J Adv Environ Health Res (2020) 8: 168-176 DOI: 10.22102/jaehr.2020.230438.1168

ORIGINAL PAPER

Investigation of artificial intelligence approaches (ANN-MLP, CAFIS) for the daily prediction of winter air pollutants (CO_2, SO_2, NO_X, O_3) in Hamedan city using meteorological data

Mohamad Parsi Mehr, Eisa Solgi ✉

Department of Environment, Faculty of Natural Resources and Environment, Malayer University, Malayer, Hamedan, Iran

Date of submission: 09 May 2020, ***Date of acceptance:*** 14 Jul 2020

ABSTRACT

Recently, several factors such as the physical growth of cities and the increased number of industries and cars, Hamedan city in Iran has faced the issue of air pollution. Due to the increased fuel consumption for heating purposes in the cold winters of this city, the pollution rate is higher in this season. Hamedan is surrounded by Alvand Mountains, which makes the air pollution control policies and air pollution management more important in this city. In the present study, the new methods of artificial neural network and meteorological data were used and compared as a tool for the prediction and warning of air pollution in Hamedan city. Highly accurate methods are available for the prediction of meteorological variables, which provide reliable data for the prediction of air pollution. In order to avoid over-training and assess the network compatibility with the lack of data, the minimum number of the data input data was used in this study. According to the results, the combined approaches of the artificial neural network were applicable in this regard, while ANN-MLP with the momentum learning rule and the TanhAxon transfer function yielded more accurate results compared to CAFIS.

Keywords: Artificial Neural Network, Early Warning, Air Pollutants

Introduction

Air pollution is currently considered to be a severe environmental hazard across the world, which is caused by several factors, such as demographics, economic growth, consumption patterns, energy options, cultural traditions, limitations to climatic conditions, urban patterns, developmental patterns, distance to industrial and mining sites, and air quality regulations.[1] The growth of urbanization has led more than half the world's population to live in cities, and urban air quality has a substantial impact on the health of human communities.[2]

Air pollution could give rise to severe diseases of the respiratory and cardiovascular systems and cause adverse changes in environmental conditions.[3] Recently, this issue has attracted the attention of researchers due to its adverse effects on the human health, urging urban managers to implement air pollution monitoring in most cities.[4] However, the momentary monitoring of air quality alone cannot meet all the needs of urban managers to control and manage environmental conditions. Therefore, an accurate and reliable model is essential to the prediction of air pollution as it could be used to predict air pollution in the early stages,

✉ *Eisa Solgi*
e.solgi@yahoo.com

Citation: Parsi Mehr M, Solgi E. Investigation of artificial intelligence approaches (ANN-MLP, CAFIS) for the daily prediction of winter air pollutants (CO_2, SO_2, NO_X, O_3) in Hamedan city using meteorological data. J Adv Environ Health Res 2020; 8(2): 168-176

preventing its harmful health and environmental effects through proper control measures.[3]

In general, air pollution is affected by factors such as the source of pollutants[5-7] and meteorological conditions of the region.[8-12] Meteorological conditions and parameters are among the controlling factors for the transmission and diffusion of the air pollutants in the environment.[3] Therefore, the use of these data and models could yield beneficial results. Recent evidence suggests that meteorological conditions play a key role in the daily fluctuations of air pollutant concentrations.

Recently, extensive research has been focused on the prediction of air pollution with the aim of the formation and development of models with meteorological data, including statistical models,[11] community multi-scale air quality model,[12] research and prediction models with chemistry,[13] neuro-fuzzy inference systems,[14] and other similar models.[15] These methods have performed well in the prediction of air pollution, thereby allowing the identification of new correlations between the collected data. Among these models, the artificial neural network (ANN), which has nonlinear mapping capabilities and self-adaptation, has proved superior and is widely used in predictive fields. Recently, various ANN structures have been developed to improve the predictive function of air pollutant concentrations,[16-20] and several studies have been conducted in this regard.

