Revista de Ciencias Tecnológicas (RECIT). Universidad Autónoma de Baja California ISSN 2594-1925
Volumen 8 (4): e412. Octubre-Diciembre, 2025. https://doi.org/10.37636/recit.v8n4e412

Research article

Predictive modeling of carbon monoxide emissions using deep learning and environmental features from a Mexican border city

Modelado predictivo de emisiones de monóxido de carbono utilizando aprendizaje profundo y variables ambientales de una ciudad fronteriza mexicana

E. Ivette Cota-Rivera[1], Abelardo Mercado-Herrera[1], Fabián N. Murrieta-Rico[1]

[1]Universidad Politécnica de Baja California, Av. Claridad, Plutarco Elías Calles, 21376 Mexicali, Baja California, México.

Corresponding autor: E. Ivette Cota-Rivera, Universidad Politécnica de Baja California, Av. Claridad, Plutarco Elías Calles, 21376 Mexicali, Baja California, México. E-mail: eicotar@upbc.edu.mx. ORCID: 0009-0002-6240-9269.

Received: March 26, 2025 **Accepted**: September 8, 2025 **Published:** October 4, 2025

Abstract. – *Carbon monoxide (CO) poisoning constitutes a critical issue with global ramifications, resulting in impacts on the changing atmospheric composition that affect air quality and lead to fatalities worldwide. The prediction of CO concentration levels is of utmost importance due to the negative impacts of CO on human health. The present work aims to advance the field of emission science and reduction strategies by introducing an enhanced neural network model. This model integrates a methodology based on a feed-forward artificial neural network with meteorological factors—specifically, wind speed (WS), wind direction (WD), and outdoor temperature (OT). Hourly measurements taken throughout a year, alongside two time-series variables (day and month), are utilized to feed the neural network during its training-testing process. The input data are sourced from an air pollutant-monitoring station situated in a Mexican border city. The proposed neural network model demonstrates its efficacy and reliability in predicting CO concentrations, affirming its potential to inform regulatory measures, protect atmospheric resources, and advance future research efforts in atmospheric science.*

Keywords: Artificial neural network; Backpropagation; Carbon monoxide.

Resumen. *La intoxicación por monóxido de carbono (CO) constituye un problema crítico con ramificaciones globales, que impacta la cambiante composición atmosférica, afectando la calidad del aire y causando muertes en todo el mundo. La predicción de los niveles de concentración de CO es crucial debido a sus efectos negativos en la salud humana. El presente trabajo busca impulsar la ciencia de las emisiones y las estrategias de reducción mediante la introducción de un modelo mejorado de red neuronal. Este modelo integra una metodología basada en una red neuronal artificial de propagación hacia adelante con factores meteorológicos, específicamente la velocidad del viento (V), la dirección del viento (DV) y la temperatura exterior (TE). Las mediciones horarias tomadas a lo largo de un año, junto con dos variables de series temporales (día y mes), se utilizan para alimentar la red neuronal durante su proceso de entrenamiento y prueba. Los datos de entrada provienen de una estación de monitoreo de contaminantes atmosféricos ubicada en una ciudad fronteriza mexicana. El modelo de red neuronal propuesto demuestra su eficacia y confiabilidad en la predicción de las concentraciones de CO, lo que confirma su potencial para fundamentar medidas regulatorias, proteger los recursos atmosféricos e impulsar futuras investigaciones en ciencias atmosféricas.*

Palabras clave: Red neuronal artificial; Retropropagación; Monóxido de carbono.

1. Introduction

Carbon monoxide (CO) is a colorless, odorless, tasteless, non-irritating and flammable gas found or formed when carbonaceous matter, deposited in faulty equipment that uses gas, solid, or oil fuel, incompletely or defectively [1,2]. Daily life examples that represent potential danger are gas heaters in enclosed rooms and bathrooms, embers, cars released gases, fire, etc.

Carbon monoxide is toxic to humans for being similar to hemoglobin; both bind decreasing the oxygen flow in the body. Once it is inhaled, it passes to the bloodstream and binds strongly to the hemoglobin and forms carboxyhemoglobin [3,4]. CO has an affinity for hemoglobin 200-250 times that of oxygen, and the symptoms and signs occurring after inhalation of carbon monoxide have been attributed to underlying tissue hypoxia [5,6]. Regarding health effects associated with CO exposure we can find are a lack of motor coordination, deterioration of cardiovascular diseases, fatigue, headache, confusion, nausea and dizziness. In extreme cases, at higher exposure over the top of high levels of what an urban zone atmosphere has, can lead to death [7,8].

Cars are the main source of CO in urban areas; consequently, nowadays most recent cars have a catalytic converter that allows them to reduce CO emissions to the atmosphere, as well as other pollutant gases [9,10].

In Mexico, CO research is done through physicochemical processes according to the ecological technician norm NTE-CCAM-001/91, which establishes the measurement methods to determine carbon monoxide concentration in the environmental air and the procedures for measurement equipment calibration [11,12].

Nevertheless, nowadays there are powerful tools to describe the atmosphere pollution dynamics in urban areas such as tridimensional mathematics models that describe its transmission and chemical transformation. In this original field we can find recent numerical models that simulate the methodology and chemical transmission processes of the pollutants in the atmosphere simultaneously [13,14].

Artificial Neural Networks (ANN) have been introduced recently as an alternative to the conventional statistics methods for pollution modeling with the following advantages: prior assumptions regarding data distribution are not performed; they are able to model nonlinear functions; they can be re-trained for a better generalization when new information or ignored data is available [15,16].

2. Literature review

Artificial Neural Network (ANN) utilization in predictive modeling is not something new; however, its usage in environmental phenomena studies is relatively recent. These kinds of studies have gained popularity due to promising results, as shown [17-27].

Currently in scientific literature it is possible to find several works about regression modeling development for environmental pollutants such as ozone, MPx, COx, NOx among others. We will analyze some of them as follows:

According to [28], accurately predicting CO_2 levels is critical for pollution control. In their study, they used CNN-LSTM and seq2seq LSTM models to predict CO concentrations up to 6 hours in advance, showing favorable results—especially for the seq2seq LSTM model, which was slightly superior to the CNN-LSTM model. For this study, they used hourly data from six

ISSN: 2594-1925

stations in Selangor, evaluating performance based on statistical measures such as RMSE, MAE, and MAPE.

