ESTIMATION OF GROUND-LEVEL FINE PARTICULATE MATTER ($PM_{2.5}$) FROM ARTIFICIAL NEURAL NETWORKS USING METEOROLOGICAL PARAMETERS OVER MAJOR CAPITAL CITIES AT INDO-GANGETIC PLAIN (IGP)

Sumit Singh[1], Amarendra Singh[2], A.K. Shukla[3], Osama Siddiqei[4]

[1,4]M.Tech. Student, Environmental Engineering, Institute of Engineering and Technology, Lucknow, Uttar Pradesh, India
[2] Ph.D. Research Scholar, Department of Civil Engineering, Institute of Engineering and Technology, Lucknow, Uttar Pradesh, India
[3]Professor, Department of Civil Engineering, Institute of Engineering and Technology, Lucknow, Uttar Pradesh, India

ABSTRACT

Fine particulate matter ($PM_{2.5}$) has received a widespread attention all over the world because of its direct association with the degradation of air quality and the related human health effects. The regular monitoring of $PM_{2.5}$ concentrations is critically necessary to analyze their aftermath effects but the limited number of ground based air quality monitoring stations acts as a barrier in the evaluation of space-time dynamics of air pollution. This study is conducted to estimate the ground-level concentrations of fine particulate matter ($PM_{2.5}$) from statistical models developed using algorithms of artificial neural networks incorporating satellite measurements. A total of 12 ANN models were developed for three major cities of Indo-Gangetic Plain viz., Delhi, Lucknow and Patna using Moderate Resolution Imaging Spectroradiometer (MODIS) and INSAT-3D aerosol optical depth (AOD) to estimate the $PM_{2.5}$ concentrations from Jan-Dec 2019. The ANN models were trained using the AOD, meteorological parameters and $PM_{2.5}$ dataset of the previous two years (2017-2018) subdivided as train, validation and test groups in the ratio of 0.7, 0.15 and 0.15. High correlation values were obtained during the training of ANN models ($R \geq 0.7$ for each group). The well trained models provided good correlation coefficients obtained from the regression analysis of estimated and observed values accompanied with low RMSE and absolute percentage error. The ANN model developed using MODIS Aqua AOD for Patna showed highest correlation value of 0.85 followed by Delhi ($R = 0.70$) and Lucknow ($R = 0.66$). Although the degree of correlation varies for different sites, this study demonstrates the potential of ANN in air quality monitoring.

Keywords: -*AOD, $PM_{2.5}$, ANN, MODIS, INSAT-3D*

---

1. INTRODUCTION

Recent advancements in technology have improved the standard of living in the developing countries like India and China but at the same time the problem of pollution has became worse. Many epidemiological researches and studies have described that the fine particulate matter and many human health problems have a certain close relationship. The atmospheric particles which have aerodynamic diameter less than 2.5 µm also referred as $PM_{2.5}$ do not settle in the air and can leads to serious respiratory, cardiovascular disorders and even pre-mature deaths [1]. Several studies have also shown that the elevated concentrations of respirable particulate matter may leads to the increase in overall mortality rate and reduction in average life expectancy [2]. Therefore, $PM_{2.5}$ has gained a widespread attention in the past recent years and been a central point of atmospheric pollutants research. However, the limited number of ground based air quality monitoring stations acts as a barrier in the evaluation of space-time dynamics of air pollution [3].

In the recent years, with the launch of satellites and the continuous improvements in data retrieval technology, estimation of ground-level fine particulate matter concentrations from remote sensed AOD data has received considerable attention [4]. However, AOD alone cannot be used for the estimation of $PM_{2.5}$ values as AOD provides the measure of columnar aerosol loadings whereas $PM_{2.5}$ values are used to represent the mass concentrations of particulate matter at the near surface of earth. With incorporating various meteorological parameters the estimation can be improved to much extent [5]. Many researches and scientists have used the linear regression models for $PM_{2.5}$ estimation but there were certain limitations due to the non-linear relationship between the parameters. To overcome the limitations of non-linearity and complex relationship of predicting parameters, a new approach is required. Artificial neural network or artificial intelligence based models can provide better results because of their ability to handle complicated, extreme non linear data [6]. This study aimed to develop $PM_{2.5}$ predicting models based on artificial neural network algorithm using remote sensed satellite aerosol optical depth (AOD), land use data and several meteorological parameters.

