	ppm
NO2	294.402	17.15	15	117	44.0951	ppb
SO2	3.331	1.825	1	15	4.2212	ppb
PM10	736.72	27.14	17	194	70.33	Ug/m ³
PM2.5	171.78	13.106	5	90	27.424	Ug/m ³

Each of the variables is explored in greater depth below. The word ozone comes from the Greek word ozein, which means "to smell." Ozone has a distinct pungent odor that allows it to be detected even in trace amounts. Ozone will rapidly react with many chemical compounds and is explosive in concentrated amounts.

The oxygen molecule has a double atom of oxygen, but the difference in an oxygen atom has created the fundamental differences between the two molecules. Ozone is produced by photochemical reactions of hydrocarbons from the exhaust of cars and nitrogen oxides in the atmosphere. Ozone is usually unstable and acts like chlorine in many reactions. It acts as an antimicrobial agent with the mechanism of combining with proteins and disabling the reducing enzyme that is important for the respiration of the cell. Carbon monoxide, or "CO" is a colorless, odorless gas but highly toxic. Carbon monoxide (CO) is a product of incomplete combustion of hydrocarbons. The main source of the production of this gas is the cars.

The carbon monoxide gas in open air and in a small amount endangers the lives of patients with cardiovascular and pulmonary diseases. In healthy people, it causes headaches, dizziness, and high fatigue and stimulates the nerves. Inhalation of this gas in enclosed and roofed environments causes suffocation and death. The materials in the soil can absorb some of the carbon monoxide and some of them are converted to carbon dioxide through interactions. However, if the amount of carbon monoxide exceeds 750 ppm, it will cause death. Particulate matters can be classified into the following main groups: fine particles with diameters generally smaller than 2.5 μm, coarse particles with diameters generally larger than 2.5 μm, fumes with a diameter of from 0.001 to 1 μm, and mist with a diameter of 0.1 to 10 μm. Fumes are particles formed by distillation, or chemical reaction, and their diameter is relatively larger than smoke.

Particulate matters with a diameter of less than 10 microns are introduced as the main indicator of suspended matter because of getting to the lower respiratory tract. According to studies, suspended particles are more hazardous

than sulfur oxides and nitrogen oxides for health, and the PM₁₀ pollution levels have an important role in exacerbating cardiopulmonary disease, reducing the immune system against diseases, lung tissue destruction, childhood asthma, early mortality and cancer. Nitrogen dioxide (NO₂), released more than other nitrogen oxides in the air, is brown and stinky gas that enters the air through motor vehicles and factories that use internal combustion engines. This gas stimulates the eyes and deep parts of the lungs and causes excessive tiredness and increase in the number of cases of a disease. Sulfur dioxide (SO₂) and sulfur trioxide are sulfur species that are found in the lower Earth's atmosphere. Sulfur dioxide is a colorless, non-flammable, and non-explosive gas. Sulfur dioxide is usually converted to sulfur trioxide (SO₃) in the atmosphere and then sulfur trioxide is rapidly converted to sulfuric acid (H₂SO₄). Sulfur oxides will have the most severe adverse effects on the environment if they are in the air where suspended particles are present and relative humidity is also relatively high. The correlation coefficient of each data relative to the other variables is shown in Table 2.

Table 2. Correlation matrix between concentrations of pollutants in the time series

	O ₃	CO	NO ₂	SO ₂	PM ₁₀	PM _{2.5}
O ₃	-----	-0.28	-0.31	-0.25	-0.03	-0.43
CO	-0.28	-----	0.37	0.50	0.52	0.55
NO ₂	-0.31	0.37	-----	0.13	0.07	0.34
SO ₂	-0.25	0.50	0.13	-----	0.49	0.57
PM ₁₀	-0.03	0.52	0.07	0.49	-----	0.67
PM _{2.5}	-0.43	0.55	0.34	0.57	0.67	-----

As can be seen from the Table 2, PM_{2.5} and PM₁₀ variables have the highest correlation to each other.

3. MODELS

Before the data are imported into the proposed model, we must preprocess them. Given that the amount of some variables was not recorded in some days, the average 4 days ago, was considered. The results of this method are plotted on one of the concentrations of pollutant (O₃) in Fig. 2. As can be seen from Fig. 2, the discontinuities segments have been connected with appropriate and acceptable accuracy by applying the proposed method.

