Proceedings of the 6th International Conference on Civil Engineering for Sustainable Development (ICCESD 2022), 10~12 February 2022, KUET, Khulna, Bangladesh (ISBN-978-984-35-1972-6)

DEVELOPMENT OF $PM_{2.5}$ CONCENTRATION PREDICTION MODEL USING MACHINE LEARNING APPROACHES

Sudipta Roy*[1], Rumpa Chowdhury[2] and Prithula Purna[3]

[1] *M.Sc. Student, Bangladesh University of Engineering and Technology, Bangladesh, e-mail: sudiptaroy1994@gmail.com*
[2] *M.Sc. Student, Bangladesh University of Engineering and Technology, Bangladesh, e-mail: rumpa.ce.buet@gmail.com*
[3] *B.Sc. Student, Bangladesh University of Engineering and Technology, Bangladesh, e-mail: prithula.twinkle@gmail.com*

***Corresponding Author**

ABSTRACT

This study develops mathematical models to predict the concentration of $PM_{2.5}$ (Particulate matter of 2.5 microns or smaller) using machine learning-based models. These models forecast monthly average $PM_{2.5}$ concentration based on PM_{10} (Particulate matter of 10 microns or smaller) concentration data considering influence of seasonal and regional difference in relative humidity. The dataset for this analysis is obtained through the monthly air quality report of the Clean Air and Sustainable Environment (CASE) project from 2016 to 2018, which collects air quality monitoring data from 11 Continuous Air Monitoring Station (CAMS) located in 8 (eight) major cities of Bangladesh. The one-way ANOVA test result verifies the influence of seasonal variation on PM concentration which is consistent with prior studies but repudiates the regional effect. The mean ratios of $PM_{2.5}$ to PM_{10} are slightly higher in the dry seasons than in the wet seasons throughout the study period. Different machine learning approaches, such as Multiple Regression (Linear and Logarithmic), Artificial Neural Network (ANN) and Random Forest (RF) methods, are used to develop the mathematical models for $PM_{2.5}$ prediction using R statistical packages. The monthly average PM_{10} concentration and the seasonal variation are used as input variables and found significant in the developed models. All the models are statistically significant with high R^2 values, where the highest R^2 (0.90) is found from the ANN model (with 2,1 hidden layer formation). This ANN model performs slightly better in forecasting spatially fluctuating $PM_{2.5}$ concentrations, which is supported well by previous corresponding studies. These models will lead to the effective prediction of $PM_{2.5}$ in absence of precise concentration measurement facilities which will eventually help to identify and reduce the pollution.

Keywords: *Particulate matter, Regression, ANN, RF, One way ANOVA*

1. INTRODUCTION

At present, Air pollution is an issue of concern worldwide as an outcome of rapid urbanization and economic growth. In developing countries of Asia, people have exposure to air pollution that is largely caused by particulate matter (PM). Fine particulate matter is considered a potential pollutant for being inhalable in nature that causes both economic loss and health implications on society (Begum, 2016). Scientists have been studying the health effects of PM for many years and careful analysis of data supported the fact that PM has adverse health effects like respiratory and cardiopulmonary diseases even at relatively low concentrations and many epidemiological studies conducted in various areas of the United States and European countries suggest that all-cause daily mortality increased by 0.5%-1.5% and 0.6% respectively for each 10 µg/m^3 increase in PM_{10} (Valavanidis et al., 2008). Airborne particulate matters are found in metropolises as well as small and sizeable towns, and these usually consist of organic materials adsorbed onto particles that can be classified into volatile or semi-volatile organic species, transition metals, ions, reactive gases, carbonaceous materials produced by vehicular combustion processes, materials of biologic origin etc. (Valavanidis et al., 2008).

Prediction of PM_{10} and $PM_{2.5}$ is significant as it will help by warning concerned authorities and governments as well as residents about the condition of the affected area and taking timely action for protecting public health and for this purpose, we need prediction models for forecasting concentration levels of particulate matters accurately which will be based on real-time data of the pollutants collected from air quality monitoring stations. Due to complications in measurement processes, sometimes the air quality monitoring systems lack $PM_{2.5}$, while they can measure PM_{10} with their existing facilities. In many cases particulate matter with aerodynamic diameter up to 10 µm (PM_{10} and $PM_{2.5}$) have common origins that lead to the assumption that their concentrations in the air are highly correlated, and those sources can be identified using the ratio of $PM_{2.5}$ to PM_{10} (Nikolova et al., 2015; Zhao et al., 2019). In some previous studies, the $PM_{2.5}/PM_{10}$ ratio proved to be significant in assessing Air Quality Index (AQI) and is analyzed statistically to determine its influence on ambient air quality (Duan et al., 2015; Zhou et al., 2019; Zhao et al., 2019).

