Global NEST Journal, Vol 27, No 3, 06865

Modeling of $PM_{2.5}$ concentrations using artificial neural networks: a case study of Islamabad

Naeem Sadiq[1],* and Zaheer Uddin[2,3]
[1]Institute of Space Science & Technology University of Karachi, Karachi, Pakistan
[2]Department of Physics, University of Karachi, Karachi, Pakistan
[3]Dawood University of Engineering & Technology, Karachi, Pakistan
Received: 04/10/2024, Accepted: 16/12/2024, Available online: 19/12/2024
*to whom all correspondence should be addressed: e-mail: nsadiq@uok.edu.pk
https://doi.org/10.30955/gnj.06865

Graphical abstract

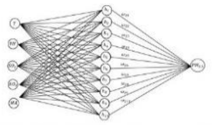

Figure 1: Model 2, which inputs Temperature, Relative humidity, Sulphar dioxide, and Nitrogen dioxide together with a five-day moving average of $PM_{2.5}$

ANN Model No.	Input Parameters
1	Temperature, Humidity, Sulphur dioxide, Nitrogen dioxide, and Five days moving average of $PM_{2.5}$
2	Temperature, Humidity, Sulphur dioxide and Nitrogen dioxide
3	Temperature and Humidity
4	Sulfur dioxide and Nitrogen dioxide
5	Temperature, Humidity, and Five days moving average of $PM_{2.5}$
6	Sulfur dioxide, Nitrogen dioxide, and Five-day moving average of $PM_{2.5}$

Table 1: Showing the configuration of input variables in different ANN models

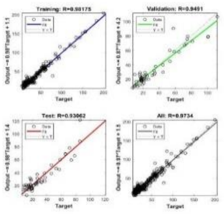

Figure 2: Regression lines for ANN model 2 with 20 neurons in the hidden layer & five variables in the Input layer

Abstract

The increasing air pollution has become a serious concern as it links with health issues. Worldwide, it has caused premature deaths. The intense research is the need of time. One of the main causes of air pollution is the existence of particulate matter, i.e. the air contains a mixture of solid particles and liquid droplets. These particles/droplets cause severe effects on health issues, especially breathing problems. This study explores the application of Artificial Neural Networks (ANNs) to forecast the level of $PM_{2.5}$ in Islamabad, addressing a critical environmental concern impacting public health. Using a diverse set of inputs Temperature, Humidity, Sulfur Dioxide (SO_2), and Nitrogen Dioxide (NO_2), alongside a five-day average of we developed models to enhance predictive accuracy. Various neural network architectures were evaluated, featuring hidden layers with neuron counts ranging from 5 to 25 and a selection of input variables. A five-day moving average of $PM_{2.5}$ was also added in two models. Model performances increased linearly with the number of model neurons, peaking at a 20-neuron configuration, yielding a correlation coefficient of 0.979 and a root mean squared error of 7.82. For the training, the values of R^2 increase from 0.949 to 0.980, as the number of neurons is increased from 5 to 25 and values of RMSE decrease from 7.53 to 5.82. But in validation and testing the RMSE values are lowest (8.20 and 8.13) for n = 20 neurons. The results obtained in this work are better than those obtained in the published work of the ANN model, in which only chemical parameters were utilized. These results highlight the importance of input selection and model complexity in capturing the intricate relationships inherent in air quality data. The study demonstrates that ANNs can be a powerful tool for air quality forecasting and provide valuable insights for policymakers and public health officials.

Keywords: Artificial neural networks, chemical inputs, environmental monitoring, Islamabad meteorological parameters, $PM_{2.5}$ prediction, urban air quality

1. Introduction

Air pollution, especially particulate matter, is one of the biggest problems concerning public health and the environment. Fine particulate matter, $PM_{2.5}$, can reach as deep as the lung and cause respiratory and cardiovascular problems since it is less than 2.5 micrometers in diameter (Kan, 2022; Pope *et al*. 2002). Few American studies also claim that Increased $PM_{2.5}$ levels lead to cardiopulmonary mortality (Laden *et al*. 2006; Kumar *et al*. 2015). The main cities have led to severe air pollution challenges, in general, with significant implications for public health (Chen *et al*. 2013). It has been also explored by research that sustained exposure to air pollution can reduce the expected life by up to 5.5 years in some areas of China (Dentener *et al*. 2006). The development of $PM_{2.5}$ concentration influencers and accurate prediction models was paramount to implementing effective mitigation strategies.