In a study, Mohebbi *et al.* evaluated the ability of dynamic neural networks in predicting the concentration of carbon monoxide in Shiraz (Iran) in the absence of traffic data using meteorological data, and the obtained results indicated that dynamic neural networks had a high performance in the modeling of carbon monoxide concentration in the absence of traffic data.[21] In another research, Gao *et al*. investigated the feasibility of using the ANN model with meteorological parameters as the input variables to predict the ozone concentration in Jinan urban area in northern China. According to the findings, the ANN model could accurately predict the ozone level of the environment. Maximum temperature, atmospheric pressure, duration of sunshine, and maximum wind speed are known as the dominant input variables, which significantly affect the ozone forecast of the environment.[22]

The air quality of Hamedan city in Iran has changed extensively in recent years, so that the city has been reported to have consecutive days of unhealthy air quality based on the relevant indicators in the winter. For this reason, it is of utmost importance to recognize and predict the daily air pollution in order to access the initial warning and management before the occurrence of air pollution.

The present study aimed to optimize and evaluate combined ANN methods for the modeling and prediction of urban air pollutants in Hamedan based on meteorological parameters in order to provide an effective tool for prediction and warning before the incidence of air pollution.

Materials and Methods

Fig. 1 depicts the schematic view of the research process.

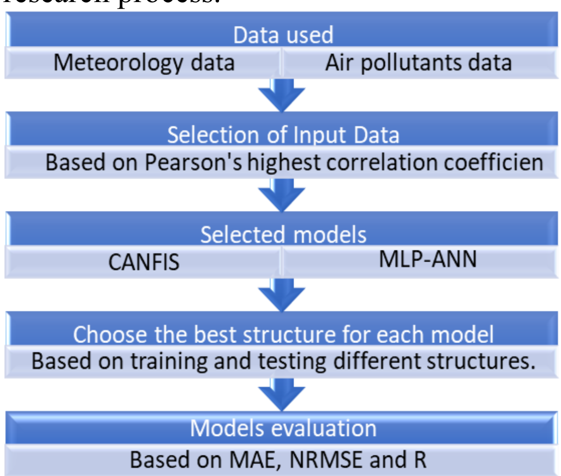

Fig 1. Grafical abestract

Study area

Hamedan province is located in an area of 20,172 square kilometers in the west of Iran on the slopes of Alvand Mountains at the altitude of 1,800 meters above the sea level. The climate of this region is cold and dry, and

Hamedan city covers an area of 2,831 square kilometers in the province[23] (Fig. 2).

Applied data

In order to model and predict air pollutants, we used the meteorological data (minimum daily temperature, maximum daily temperature, average daily temperature, total daily rainfall, sunny hours, cloudy hours, maximum wind speed, wind direction, daily average wind speed, maximum humidity, minimum humidity, and average humidity) and the data on the particulate matters ($PM_{2.5}$ and PM_{10}). In addition, the air pollutant data on O_3, CO, NO_x, and SO_2 were obtained from an air pollution monitoring station affiliated to the Environment Organization of Hamedan Province in Hamedan city. We also used the data collected from both the meteorological and environmental stations for the winter of 2017-2018.

Statistical analysis

Data analysis was performed in SPSS version 25 using Pearson's correlation-coefficient to assess the correlations between the meteorological and air pollution parameters; the correlation-coefficient indicates the degree to which the parameters were affected by each other. In addition, various ANN models were examined to predict and model the air pollution parameters. For use in the optimal models, the data with the most significant correlation with *the desired parameters were used.*

Multilayer Perceptron Artificial Neural Network (MLP-ANN)

The ANN design is inspired by the structure of the human brain and relies on advanced learning processes.[24] The overall structure of ANN has three layers with specific tasks, including the data input layer to ANN, information processing layer (middle layer), and output layer, which showed the results and outputs in addition to the processing of each network input parameter. In the current research, we used the multilayer perceptron network with the back propagation algorithm.