In their study, [29] evaluated six regression techniques: linear regression, decision trees, Gaussian processes, tree ensembles, support vector machines (SVMs), and artificial neural networks. Of these, the 5/2 Gaussian regression model demonstrated superior performance in all scenarios, achieving high R^2 values (0.97) and low RMSE values (0.084–0.088). It was also observed that time lag influences the accuracy of the predictions, with greater accuracy being achieved with longer lags.

Traditional air quality models are computationally intensive, which limits their use for rapid applications. To address this problem, [30] proposed CoNOAir, a machine learning model based on complex neural operators. This model outperforms more advanced models such as the Fourier Neural Operator (FNO). It stands out for its ability to identify extreme pollution events and maintains high and consistent performance in several Indian cities, with R^2 values above 0.95 in all evaluated locations. CoNOAir offers authorities an effective tool for issuing early warnings and designing intervention strategies, representing an important step toward reliable, real-time predictions of CO pollution in densely populated urban areas.

In [31], a neural network model was developed to predict the ozone maximum level in Istanbul using a multilayer perceptron neural network using meteorological and pollution variables. He found that there's no significant difference between the neural network regression model and regression techniques for predicting ozone concentrations in Istanbul.

The authors in [32] developed a neural network for predicting the daily maximum level of ozone using pollution and meteorological variables, which later was compared with two traditional statistic models, regression model, and box-jenkims ARIMA, the results showed that the neural network model was higher than the regression one and box-jenkims ARIMA one.

In [33], a neural network model was developed, which combines adaptable radial base functions with statistic characteristics of ozone. It was used to predict the daily highest level of ozone concentration in Hong Kong during 1999 and 2000, this simulation showed effectiveness and reliability.

In their work [34], the authors proposed a deep learning framework to predict air quality over a 24-hour time lapse. The framework utilizes temporal features derived from spatiotemporal correlations of air quality monitoring stations, including $PM_{2.5}$ concentrations, meteorological data, and temporal data that capture the dynamic nature of air quality changes over time. The proposed model demonstrates better stability and performance compared to traditional methods such as multiple linear regression (MLR) and support vector regression (SVR), as well as other LSTM-based models. It achieves high correlation coefficients and accuracy in predicting $PM_{2.5}$ concentrations over different time horizons.

A semi-experimental regression model was proposed in [35], which is a nonlinear multivariate regression model that incorporates past contamination levels to predict future concentrations. The developed model demonstrated superior performance in terms of precision and capability compared to existing models. Evaluation indices such as RMSE, R^2, and MAPE consistently indicated superior results when compared to other models. Furthermore, it offers a robust and accurate approach that can enhance decision-making processes, support

ISSN: 2594-1925

environmental management efforts, and ultimately contribute to better air quality control and public health protection.

In [36], the authors developed machine learning models to analyze air pollution data from the Taiwan Air Quality Monitoring Network. The models were used to predict $PM_{2.5}$ concentrations based on statistical metrics such as MAE, MSE, RMSE, and R^2. The results indicated that the proposed machine learning models outperformed previous models in forecasting $PM_{2.5}$ concentrations. The actual values and predicted values were found to be very close to each other, demonstrating the effectiveness of the models. The study concluded that the gradient boosting regressor model was the most suitable for forecasting air pollution on the Taiwan Air Quality Monitoring Network data in Taiwan.

The linear regression method used in [37] for calculating CO_2 emissions from coal-fired power plants is more accurate compared to the IPCC guidelines and the coal consumption rate method. The linear regression method considers proximate analysis data to establish a fuel characteristic coefficient, resulting in more accurate calculations.

In this paper, a border city with specific geographical, political, economic, sociological, and environmental characteristics is analyzed. Over the last decade, efforts to improve environmental quality have taken on a bi-national sense of responsibility, similar to other border zones between Mexico and the United States. This initiative has been reinforced by a continuous bilateral economic and social exchange—including jobs, traditions, goods, and services—that has directly impacted pollutant levels. The rapid development of the city has led to increasing activity in the industrial, commercial, and service sectors, and consequently, a growing number of motor vehicles. This increase, especially in malfunctioning second-hand cars, has deteriorated air quality in Mexicali [38].

Mexicali is considered the city with the highest levels of pollution in the country. It's established that in Baja California 202 thousand tons of atmospheric pollutants are generated per year, which leads to a 3% mortality rate. Mexicali is located in the first place with the highest MP_{10} pollution levels, which are solid or liquid dust particles, ashes, and other aggressive elements e.g. carbon monoxide (CO). This gas pollutant is located in the third place in the country with the highest density levels in the city and it is originated from gas, kerosene, coal, petroleum and wood, among others [39].

Fig. 1 shows the graphical representation of how 8-hour norm is fulfilled for 2005 and the temporal tendency for the average concentration per year considered by the population for 2000 and 2005. Cities classification in terms of CO pollution levels was determined by using the number of days above the norm value, the maximum second per year mobile average of 8 hours, and the average concentration per year considered by the population, all for 2005. As It can be seen, the most contaminated city with CO is Mexicali, then Ciudad Juarez, these cities do not fulfill the 8-hour norm [39].

ISSN: 2594-1925

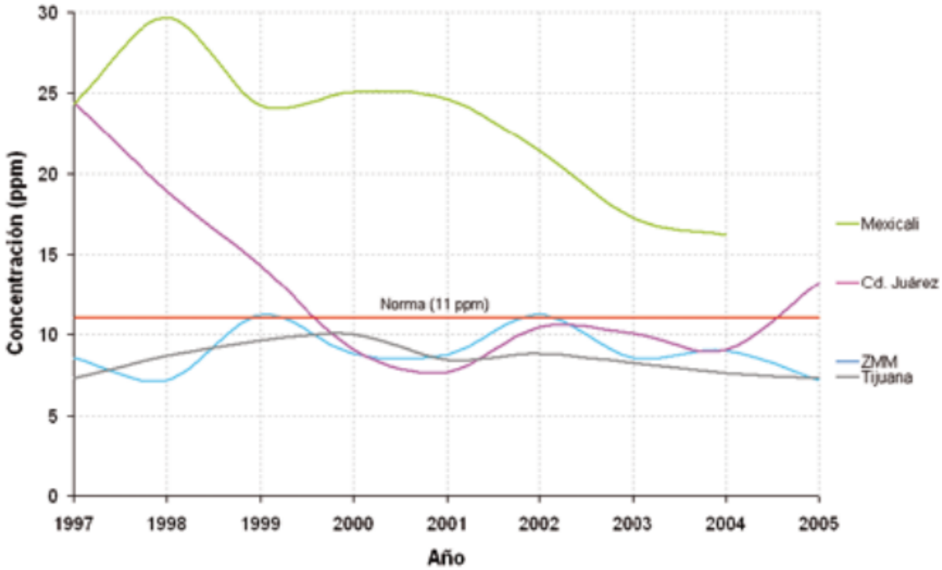

Figure 1. Fulfillment of the 8-hour norm for the CO emissions graph. Source: Third calendar of air quality data and tendency in nine Mexican cities.

