

Predicting Air Pollution In Advance Using Feed Forward Neural Networks: A Case Study on Delhi

Asha Unnikrishnan¹, Dr. S Rajeswari²

¹Research Scholar- Department of Computer Science, Sree Saraswathi Thyagaraja College, Pollachi, Tamil Nadu, India.

²Associate Professor- Department of Computer Science, Sree Saraswathi Thyagaraja College, Pollachi, Tamil Nadu, India.

Author Orcid ID : 0000-0003-2998-9126

ABSTRACT

Around the world, outdoor air pollution is becoming a major hazard to human health. The policy makers need to have a solid scientific foundation for developing a comprehensive plan for reducing air pollution, which comes from air quality monitoring and forecasts. Furthermore, the people can minimise their exposure to dangerously high levels of air pollutants by taking preventive action if air pollution projections are made available to them. This work proposes an intelligent air pollution prediction system that uses an Extreme Learning Machine (ELM) to anticipate the following day's air quality index for five different pollutants (PM10, PM2.5, NO2, CO, and O3). It is discovered that the suggested ELM-based system outperforms the current air pollution prediction systems in terms of prediction accuracy.

Keywords—Artificial neural networks, Extreme Learning Machine , Air Quality Index, Regression

1. Introduction

When the concentration of certain gases or particulates in the air is higher than the established acceptable limits, it is referred to as air pollution.

Both natural and man-made factors contribute to this, with the latter having more noticeable consequences. This is a result of the tendency towards industrialization and rapid, disorganised urbanisation. Concerns about ecological sustainability are frequently subordinated to those about economic development, particularly in large cities like Delhi that must sustain large populations. Furthermore, the city produces enormous volumes of municipal solid garbage every day. The most important sources of air pollution are the pollutants produced by factories, cars, and the burning of solid waste. However, natural occurrences do not cause air pollution [1].

The amount of air pollutants present at any given time and location is also influenced by meteorological factors such as temperature, rainfall, wind direction, speed, and relative humidity. For example, because of the predominant weather circumstances in that environment, it is determined that the quality of the air is lowest during the winter. It is imperative to underscore the detrimental effects of air pollution on human health. Respiratory and cardiovascular disorders can be brought on by pollutants such as carbon monoxide, sulphur dioxide, nitrogen dioxide, and particle matter. Lead exposure may damage brain function, and ozone exposure may have an impact on normal liver function [2].

There are numerous additional techniques for predicting air pollution in the literature.

These approaches can be divided into two groups: soft computing techniques and classical methodologies. For the purpose of forecasting pollutants, traditional methods employ statistical models, while soft computing uses artificial intelligence models. When given new data, artificial neural networks can be trained to learn from datasets and provide precise predictions. It has been

suggested to use neural networks to anticipate the +1, +2, and +3 pollution levels of SO₂, PM₁₀, and CO [3]. A feed-forward backpropagation network was utilised in this system. Despite being a popular learning method in artificial neural networks, back propagation has a sluggish convergence because of gradient descent and iterative tuning.

NO₂, SO₂, SPM, and RSPM were also forecasted using statistical methods [4] including SRS (single exponential smoothing) and ARIMA (auto regression integrated moving averages). They looked for trends in the historical pollutant data and made predictions based only on past pollutant values, ignoring the prevailing weather conditions. A site-optimized semi-empirical model, or SOSE model, has been developed in another study to forecast CO concentrations [5]. SOSE is a statistical model that just uses wind direction and speed to estimate pollution concentrations. By accounting for the dispersion of pollutants caused by wind in the atmosphere, an analytical dispersion model has also been utilised to estimate PM₁₀ [6]. An other method for predicting PM₁₀ levels employs a three-layer feedforward back propagation mechanism.

After comparing the two suggested networks, the authors concluded that FFNN was superior because Elman networks needed longer training time. After comparing statistical regression with methods like PCA (Principal Component Analysis), LRA (Linear Regression Analysis), and ARIMA (Auto Regressive Integrated Moving Averages) for ozone level prediction, the authors [1] concluded that classification techniques worked better for their data.

An intelligent air pollution prediction system is suggested in this study to forecast the different pollutant levels for the following day.

2. Data Source And Experimental Methods

The data source for this study is the already-existing SAFAR air pollution monitoring and forecasting system. ELM is used to anticipate Delhi's air pollution, and SAFAR is also used for performance evaluation.