Network design is based on a combination of data on the influential parameters in the air quality over time in various structures from the number of the data in the input layer. In each structure, the input data are through the output of the first layer neurons after processing, moving to the neurons of the next layers, and finally transmitting to the network output if acceptable. Otherwise, they return to the previous layers by calculating the computational error, and the calculations are repeated to obtain acceptable results.[25] In our study, the normalized data were used as a network input to increase the data processing speed and prevent network interruptions in the local minimums.[26]

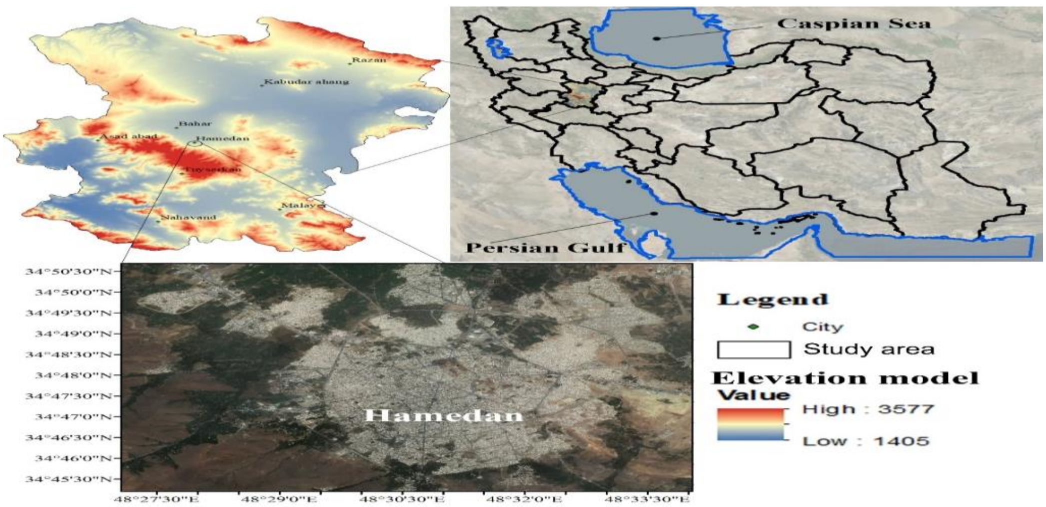

Fig. 2. Study area

Table 1. Dispersion of used data

	N	Mean	Std. Deviation	Min	Max
O_3 (ppb)	84	13.4843	4.53673	5.93	23.59
CO (ppm)	84	1.4627	0.72311	0.37	3.76
NO_x (ppb)	84	32.2070	16.30880	16.98	90.18
SO_2 (ppb)	84	12.3190	1.61921	9.15	17.85
PM_{10}	84	78.4418	43.51749	15.04	197.68
$PM_{2.5}$	84	27.3755	15.05702	9.75	86.21
Minimum daily temperature	84	-2.065	6.2733	-17.9	11.6
Maximum daily temperature	84	10.836	5.3316	-4.7	19.0
Average daily temperature	84	3.829	5.6729	-11/8	14/4
Total daily rainfall	84	0.6455	1.69519	0.00	7.10
Sunny hours	84	6.446	2.7069	0.0	11.0
Cloudy hours	84	3.2002	1.97923	0.00	7.25
Maximum daily wind speed	84	7.62	3.712	3	20
wind direction	84	210.00	90.833	10	360
Average wind speed (m/s)	84	2.6106	1.47389	0.50	7.50
Maximum humidity	84	82.74	14.564	19	100
Minimum humidity	84	38.26	16.891	10	95
Average humidity	84	61.3144	16.39141	16.25	97.75

Table 1 shows different values of maximum, minimum, mean, and standard deviation using the SPSS version 25 and summary of the results.