Several studies have been performed in Mexicali regarding pollution. In [40], 12 prediction models were developed with the objective of determining the daily highest level of ozone using meteorological data and previous concentrations showing that in the atmospheric pollutants prediction such ozone, the results of the non-parametric models were better than those obtained with parametrical and semi-parametrical techniques.

Due to the adverse effects in health caused by the high levels of Carbon Monoxide, it is necessary to have an accurate model to predict CO concentrations. In this paper a machine learning based model is presented in order to predict the 8-h average of CO daily concentrations in Mexicali. An artificial neural network model is designed, trained and tested as a CO regression model, using environmental conditions as temperature and wind.

The article is organized as follows: in methodology section, a detailed description of the city under analysis is given, as long as variables description, some basic concepts of ANN and the methods for its design, training and testing. Following to this, proper results and discussion are given, supported with statistics, tables and charts; also the ANN performance is compared with other ANN model found in literature. Finally, a conclusion section with the most important findings about the model and the obtained results is given. Observations regarding the importance of this kind of research are given by presenting a critical analysis about ANN advantages and disadvantages in the environmental modeling scenario.

3. Study Area and Methodology

3.1. Site description and data

Mexicali city is located along the international border at 32° 40’N and 115° 28’W as shown in Fig. 2. The region is part of the Colorado River basin. It is bounded by United States of America to the north; to the east with the Mexican State of Sonora and Gulf of California; Mexican

ISSN: 2594-1925

Ensenada city to the South; and to the west with Mexicans Ensenada and Tecate municipalities [38]. Mexicali's population is 1,049,792. Nowadays it has an approximated extension of 11,372 hectares, out of which 795 are considered rural locations. In 1996, 504 hectares were allocated to urban development [39].

Figure 2. Geographical location of Mexicali, Baja California. Source: INEGI. Marco Geoestadístico, 2017

In the following, the features or variables used to feed the ANN prediction model are described. Data acquisition, statistics and other relevant information are given to the reader.

3.1.1. Wind

In Fig. 3 the average wind rose per year in surface, with information of the local airport is shown. A differentiated flow pattern can be observed basically in two most frequent directions. Winds located in the north-west quadrant occur with a 45% frequency, being west- northwest and north-northeast the main directions. This particular situation takes place mainly from October to May. Southwest quadrant winds pattern is presented with a 39% frequency, being south and southeast the most frequent, happening mostly during June and September [39].

3.1.2. Temperature

Temperature in the region has a notorious annual variability. In summer 50°C can be often reached, while in winter temperatures under 0°C are registered [40].

3.1.3. Carbon Monoxide

The estimated carbon monoxide readings (CO), in the emissions inventory, were more than 78

ISSN: 2594-1925

thousand tons. The contribution to this value is distributed regarding its source as follows: 70% correspond to the transportation sector, 4% to the federal industrial sector and 22% to local sources [40].

In the following section we can see the fundamental concepts of an ANN model, its construction process and predictive training. Going through: what is a neural network, different kinds of ANN, implemented model scheme, data set formatting, applied training methods, and model improvement.

3.2. Artificial Neural Network model

The Artificial Neural Network (ANN) is a mathematical model whose primary unit is called neuron. It consists of an interconnected group of these units, and it processes data using the following function:

$$Y(X) = w^t X + w_0 \tag{1}$$

where X represents the ANN inputs, w the interconnections weights, and Y is the ANN output.

By using a supervised or unsupervised learning algorithm, the ANN is trained by adjusting their structure –i.e. weights— and parameters by minimizing some error function that evaluates the data adjustment degree between observed and ANN predicted values –i.e. Y(X) [41].

In Computer Science literature, the ANN is a well-known and studied mathematical model, having a wide spectrum of structures and configurations. One of the most solicited versions is the Multilayer Feedforward-Backpropagation Network [42]. In this study, this later ANN configuration is used to proximate the 8-h average of daily CO concentrations in Mexicali City. In Fig. 4, it is depicted how the input and outputs variables are glued together in an ANN structure.

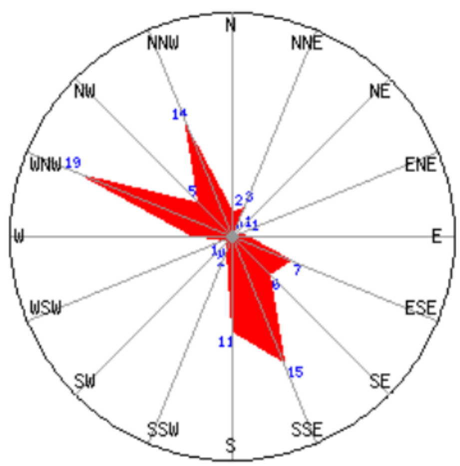

Figure 3. Annual distribution of wind direction in Mexicali. Source: Mexicali's air quality improvement program 2011-2020, Secretaría de protección al ambiente (Environment protection office).

ISSN: 2594-1925

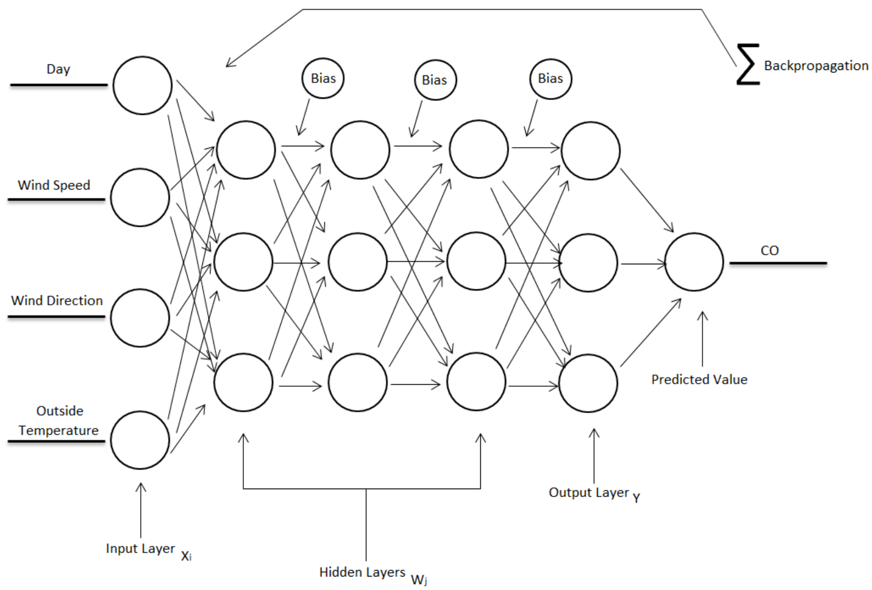

Figure 4. A Multilayer Feedforward-Backpropagation ANN used to predict CO. Source: own elaboration.