2.1 SAFAR

India is among the most polluted nations in the world; in the WHO's 2014 report "Ambient air pollution in cities," Delhi, the country's capital, was named the most contaminated city [8]. The Ministry of Earth Sciences in India oversees the SAFAR (System of Air Quality Forecasting and Research) programme [9]. It was established by the Indian government to evaluate the results of emission reduction efforts made in advance of Delhi's hosting of the 2010 Commonwealth Games [10]. SAFAR maintains ten monitoring stations throughout Delhi, where specialist equipment is used to track the concentrations of five air pollutants (PM₁₀, PM_{2.5}, NO₂, CO, and O₃) as well as meteorological variables (temperature, relative humidity, rainfall, wind direction, and speed).

The information is then processed at the Indian Institute of Tropical Meteorology (IITM), located in Pune. There, it uploads the anticipated air quality indices for the current and upcoming days, along with pertinent health advisories, for each of the pollutants under observation to its website (<http://www.safar.tropmet.res.in>).

WRF-Chem is used by SAFAR to forecast pollution for one day [11]. On its website, the current state of the air quality as well as the forecast for the following day are provided, although not in terms of absolute concentrations. Instead, the raw concentration values of pollutants are transformed into illustrative figures known as air quality indices.

This is done because the air quality index goes from 0 to 500, with each band of the index representing a different level of air quality, making it easier for the general public to understand.

2.2 Extreme Learning Machine

A learning approach for single hidden layer feedforward neural networks is called Extreme Learning Machine [12]. The input weights and biases of the hidden layer neurons in ELM are allocated at random, and there is just one hidden layer. Moreover, ELM learns far more quickly than backpropagation-based FFNN as it doesn't require iterative adjustment. It has been discovered that ELM exhibits strong generalisation across numerous applications. Many machine learning tasks, including feature selection, clustering, regression, and classification, can be accomplished

with ELM [13].

What follows is a summary of the ELM algorithm:

The ELM algorithm is described as follows if the activation function is $g(x)$, the number of hidden layer neurons is m , and the training set S of N readings is $S = \{(x_i, t_i)\}$, for $i = 1, 2, \dots, N$.

Step 1 Divide the input weight and bias into two groups and allocate them to each hidden layer neuron at random.

Step 2: Determine the hidden layer, H , 's output matrix.

Step 3: To get the output weight, use the matrix H that was generated in step 2 and the formula $\beta = H^\dagger T$, where H^\dagger is the hidden layer output matrix H 's Moore–Penrose generalised inverse matrix.

The ELM algorithm generates values for w_i and b_i using a random function, and then utilises the Moore–Penrose generalised inverse matrix to get the corresponding β_i .

2.3 Proposed Methodology

The following is a summary of the suggested methodology:

Data Gathering The SAFAR web portal (<http://safar.tropmet.res.in>) provided the Delhi University monitoring station with daily meteorological and air pollution data. This included the current values of weather conditions such as relative humidity, wind speed, temperature, rainfall, and wind direction, as well as the air quality index for PM10, PM2.5, NO2, CO, and O3. Data from SAFAR has been gathered for 37 days. The network was trained using the first thirty days' worth of data, and it was tested using the remaining seven days' worth of data.

Choosing Variables and Building the ELM Model With the exception of CO, all other air pollution indices in the data gathered from the SAFAR portal showed continuous values in the range [0, 500]. CO, however, only took two values from the air quality index, which were 59 and 104. Therefore, the ELM binary classification model is used to predict the CO air quality index, and the linear regression model is used to predict the PM10, PM2.5, NO2, and O3 air quality indexes. Eight input factors were selected following a series of studies, including temperature, humidity, wind speed, PM10, PM2.5, NO2, CO, and O3. The ELM therefore has eight input neurons. In order to forecast the five distinct pollutants (PM10, PM2.5, NO2, CO, and O3), five distinct ELM networks were constructed.

The network depicted in Figure 1 is utilised to forecast the following day's PM10 levels. In a similar vein, networks were built to predict PM2.5, NO2, and O3.

ELM Model Training and Testing For every pollutant, different activation functions and varying numbers of hidden layer neurons were used in training studies. Testing was done using the parameters for which the best results were obtained during the training phase.