Many studies have been conducted on the relationship with $PM_{2.5}$, driven by various environmental factors. Meteorological factors, such as temperature, relative humidity, and wind speed, were considered key variables concerning $PM_{2.5}$ dispersion and accumulation (Jing *et al*. 2020; Xu *et al*., 2018). For instance, studies in Beijing, China, have pointed out that meteorological factors contributed significantly to $PM_{2.5}$ in different seasons. Moreover, many studies have pointed out the contribution of gaseous pollutants, including nitrogen dioxide and sulfur dioxide, to $PM_{2.5}$ formation (Seinfeld and Pandis, 2016; Blanchard and Hidy, 2005). These precursors of secondary aerosols comprise a large fraction of $PM_{2.5}$ (Seinfeld and Pandis, 2016).

These studies resorted to many statistical and machine learning techniques to model $PM_{2.5}$ concentrations, including multiple linear regressions, support vector machines, and ANNs. Traditional statistical models in wide use, such as multiple linear regression, have been applied, but most fail to capture the complex nonlinear relationship between $PM_{2.5}$ and its influencing factors (Christopher and Nasrabadi, 2006). In recent years, ANN has emerged as a powerful tool for predicting complex environmental phenomena. ANNs are increasingly important in predicting complex environmental phenomena because they can learn nonlinear relationships and handle large volumes of data. Some studies have used ANN to forecast $PM_{2.5}$ concentration; the results proved much better than those of traditional statistical models. For instance, the Iran study used ANNs to forecast $PM_{2.5}$ concentration with meteorological input parameters (Goudarzi *et al*. 2021). It thus proved to be a reliable method. Other studies conducted in some parts of India have also shown ANNs to be quite effective in predicting air quality (Ravindra *et al*., 2016; Tripathi *et al*., 2019; Rao *et al*., 2021).

Islamabad is the capital of Pakistan and is considered one of the most polluted cities, particularly regarding air quality in winter (Rasheed *et al*. 2014). The very same can be said about many other cities within Pakistan, as research dealing with this alarming extent of air pollution has been documented in major cities like Lahore (Ahmad *et al*. 2020) and, Karachi (Moyebi *et al*. 2023) and several minor towns (Anjum *et al*. 2021). An accurate model for forecasting $PM_{2.5}$ concentrations will lead to the appropriate mitigation measures to curtail this increasing issue.

2. Methodology

The air quality index is mainly dependent on the value of $PM_{2.5}$. To predict the concentration levels of $PM_{2.5}$ in Islamabad, various models of Artificial Neural Networks (ANNs) have been developed and employed by utilizing a comprehensive dataset comprising meteorological and chemical parameters. The architecture of the ANN is shown in Figure 1.

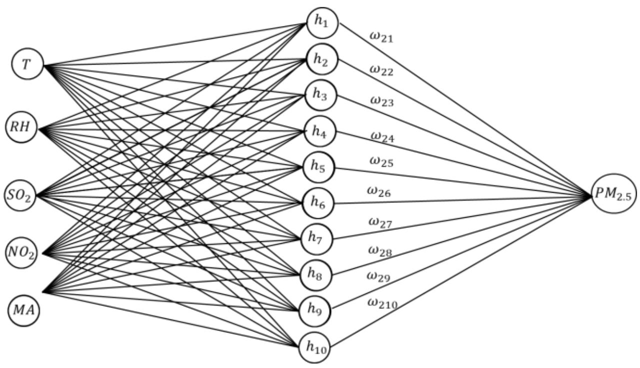

Figure 1. Model 2, which inputs Temperature, Relative humidity, Sulphar dioxide, and Nitrogen dioxide together with a five-day moving average of $PM_{2.5}$

The activation function in ANN architecture was a sigmoid function that relates input and hidden layers. The sigmoid function is given by

$$f(u) = \frac{1}{1+e^{-u}} \tag{1}$$

Where u is the weighted input to be fed in the activation function to find the values in the hidden layer and may be found by the following equation

$$u = \sum X_i \times W_{1_{ij}} \tag{2}$$

Here X_i are the input parameters and $W1_{ij}$ are the random weights. The neurons in the hidden layers acquire the values given by Y_{hidden} by substituting equation 2 in the activation function.