The use of ANN in CO prediction is not a new field in environmental studies. In [43], a neural network model was developed to estimate the hourly average CO concentrations in urban areas of Rosario. Experimental results indicate that the neural network predicted the CO concentrations accurately in comparison to data observed. In [44], neural network models were created to predict air quality in terms of CO, using meteorological and traffic variables. Experiments showed that ANN models that combine this information performs in an acceptable matter. The study also showed that taking away the traffic variables doesn't affect the model's performance; nevertheless, when the meteorological characteristics were taken away and the traffic ones remained a poor performance is obtained.

The difference between this two models and the model here proposed is the simplicity of the proposed model. In this work we only used four meteorological variables and two time variables and satisfactory results were obtained. This is the first work concerning this topic carried out in this particular location, and according to the consulted literature, in our country.

3.2.1. Data Sources

Environmental pollutant data were obtained from the weather measurement station located at University of Baja California (UABC), link available: http://aire.bajacalifornia.gob.mx that is a public institution devoted to higher education and research. This data was gathered with daily time records. The variable containing these records are: Date, duration, (8 hours average), 8-h average (ppm), number of observations, daily maximum value (ppm), daily rate, Carbon Monoxide (CO), Ozone (O_3), Nitrogen Dioxide (NO_2), Sulfur Dioxide (SO_2), $MP_{2.5}$ and MP_{10} in daily records.

Atmospheric data used in the ANN training phase were obtained from a Mexican Federal Agency, CONAGUA link available: http://smn.cna.gob.mx/emas/catalogo/MEXBN.htm, whose primary goal is to collect and analyze information about consumption and uses of water in Mexico. This data is gathered in a 10-minute

ISSN: 2594-1925

interval fashion and consist of the following: Time (24 hrs), Date, Wind Direction (360º), Maximum Wind Direction (360º), Wind Speed (kph), Maximum Wind Speed (kph), Average Temperature (Celsius), Relative Humidity (%), Barometric Pressure (mbar), Rain (mm), Solar Radiation(W/m^2).

3.2.2. Data Pre-processing

Before input data is presented to the ANN, a preprocessing step takes place. Most of the data is recorded at stations in different time scales. Therefore, all input variables were converted to a daily basis, to match environmental pollutant data, as shown in Table 1. A resulting data matrix of 302 registers per 4 variables and 1 prediction variable was finally obtained as a training data.

Table 1. Input Variables.

Parameter	Time	Units
Input		
Day (D)	Daily	Numeric
Month (M)	Monthly	Numeric
Wind Speed (WS)	daily average	m/s
Wind Direction (WD)	daily average	Degrees
Outside Temperature (OT)	daily average	Celsius
Output		
CO	daily average	Ppm

A data random separation was done to create two data sets, a training set with the 80% of data, and a test set with the 20% left. The former set was internally separated in training-validation-test sets by the MATLAB Neural Network Toolbox. The later set was used only to assess the ANN final performance. The parameters used during the training phase are listed in Table 2.

Table 2. ANN training parameters.

Name	Formula	
Data normalization [-1,1]	$y = \dfrac{(ymax - ymin)(x - xmin)}{(xmax - xmin)} + ymin$	(2)
Internal activation function Hyperbolic Tangent Sigmon	$\theta(r) = \dfrac{2}{1 + e^{-r}} - 1$	(3)
Output activation function Linear Transfer	$\theta(i) = \dfrac{Y_i}{X_i}$	(4)
Weight Change Gradient descent weight	$\delta(z) = \theta(\beta \cdot (\sum_i w_{ij} \cdot \alpha_i))$	(5)
Learning rate Bias learning	0.01	
Performance measure	Mean Square Error (MSE)	

ISSN: 2594-1925

Once the network was trained, we proceeded to use the test set –i.e., the 20% left data— to evaluate its generalization capacity. A total of 50 random repetitions of the whole training and testing were run. Two scenarios were analyzed in the experimental phase: Model 1, integrated D, M, WS and WD variables, having CO as target variable; and Model 2, integrated by D, M, WS, WD and T variables.

A dimensionality reduction stage was included in Model 2, using a Backward Search strategy, which consists in discarding variables in a one by one mode [45]. Hence, M variable was discarded from Model 2.

4. Discussion and results

Table 3 shows MSE performance measure over 50 runs for the two models. Test MSE readings correspond to the 20% test set. It is seen that both models yield low errors with almost similar values, being Model 2 the one that shows the lowest MSE.

Table 3. MSE performance measure and its standard error.

MSE	Training	Validation	Test
Model 1	0.9244 ± 0.0891	0.9302 ± 0.0850	0.9250 ± 0.0977
Model 2	0.9396 ± 0.0893	0.9403 ± 0.0919	0.9447 ± 0.0898

In order to assess the results presented in Table 3, the Wilcoxon signed-rank test was applied, as shown in Table 4, which is a non-parametrical test that compares the median of two selected samples to determine significant differences between them [46]. With a significance level of 5%, the p-value obtained from these three groups indicates that there's no significant difference between the models.

Table 4. Wilcoxon signed-rank test comparing Model 1 and Model 2.

MSE	Training	Validation	Test
p-value	0.8713	0.6716	0.4670

ISSN: 2594-1925

Figure 5. Correlation coefficient calculated for the training, validation y test data for each of the 50 random attempts, of models 1 and 2.

A simple correlation analysis was done between the predicted and observed target value for each model [47]. Fig. 5 depicts the correlation coefficient calculated for training, validation and test data for each of the 50 runs. It is observed that the network predictions follow the observed data satisfactory and consistently, i.e. there is a correlation R close to one in both models; however, model 2 represents less variability in the prediction.