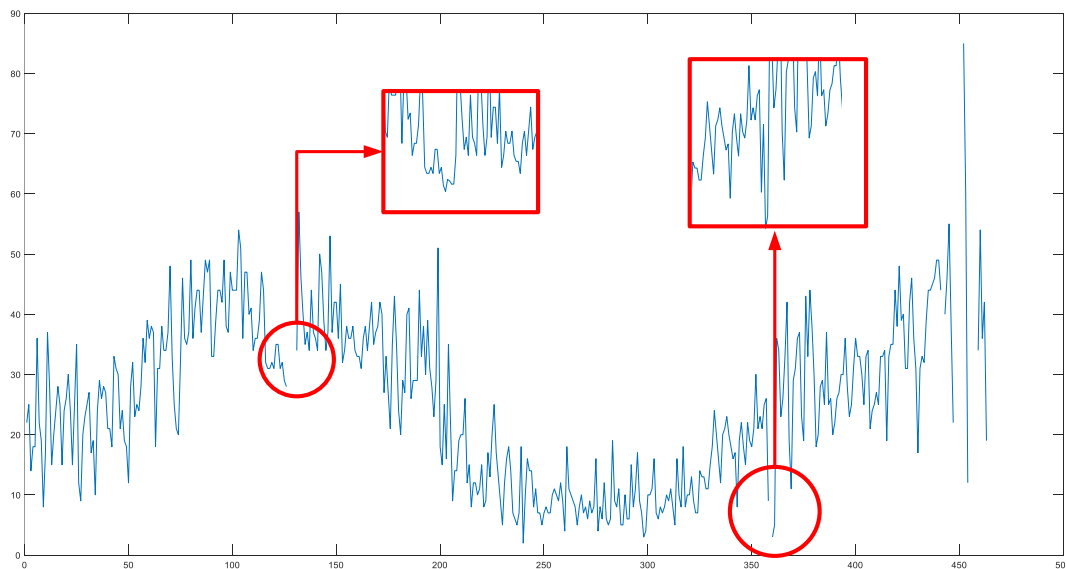


Fig. 2. Compensation of empty values using an average of 4 days ago

After that the discontinuities segments have been corrected, the next step of preprocessing of the data before entering the data into the neural network is normalization. Data normalization is done in order to increase the

accuracy and speed of the network, as well as to minimize the changes of the neuron's weight and to enhance neurons quicker response to input signals. The equation used to normalize is as follows:

$$x_{norm} = \frac{x - \bar{x}}{\sigma} \tag{1}$$

Where \bar{x} mean of data, σ is the standard deviation of the data. We now come to the prediction phase of the time series, which is based on a deep recurrent neural network with BI-LSTM model. Deep learning is a subfield of machine learning methods based on artificial neural networks used to learn how to describe and represent data. Deep learning can be supervised, semi-supervised or unsupervised. Employing different methods of deep learning have led to very significant changes in a variety of areas such as machine vision, natural language processing, speech recognition, audio recognition, social network filtering, machine translation, bioinformatics, drug design, medical image analysis, and material inspection. One of the types of deep learning is recurrent neural networks. Sometimes, for output calculation, it is not only necessary to input value at the current moment, but also the output value is dependent on the information seen at the earlier times.

The main idea in the recurring neural networks is to consider a hidden state for the network, which as a network memory; it has a duty to keep information obtained from the past inputs. In a recurrent neural network, this is also referred to as a recurrent cell; the output value of each moment is calculated based on the input value of that moment and hidden state vector of network. Also, the hidden state of the cell is updated on the basis of the new input and the current hidden state to accommodate the new input information for later moments.

As shown in Figure 3, each node contains a function that calculates the current hidden state h_t , and the output y_t is derived using the current input x_t and the prior hidden state h_{t-1} using the following equations:

$$h_t = f(w_{hh}h_{t-1} + w_{hx}x_t) \tag{2}$$

$$y_t = w_{yh}h_t \tag{3}$$

where w_{hh} , w_{hh} , and w_y (also known as weight matrices) are the weights for the hidden-to-hidden, input-to-hidden, and hidden-to-output connections, respectively. Furthermore, each node is coupled with an activation function f . This is an element-wise non-linearity function that is frequently selected from a set of existing functions such as the sigmoid and hyperbolic tangent. The current hidden state is h_t , while the current output is y_t .

It should be noted that the weights of a recurrent cell are shared between steps of different times [20].

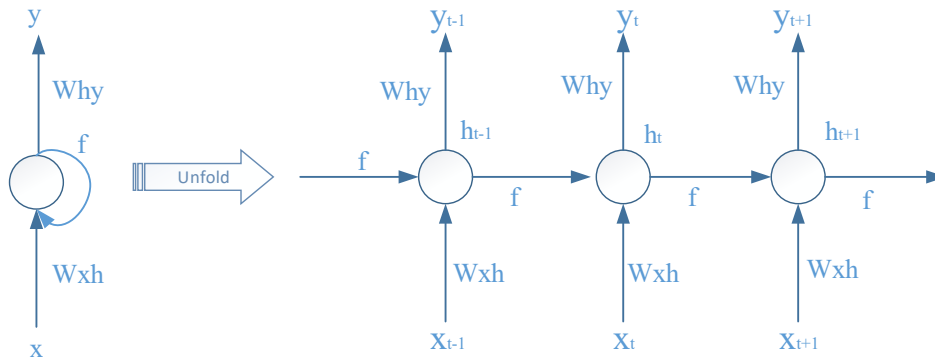


Fig. 3. Diagram of an RNN node through time, where h_{t-1} represents the previous hidden state, x_t represents the current input sample, h_t represents the current hidden state, y_t represents the current output, and f represents the activation function.