$$Y_{hidden} = f(u) \tag{3}$$

The activation function that links the output and hidden layer is the ReLu function.

$$Z_{out} = \sum Y_{hidden} \times W_{2_{ij}} \tag{4}$$

Where Z_{out} is fed to the activation function to obtain the output values.

$$Y_{out} = F\,(Z_{out}) \tag{5}$$

1. Data collection

The dataset for this study was obtained from the Pakistan Environmental Protection Agency, Islamabad. The meteorological data for the year 2021 was used in the calculation. It includes the variables temperature (T) and Humidity (H), together with the chemical variables, Sulphur dioxide (SO_2) and Nitrogen dioxide (NO_2).

2. Preprocessing of the Data

Before model development, the dataset was preprocessed with due care, which included:

Normalization: Data input features have been normalized to enhance model convergence and performance.

Handling Missing values: To maintain data integrity, the missing data entries were filled out by taking averages.

3. Model Architecture

Six different ANN architectures were designed depending on the choices of input parameters (see Table 1). The configuration of each model was

Input Layer: This contains the selected input parameters, ranging from two to five inputs. Figure 1 shows the architecture of model 1, the other models have the same architecture, the only difference is the choice of input parameters.

Hidden Layer: One hidden layer containing different numbers of neurons to obtain the best performance of the model. Each model was trained with 5, 10, 15, 20, and 25 neurons in the hidden layer.

Output Layer: includes the target variable, $PM_{2.5}$.

The performance of the training models was based on the following metrics: (1) Mean Squared Error (MSE) to quantify the prediction accuracy and (2) Correlation Coefficient (R): to estimate the strength of the relationship between actual and predicted values. These varying configurations were carried out methodically to reach the most effective architecture that could more closely predict $PM_{2.5}$ levels.

Table 1. Showing the configuration of input variables in different ANN models

ANN Model No.	Input Parameters
1	Temperature, Humidity, Sulphur dioxide, Nitrogen dioxide, and five-day moving average of $PM_{2.5}$
2	Temperature, Humidity, Sulphur dioxide and Nitrogen dioxide
3	Temperature and Humidity
4	Sulfur dioxide and Nitrogen dioxide
5	Temperature, Humidity, and Five-day moving average of $PM_{2.5}$
6	Sulfur dioxide, Nitrogen dioxide, and Five-day moving average of $PM_{2.5}$

2.1.1. Study area

Islamabad is the capital of Pakistan, situated at a latitude of approximately 33.68° N and a longitude of 73.04° E. The city is nestled at the foothills of the Margalla Hills and sprawls within an area of approximately 906 square kilometers with an amalgamation of urban and green spaces.

The city climatically falls into a humid subtropical climate category, experiencing distinct seasonal variations influencing air quality. Increasing air pollution increasingly affected the city, particularly particulate matter ($PM_{2.5}$), which poses significant health risks to its residents. Rapid urbanization, increased vehicle emissions, and industrial activities contribute to elevated $PM_{2.5}$ levels, making it crucial to study these environmental challenges.

3. Results and discussions

The polluted air may contain PM_1, $PM_{2.5}$, and PM_{10}, each of these pollutants has a different impact on human depending upon their size and interaction ability. $PM_{2.5}$, however, can penetrate the lungs more deeply than the other two pollutants and has a greater probability of causing severe health issues by interacting with blood cells. Due to their small size, $PM_{2.5}$ can stay longer in the environment and play a role in the formation of smog, and acid rain (MacDonald, 1989; Mehta, 2010; Gaffney *et al*. 1987; Gaffney *et al*. 2009). The most probable sources of $PM_{2.5}$ are industrial waste and heavy traffic. The people living in densely populated areas may be affected seriously due to high concentrations of $PM_{2.5}$ in the environment, they may suffer from skin cancer to lung cancer. Air quality mainly depends on the concentration of $PM_{2.5}$.