Co-Active Neuro-Fuzzy Inference System (CANFIS)

CANFIS was introduced by Jang *et al*. in 1997 as a general form of adaptive neuro-fuzzy inference systems (ANFIS).[27] CANFIS could be considered a global estimate of any nonlinear function, an important feature of which is the benefits of integrating an artificial neural network with a fuzzy inference system in one format. CANFIS consists of five layers, including the fuzzification layer, rule layer, normalization layer, defuzzification layer, and summation layer.[28] The function of each layer has been described by Aytek *et al*.[29]

Evaluation of the models

Error values should be at the minimum; for this purpose, training and testing should be repeated with various structures, so that the error would be minimized and better structures could be found. In order to evaluate and compare the results of various structures in the present study, the coefficient of determination (R), mean absolute error (MAE), and normalized root mean square error (NRMSE) were calculated.

$$\mathrm{MAE:}\frac{1}{n}\sum_{i}^{N}=1\,|xi-yi| \tag{1}$$

In the Eq. 1, xi is the actual data, yi represents the estimated data, and n shows the total number of the data.

$$R=\frac{\sum_{i=1}^{n}(\mathrm{Y\ act}-\hat{\mathrm{Y}}\ \mathrm{act})\,(\mathrm{Y\ est}-\hat{\mathrm{Y}}\ \mathrm{est})}{\sqrt{\sum_{i=1}^{n}}\ (\mathrm{Y\ act}-\hat{\mathrm{Y}}\ \mathrm{act})^2\sum_{i=1}^{n}(\mathrm{Y\ est}-\hat{\mathrm{Y}}\ \mathrm{est})} \tag{2}$$

In the Eq. 2 $Y\ act$ shows the actual values, $\hat{Y}\ act$ is the average of the actual values, $Y\ est$ represents the estimated values, and $\hat{Y}\ est$ is the average of the estimated values.

$$\mathrm{NRMSE}=\frac{\mathrm{RMSE}}{\mathrm{Pollution\ average}} \tag{3}$$

In the Eq. 3, RMSE is the root mean square error.

Results and Discussion

Statistical analysis

In the current research, Pearson's correlation-coefficient was used to evaluate the correlations between the values of the used parameters. High correlation-coefficients indicated the common control factor between the parameters, based on which the parameters with the highest correlation-coefficient were selected as the most influential factors for the ANN structure. Table 2 shows the results of the correlation study.

In order to prevent the ANN over-learning and adapting to the lack of the input data, only three network input parameters were used for the prediction of each variable. The variables were selected based on the highest correlation-coefficients between the variable and predicted parameter. For the O_3 parameter, the variables included the mean daily temperature, minimum daily temperature, and mean daily wind speed.

For the NO_x parameter, the input variables included the minimum daily temperature, mean daily wind speed, and $PM_{2.5}$. For the SO_2 parameter, the variables of mean daily wind speed, $PM_{2.5}$, and PM_{10} were used. As can be seen, the air pollutants were significantly correlated with the meteorological parameters due to the fact that the meteorological parameters had a direct impact on the air pollutants. For instance, the wind variables could reduce the concentration of the pollutants if present or, and if not, they accelerated the production process of the secondary pollutants, which justified the high dependence.

Table 2. Pearson's correlation-coefficient

	O_3 (ppb)	CO (ppm)	NO_x (ppb)	SO_2 (ppb)
PM_{10}	-0.340**	0.503**	0.479**	0.538**
$PM_{2.5}$	-0.549**	0.684**	0.674**	0.620**
Minimum daily temperature	0.653**	-0.616**	-0.569**	-0.294**
Maximum daily temperature	0.425**	-0.327**	-0.210	0.005
Average daily temperature	0.625**	-0.567**	-0.464**	-0.206
Total daily rainfall	0.109	-0.156	-0.213	-0.256*
Sunny hours	0.150	-0.040	0.076	0.054
Cloudy hours	0.268*	-0.385**	-0.494**	-0.251*
Maximum daily wind speed	0.534**	-0.494**	-0.394**	-0.274*
wind direction	0.404**	-0.445**	-0.262*	-0.440**
Average wind speed (m/s)	0.820**	-0.739**	-0.525**	-0.501**
Maximum humidity	-0.591**	0.456**	0.234*	0.201
Minimum humidity	-0.468**	0.355**	0.138	0.095
Average humidity	-0.650**	0.515**	0.249*	0.211

MLP-ANN

Based on the selection of the inputs and outputs of four models for the modeling and prediction of the four parameters of air pollution using Pearson's correlation-coefficient with various efforts and errors, the optimal results for the number of the hidden layers, transfer function, and learning rule were obtained (Table 3). Accordingly, the momentum learning rule and TanhAxon transfer function for the four models had the optimal results. In general, the MLP-ANN approach yielded accurate and acceptable results, which indicated the reliability of the tool for the management and prediction of air quality.