The length of the input sequence to a recursive cell can be limited or unlimited. Because of the sharing of cell weights between different time steps, increasing the length of the input sequence does not make any problem in increasing the number of parameters for using these networks. However, in practice, a simple recurrent network operates poorly in modeling long-term dependencies. The reason can be found in the way of training and updating the weights of the recurrent network. In order to train the recurrent network, at each time step, the calculated gradient

from the cost function of the network output is propagated to all moments and updates the weights (which are shared between all the moments). In single-output applications, the gradient is calculated at the last time step. As the length of the sequence increases, the gradient that reaches out to the weights of the distant past times from the output may be very small or very large that this phenomena, called the gradient vanishing and gradient exploding, respectively [20]. Therefore, the gradient that reaches out to the distant past times either will not affect the update of weights or will disrupt the training process.

The Long Short-Term Memory (LSTM) cells were introduced to resolve this problem. Fig. 4 shows the architecture of a common LSTM unit.

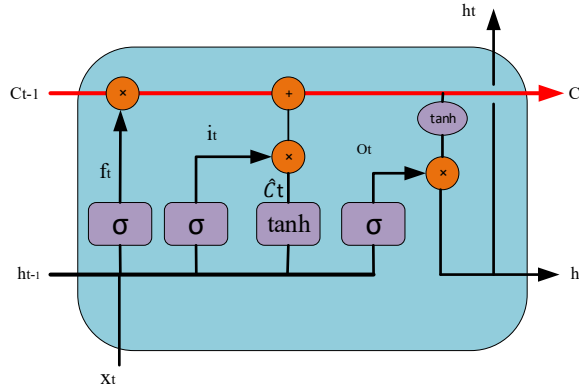


Fig. 4. Architecture of a common LSTM cell

The LSTM unit may process input sequentially while remaining concealed over time. As illustrated in Figure 4, a common LSTM unit incorporates a variety of gate units. These gates determine when prior concealed states are forgotten and when states are updated with new information. These gates, in general, control the flow of information into and out of the cell.

The most important part of the LSTM cell architecture that eliminates the problem of the gradient vanishing is the existence of a relatively straight forward path (indicated by red) for the flow of gradients from the output to the cell input. The function of each LSTM unit is as follows [21]:

$$f_t = \sigma(w_f \cdot [h_{t-1}, \Phi_{t-1}] + b_f) \tag{4}$$

$$i_t = \sigma(w_i \cdot [h_{t-1}, \Phi_{t-1}] + b_i) \tag{5}$$

$$o_t = \sigma(w_o \cdot [h_{t-1}, \Phi_{t-1}] + b_o) \tag{6}$$

$$\hat{C}_t = \tanh(w_c \cdot [h_{t-1}, \Phi_{t-1}] + b_c) \tag{7}$$

where b_t , h_t , and Φ_t are network inputs. $b_c, b_o, b_i, b_f, w_c, w_o, w_i$ and w_f are the corresponding weight matrix. σ represents the sigmoid activation function. i_t, f_t and o_t are input gate, forget gate and output gate, respectively. The output of the LSTM cell is h_t which is calculated from the following equations:

$$C_t = f_t C_{t-1} + i_t \hat{C}_t \tag{8}$$

$$h_t = o_t \tanh(C_t) \tag{9}$$

According to the studies conducted in the time series, we know that the proper output of the recurrent cell at each moment is dependent on both the previous moments and the subsequent components of the sequence. Bidirectional recurrent networks are the solution for modeling time information about before and after each moment.

These networks use two sub-layers to account for the complete input context by having input sequences in both directions. These two sublayers create the output sequence y by computing the forward and backward hidden

sequences \vec{h} , \overleftarrow{h} respectively (see Fig. 5). Figure 5 depicts the architecture of a BLSTM, where each output is dependent on each individual input sequence (Zhou et al. 2016).

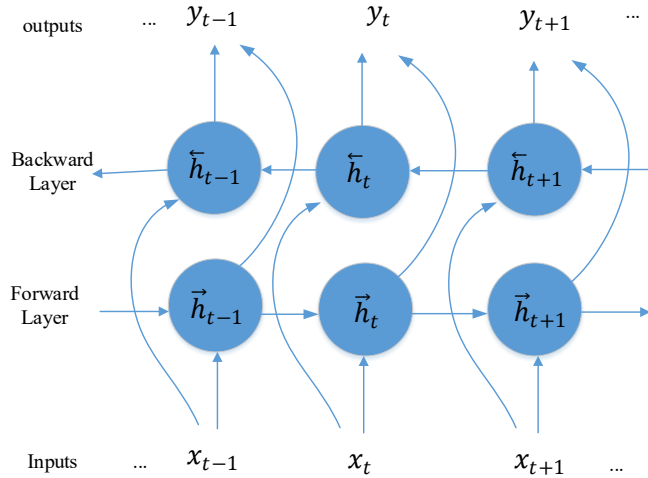


Fig. 5. Architecture of a BLSTM

A notable point in the way of the training of the proposed neural network is that the time delays of 1, 2, 3, 4, 5, 6, 12, 16, 18, 20 and 24 days ago were given to the network input. Otherwise the results will be much lower than the desired values.