Machine learning (ML) is an effective data analysis and modeling tool which is closely associated with artificial intelligence and widely accepted by researchers and scientists for different types of classification and prediction applications. In case of supervised learning, ML model can learn and train from different empirical data and used to predict or classify unknown data based on the supervision. Some popular supervised ML algorithms are Linear Regression, Artificial Neural Network (ANN), Random Forest (RF) etc. which are used frequently in mathematical model development (Deka, 2019). Nikolova et al. (2015) established a mathematical model for calculating $PM_{2.5}$ concentration using that of PM_{10} for purpose and it was conducted in Bulgaria where the number of $PM_{2.5}$ monitor stations was limited. Again, Machine learning was used for forecasting PM concentrations in some past studies (Asadollahfardi et al., 2016; Evanov et al., 2018). McKendry (2002) compared ANN models to conventional linear models for PM forecasting.

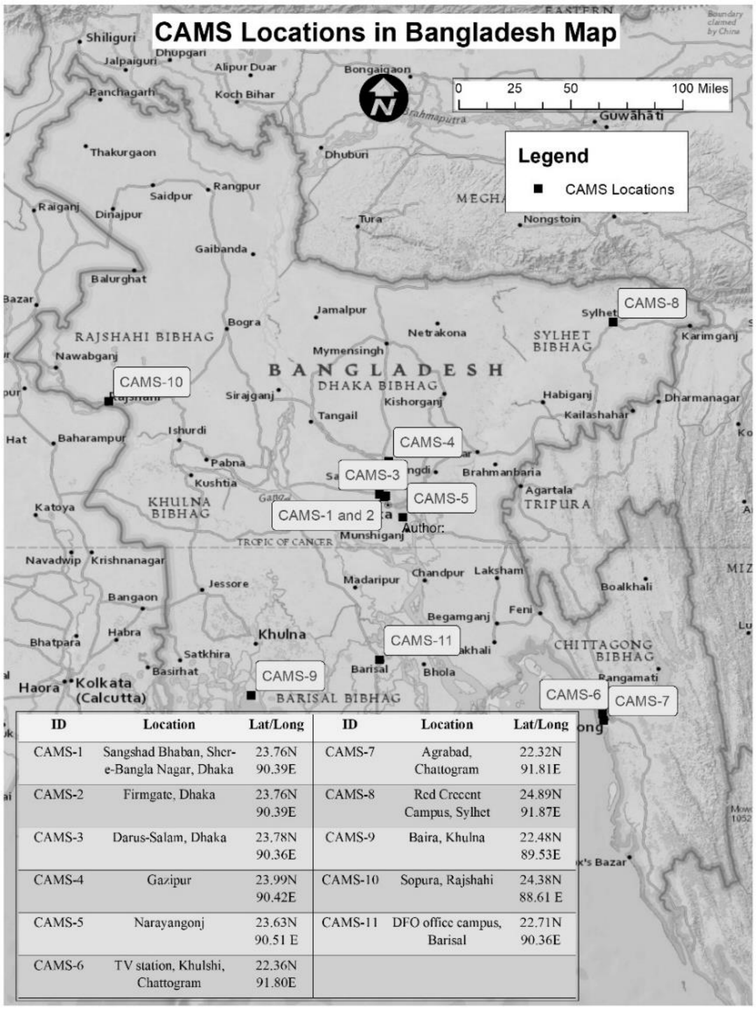
In this study, data on PM_{10} and $PM_{2.5}$ concentration are collected from 11 Continuous Air Monitoring Stations located in 8 (eight) major cities of Bangladesh and are statistically analyzed. Firstly, the concentration of $PM_{2.5}$ is forecasted by establishing multiple regression, ANN and RF models based on that of PM_{10}. Secondly, the significance of seasonal and regional variations are assessed through the one-way ANOVA test. Finally, developed models are compared and checked for their efficiency.

2. METHODOLOGY

The methodology of this study is divided into sections like selection of study area, data collection and processing and statistical analysis which are described hereafter. “R Statistical Package” is used in this study for all sorts of analysis.

2.1 Study Area

The dataset required for this study is collected through the Clean Air and Sustainable Development (CASE) project which is associated by Department of Environment (DoE) as a part of thoroughly monitoring air quality of Bangladesh. As a part of this project, real-time measurement of particulate matter ($PM_{2.5}$ and PM_{10}) and other ambient level pollutants are recorded in eleven (11) Continuous Air Monitoring Stations (CAMS) at eight (8) major cities (Dhaka, Gazipur, Narayangonj, Chattogram, Khulna, Rajshahi, Sylhet and Barisal) of Bangladesh which is mapped in Figure 1 (DoE, 2018).