Table 3. Structures and Results of MLP-ANN Models

Parameter	Number of inputs	Learning rule	Transfer function	Number of hidden layers	Processing Elements	NRMSE	MAE	R
O_3	3	Momentum	TanhAxon	1	4	0.1246	1.343	0.9392
CO	3	Momentum	TanhAxon	1	4	0.5104	0.4793	0.9475
NO_x	3	Momentum	TanhAxon	1	4	0.2207	5.8308	0.9320
SO_x	3	Momentum	TanhAxon	1	4	0.4225	0.8633	0.8762

Fig. 3 shows the overlap of the predicted values with the actual values. As can be seen, the horizontal axis was the sample number, the vertical axis showed the air pollution values, the continuous line indicated the observational values, and the dotted line showed the predicted values. Notably, when the two lines were closer, the model output was closer to reality, and the accuracy of the model was higher, while the most accurate results were obtained when the two lines completely overlapped.

CANFIS

In the present study, the selected CANFIS structure in all the four models encompassed the bell for membership function, TSK for the fuzzy model, processing elements with Transfe Axon, and learning rule with the momentum as the proposed structure by the software. In addition, the superiority was confirmed by the repeated training and testing of the data. The results of the CANFIS method (Table 4) demonstrated that in all the parameters (except SO_2), the CANFIS model has a weaker performance compared to the MLP-ANN method. However, the results showed that the CANFIS approach also had significant capabilities for the modeling of the air quality in this range.