4. RESULTS

Since no neural network with specific information structure can be considered as the most appropriate network, networks are selected based on repeated trials, errors method and are evaluated by various criteria. The network settings used in optimal conditions are presented in the Table 2.

Table 2. Optimal parameters of the Bi-LSTM Deep Neural Network

Variable	value	Description
numHiddenUnits1	250	Count of hidden layers in bi-LSTM unit One.
numHiddenUnits2	250	Count of hidden layers in bi-LSTM unit Two.
maxEpochs	2500	Maximum number of training Epoch in bi-LSTM.
miniBatchSize	120	Minimum batch size in bi-LSTM.

We trained our models on total available data using 80% of the data for training and the remaining 20% for testing the model results. The result of the prediction of the concentration of O3 pollutant is shown in Fig. 6.

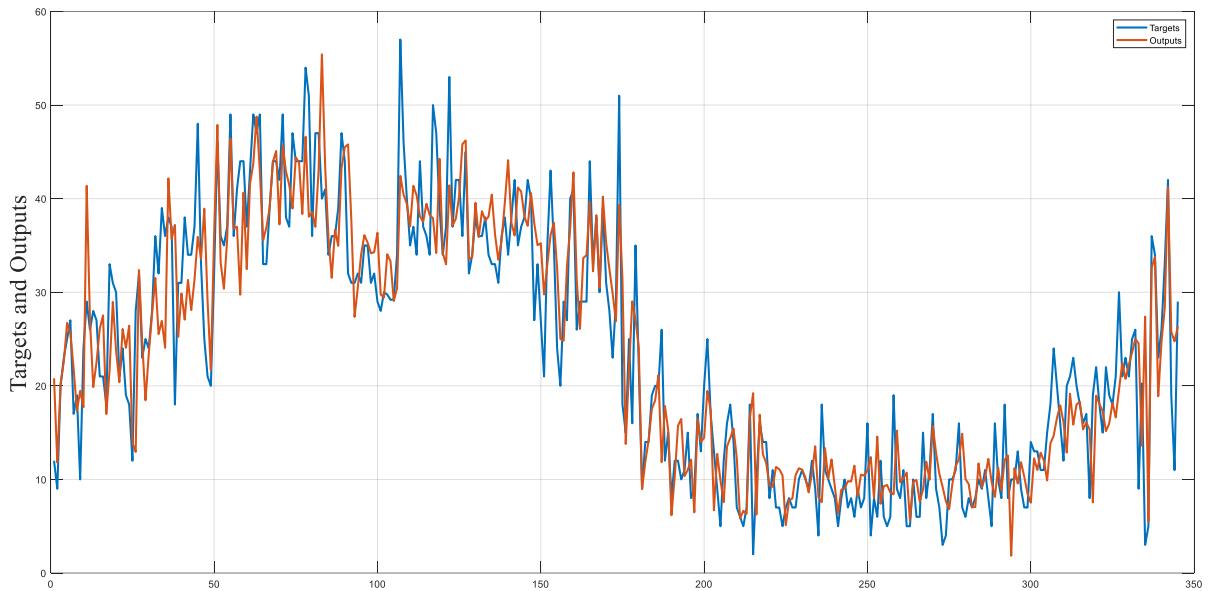


Fig. 6. The performance of the LSTM neural network in predicting the concentration of O₃ pollutant over its actual value

As it can be seen from Fig. 6, the proposed network has been able to pursue the concentration of O₃ pollutant with appropriate and acceptable precision. The standard deviation of the predicted error of the concentration of O₃ pollutant over the actual value is shown in Fig. 7.