The primary purpose of this study was to calculate the air quality index using particulate matter ($PM_{2.5}$). The secondary purpose of the study is to develop an artificial neural network model to predict $PM_{2.5}$ using chemical or / and environmental parameters. The prediction of $PM_{2.5}$ was made by chemical parameters (carbon dioxide and sulphur dioxide) and temperature and relative Humidity. Six different models were developed, and $PM_{2.5}$ was predicted using these models. Table 1 mentions six different ANN models developed to study and predict $PM_{2.5}$ for Islamabad. The input parameters for this study were Temperature, Humidity, Sulfur dioxide, and Nitrogen dioxide. Additionally, we introduced a five-day average of $PM_{2.5}$ as a fifth input.

Initially, we tested a variety of hidden neuron configurations, as earlier mentioned, to analyze their impact on the model's output.

In Model 1, all four parameters mentioned in Table 1 and a five-day moving average of $PM_{2.5}$ were taken as input variables. Again, the model was trained for various numbers of neurons (5, 10, 15, 20 and 25). A comparison was made between the training validation and testing results based on Root Mean Squared Error (RMSE) and correlation coefficient (R). It was found that model 2 also showed the best performance for 20 neurons. The values of RMSE for training validation and testing for 20 neurons are 5.91, 8.20, and 8.42, and the correlation coefficient values are 0.98, 0.95, and 0.93, respectively. Figure 2 shows the regression lines for training, validation, testing, and all data. The complete results are shown in Table 2.

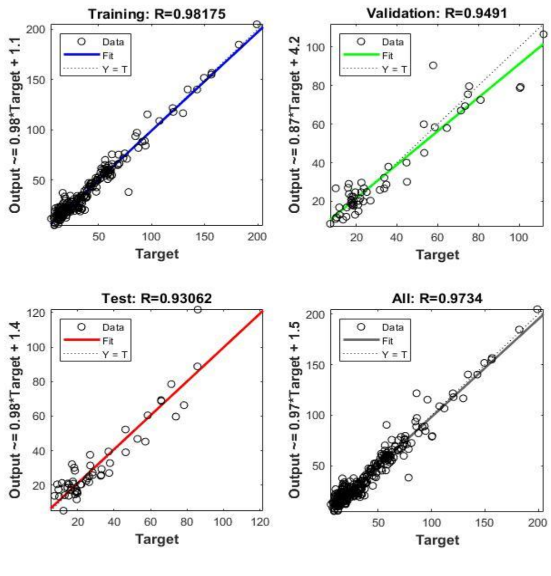

Figure 2. Regression lines for ANN model 2 with 20 neurons in the hidden layer & five variables in the Input layer

In Model 2, all four parameters mentioned in the Table 1 were taken as input variables. The data was divided into three parts: i.e. 70%, 15%, and 15% for training, validation, and testing. The model was trained for various neurons (5, 10, 15, 20 and 25). A comparison was made between the training validation and testing results based on Root Mean Squared Error (RMSE) and correlation coefficient (R). This model shows the best performance for 20 neurons. The values of RMSE for training validation and testing for 20 neurons are 9.37, 9.87, and 8.42, and the values of the correlation coefficient are 0.94, 0.93, and 0.97, respectively. Figure 3 shows the regression lines for training, validation, testing, and all data. The complete results are shown in Table 3.

Table 2. Performance Metrics of ANN with Varying Neuron Counts for $PM_{2.5}$ Prediction (five Inputs, one output)

Neurons	5		10		15		20		25	
	RMSE	R	RMSE	R	RMSE	R	RMSE	R	RMSE	R
	9.30	0.949	7.53	0.966	6.96	0.975	5.91	0.982	5.82	0.980
	9.31	0.957	8.65	0.965	9.23	0.933	8.20	0.949	8.55	0.962
	9.39	0.932	7.54	0.964	8.78	0.917	8.13	0.931	8.61	0.942

Table 3. Performance Metrics of ANN with Varying Neuron Counts for $PM_{2.5}$ Prediction (four Inputs, one output)

Neurons	5		10		15		20		25	
	RMSE	R	RMSE	R	RMSE	R	RMSE	R	RMSE	R
	9.92	0.929	9.30	0.950	9.97	0.944	9.37	0.945	8.82	0.958
	9.80	0.953	9.35	0.926	9.74	0.895	9.87	0.933	9.88	0.900
	10.58	0.953	9.98	0.931	11.14	0.921	8.42	0.969	9.82	0.909