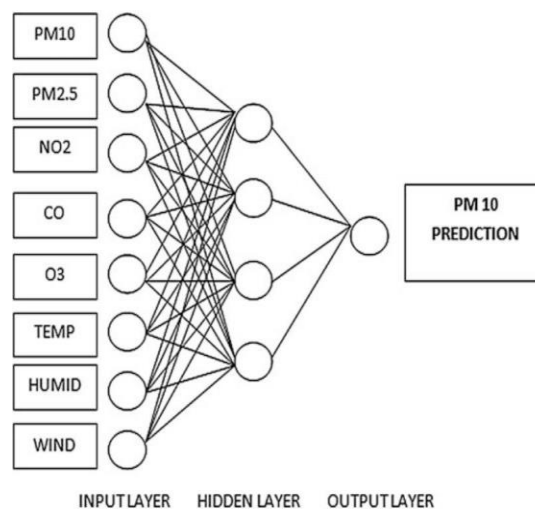


Fig. 1: Suggested ELM network for PM10 prediction

Performance Assessment of the Suggested Model :The SAFAR prediction and actual values were used to assess the predicting ability of the suggested ELM model. The regression model of ELM was utilised to predict PM10, PM2.5, NO2, and O3, hence the "root mean square error" was utilised to compare their prediction performance. On the other hand, as CO is predicted using the ELM classification model, "testing accuracy" was utilised to compare CO prediction performance.

3. Experiments

For the Delhi University monitoring station, the data was gathered in February and March 2016 via the SAFAR portal. Subsequently, the dataset was sorted to eliminate duplicate measurements taken on successive days, particularly on weekends when the portal did not consistently update the information. All of the contaminants had values between 0 and 500, according to the portal. By dividing by an appropriate power of 10, all columns in the dataset were normalised between [0, 1], with the exception of the desired output. According to the results of the studies, the first 30 readings were supplied as a training file, and the following 7 readings were given neurons for each of the five contaminants executed as a testing file in a text file format for the ELM application. Table 1 lists the ideal ELM values in summary form.

Pollutant	No. of hidden layer neurons	Activation function
PM ₁₀	4	radbas
PM _{2.5}	6	sig
NO ₂	4	sin
CO	3	sig
O ₃	3	sig

Table 1: ELM parameters for various contaminants

4. Findings And Discussion

Below are the experimental findings from each of the five suggested ELM networks. The pollutant air quality index value is shown on the y-axis in these bar graphs, while the prediction day is shown on the x-axis. In relation to the actual value for that day, as reported on the SAFAR portal, the graph compares the predictions made by SAFAR and the proposed ELM network.

4.1 PM10 Prediction

Fig. 2 displays the PM10 prediction for the next seven days. It is discovered that for each of the seven days in the testing data set, the air quality index values that ELM predicted are fairly near to the actual values.

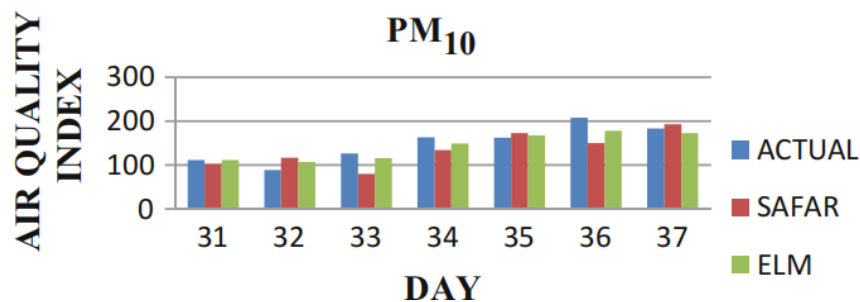


Fig. 2: Comparison of PM10 predictions

4.2 PM2.5 Prediction

Fig. 3 displays the PM2.5 projection for the next seven days. It is discovered that for each of the seven days in the testing data set, the air quality index values that ELM predicted are fairly near to the actual values.