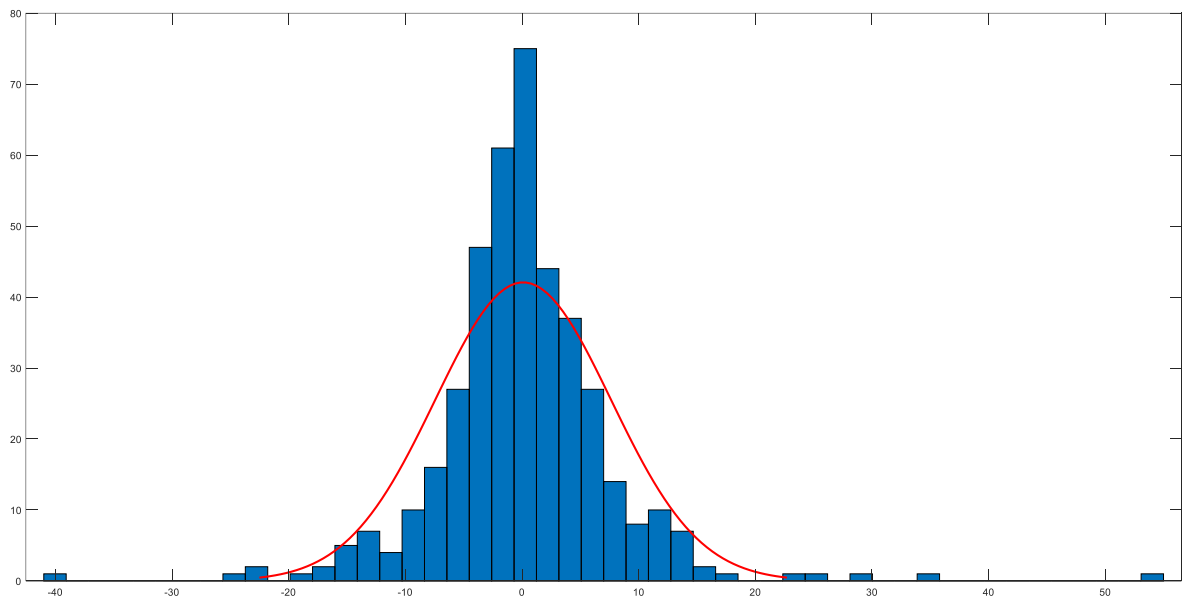


Fig. 7. The standard deviation of the error obtained from the proposed model over the actual value of the concentration of O₃ pollutant.

One of the values measured in the results is the calculation of the regression coefficient. The regression coefficient of the concentration of O₃ pollutant is shown in Fig. 8. As the correlation coefficient is closer to 1 or -1, the strength of the linear relationship between the actual value and the predicted value is more severe. In fact, as the data is closer to the one-to-one graph, it will indicate more ability of the model to predict the concentration of pollutants.

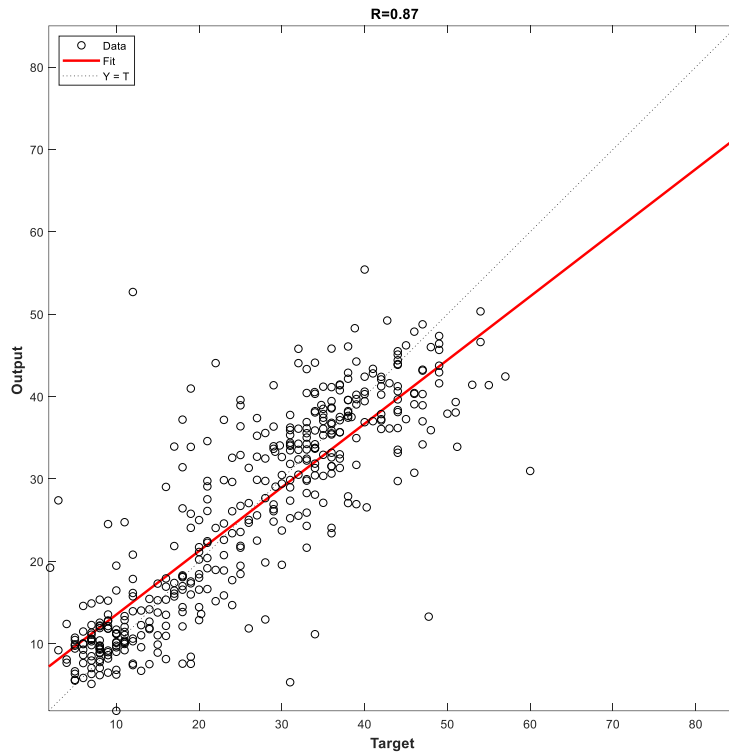


Fig. 8. Scatter plot diagram of the predicted data versus the measured data for O₃ concentration

The criteria for accepting the predicted results in each of the neural network structures were based on the low error rate obtained from computational values including of the mean square error (MSE), the root mean square error (RMSE), the maximum absolute error, the standard deviation, and the correlation coefficient of R² which are calculated according to the following equations:

$$MSE = \frac{1}{N} \sum (P_i - O_i)^2 \tag{10}$$

$$RMSE = \left[\frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2 \right]^{1/2} \tag{11}$$

$$R^2 = 1 - \frac{\sum_{i=1}^N [O_i - P_i]^2}{\sum_{i=1}^N [O_i - \bar{O}]^2} \tag{12}$$

That N is the total number of data, O_i is the original data, P_i is the prediction values, and \bar{O} is the average of the original data. The results of these criteria are shown in Table 3.