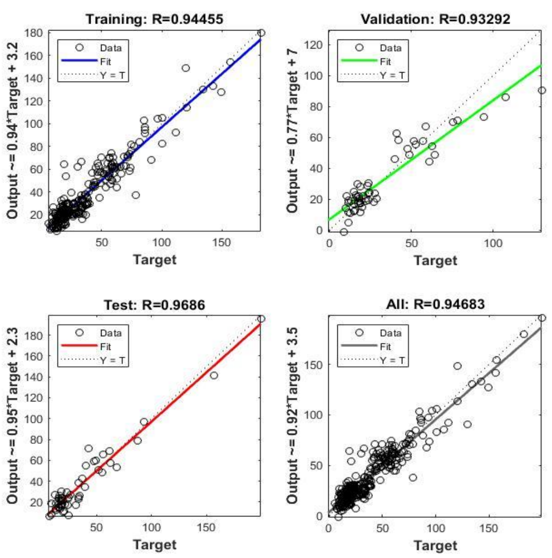

Figure 3. Regression lines for ANN model 1 with 20 neurons in the hidden layer & four variables in the Input layer

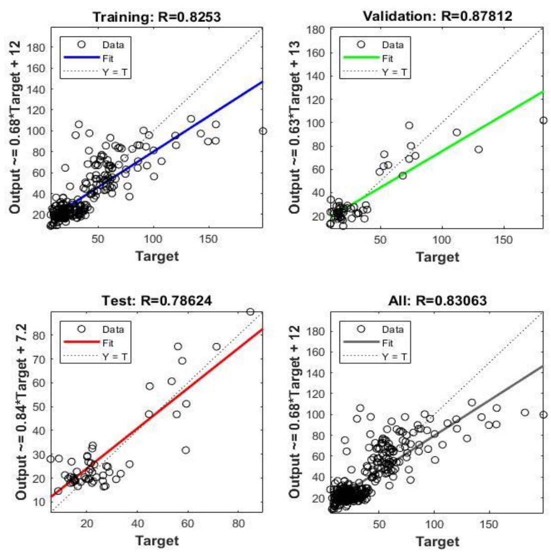

Figure 4. Regression lines for ANN model 3 with 10 neurons in the hidden layer & two Meteorological variables in the Input layer

Table 4. Performance Metrics of ANN for Meteorological Inputs in $PM_{2.5}$ Prediction (two meteorological Inputs, one output)

Neurons	10	
	RMSE	**R**
	16.77	0.825
	16.35	0.878
	11.76	0.786

Table 5. Performance Metrics of ANN for Chemical Inputs in $PM_{2.5}$ Prediction (two chemical Inputs, one output)

Neurons	10	
	RMSE	**R**
	13.13	0.901
	10.23	0.882
	12.13	0.908

In Model 3, we only utilized the two meteorological parameters mentioned in Table 1 as input variables. Again, the model was trained for various numbers of neurons (5, 10, 15, 20 and 25). It was found that model 3 shows the best performance for ten neurons in the hidden layer; hence, the results of the training validation and testing based on Root Mean Squared Error (RMSE) and correlation coefficient (R) for ten neurons in the Gidden layer are summarized in Table 4. The values of RMSE for training validation and testing for ten neurons are 16.77, 16.35, and 11.76, and the correlation coefficient values are 0.82, 0.87, and 0.78, respectively. Figure 4 shows the regression lines for training, validation, testing, and all data. The complete results are shown in Table 4.

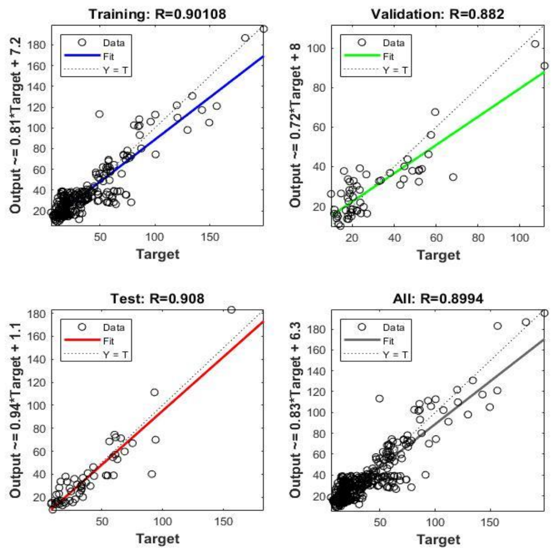

Figure 5. Regression lines for ANN model 4 with 10 neurons in the hidden layer & two Chemical variables in the Input layer