ID	Location	Lat/Long	ID	Location	Lat/Long
CAMS-1	Sangshad Bhaban, Sher-e-Bangla Nagar, Dhaka	23.76N 90.39E	CAMS-7	Agrabad, Chattogram	22.32N 91.81E
CAMS-2	Firmgate, Dhaka	23.76N 90.39E	CAMS-8	Red Crecent Campus, Sylhet	24.89N 91.87E
CAMS-3	Darus-Salam, Dhaka	23.78N 90.36E	CAMS-9	Baira, Khulna	22.48N 89.53E
CAMS-4	Gazipur	23.99N 90.42E	CAMS-10	Sopura, Rajshahi	24.38N 88.61 E
CAMS-5	Narayangonj	23.63N 90.51 E	CAMS-11	DFO office campus, Barisal	22.71N 90.36E
CAMS-6	TV station, Khulshi, Chattogram	22.36N 91.80E			

Figure 1: Location of CAMS in Bangladesh

2.2 Data Collection and Processing

Monthly average $PM_{2.5}$ (24 hrs.) and average PM_{10} (24 hrs.) data of a three (3) year-long period i.e., from January 2016 to December 2018 collected from CASE monthly reports are used in this study. During this period, in some CAMSs, the data was unavailable for a few months, so the data for those periods and CAMSs are excluded in the analysis. In this study, the dry period is considered from October to March and the wet season is considered from April to September. The significance of the PM concentrations for different seasons are measured with one-way ANOVA test.

2.3 Mathematical Modelling for Correlation Establishment

The establishment of correlation between PM_{10} and $PM_{2.5}$ required an assumption of generic form of correlation equation which is assumed in both linear form and logarithmic form. The equations are given below:

$$PM_{2.5} = C + K_0 PM_{10} + K_i\beta_i \quad (1)$$

$$log_{10} PM_{2.5} = C + K_0 \, log_{10} PM_{10} + K_i\beta_i \quad (2)$$

Where, C, K_0 and Ki are constant terms and βi are the other independent factors affecting the relationship between PM_{10} and $PM_{2.5}$. The value of i starts from 1 to n and β is a binary variable having only two values, 0 and 1. In this study, locations of three (3) CAMSs out of eleven (11) are in coastal regions like Chattogram and Khulna, where the relative humidity is higher. The data collection period is divided into two seasons in a year: dry and wet seasons. In wet seasons, the relative humidity is higher. So, taking these factors in consideration during correlation establishment, two (2) binary variables are introduced in trial phase of modelling (Table 1) to assess the influence of relative humidity in PM concentration.

Table 1: Binary variables/ factors used in model building

Variables	Will be 1 when	Will be 0 when
Season	Wet	Dry
Region	Coastal	Others

Stepwise multiple regression is used to find the significant variables that can be used to build the correlation equation. The established best-fitted empirical model is selected based on their goodness-of-fit parameter (R^2 value) and the significance of model parameters.

2.4 Artificial Neural Network Model

In this study Multi-Layer Feed-Forward (MLF) Network, which is one of the most popular forms of Artificial Neural Network (ANN), is used (Figure 2). The network is divided into three (3) layers; input layer, hidden layer(s), and output layer. The structure of the ANN is decided by a trial-and-error method to minimize the root mean square error (RMSE) and mean absolute error (MAE). The statistically significant variables in multiple regression model are used as an input in this network. The number of nodes in hidden layers and the number of hidden layers is assumed intuitively and then the network is improved by trial-and-error method. To avoid network overflowing due to very large or small weights the inputs of the network are transformed in the range of 0-1. For this reason, the PM concentrations are divided by a number greater than the largest PM concentration in the database.

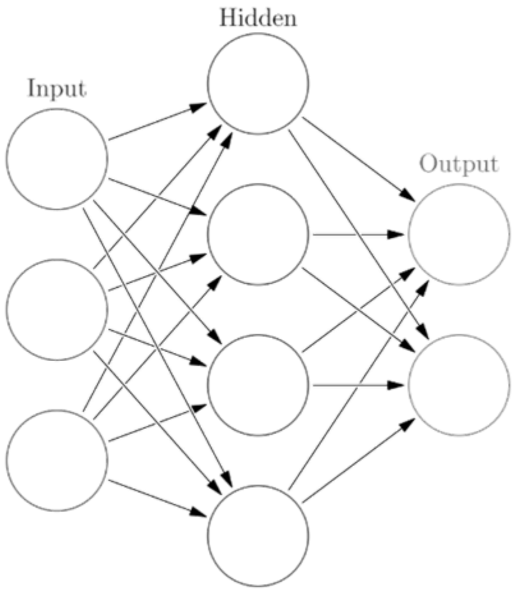


Figure 2: Standard Multi-Layer Feed-Forward (MLF) Network

2.5 Random Forest Model

Random Forest (RF) is one of the most frequently used supervised machine learning algorithm using ensemble method for regression, which is combining prediction from multiple models to make a more accurate prediction. It is the association of some decision trees which predicts the target value independently and then all of them are averaged to get the actual prediction. It is a powerful and accurate model for linear and non-linear prediction. The structure of the model is decided by trial-and-error method, which is like the ANN process stated above. The scaled dataset is used for this model also.