Figs. 6 and 7 show the between observed and predicted data for both models. To create these graphs, the best and worst runs out of the 50 random run were considered, i.e. the best and worst MSE.

Considering this evidence, it is fair to say that there is an acceptable prediction of CO level in both models; however, Model 2 yielded the best results.

ISSN: 2594-1925

Table 6 shows CO concentrations for Models 1 and 2 corresponding to the best runs out of a total of 50, along with the corresponding observed value. Also the Wilcoxon signed-rank test p-value regarding differences between observed and predicted values are given. According to these readings, there are no statistical differences between predicted and observed target values. Therefore, all findings about the CO prediction ability are supported through the proposed models in this study.

Table 6. Standard error and average of the CO values observed (μo) and predicted (μp) for the proposed models, and p-value of the comparison between observed and predicted.

Best run	μo(ppm)	μp(ppm)	p<0.05
Model 1	0.8557± 0.1059	0.8702± 0.1059	0.1015
Model 2	0.8325± 0.0319	0.8188± 0.0319	0.5346

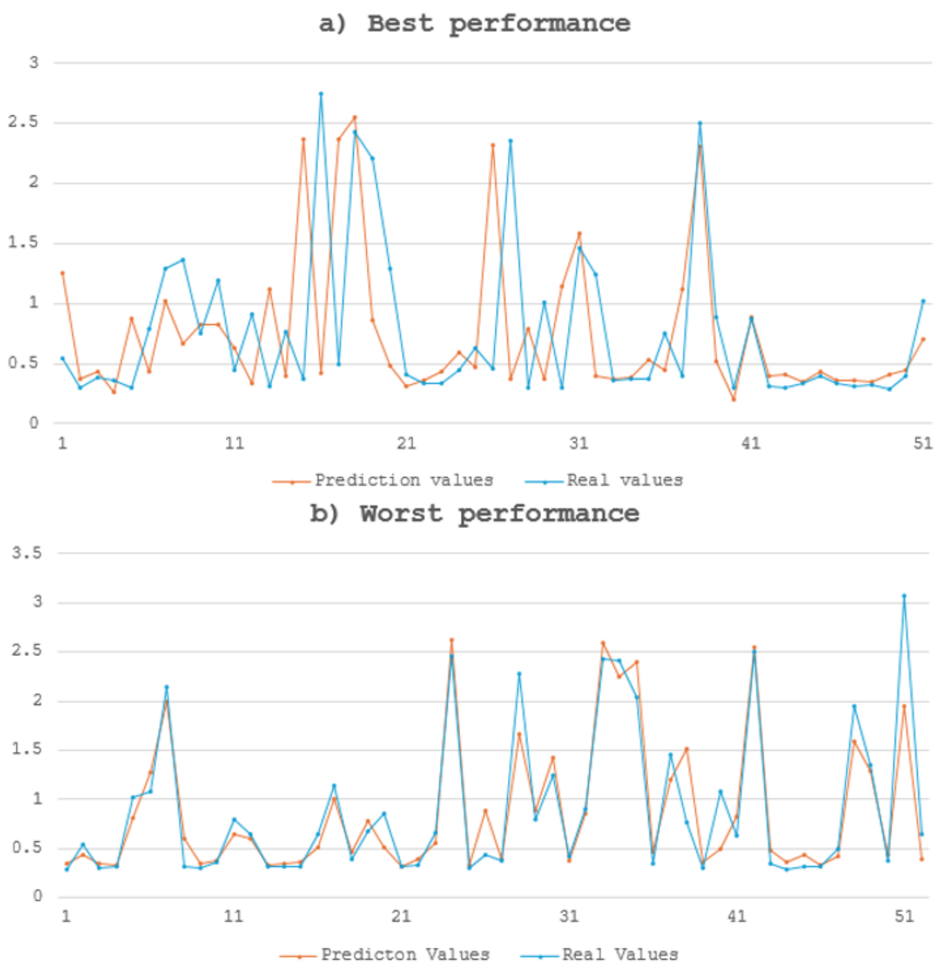

Figure 6: Model 1 observed vs. Predicted data. a) Best and b) Worst.