In Model 4, we only utilized the two chemical parameters mentioned in Table 1 as input variables. Again, the model was trained for various numbers of neurons (5, 10, 15, 20, and 25). It was found that model 4 also shows the best performance for ten neurons in the hidden layer; hence, the results of the training validation and testing based on Root Mean Squared Error (RMSE) and correlation coefficient (R) based on ten neurons are summarized in Table 4. The values of RMSE for training validation and testing for ten neurons are 13.13, 10.23, and 12.13, and the values of the correlation coefficient are 0.90, 0.88, and 0.91, respectively. Figure 5 shows the regression lines for training, validation, testing, and all data. The complete results are shown in Table 5.

In Model 5, we only utilized the two meteorological parameters along with a five-day moving average of $PM_{2.5}$, mentioned in Table 1, which were taken as input variables. Again, the model was trained for various numbers of neurons (5, 10, 15, 20 and 25). It was found that model 5 shows the best performance for 20 neurons in the hidden layer; hence, the results of the training validation and testing based on Root Mean Squared Error (RMSE) and correlation coefficient (R) based on 20 neurons are summarized in Table 5. The values of RMSE for training validation and testing for 20 neurons are 12.25, 11.59, and 10.63, and the correlation coefficient values are 0.92, 0.88, and 0.91, respectively. Figure 6 shows the regression lines for training, validation, testing, and all data. The complete results are shown in Table 6.

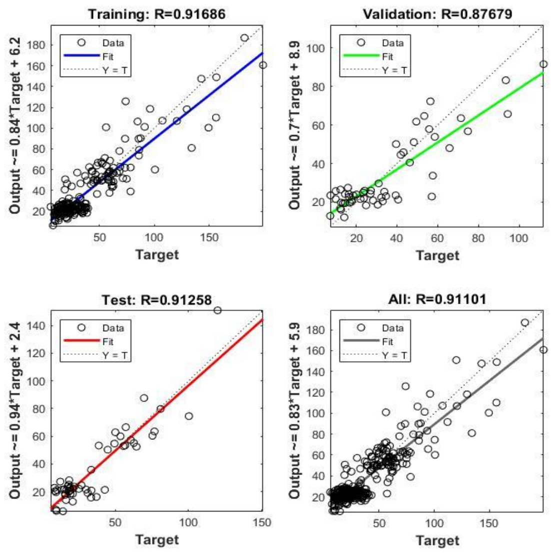

Figure 6. Regression lines for ANN model 5 with 20 neurons in a hidden layer with two meteorological with one Moving Average of $PM_{2.5}$

Table 6. Performance Metrics of ANN for Meteorological Inputs with 20 Neurons in $PM_{2.5}$ Prediction(two meteorological Inputs, one output)

Neurons	20	
	RMSE	R
	12.25	0.917
	11.59	0.877
	10.63	0.913

Table 7. Performance Metrics of ANN for Chemical Inputs with 20 Neurons in $PM_{2.5}$ Prediction (two chemical Inputs, one output)

Neurons	20	
	RMSE	R
	8.25	0.952
	7.82	0.979
	8.29	0.953

These results reflect that the 20-neuron configuration gives the right level of complexity to yield optimal predictive accuracy.

In Model 6, we only utilized the two chemical parameters and a five-day moving average of $PM_{2.5}$, as mentioned in Table 1, which were taken as input variables. Again, the model was trained for various numbers of neurons (5, 10, 15, 20 and 25). It was found that model 5 shows the best performance for 20 neurons in the hidden layer; hence, the results of the training validation and testing based on Root Mean Squared Error (RMSE) and correlation coefficient (R) based on 20 neurons are summarized in Table 5. The values of RMSE for training validation and testing for 20 neurons are 8.25, 7.82, and 8.29, and the correlation coefficient values are 0.95, 0.98, and 0.95, respectively. Figure 7 shows the regression lines for training, validation, testing, and all data. The complete results are shown in Table 7.