3. RESULTS AND DISCUSSION

3.1 ANOVA – One Way Test Result

The seasonal variation between the means of PM_{10} and $PM_{2.5}$ concentration in two different seasons (dry and wet) are found statistically significant in One-way ANOVA test at 95% confidence interval which is depicted by the F values with $Pr < 0.05$ in Table 2. Dash and Dash (2015) found the PM_{10} and $PM_{2.5}$ concentrations are significantly varied based on the seasonal differences (summer, monsoon, and winter) in their one-way ANOVA test considering data from March 2014 to February 2015 in India. This corresponding test result supports our one-way ANOVA test result. But the seasonal variation between the means of PM_{10} and $PM_{2.5}$ concentration in different region types (coastal and others) are found statistically insignificant which indicates the influence of coastal humidity variation doesn't have any statistical impact on PM concentration.

Table 2: One-Way ANOVA test results for PM_{10} and $PM_{2.5}$ concentration based on regional and seasonal variation

Parameters	Variance source	Sum of Squares	Degree of Freedom	Mean Squares	F - values
PM_{10}	Between seasons	1135011	1	1135011	291.2***
	Within seasons	1025169	263	3898	
	Total	2160180	264	-	
	Between regions	29138	1	29138	3.596
	Within regions	2131043	263	8103	
	Total	2160181	264	-	
$PM_{2.5}$	Between seasons	421893	1	421893	295.3***
	Within seasons	375799	263	1429	
	Total	797692	264	-	
	Between regions	9857	1	9857	3.291
	Within regions	787835	263	2996	
	Total	797692	264	-	

3.2 Mathematical Model Development

The PM_{10} and $PM_{2.5}$ concentrations in the atmosphere are strongly co-related due to their common origin, stated by Nikolova et al. (2015). One of the objectives of this study is to develop mathematical model for predicting the concentration of $PM_{2.5}$ using monitored data of PM_{10} concentrations. Two different machine learning approaches such as multiple regression and artificial neural network are used in the modelling purpose which are briefly described in methodology portion.

3.2.1 Regression Model

The following best-fitted (3) linear and (4) logarithmic equations have been found through the multiple regression approach using three-year PM_{10} and $PM_{2.5}$ concentration dataset from 2016 to 2018.

$$PM_{2.5} = 4.54 + 0.53\times PM_{10} - 9.98\times\beta_{season} \quad (3)$$

$$log(PM_{2.5}) = 0.037 + 0.866\times log(PM_{10}) - 0.134\times\beta_{season} \quad (4)$$

Where, β seasonal binary variable (Table 1 in methodology is referenced). During the trial session of regression procedure, the regional binary variable is included but as an insignificant parameter, it is removed during the final model formation. The values of R^2 for equations (3) and (4) are 0.89 and 0.88

respectively. Both the linear and the logarithmic models are found significant (Pr. < 0.05) with satisfactory R^2 values indicating healthy correlation between dependent and independent variables.

3.2.2 ANN Model

The input variables of the ANN model trial phase are determined by the one-way ANOVA test conducted earlier. According to this test result, PM_{10} concentration and season of measurements are given as an input parameter along with (1) hidden layer formation in ANN model to predict $PM_{2.5}$ concentrations as output. A comparative analysis is carried out to find the best architecture of the hidden layers with least error. The dataset is firstly scaled as described in methodology and then divided into two parts such as training data (75% of dataset) and test data (25% of dataset) by random criteria. The model network is first trained with the training data. The best result in the training of the network is achieved after 132 iterations, as the mean square error (performance) is 0.12, where two hidden layers with (2,1) neuron formation gives the least RMSE and MAE error with best output, as shown in figure 3. Several studies recommend that to prove the validity of the model (Feng et al., 2015; Voukantsis et al., 2011). After that the generated model is tested to find the desired output with test data.