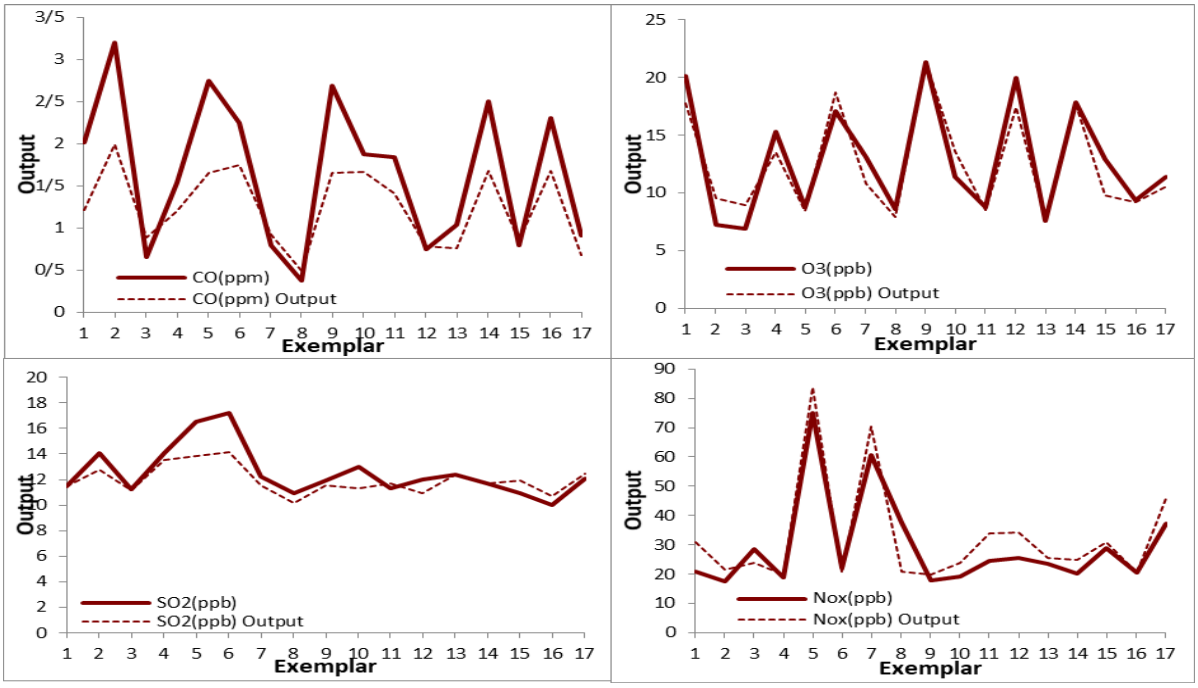

Fig. 3. Desired network output and actual data

Table 4. Structures and results of CANFIS models

Parameter	MF -FM	Transfer-LR	Structure (In put, PEs, Out put)	NRMSE	MAE (mg/L)	R
O_3	Bell-TSK	Axon-Momentum	3,1,1	0.2428	1.7815	0.8773
CO	Bell-TSK	Axon-Momentum	3,1,1	0.2118	0.2670	0.9106
NO_3	Bell-TSK	Axon-Momentum	3,1,1	0.321	8.0504	0.9021
SO_2	Bell-TSK	Axon-Momentum	3,1,1	0.2689	0.3203	0.8728

Fig. 4 shows the comparison chart of the data predicted by the CANFIS models with the actual data. In general, the CANFIS method had lower accuracy in all the parameters, with the exception of SO_2, for which both methods yielded similar results.

Conclusion

With the physical development of urbanism, population growth, industrial development, and increased number of cars, the concerns regarding air pollution in Hamedan have increased, especially since the phenomenon of temperature inversion has occurred several times in winter, intensifying air pollution. Meteorological parameters play a key role in the moderation or intensification of air pollution. Accurate methods are available for the prediction of meteorological variables. Therefore, the prediction of air pollutants based on these data could largely contribute to urban managers and planners in face of the adverse effects of air pollution on cities.