Our results underscore the effectiveness of employing a 20-neuron configuration for chemical modeling

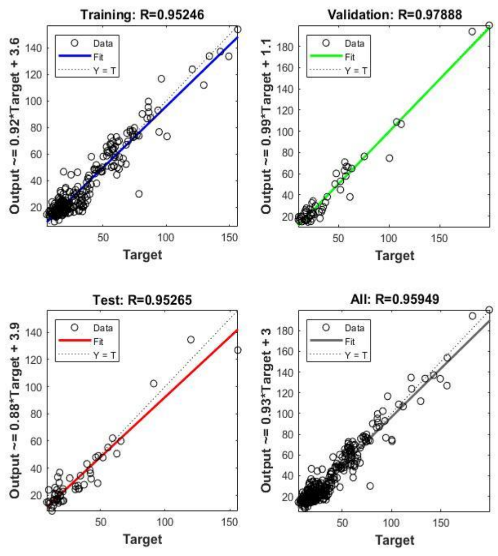

Figure 7. Regression lines for ANN model 6 with 20 neurons in a hidden layer with two chemicals with 1 Moving Average of $PM_{2.5}$

4. Conclusion

This study successfully demonstrates the effectiveness of Neural Networks in predicting $PM_{2.5}$ concentrations in Islamabad using different meteorological and chemical variables. The best predictions correspond to configurations of different numbers of neurons for different models. To predict $PM_{2.5}$, six different ANN models were developed using various combinations of meteorological and environmental variables. Including diverse input parameters comprising temperature, Humidity, sulfur dioxide, nitrogen dioxide, and a 5-day moving average of $PM_{2.5}$ enhanced the model's ability to capture intricate relationships within data sets. The significant reduction in mean squared error across various input configurations underscores the importance of input selection and architecture structure, which can play the most crucial role in reliable forecasting.

It is observed that by considering only two parameters, either meteorological or chemical, the results are not so promising. However, we get relatively better results at ten neurons, as in the case of Model 3 and Model 4. RMSE values are higher than those for the other four models. It is also why the RMSE values in reference (Ahmed *et al*. 2022) are higher than those obtained in this work with the combination of environmental and chemical parameters as in model 1 and model 2. While adding a parameter, that is, the five-day moving average, with either meteorological or chemical variables (model 5 and model 6), results significantly improved, especially in the case of chemical parameters (see table 7). It is also observed that, overall, Model 2 and Model 6 perform well compared to the other four models. Both RMSE and R are the least and maximum, respectively.

These results not only underline the potential of ANN as a strong tool in air quality modeling but also provide a foundation for future research to optimize predictive models for urban environments. As cities like Islamabad confront the challenges of air pollution, it therefore calls for proper and effective policy in addition to public health strategies.

Acknowledgment

There is no funding for the publication, but one of the authors received a stipend from the Sindh Higher Education Commission.

References

Ahmad M., Cheng S., Yu Q., Qin W., Zhang Y. and Chen J. (2020). Chemical and source characterization of $PM_{2.5}$ in summertime in severely polluted Lahore, Pakistan. *Atmospheric Research*, **234**, 104715.

Ahmed M., Xiao Z. and Shen Y. (2022). Estimation of ground $PM_{2.5}$ concentrations in Pakistan using convolutional neural network and multi-pollutant satellite images. *Remote Sensing*, **14**(7), 1735.

Anjum M.S., Ali S.M., Subhani M.A., Anwar M.N., Nizami A.S., Ashraf U. and Khokhar M.F. (2021). An emerged challenge of air pollution and ever-increasing particulate matter in Pakistan; a critical review. *Journal of Hazardous Materials*, **402**, 123943.

Blanchard C.L. and Hidy G.M. (2005). Effects of SO_2 and NO_x emission reductions on $PM_{2.5}$ mass concentrations in the southeastern United States. *Journal of the Air and Waste Management Association*, **55**(3), 265–272.

Chen Y., Ebenstein A., Greenstone M. and Li H. (2013). Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy. *Proceedings of the National Academy of Sciences*, **110**(32), 12936–12941.

Christopher M. and Nasrabadi N.M. (2006). Pattern Recognition and Machine Learning. Springer. *vol. Bishop*, **183**.

Dentener F., Stevenson D., Ellingsen K.V., Van Noije T., Schultz M., Amann M. and Zeng G. (2006). The global atmospheric environment for the next generation. *Environmental Science & Technology*, **40**(11), 3586–3594.