Table 3. The results of network performance in predicting concentrations of O₃, CO, NO₂, SO₂, PM₁₀, PM_{2.5}

	MSE			RMSE			Error Mean			Error Std			Regression		
	Train Data	Test Data	All Data	Train Data	Test Data	All Data	Train Data	Test Data	All Data	Train Data	Test Data	All Data	Train Data	Test Data	All Data
O3	34.3	114	47.5	5.85	10.6	6.89	0.088	0.24	0.11	5.86	10.7	6.90	0.90	0.42	0.87
CO	0.37	0.60	0.41	0.61	0.77	0.64	-0.004	0.02	-0.0	0.61	0.78	0.64	0.66	0.29	0.62
NO2	64.58	223	91.0	8.03	14.9	9.54	-0.239	0.49	-0.11	8.04	15.0	9.55	0.89	0.38	0.84
SO2	2.29	2.67	2.36	1.51	1.63	1.53	-0.021	-0.01	-0.02	1.51	1.64	1.53	0.71	0.31	0.67
PM10	369.6	419	377	19.2	20.4	19.4	2.023	7.83	2.99	19.1	19.0	19.2	0.76	0.66	0.75
PM2.5	105	57.4	97.1	10.2	7.57	9.85	-0.491	-2.78	-0.87	10.2	7.10	9.83	0.72	0.46	0.72

By considering the results of the network efficiency in prediction, it is obvious that the highest and lowest correlation coefficients between values are related to O3 and CO respectively.

5. DISCUSSION

In order to evaluate the proposed model, we compared the obtained results with two Perceptron and Shallow Learning models. One of the simplest and, at the same time, the most effective proposed layout for use in actual neural modeling, is a multilayer perceptron (MLP) neural model, which consists of an input layer, one or more hidden layers and an output layer. The results of the comparison of the proposed model with these two models are shown in Table 4.

Table 4. Comparison of the results of the proposed method with MLP and Shallow Learning methods

	MSE			RMSE			Regression		
	This work	Shallow Learning	MLP	This work	Shallow Learning	MLP	This work	Shallow Learning	MLP
O3	47.5	107.83	441.16	6.89	10.38	21.004	0.87	0.76	0.87
CO	0.41	0.409	0.4325	0.64	0.640	0.657	0.62	0.67	0.62
NO2	91.0	156.78	347.46	9.54	12.52	18.640	0.84	0.76	0.79
SO2	2.36	13.21	6.3129	1.53	3.63	2.5126	0.67	0.36	0.66
PM10	377	381.07	821.41	19.4	19.52	28.660	0.75	0.73	0.72
PM2.5	97.1	516.95	411.56	9.85	22.73	20.287	0.72	0.72	0.77

In order to better understanding and comparing the results to each other, the graphical model of the MSE results is shown in Fig. 9. As can be observed from the Fig.9, the MSE value in the proposed model is lower than the other two models in predicting the concentration of pollutants.

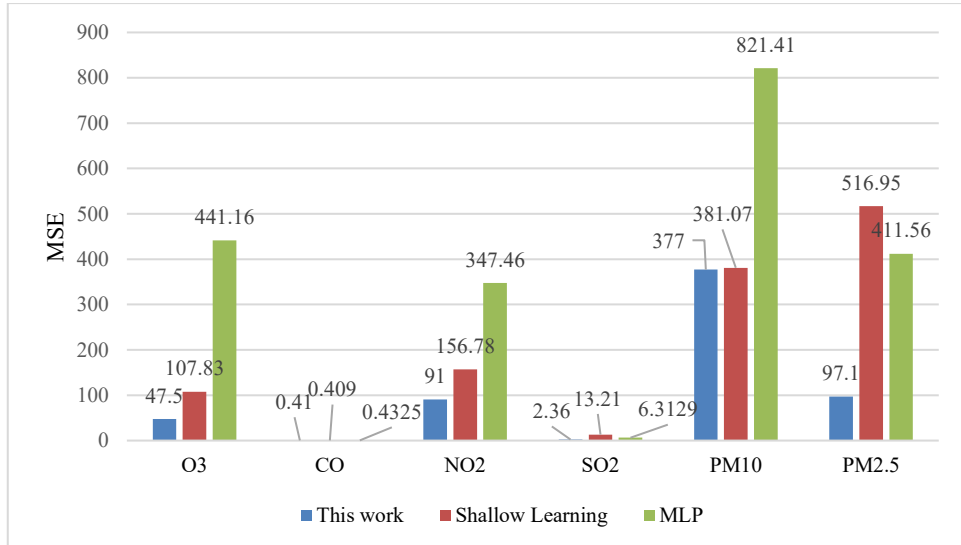


Fig. 9. Comparison of MSE for proposed method over MLP and SHALLOW LEARNING methods for predicting concentration of pollutants

The point is that the R and MSE should be compared with each other simultaneously. The graphical model of R is also shown in Fig. 10. By default, if the R value corresponding to the concentration of the PM2.5 pollutant with the MLP method is higher than the proposed method. However, the MSE or even its RMSE is higher than the proposed method. Therefore, in choosing a good method, all aspects must be considered.