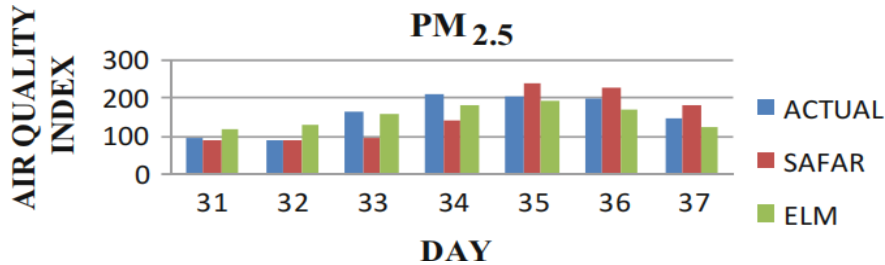


Fig. 3: Comparison of PM2.5 predictions

4.3 NO2 Prediction

Figure 4 displays the NO2 prediction for the next seven days. It is discovered that for six of the seven days in the testing data set, the air quality index values that ELM predicted are rather similar to the actual values.

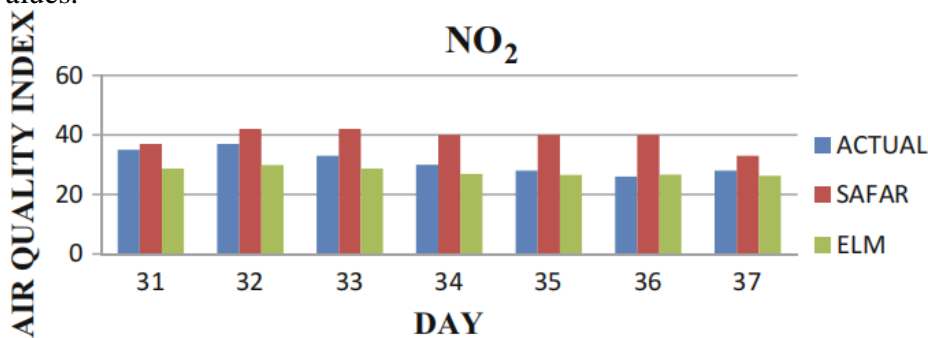


Fig. 4: Comparison of NO2 predictions

4.4 Carbon Monoxide Prediction

Fig. 5 displays the CO prediction for the next seven days. It is discovered that for six of the seven days in the testing data set, the air quality index values predicted by ELM match the actual values.

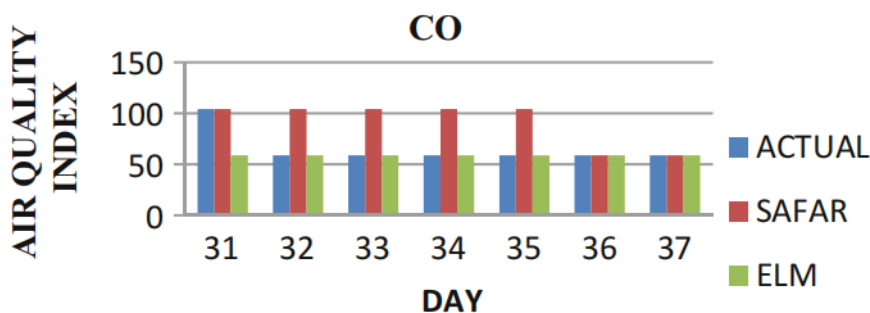


Fig. 5: Comparison of CO predictions

4.5 OZONE Prediction

Fig. 6 displays the O3 prediction for the next seven days. For five of the seven days in the testing data set, it is discovered that the air quality index values predicted by ELM are rather close to the actual values.

The performance of the proposed ELM-based system and the current SAFAR system are compared using the performance assessment metric Root Mean Square Error (RMSE), which is displayed in Fig. 7.

Fig. 8 presents a comparison of CO classification accuracy.