Gaffney J.S., Marley N.A. and Frederick J.E. (2009). Formation and effects of smog. Environmental and Ecological Chemistry; Sabljic, A., Ed.; Eolss Publishers Co., Ltd.: Oxford, UK, 2, 25–51.

Gaffney J.S., Streit G.E., Spall W.D. and Hall J.H. (1987). Beyond acid rain. Do soluble oxidants and organic toxins interact with SO_2 and NO_x to increase ecosystem effects? *Environmental science & technology*, **21**(6), 519–524.

Goudarzi G., Hopke P.K. and Yazdani M. (2021). Forecasting $PM_{2.5}$ concentration using artificial neural network and its health effects in Ahvaz, Iran. *Chemosphere*, **283**, 131285.

Haykin S. (2009). Neural networks and learning machines. Pearson Education.

Jing Z., Liu P., Wang T., Song H., Lee J., Xu T. and Xing Y. (2020). Effects of meteorological factors and anthropogenic precursors on $PM_{2.5}$ concentrations in cities in China. *Sustainability*, **12**(9), 3550.

Kan H. (2022). World Health Organization air quality guidelines 2021: implication for air pollution control and climate goal in China. *Chinese Medical Journal*, **135**(5), 513–515.

Kumar P., Khare M., Harrison R.M., Bloss W.J., Lewis A.C., Coe H. and Morawska L. (2015). New directions: Air pollution challenges for developing megacities like Delhi. *Atmospheric Environment*, **122**, 657–661.

Laden F., Schwartz J. and Speizer F.E. (2006). Reduction in fine particulate air pollution and mortality: Extended analysis of the Harvard Six Cities Study. *American Journal of Respiratory and Critical Care Medicine*, **173**(6), 667–672.

Li M., Wang W.L., Wang Z.Y. and Xue Y. (2018). Prediction of $PM_{2.5}$ concentration based on the similarity in air quality monitoring network. *Building and Environment*, **137**, 11–17.

MacDonald G.J. (1989). Climate Change, Smog and Acid Rain: Linkages between Pollutants, Effects and Controls. Global Climate Change Linkages. New York, NY: *Elsevier Science Publishing Company, Inc*, 95–120.

Mehta P. (2010). Science behind acid rain: analysis of its impacts and advantages on life and heritage structures. *south Asian journal of tourism and heritage*, **3**(2), 123–132.

Moyebi O.D., Sannoh F., Fatmi Z., Siddique A., Khan K., Zeb J. and Khwaja H.A. (2023). State of gaseous air pollutants and resulting health effects in Karachi, Pakistan. *Environmental monitoring and assessment*, **195**(2), 266.

Pope C.A., Burnett R.T., Thun M.J., Calle E.E., Krewski D., Ito K. and Thurston G.D. (2002). Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *Journal of the American Medical Association*, **287**(9), 1132–1141.

Rao N.D., Kiesewetter G., Min J., Pachauri S. and Wagner F. (2021). Household contributions to and impacts from air pollution in India. *Nature Sustainability*, **4**(10), 859–867.

Rasheed A., Aneja V.P., Aiyyer A. and Rafique U. (2014). Measurements and analysis of air quality in Islamabad, Pakistan. *Earth's future*, **2**(6), 303–314.

Ravindra K., Sidhu M.K., Mor S., John S. and Pyne S. (2016). Air pollution in India: bridging the gap between science and policy. *Journal of Hazardous, Toxic, and Radioactive Waste*, **20**(4), A4015003.

Seinfeld JH. and Pandis S.N. (2016). Atmospheric chemistry and physics: From air pollution to climate change. *John Wiley & Sons*.

Tripathi C.B., Baredar P. and Tripathi L. (2019). Air pollution in Delhi. *Current Science*, **117**(7), 1153–1160.

Xu Y., Xue W., Lei Y., Zhao Y., Cheng S., Ren Z. and Huang Q. (2018). Impact of meteorological conditions on $PM_{2.5}$ pollution in China during winter. *Atmosphere*, **9**(11), 429.

Zhang Z., Zeng Y. and Yan K. (2021). A hybrid deep learning technology for $PM_{2.5}$ air quality forecasting. *Environmental Science and Pollution Research*, **28**, 39409–39422.