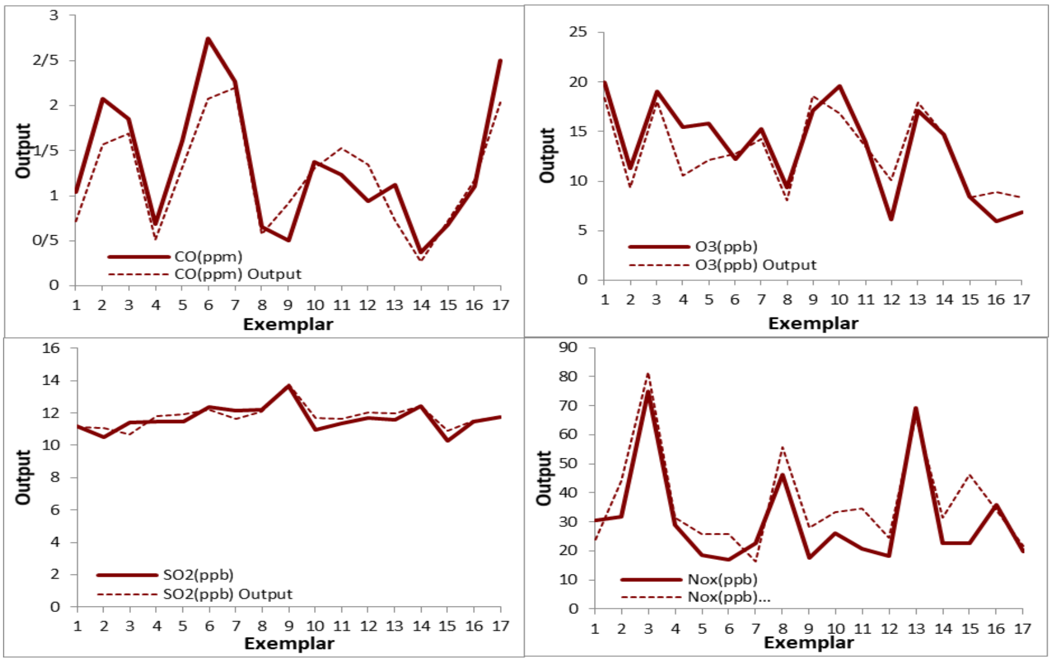

Fig. 4. Desired output and actual network output

In this study, artificial intelligence approaches were used to model and predict urban air pollutants (O_3, CO, NO_x, and SO_2), which are the most important influential factors in air pollution. According to the results, the MLP-ANN method was more accurate compared to the CANFIS method in the modeling and prediction of the air quality parameters. This model had the momentum learning rule and TanhAxon transfer function with three inputs, four neurons in one hidden layer, and coefficient of determination of higher than 90% and yielded remarkable outcomes. Therefore, it could be concluded that the selected parameters were appropriate for the use and development of network structures, as well as the selection of the input variables based on the correlations between the variables, with the air pollutants reducing the number of the input variables and producing acceptable results. Based on the results of the correlation-coefficients between the variables and output accuracy of the models, it could be inferred that daily temperature, wind speed, and PM concentrations were the most important influential factors in the studied pollutants in Hamedan, which could be due to the time-space conditions of the study area as in the cold winters of Hamedan, fuel consumption for heating purposes increases noticeably, thereby leading to a significant increase in the air pollution in winter. Since Hamedan city is surrounded by Alvand Mountains, the possibility of the accumulation of pollutants increases. Therefore, the effect of temperature is considered important in this range both in terms of pollutant production and acceleration of the processes between the pollutants. In addition, wind speed played a key role in the dispersion of the pollutants.

Consistent with the previous studies, our findings emphasized on the high ability of ANN to model and predict parameters in complex natural environments, such as air pollutant forecasting. Therefore, this tool could replace the conventional deterministic models that have proven incapable in complex environments. In conclusion, ANN could be used as an early warning system before the occurrence of pollution in order to prevent or reduce the destructive effects of air pollution on Hamedan city.

References

1. Jorquera H, Montoya LD, Rojas NY. Urban Air Pollution. In: Urban Climates in Latin America: Springer International Publishing; 2019, pp. 137–65.