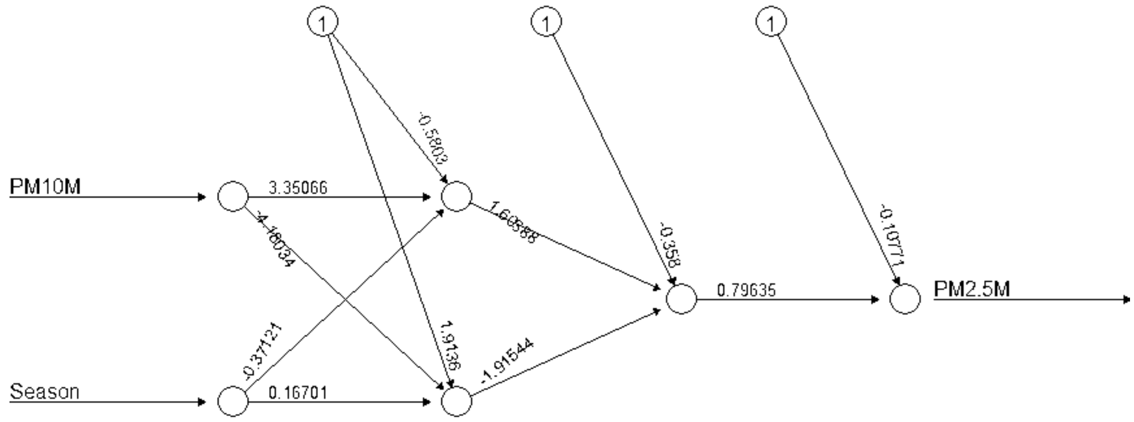


Figure 3: The ANN model for forecasting $PM_{2.5}$ concentration (The coefficients are for scaled dataset)

Table 3 shows the statistical test results found from the ANN model that exhibit an effective performance for predicting $PM_{2.5}$ concentrations. The RMSE, MAE values of the training, test and combined dataset are in a very close range. The correlation (R^2 value) of the predicted $PM_{2.5}$ concentration and actual concentration of these datasets are also good. The Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) values are used together to identify the variation in the errors in a set of forecasts as well as to assess the performance of our ANN model. The mean absolute error gives an estimation of the average magnitude errors in a set of forecasts and the root mean square error provides information about the variance in the individual errors in the sample. The RMSE will always be greater or identical to the MAE and the higher difference between them indicates the higher variance in the individual errors in the sample. When the RMSE is equal to the MAE, the magnitude of all errors is same (Bhatt et al., 2014; Lefèvre et al., 2014; Willmott et al., 1985). After evaluating the above-mentioned statistics of MAE and RMSE, we found from our ANN model that the maximum difference between the RMSE and MAE values of our dataset is 5.33 and thus, the variance in the individual errors in the sample are insignificant which proves that our ANN model can predict the $PM_{2.5}$ concentration efficiently.

Table 3: The statistical test results of the ANN model

Dataset	R^2 (Predicted vs Actual $PM_{2.5}$)	Root Mean Square Error (RMSE)	Mean Absolute Error (MAE)
Training Data	0.90	17.55	12.22
Test Data	0.88	17.12	12.53
Combined Data	0.90	17.44	12.30

3.2.3 Random Forest Model

PM_{10} concentration and season of measurements are given as an input in RF model trial to predict $PM_{2.5}$ concentrations as output. After repeated trial-and-error method, number of variables available for splitting at each tree node value for the trial, m = 1 and number of trees, n = 1000 gives the least RMSE and MAE error with best output. The mean of squared residual, i.e., error is found 481.87 for this model (Figure 4).

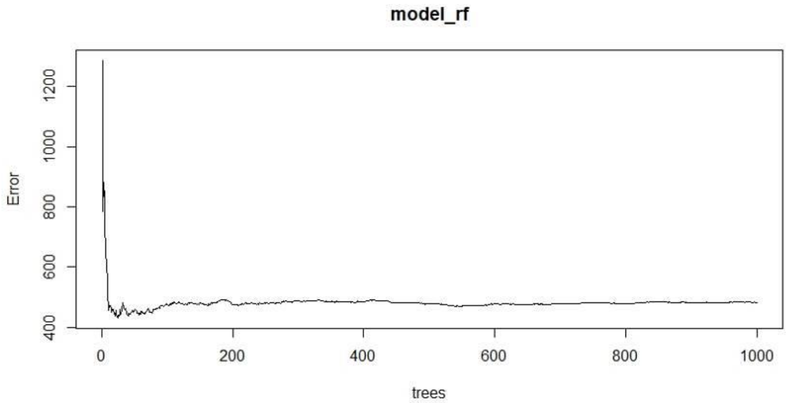


Figure 4: The error vs number of trees diagram for RF model

The statistical test results found from the RF model exhibit an effective performance for predicting $PM_{2.5}$ concentrations (Table 4). The RMSE, MAE values of the test dataset are little higher than the training and combined dataset because the performance of the model is relatively more error-prone to predict the test data. The correlation (R^2 value) of the predicted $PM_{2.5}$ concentration and actual concentration of these datasets are good for both the training and test set (> 0.80) which indicates this model is good enough to predict $PM_{2.5}$.