ISSN: 2594-1925

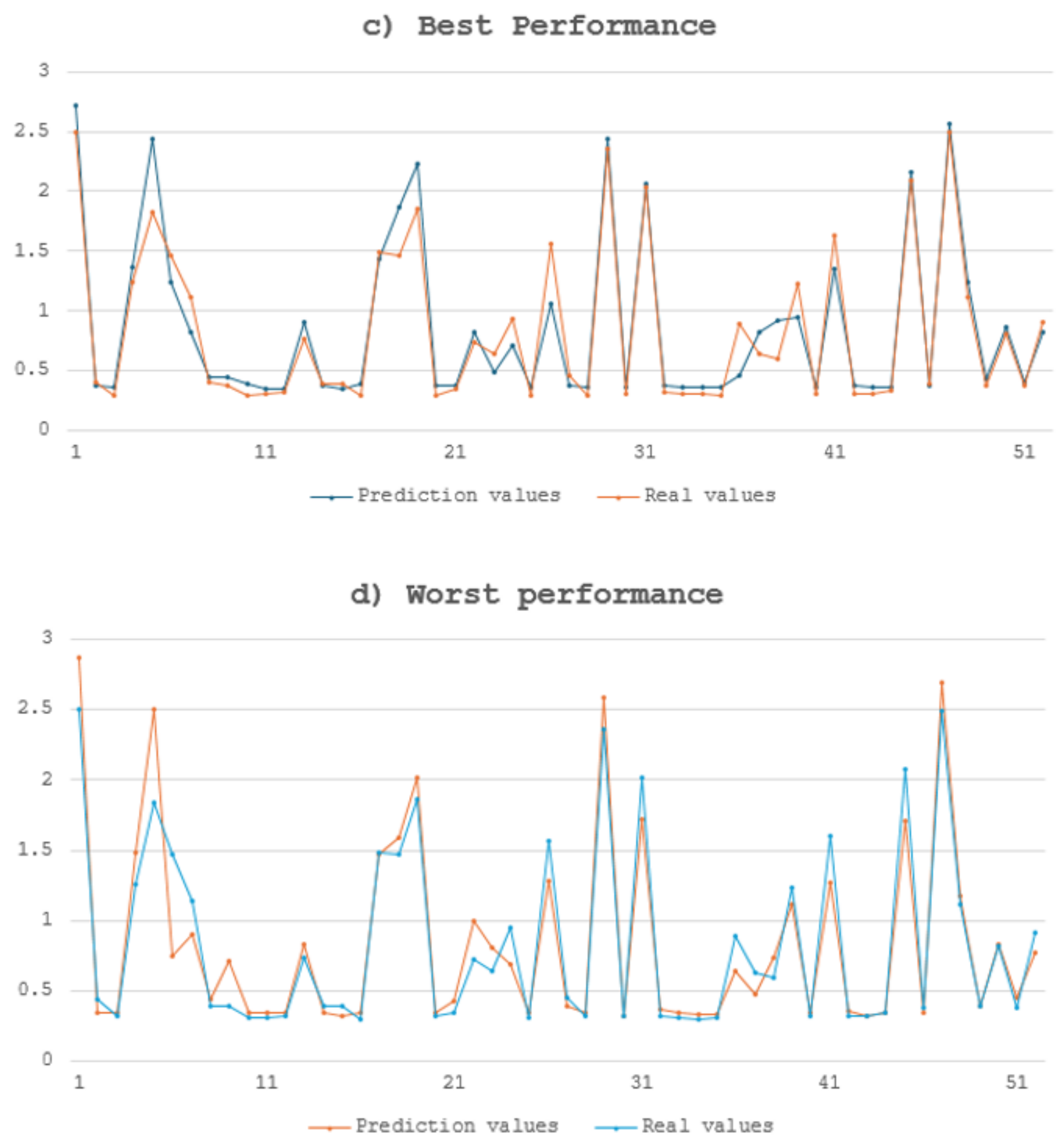

Figure 7. Model 2 observed vs. Predicted data. c) Best and d) Worst.

Once the results were obtained, they were compared with other studies reported in the international literature. For example, [24] SO_2 concentrations in Teheran were predicted using artificial neural networks (ANNs) and multiple linear regression (MLR). The ANN model showed better accuracy, with a correlation coefficient of R = 0.72 and an RMSE of 0.69 [34], a linear regression model was developed in China to predict CO_2 emissions from thermal power plants, achieving an error of only 1.72% compared to actual values. Furthermore, a NARX (Nonlinear Autoregressive Network with Exogenous Inputs) model to predict carbon monoxide (CO) concentrations in Islamabad [48], demonstrating that recurrent neural

ISSN: 2594-1925

network-based models can effectively adapt to the short-term prediction of air pollutants.

In comparison, the model proposed in this work is based on a feedforward neural network with backpropagation, trained with meteorological (wind speed and direction, outside temperature) and temporal (day and month) variables, obtaining competitive performance with low MSE values and a correlation coefficient above 0.90 on average. Unlike other studies that use more complex architectures or require large volumes of data, the present model stands out for its simplicity, lower number of variables, and its applicability in highly polluted urban contexts such as Mexicali, making it a viable alternative for early warning systems and environmental decision support.

Although the neural network model achieved high correlation coefficients, it is important to note that some sources of error can influence the prediction accuracy, such as variability in sensor calibration, missing data at the time of data collection, environmental noise, and the exclusion of potentially relevant variables such as vehicular traffic and industrial activity levels. Moreover, although the model shows promise for real-time applications, further testing with live environmental data and adaptive retraining strategies would be essential for practical implementation in urban air quality monitoring systems.

5. Conclusions

The regression model by means of an ANN approach was successfully applied to predict maximum concentrations levels of monoxide using atmospheric characteristics in a border city. A good model for maximum CO concentrations levels was developed in terms of a high correlation coefficient and low MSE. The most relevant input variables or features were found by means of a feature selection process guided by the correlation coefficient performance. Four meteorological variables (day, wind speed, wind direction and outside temperature) are enough to predict the maximum level of CO without a significant loss. This fact indicates that the predicted CO is influenced not only by the one generated in the monitored surrounding areas, but also the transmitted CO in the wind from other places.

This methodology only requires a few input variables, and it can be considered as an option to support a decision making process. This possibility can serve as a complement to physicochemical features analysis which can required a big amount of data.

Despite that the Wilcoxon non parametric test showed that there was no significant difference between both models, we conclude that model 2 is better than model 1, since it uses more meteorological variables than temporary, this allowed a closer approximation R=1, in other words, a better approximation to the real value. It could also be observed that model 2 shows a lower variability when predicting data over the 50 random, with ranges between 0.9694 and 0.9313. It is seen that the low peaks in model 2 are mostly ranged in 0.8582 and 0.8332, and the lower peak are in R=0.7984 which could be ascribed to atypical data in the measurements obtained from the measurement stations.

One of the main constraints of this neural network model is the limited availability of data, which is low due to the existent limitations found in the tested sites. The more data used for learning —i.e. the more variables available to be added— the higher precision will be reflected. Future research in this tenor includes the testing of other validation techniques such as cross validation and bootstrap for testing and analyzing the model.

ISSN: 2594-1925

6.- Authorship acknowledgment

E. Ivette Cota-Rivera: Writing – Original Draft Preparation; Conceptualization; Investigation; Methodology; Formal Analysis. *Abelardo Mercado-Herrera:* Writing – Review & Editing; Methodology; Investigation; Formal Analysis. *Fabián N. Murrieta-Rico:* Writing – Review & Editing; Conceptualization; Investigation; Formal Analysis.

References

[1] E. Mencías Rodriguez and L. M. Mayero Franco, *Manual de toxicología básica*, 1st ed. Madrid, Spain: Editorial Díaz de Santos 2000. Accessed: Aug. 25, 2025. [Online]. Available: https://www.editdiazdesantos.com/libros/mencias-rodriguez-emilio-manual-de-toxicologia-basica-L03004360601.html?articulo=03004360601

[2] World Health Organization, *WHO global air quality guidelines: Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*. Geneva, Switzerland: WHO, 2021. [Online]. Available: https://www.who.int/publications/i/item/9789240034228 (accessed Aug. 25, 2025).

[3] A. Sibón Olano, P. Martínez-García, M. A. Vizcaya Rojas, and J. L. Romero Palanco, "Intoxicación por monóxido de carbono," *Cuad. Med. Forense*, no. 47, pp. 65–69, 2007.

[4] R. N. Yadava and V. Bhatt, "Carbon monoxide: Risk assessment, environmental, and health hazard," in *Hazardous Gases*, J. Singh, R. D. Kaushik, and M. Chawla, Eds. Cambridge, MA, USA: Academic Press, 2021, ch. 8, pp. 83–96, doi: 10.1016/B978-0-323-89857-7.00030-X.