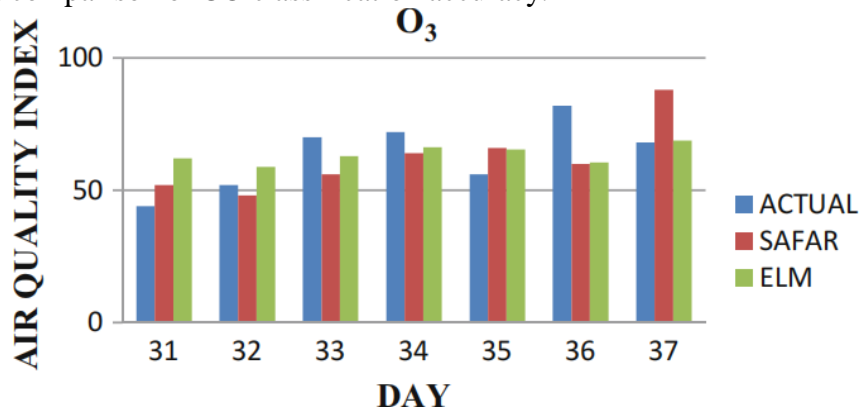


Fig. 5: Comparison of O3 predictions

5. Conclusion

In the course of this study, the focus was on predicting the air quality index for various pollutants, namely PM10, PM2.5, NO2, CO, and O3, utilizing a multivariable linear regression model known as ELM. What sets this predictive model apart is its incorporation of the previous day's pollution air quality index and weather conditions, providing a comprehensive approach to forecasting. The performance of the proposed ELM-based model underwent rigorous evaluation, with comparisons drawn between its predictions and both the actual air quality index values observed the following day and those generated by the existing SAFAR prediction system currently in operation. Notably, the findings revealed that the ELM-based predictions exhibited a higher degree of accuracy when compared to the predictions generated by the established SAFAR system, showcasing the potential of this innovative model in advancing air quality forecasting precision.

References

1. Athanasiadis, Ioannis N., Kostas D. Karatzas, Pericles A. Mitkas.: Classification techniques for air quality forecasting." Fifth ECAI Workshop on Binding Environmental Sciences and Artificial Intelligence, 17th European Conf on Artificial Intelligence. 2006.
2. Niharika, Venkatadri M., and Padma S. Rao. "A survey on Air Quality forecasting Techniques." International Journal of Computer Science and Information Technologies 5.1 (2014): 103–107.
3. Kurt, Atakan, Betul Gulbagci, Ferhat Karaca, and Omar Alagha. "An online air pollution forecasting system using neural networks." Environment International 34, no. 5 (2008): 592–598.
4. Kandy, Anurag, and Manju Mohan. "Forecasting the urban air quality using various statistical techniques." In Proceedings of the 7th International Conference on Urban Climate. 2009.
5. Peace, Margaret, Kim Dirks, and Geoff Austin. "The Prediction of Air Pollution Using A Site Optimised Model And Mesoscale Model Wind Forecasts." (2005).
6. Kumar, Anikender, and P. Goyal. "Air Quality Prediction of PM10 through an Analytical Dispersion Model for Delhi." Aerosol Air Qual. Res 14 (2014): 1487–1499.
7. Ababneh, Mohammad F., and Mohammad Hjouj Btoush. "PM10 Forecasting Using Soft Computing Techniques." Research Journal of Applied Sciences, Engineering and Technology 7, no. 16 (2014): 3253–3265.
8. WHO pollution statistics report. (accessed on 30.04.16):http://www.who.int/phe/health_topics/outdoorair/databases/cities/en/.



9. SAFAR web portal: <http://safar.tropmet.res.in>.
10. Beig, Gufran, Dilip M. Chate, Sachin D. Ghude, A. S. Mahajan, R. Srinivas, K. Ali, S. K. Sahu, N. Parkhi, D. Surendran, and H. R. Trimbake. “Quantifying the effect of air quality control measures during the 2010 Commonwealth Games at Delhi, India.” *Atmospheric Environment* 80 (2013): 455–463.
11. Srinivas, Reka, Abhilash S. Panicker, Neha S. Parkhi, Sunil K. Peshin, and Gufran Beig. “Sensitivity of online coupled model to extreme pollution event over a mega city Delhi.” *Atmospheric Pollution Research* 7, no. 1 (2016): 25–30.
12. Huang, Guang-Bin, Qin-Yu Zhu, and Chee-Kheong Siew. “Extreme learning machine: theory and applications.” *Neurocomputing* 70, no. 1 (2006): 489–501.
13. Huang, Gao, Guang-Bin Huang, Shiji Song, and Keyou You. “Trends in extreme learning machines: a review.” *Neural Networks* 61 (2015): 32–48.
14. Ekaterina Gladkova, Liliya Saychenko, “Applying machine learning techniques in air quality prediction,” *Transportation Research Procedia*, Volume 63, Pages 1999-2006, 2022.
15. Vidit Kumar, Air Pollution Prediction using Machine Learning Algorithms: A Systematic Review, *International Journal Of Engineering Research & Technology (Ijert)*, 11(12); 2022.