Table 4: The statistical test results of the RF model

Dataset	R^2 (Predicted vs Actual $PM_{2.5}$)	Root Mean Square Error (RMSE)	Mean Absolute Error (MAE)
Training Data	0.88	20.11	14.03
Test Data	0.80	22.75	15.88
Combined Data	0.86	20.81	14.50

3.3 Comparison Between Models

The statistical summary of the built models is shown in Table 5. As the R^2 values of these models are greater than 0.80, it clinches a strong relationship between the concentration of PM_{10} and $PM_{2.5}$. The higher R^2 and lower RMSE, MAE and RSE values are the criteria for a better model (Bhatt et al., 2014; Lefèvre et al., 2014; Willmott et al., 1985; Rouf et al., 2011).

Table 5: The statistical summary of the models

Model	R^2 (Predicted vs Actual $PM_{2.5}$)	Root Mean Square Error (RMSE)	Mean Absolute Error (MAE)	Residual Standard Error (RSE)	F value
Linear	0.89	18.06	13.05	0.32	1113***
Logarithmic	0.88	18.51	12.52	0.11	1011***
ANN	0.90	17.44	12.30	0.10	2341***
RF	0.86	20.81	14.50	0.14	1669***

*** is 'F' Significant at $Pr < 0.001$

In figure 5, both the graphs show the relationship between the measured PM_{10} values with the $PM_{2.5}$ values found from actual data and these mentioned models in dry and wet seasons, respectively. Several studies have recommended the ANN model over the linear regression model for determining the concentration of particulate materials (Özdemir and Taner, 2014; McKendry, 2002; Elbayoumi et al., 2015). The statistical results of the ANN model are analytically better compared to the other models. Among these models, the highest R^2 value is found in the ANN model, having a comparatively lower RMSE, MAE and RSE value. Thus, the proposed ANN model can be more effective for predicting $PM_{2.5}$ values than the other models for operational use.

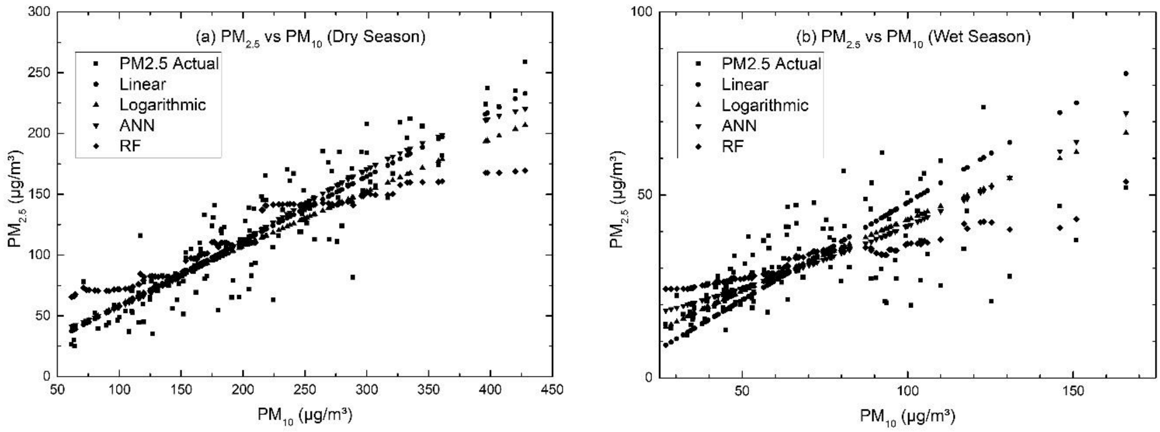


Figure 5: Prediction of $PM_{2.5}$ using different models in (a)dry and (b)wet season

3.4 $PM_{2.5}$ to PM_{10} Ratio

The $PM_{2.5}$ / PM_{10} ratio maintained a systematic pattern from season to season over the monitoring period (Rouf et al., 2011). Several studies on the concentration ratio of $PM_{2.5}$ to PM_{10} have revealed that the average ratio varies between 0.4 to 0.8, considering spatial and temporal factors. (Rouf et al., 2011; Chaloulakou et al., 2003; Harrison et al., 1997; Rajšić et al., 2004). The annual average ratios are higher in the dry season than the wet season in the three consecutive years, as shown in Table 6. Rouf et al. (2011) found an average ratio value of 0.64 during the dry season, whereas the ratio fell to 0.45 in the monsoon season due to meteorological factors, which supports our findings.

Table 6: Annual average $PM_{2.5}/PM_{10}$ ratio seasonal variation

Year	$PM_{2.5}/PM_{10}$ ratio	
	Dry season	Wet season
2016	0.58	0.47
2017	0.57	0.49
2018	0.53	0.47

Figure 6 present the relationship between PM_{10} and the ratio of $PM_{2.5}$ to PM_{10} in dry and wet seasons, respectively. One ratio value in figure 5 has found more than 1, which is not theoretically possible as $PM_{2.5}$ belongs to PM_{10}. Some errors such as measurement or instrumental faults are behind this ratio (Rouf et al., 2011). From the above analysis, the seasonal variance has a significant impact on the ratio of $PM_{2.5}$ to PM_{10} in Bangladesh.