[5] Z. Yuan, L. K. De La Cruz, X. Yang, and B. Wang, "Carbon monoxide signaling: Examining its engagement with various molecular targets in the context of binding affinity, concentration, and biologic responses," *Pharmacol. Rev.*, vol. 74, no. 3, pp. 825–875, Jul. 2022, doi: 10.1124/pharmrev.121.000564.

[6] C. G. Douglas, J. S. Haldane, and J. B. S. Haldane, "The laws of combination of hemoglobin with carbon monoxide and oxygen," *J. Physiol.*, vol. 44, no. 4, pp. 275–304, Jun. 1912, doi: 10.1113/jphysiol.1912.sp001517.

[7] L. W. Kao and K. A. Nañagas, "Carbon monoxide poisoning," *Emerg. Med. Clin. North Am.*, vol. 22, no. 4, pp. 985–1018, Nov. 2004, doi: 10.1016/j.emc.2004.05.003.

[8] U.S. Environmental Protection Agency, *What is carbon monoxide?* [Online]. Available: https://www.epa.gov/indoor-air-quality-iaq/what-carbon-monoxide (accessed Aug. 25, 2025).

[9] U.S. Environmental Protection Agency, *Automobiles and carbon monoxide*. Office of Mobile Sources, 1993. [Online]. Available: https://nepis.epa.gov/Exe/ZyNET.exe/P1002J4W.TXT (accessed Aug. 25, 2025).

[10] S. Dey and N. S. Mehta, “Automobile pollution control using catalysis,” *Resour. Environ. Sustain.*, vol. 2, p. 100006, Dec. 2020, doi: 10.1016/j.resenv.2020.100006.

[11] Secretaría de Desarrollo Urbano y Ecología (SEDUE), *Gaceta Ecológica*, no. 17, Ciudad de México, 1991. [Online]. Available: https://paot.org.mx/centro/ine-semarnat/gaceta.html (accessed Aug. 25, 2025).

[12] Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAT), *Norma Oficial Mexicana NOM-172-SEMARNAT-2019, Lineamientos para la obtención y comunicación del Índice de Calidad del Aire y Riesgos a la Salud.* Ciudad de México, México: SEMARNAT, 2019.

[13] A. Mendoza and M. R. García, "Application of a second-generation air quality model to the Guadalajara metropolitan area, Mexico," *Rev. Int. Contam. Ambient.*, vol. 25, no. 2, pp. 73–85, 2009.

[14] X. Li, S. A. Hussain, S. Sobri, and M. S. Md Said, "Overviewing the air quality models on air

ISSN: 2594-1925

pollution in Sichuan Basin, China," *Chemosphere*, vol. 271, p. 129502, May 2021, doi: 10.1016/j.chemosphere.2020.129502.

[15] M. W. Gardner and S. R. Dorling, "Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences," *Atmos. Environ.*, vol. 32, no. 14, pp. 2627–2636, Aug. 1998, doi: 10.1016/S1352-2310(97)00447-0.

[16] D. Seng, Q. Zhang, X. Zhang, G. Chen, and X. Chen, "Spatiotemporal prediction of air quality based on LSTM neural network," *Alex. Eng. J.*, vol. 60, no. 2, pp. 2021–2032, Apr. 2021, doi: 10.1016/j.aej.2020.12.009.

[17] A. Coman, A. Ionescu, and Y. Candau, "Hourly ozone prediction for a 24-h horizon using neural networks," *Environ. Model. Softw.*, vol. 23, no. 12, pp. 1407–1421, Dec. 2008, doi: 10.1016/j.envsoft.2008.04.004.

[18] G. Chattopadhyay and S. Chattopadhyay, "Autoregressive forecast of monthly total ozone concentration: A neurocomputing approach," *Comput. Geosci.*, vol. 35, no. 9, pp. 1925–1932, Sep. 2009, doi: 10.1016/j.cageo.2008.11.007.

[19] I. Chairez, R. Fuentes, T. Poznyak, M. Franco, and A. Poznyak, "Numerical modeling of the benzene reaction with ozone in gas phase using differential neural networks," *Catal. Today*, vol. 151, no. 1, pp. 159–165, Apr. 2010, doi: 10.1016/j.cattod.2010.02.057.

[20] S. M. Al-Alawi, S. A. Abdul-Wahab, and C. S. Bakheit, "Combining principal component regression and artificial neural networks for more accurate predictions of ground-level ozone," *Environ. Model. Softw.*, vol. 23, no. 4, pp. 396–403, Apr. 2008, doi: 10.1016/j.envsoft.2006.08.007.

[21] H. R. Maier and G. C. Dandy, "Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications," *Environ. Model. Softw.*, vol. 15, no. 1, pp. 101–124, Jan. 2000, doi: 10.1016/S1364-8152(99)00007-9.

[22] S. Salcedo-Sanz, J. L. Camacho, Á. M. Pérez-Bellido, E. G. Ortiz-Garcia, A. Portilla-Figueras, and E. Hernández-Martín, "Improving the prediction of average total ozone in column over the Iberian Peninsula using neural networks banks," *Neurocomputing*, vol. 74, no. 9, pp. 1492–1496, Apr. 2011, doi: 10.1016/j.neucom.2011.01.003.

[23] M. Kolehmainen, H. Martikainen, and J. Ruuskanen, "Neural networks and periodic components used in air quality forecasting," *Atmos. Environ.*, vol. 35, no. 5, pp. 815–825, Jan. 2001, doi: 10.1016/S1352-2310(00)00385-X.

[24] S. R. Shams, A. Jahani, S. Kalantary, M. Moeinaddini, and N. Khorasani, "The evaluation of artificial neural networks (ANN) and multiple linear regression (MLR) models for predicting SO2 concentration," *Urban Clim.,* vol. 37, p. 100837, May 2021, doi: 10.1016/j.uclim.2021.100837.

[25] L. Zhang, P. Liu, L. Zhao, G. Wang, W. Zhang, and J. Liu, "Air quality predictions with a semi-supervised bidirectional LSTM neural network," *Atmos. Pollut. Res.*, vol. 12, no. 1, pp. 328–339, Jan. 2021, doi: 10.1016/j.apr.2020.09.003.

[26] R. Navares and J. L. Aznarte, "Predicting air quality with deep learning LSTM: Towards comprehensive models," *Ecol. Inform.,* vol. 55, p. 101019, Jan. 2020, doi: 10.1016/j.ecoinf.2019.101019.