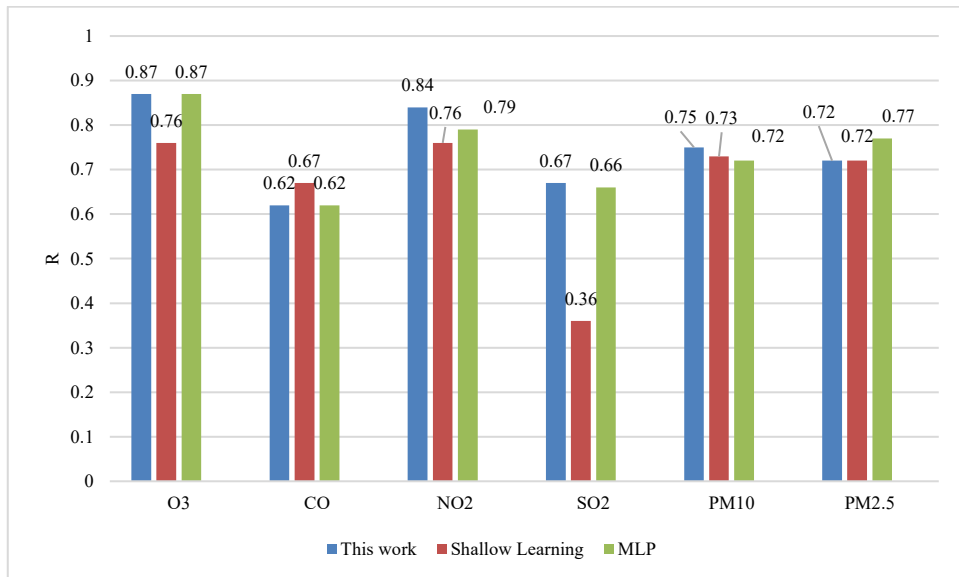


Fig. 10. Comparison of regression coefficient for the proposed method over MLP and SHALLOW LEARNING methods in predicting the concentration of pollutants

The process of updating network weights is repeated over and over again on training events so that the network reaches reasonable performance. However, the excessive repetition of this process may lead to over fitting the network training data. In order to investigate the over-fitting of the network, it is possible to isolate some of the data for validation and measure the performance of the network on validation data during the training process, without the network being trained on those data. In the event of the network performance decreases on validation dataset, while the performance increases on training dataset, over-fitting is occurred. In order to determine the correct number of epochs for the network optimization to avoid over-fitting or under-fitting, it can be used a technique called early stopping. In this method, the training process will continue to a degree that the network performance does not start to decrease on validation dataset. Otherwise, after some epochs, the training process will be stopped.

6. CONCLUSION

Air quality directly affects the quality of life and the human respiratory system. As the climate conditions can change from day to day or even hour to hour, air quality can also vary depending on the time of the weather conditions. Refining management and air quality monitoring in mega cities convert the information corresponding to the air quality into air quality index (AQI) and provide the required information to the general public. Therefore, the AQI is a key tool to understand about air quality and its relationship to human health.

The time series of air quality simultaneously include linear and nonlinear complex patterns that make difficult the prediction process. Regression techniques and artificial neural network models have been widely used to predict air quality so far. However, the default ‘linear regression’ linear in the regression models limits the possibility of the detection and prediction of the severe events and the specific conditions of pollute. Moreover, the neural network models are capable to detect the nonlinear patterns in the real-time behavior of pollutants, but the correct training of such models is obtained with trial and error. Therefore, there is no specific logic for training of these models. The main purpose of this research is to demonstrate the effectiveness of BI-LSTM deep neural networks in predicting the air pollution in Tehran according to the parameters of measured concentrations of air quality monitoring stations in one of the stations in Tehran. The results of time modeling showed the appropriate performance of the proposed hybrid model in geophysical station at a very high level.

CONFLICT OF INTEREST

There is not any conflict of interest in this manuscript.