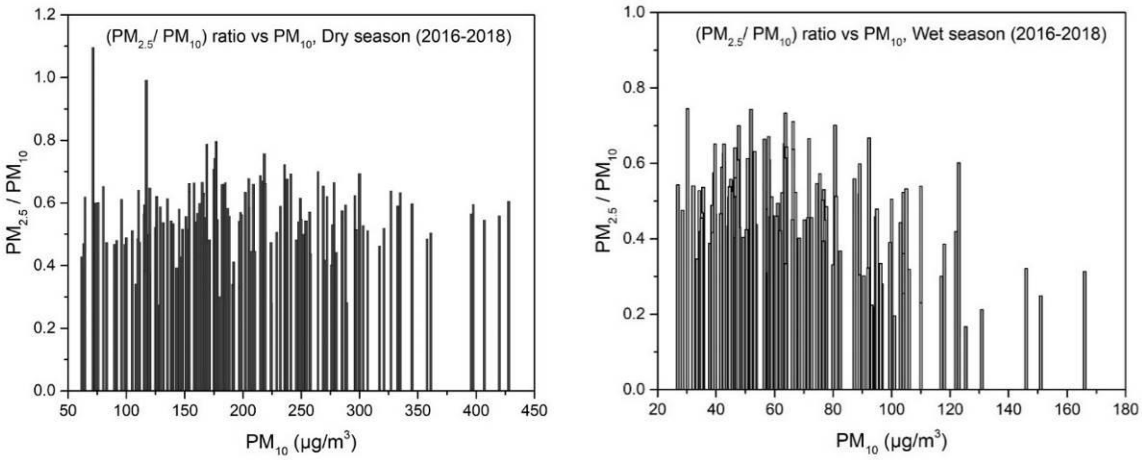


Figure 6: Relationship between PM_{10} and the ratio of $PM_{2.5}$ to PM_{10} in different seasons (2016-2018)

4. CONCLUSIONS

In this study, multiple regression (linear and logarithmic), ANN and RF mathematical models are used to forecast $PM_{2.5}$ concentration as a function of PM_{10} concentration using R statistical packages. The mean ratios of $PM_{2.5}$ to PM_{10} obtained from 11 Continuous Air Monitoring Station (CAMS) located in some major cities of Bangladesh are found slightly higher in the dry seasons than the wet seasons from 2016 to 2018. The impact of seasonal variation on PM concentration is found significant from the one-way ANOVA test results. In contrast, the regional effect is found negligible in the one-way ANOVA test results. All the models are statistically effective, but the ANN model performed slightly better in predicting $PM_{2.5}$ concentrations compared to other models. These models will help to estimate the actual concentration of $PM_{2.5}$ in the absence of exact concentration measurement amenities, which will eventually aid in detecting and lessening the pollution.

REFERENCES

Asadollahfardi, G., Zangooei, H., & Aria, S. H. (2016). Predicting $PM_{2.5}$ Concentrations Using Artificial Neural Networks and Markov Chain, a Case Study Karaj City. *Asian Journal of Atmospheric Environment*, *10*(2), 67–79. https://doi.org/10.5572/ajae.2016.10.2.067.

Begum, B. A. (2016). Dust Particle (PM_{10} and $PM_{2.5}$) Monitoring for Air Quality Assessment in Naryanganj and Munshiganj, Bangladesh. *Nuclear Science And Applications*, *25*(2). www.worldbank.org.bd/bds.

Bhatt, A.K., Pant, D., & Singh, R. (2014). An analysis of the performance of Artificial Neural Network technique for apple classification. *AI & society*, 29(1), 103-111

Chaloulakou, A., Kassomenos, P., Spyrellis, N., Demokritou, P., & Koutrakis, P. (2003). Measurements of PM_{10} and $PM_{2.5}$ particle concentrations in Athens, Greece. *Atmospheric Environment*, *37*(5), 649-660.

CM, S. (1979). Epidemiologic evidence and the United States air quality standards. *American Journal of Epidemiology*, *110*(6), 661–671. https://doi.org/10.1093/OXFORDJOURNALS.AJE.A112847

Dash, S.K. and Dash, A.K., (2015). Assessment of ambient air quality with reference to particulate matter (PM_{10} and $PM_{2.5}$) and gaseous (SO_2 and NO_2) pollutant near Bileipada, Joda area of Keonjhar, Odisha, India. *Pollut. Res*, *34*(4), 817-824.

Deka, P. C. (2019). *A Primer on Machine Learning Applications in Civil Engineering.* CRC Press.