[27] A. Masood and K. Ahmad, "A review on emerging artificial intelligence (AI) techniques for air pollution forecasting: Fundamentals, application and performance," *J. Clean. Prod.,* vol. 322, p. 129072, Nov. 2021, doi: 10.1016/j.jclepro.2021.129072.

[28] N. Zaini, L. W. Ean, and A. N. Ahmed, "Forecasting of carbon monoxide concentration based on sequence-to-sequence deep learning approach," in *Advances in Visual Informatics*, Nov. 2021, pp. 518–529, doi: 10.1007/978-3-030-90235-3_45.

ISSN: 2594-1925

[29] M. A. Almubaidin *et al.*, "Machine learning predictions for carbon monoxide levels in urban environments," *Results Eng.,* vol. 22, p. 102114, Jun. 2024, doi: 10.1016/j.rineng.2024.102114.

[30] S. Bedi, K. Tiwari, P. A. P., S. H. Kota, and N. M. A. Krishnan, "A neural operator for forecasting carbon monoxide evolution in cities," *npj Clean Air*, vol. 1, no. 1, p. 2, Mar. 2025, doi: 10.1038/s44407-024-00002-5.

[31] F. Inal, "Artificial neural network prediction of tropospheric ozone concentrations in Istanbul, Turkey," *CLEAN – Soil Air Water*, vol. 38, no. 10, pp. 897–908, 2010, doi: 10.1002/clen.201000138.

[32] J. Yi and V. R. Prybutok, "A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area," *Environ. Pollut.,* vol. 92, no. 3, pp. 349–357, Jan. 1996, doi: 10.1016/0269-7491(95)00078-X.

[33] W. Wang, W. Lu, X. Wang, and A. Y. T. Leung, "Prediction of maximum daily ozone level using combined neural network and statistical characteristics," *Environ. Int.,* vol. 29, no. 5, pp. 555–562, Aug. 2003, doi: 10.1016/S0160-4120(03)00013-8.

[34] W. Mao, W. Wang, L. Jiao, S. Zhao, and A. Liu, "Modeling air quality prediction using a deep learning approach: Method optimization and evaluation," *Sustain. Cities Soc.,* vol. 65, p. 102567, Feb. 2021, doi: 10.1016/j.scs.2020.102567.

[35] M. Zeinalnezhad, A. G. Chofreh, F. A. Goni, and J. J. Klemeš, "Air pollution prediction using semi-experimental regression model and adaptive neuro-fuzzy inference system," *J. Clean. Prod.*, vol. 261, p. 121218, Jul. 2020, doi: 10.1016/j.jclepro.2020.121218.

[36] Doreswamy, H. Ks, Y. Km, and I. Gad, "Forecasting air pollution particulate matter (PM2.5) using machine learning regression models," *Procedia Comput. Sci.,* vol. 171, pp. 2057–2066, Jan. 2020, doi: 10.1016/j.procs.2020.04.221.

[37] Y. Libao, Y. Tingting, Z. Jielian, L. Guicai, L. Yanfen, and M. Xiaoqian, "Prediction of CO2 emissions based on multiple linear regression analysis," *Energy Procedia*, vol. 105, pp. 4222–4228, May 2017, doi: 10.1016/j.egypro.2017.03.906.

[38] Gobierno del Estado de Baja California, *Programa para mejorar la calidad del aire de Mexicali: 2000–2005*. Mexicali, B.C., Mexico: ProAire, 2010. [Online]. Available: https://www.gob.mx/cms/uploads/attachment/file/69316/12_PROAIRE_MEXICALI_2000-2005.pdf (accessed Aug. 25, 2025).

[39] Gobierno del Estado de Baja California, *Programa de gestión para mejorar la calidad del aire del Estado de Baja California (ProAire BC 2011–2020).* Mexicali, B.C., Mexico: Gobierno de B.C., 2011. [Online]. Available: https://www.gob.mx/cms/uploads/attachment/file/310361/24_ProAire_Baja_California.pdf (accessed Aug. 25, 2025).

[40] E. Salazar-Ruiz, J. B. Ordieres, E. P. Vergara, and S. F. Capuz-Rizo, "Development and comparative analysis of tropospheric ozone prediction models using linear and artificial intelligence-based models in Mexicali, Baja California (Mexico) and Calexico, California (US)," *Environ. Model. Softw.,* vol. 23, no. 8, pp. 1056–1069, Aug. 2008, doi: 10.1016/j.envsoft.2007.11.009.

[41] H. Allende, C. Moraga, and R. Salas, "Artificial neural networks in time series forecasting: A comparative analysis," *Kybernetika*, vol. 38, no. 6, pp. 685–707, 2002.

[42] K. J. Cios, W. Pedrycz, and R. W. Swiniarski, *Data Mining Methods for Knowledge Discovery*. New York, NY, USA: Springer, 2012.

[43] B. Drozdowicz, S. J. Benz, A. S. M. Santa Cruz, and N. J. Scenna, "A neural network based model for the analysis of carbon monoxide contamination in the urban area of Rosario," *WIT Trans. Ecol. Environ.,* vol. 21, p. 8, 1997, doi: 10.2495/AIR970641.

ISSN: 2594-1925

[44] S. M. S. Nagendra and M. Khare, "Modelling urban air quality using artificial neural network," *Clean Technol. Environ. Policy*, vol. 7, no. 2, pp. 116–126, Feb. 2005, doi: 10.1007/s10098-004-0267-6.

[45] S. M. LaValle, *Planning Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 2006.

[46] B. Rosner, *Fundamentals of Biostatistics*, 6th ed. Belmont, CA, USA: Thomson-Brooks/Cole, 2006. [Online]. Available: http://catdir.loc.gov/catdir/enhancements/fy1514/2004117046-t.html (accessed Aug. 26, 2025).

[47] R. R. Lutz and R. M. Woodhouse, "Requirements analysis using forward and backward search," *Ann. Softw. Eng.,* vol. 3, no. 1, pp. 459–475, Jan. 1997, doi: 10.1023/A:1018929719710.

[48] M. R. Mohebbi, A. Karimi Jashni, M. Dehghani, and K. Hadad, "Short-term prediction of carbon monoxide concentration using artificial neural network (NARX) without traffic data: Case study: Shiraz City," *Iran. J. Sci. Technol. Trans. Civ. Eng*., vol. 43, no. 3, pp. 533–540, Nov. 2018, doi: 10.1007/s40996-018-0210-4.

ISSN: 2594-1925