REFERENCES

- [1] Beamish, L. A., Osornio-Vargas, A. R., & Wine, E. (2011). Air pollution: An environmental factor contributing to intestinal disease. *Journal of Crohn's & Colitis*, 5(4), 279–286. <https://doi.org/10.1016/j.crohns.2011.02.017>
- [2] Garcia, J., Cerdeira, R., Coelho, L., Kumar, P., & Carvalho, M. D. G. (2014). Influence of pedestrian trajectories on school children exposure to PM₁₀. *Journal of Nanomaterials*, 2014, 1–8. <https://doi.org/10.1155/2014/505649> ACM Digital Library+1 Wiley Online Library+1
- [3] Arhami, M., Kamali, N., & Rajabi, M. M. (2013). Predicting hourly air pollutant levels using artificial neural networks coupled with uncertainty analysis by Monte Carlo simulations. *Environmental Science and Pollution Research*, 20(7), 4777–4789. <https://doi.org/10.1007/s11356-012-1451-6> periodicos.capes.gov.br+1ResearchGate+1
- [4] Shakerkhatibi, M., Mohammadi, N., Zoroufchi Benis, K., Behrooz Sarand, A., Fatehifar, E., & Asl Hashemi, A. (2015). Using ANN and EPR models to predict carbon monoxide concentrations in urban area of Tabriz. *Environmental Health Engineering and Management Journal*, 2(3), 117–122. <http://chemj.com/article-1-90-en.htm> lehemj.com
- [5] Kumar, A., & Goyal, P. (2011). Forecasting of daily air quality index in Delhi. *Science of the Total Environment*, 409(24), 5517–5523. <https://doi.org/10.1016/j.scitotenv.2011.08.069>
- [6] Coman, A., Ionescu, A., & Candau, Y. (2008). Hourly ozone prediction for a 24-h horizon using neural networks. *Environmental Modelling & Software*, 23(12), 1407–1421. <https://doi.org/10.1016/j.envsoft.2008.04.004> ACM Digital Library+1ScienceDirect+1
- [7] Finardi, S., De Maria, R., D'Allura, A., Cascone, C., Calori, G., & Lollobrigida, F. (2008). A deterministic air quality forecasting system for Torino urban area, Italy. *Environmental Modelling & Software*, 23(3), 344–355. <https://doi.org/10.1016/j.envsoft.2007.04.001>
- [8] Polydoros, G. N., Anagnostopoulos, J. S., & Bergeles, G. C. (1998). Air quality predictions: dispersion model vs Box-Jenkins stochastic models. An implementation and comparison for Athens, Greece. *Applied Thermal Engineering*, 18(11), 1037–1048. [https://doi.org/10.1016/S1359-4311\(98\)00016-7](https://doi.org/10.1016/S1359-4311(98)00016-7)
- [9] Shi, J. P., & Harrison, R. M. (1997). Regression modelling of hourly NO_x and NO₂ concentrations in urban air in London. *Atmospheric Environment*, 31(24), 4081–4094. [https://doi.org/10.1016/S1352-2310\(97\)00282-3](https://doi.org/10.1016/S1352-2310(97)00282-3)
- [10] Hubbard, M. C., & Cobourn, W. G. (1998). Development of a regression model to forecast ground-level ozone concentration in Louisville, KY. *Atmospheric Environment*, 32(14–15), 2637–2647. [https://doi.org/10.1016/S1352-2310\(97\)00444-5](https://doi.org/10.1016/S1352-2310(97)00444-5)
- [11] Dueñas, C., Fernández, M. C., Cañete, S., Carretero, J., & Liger, E. (2002). Assessment of ozone variations and meteorological effects in an urban area in the Mediterranean Coast. *Science of the Total Environment*, 299(1–3), 97–113. [https://doi.org/10.1016/S0048-9697\(02\)00251-6](https://doi.org/10.1016/S0048-9697(02)00251-6)
- [12] Singh, K. P., Gupta, S., Kumar, A., & Shukla, S. P. (2012). Linear and nonlinear modeling approaches for urban air quality prediction. *Science of the Total Environment*, 426, 244–255. <https://doi.org/10.1016/j.scitotenv.2012.03.076> PubMed+2ScienceDirect+2ScienceDirect+2
- [13] Cai, M., Yin, Y., & Xie, M. (2009). Prediction of hourly air pollutant concentrations near urban arterials using artificial neural network approach. *Transportation Research Part D: Transport and Environment*, 14(1), 32–41. <https://doi.org/10.1016/j.trd.2008.10.004>
- [14] Cheng, S., Li, L., Chen, D., & Li, J. (2012). A neural network based ensemble approach for improving the accuracy of meteorological fields used for regional air quality modeling. *Journal of Environmental Management*, 112, 404–414. <https://doi.org/10.1016/j.jenvman.2012.08.020>
- [15] Dutot, A.-L., Rynkiewicz, J., Steiner, F. E., & Rude, J. (2007). A 24-h forecast of ozone peaks and exceedance levels using neural classifiers and weather predictions. *Environmental Modelling & Software*, 22(9), 1261–1269. <https://doi.org/10.1016/j.envsoft.2006.08.005>
- [16] Al-Alawi, S. M., Abdul-Wahab, S. A., & Bakheit, C. S. (2008). Combining principal component regression and artificial neural networks for more accurate predictions of ground-level ozone. *Environmental Modelling & Software*, 23(4), 396–403. <https://doi.org/10.1016/j.envsoft.2007.06.005>
- [17] Pummakarnchana, O., Tripathi, N., & Dutta, J. (2005). Air pollution monitoring and GIS modeling: a new use of nanotechnology based solid state gas sensors. *Science and Technology of Advanced Materials*, 6(3–4), 251–

255. <https://doi.org/10.1016/j.stam.2005.02.003>

- [18] Shad, R., Mesgari, M. S., Abkar, A., & Shad, A. (2009). Predicting air pollution using fuzzy genetic linear membership kriging in GIS. *Computers, Environment and Urban Systems*, 33(6), 472–481. <https://doi.org/10.1016/j.compenvurbsys.2009.10.004>
- [19] Golbaz, S., Farzadkia, M., & Kermani, M. (2008). Determination of Tehran air quality with emphasis on air quality index (AQI); 2008–2009. *Iran Occupational Health*, 6(4), 62–68.
- [20] 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>
- [21] Zhou, P., Shi, W., Tian, J., Qi, Z., Li, B., Hao, H., & Xu, B. (2016). Attention-based bidirectional long short-term memory networks for relation classification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 207–212). <https://doi.org/10.18653/v1/P16-2034>