DoE. (2018). Ambient Air Quality in Bangladesh. Clean Air and Sustainable Environment Project, Department of Environment, Ministry of Environment, Forest and Climate Change of the Government of Bangladesh. t.ly/iQ6o

Duan, J., Chen, Y., Fang, W., & Su, Z. (2015). Characteristics and relationship of PM, PM_{10}, $PM_{2.5}$ concentration in a polluted city in northern China. *Procedia Engineering*, *102*, 1150–1155. https://doi.org/10.1016/j.proeng.2015.01.239

Elbayoumi, M., Ramli, N.A. and Yusof, N.F.F.M., (2015). Development and comparison of regression models and feedforward backpropagation neural network models to predict seasonal indoor PM2. 5–10 and $PM_{2.5}$ concentrations in naturally ventilated schools. *Atmospheric Pollution Research*, *6*(6), 1013-1023.

Feng, X., Li, Q., Zhu, Y., Hou, J., Jin, L., & Wang, J. (2015). Artificial neural networks forecasting of $PM_{2.5}$ pollution using air mass trajectory based geographic model and wavelet transformation. *Atmospheric Environment*, *107*, 118-128.

Harrison, R. M., Deacon, A. R., Jones, M. R., & Appleby, R. S. (1997). Sources and processes affecting concentrations of PM_{10} and $PM_{2.5}$ particulate matter in Birmingham (UK). *Atmospheric environment*, *31*(24), 4103-4117.

Ivanov, A., Voynikova, D., Stoimenova, M., Gocheva-Ilieva, S., & Iliev, I. (2018). Random forests models of particulate matter PM_{10}: A case study. *AIP Conference Proceedings, 2025(1), 030001.* https://doi.org/10.1063/1.5064879

Lefèvre, S., Sun, C., Bajcsy, R., & Laugier, C. (2014, June). Comparison of parametric and non-parametric approaches for vehicle speed prediction. In *2014 American Control Conference* (pp. 3494-3499). IEEE.

McKendry, I. G. (2002). Evaluation of artificial neural networks for fine particulate pollution (PM_{10} and $PM_{2.5}$) forecasting. *Journal of the Air & Waste Management Association*, *52*(9), 1096-1101.

Nikolova, Y., Ilieva, N., & Sokolovski, E. (2015). Modelling of ambient air $PM_{2.5}$ concentration for air quality assessment. *Journal of Chemical Technology and Metallurgy*, *50*(1), 89-96.

Özdemir, U., & Taner, S. (2014). Impacts of meteorological factors on PM_{10}: artificial neural networks (ANN) and multiple linear regression (MLR) approaches. *Environmental Forensics*, *15*(4), 329-336.

Rajšić, S. F., Tasić, M. D., Novaković, V. T., & Tomašević, M. N. (2004). First assessment of the PM_{10} and $PM_{2.5}$ particulate level in the ambient air of Belgrade City. *Environmental Science and Pollution Research*, *11*(3), 158-164.

Rouf, M. A., Nasiruddin, M., Hossain, A. M. S., & Islam, M. S. (2011). Trend of Particulate Matter $PM_{2.5}$ and PM_{10} in Dhaka City. *Bangladesh Journal of Scientific and Industrial Research*, *46*(3), 389-398.

Valavanidis, A., Fiotakis, K., & Vlachogianni, T. (2008). Airborne particulate matter and human health: Toxicological assessment and importance of size and composition of particles for oxidative damage and carcinogenic mechanisms. *Journal of Environmental Science and Health - Part C Environmental Carcinogenesis and Ecotoxicology Reviews*, *26*(4), 339–362. https://doi.org/10.1080/10590500802494538

Voukantsis, D., Karatzas, K., Kukkonen, J., Räsänen, T., Karppinen, A., & Kolehmainen, M. (2011). Intercomparison of air quality data using principal component analysis, and forecasting of PM_{10} and $PM_{2.5}$ concentrations using artificial neural networks, in Thessaloniki and Helsinki. *Science of the Total Environment*, *409*(7), 1266-1276.

Willmott, C. J., Ackleson, S. G., Davis, R. E., Feddema, J. J., Klink, K. M., Legates, D. R., O'donnell, J. & Rowe, C. M. (1985). Statistics for the evaluation and comparison of models. *Journal of Geophysical Research: Oceans*, *90*(C5), 8995-9005.

Zhao, D., Chen, H., Yu, E., & Luo, T. (2019). $PM_{2.5}$ /PM_{10} ratios in eight economic regions and their relationship with meteorology in China. *Advances in Meteorology*, *2019*. https://doi.org/10.1155/2019/5295726

Zhou, X., Han, J., Zhao, Y., Wu, B., Cheng, G., Xu, S., & Jin, X. (2019). Characteristics of Vibration and Sound Radiation of Metro Resilient Wheel. *Chinese Journal of Mechanical Engineering (English Edition)*, *32*(1), 1–12. https://doi.org/10.1186/s10033-019-0383-1