2. Chen M, Yang J, Hu L, Hossain MS, Muhammad G. urban healthcare big data system based on crowdsourced and cloud-based air quality indicators. IEEE Communications Magazine 2018; 56(11): 14–20.
3. Bai Y, Li Y, Wang X, Xie J, Li C. Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions. Atmos Pollut Res 2016; 7(3): 557–66.
4. Cheng W, Shen Y, Zhu Y, Huang L. A neural attention model for urban air quality inference: Learning the weights of monitoring stations. Available from: URL: https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/download/16607/15925.
5. Gantt B, Meskhidze N, Zhang Y, Xu J. The effect of marine isoprene emissions on secondary organic aerosol and ozone formation in the coastal United States. Atmos Environ 2010; 44(1): 115–21.
6. Urbanski SP, Hao WM, Nordgren B. The wildland fire emission inventory: Western United States emission estimates and an evaluation of uncertainty. Atmospheric Chemistry and Physics 2011; 11(24): 12973–3000.
7. Gao XL, Hu TJ, Wang K. Research on motor vehicle exhaust pollution monitoring technology. Applied Mechanics and Materials 2014; 620: 244–7.
8. Wang H, Xing C, Yu F. Study of the hydrological time series similarity search based on Daubechies wavelet transform. Unifying Electrical Engineering and Electronics Engineering. Lecture Notes in Electrical Engineering 2014, vol 238. Springer, New York, NY.
9. Wang J, Wang Y, Liu H, Yang Y, Zhang X, Li Y, *et al.* Diagnostic identification of the impact of meteorological conditions on $PM_{2.5}$ concentrations in Beijing. Atmos Environ 2013; 81: 158–65.
10. Wu Q, Xu W, Shi A, Li Y, Zhao X, Wang Z, *et al*. Air quality forecast of PM_{10} in Beijing with Community Multi-scale Air Quality Modeling (CMAQ) system: Emission and improvement. Geosci Model Dev 2014; 7(5): 2243–59.
11. Ozel G, Cakmakyapan S. A new approach to the prediction of PM_{10} concentrations in Central Anatolia Region, Turkey. Atmos Pollut Res 2015; 6(5): 735–41.
12. Djalalova I, Delle Monache L, Wilczak J. $PM_{2.5}$ analog forecast and Kalman filter post-processing for the Community Multiscale Air Quality (CMAQ) model. Atmos Environ 2015; 108: 76–87.
13. Saide PE, Carmichael GR, Spak SN, Gallardo L, Osses AE, Mena-Carrasco MA, *et al*. Forecasting urban PM_{10} and $PM_{2.5}$ pollution episodes in very stable nocturnal conditions and complex terrain using WRF–Chem CO tracer model. Atmos Environ 2011; 45(16): 2769–80.
14. Domańska D, Wojtylak M. Application of fuzzy time series models for forecasting pollution concentrations. Expert Syst Appl 2012; 39(9): 7673–9.
15. Yahya K, Zhang Y, Vukovich JM. Real-time air quality forecasting over the southeastern United States using WRF/Chem-MADRID: Multiple-year assessment and sensitivity studies. Atmos Environ 2014; 92: 318–38.
16. Feng Y, Zhang W, Sun D, Zhang L. Ozone concentration forecast method based on genetic algorithm optimized back propagation neural networks and support vector machine data classification. Atmos Environ 2011; 45(11): 1979–85.
17. Wu S, Feng Q, Du Y, Li X. Artificial neural network models for daily PM_{10} air pollution index prediction in the urban area of Wuhan, China. Environ Eng Sci 2011; 28(5): 357–63.
18. Paschalidou AK, Karakitsios S, Kleanthous S, Kassomenos PA. Forecasting hourly PM_{10} concentration in Cyprus through artificial neural networks and multiple regression models: Implications to local environmental management. Environ. Sci Pollut Res 2011; 18(2): 316–27.
19. Antanasijević DZ, Ristić MĐ, Perić-Grujić AA, Pocajt VV. Forecasting human exposure to PM_{10} at the national level using an artificial neural network approach. J Chemom 2013; 27(6): 170–7.
20. Li C, Liang M, Wang T. Criterion fusion for spectral segmentation and its application to optimal demodulation of bearing vibration signals. Mech Syst Signal Pr 2015; 64-65: 132–48.
21. Mohebbi MR, Jashni AK, Dehghani M, Hadad K. 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 2019; 43: 533-40.
22. Gao M, Yin L, Ning J. Artificial neural network model for ozone concentration estimation and Monte Carlo analysis. Atmospheric Environment 2018; 184: 129–39.
23. Kalvandi R, SafiKhani K, Najafi Gh, Babakhanlou P. Identification of medicinal plants of Hamedan Province. Iranian Journal of Medicinal and Aromatic Plants 2007; 23(3): 350-74. [In Persian]
24. Ostad-Ali-Askari K, Shayannejad M, Ghorbanizadeh-Kharazi H. Artificial neural network for modeling nitrate pollution of groundwater in marginal area of Zayandeh-rood River, Isfahan, Iran. KSCE J Civ Eng 2017; 21(1): 134–40.
25. Parsimehr M, Shayesteh K, Godini K, Bayat Varkeshi M. Using multilayer perceptron artificial neural network for predicting and modeling the chemical oxygen demand of the Gamasiab River. Avicenna J Environ Health Eng 2018; 5(1): 15–20.
26. Asadpour G, Nasrabadi T. Municipal and medical solid waste management in different districts of Tehran, Iran. Fresenius Environ Bull 2011; 20(12): 3241–5.
27. Aziz K, Rahman A, Shamseldin A Y, Shoaib M. Co-active neuro fuzzy inference system for regional flood estimation in Australia. J Hydrol Environ Res 2013; 1(1): 11–20.
28. Malik A, Kumar A, Kisi O. Monthly pan-evaporation estimation in Indian central Himalayas using different heuristic approaches and climate based models. Computers and Electronics in Agriculture 2017; 143: 302–13.
29. Aytek A. Co-active neurofuzzy inference system for evapotranspiration modeling. Soft Computing 2009; 13(7): 691-700.