2. MATERIALS AND METHODS

2.1 Study Area

Indo-Gangetic Plain (IGB) region is one of the most polluted regions in terms of fine particulate matter ($PM_{2.5}$) pollution in India. IGB region is always considered very important in terms of research interest because of its special and unique topographical nature and high population and industrial density [7]. The study area comprises three major capital cities located on the IGB region: Delhi, Lucknow and Patna as shown in Fig -1. The selected air quality monitoring stations of these cities are given in the Table -1.

Table -1: Geo-locations of the selected air-quality monitoring stations

State	City	Monitoring Station Name	Latitude (°N)	Longitude(°E)
Delhi	Delhi	NSIT Dwarka	28.61	77.03
Uttar Pradesh	Lucknow	Central School	26.84	81.00
Bihar	Patna	IGSC Planetarium Complex	25.61	85.13

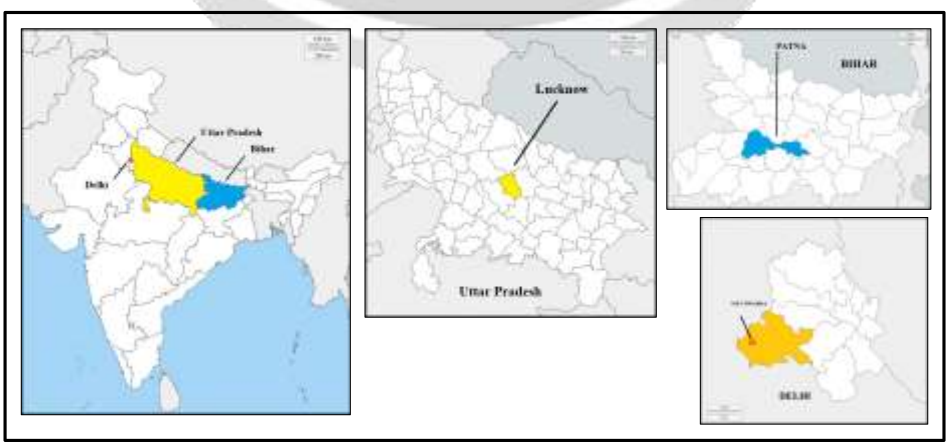

Fig -1: Study area locations – Delhi, Lucknow and Patna

2.2 Data Description and Collection

(a) Aerosol Optical Depth (AOD)

Aerosol Optical Depth (AOD) provides a quantitative measure of the aerosol loading (e.g., smoke particles, urban haze, sea salt, desert dust) in the atmosphere and it can be used as an estimate for ground-level particulate matter ($PM_{2.5}$ and PM_{10}). The Moderate Resolution Imaging Spectroradiometer (MODIS) is a payload imaging sensor worked by Santa Barbara Remote Sensing that was propelled into Earth circle by NASA in 1999 on board the Terra Satellite, and in 2002 on board the Aqua satellite operating at an altitude of 700 kilometers. These polar synchronous satellites measures reflected sun radiation as well as terrestrial emissions in 36 different channels at working wavelength range of 0.41-14.4 μm [8].In order to see the applicability of the developed models, MODIS Mean (average of Aqua and Terra AOD) is also included in the study. INSAT-3D is a geostationary weather satellite launched in 2013 by India at 82° East longitude provides Aerosol Optical Depth (AOD) data at 660 nm over both land and ocean. The INSAT-3D measures top of the atmosphere (TOA) radiance in six channels: three at infrared wavelengths, one in the visible wavelength and one sensitive to both solar and Earth-emitted radiance [9]. The daily AOD values were retrieved for the study locations for the time duration of Jan 2017 – Dec 2019.

(a) Meteorological Parameters and $PM_{2.5}$ data

The meteorological parameters used in the study are ambient temperature, relative humidity, wind speed, barometric pressure, and planetary boundary layer height. The first four parameters and the $PM_{2.5}$ concentration daily data is available on the CPCB website for all the monitoring stations selected in this study. The planetary boundary layer height (PBLH) and data is important in the dispersion analysis of particulate matter. PBLH monthly data is remotely sensed from the MERRA satellite of NASA and available on the Giovanni platform. In order to analyze the affect of land cover type on the AOD-$PM_{2.5}$ relationship, the Normalized Difference Vegetation Index (NDVI) vegetation product (MODIS 3) was also retrieved during the same period of the study.

2.3Artificial Neural Networks

ANN is regarded as a great and powerful model in handling many different types of information processing problems in the same way biological nervous system, such as brain processes the information. Back propagation algorithm is the most commonly used training algorithm in which the input training data is repeatedly sent to ANN [10]. The $PM_{2.5}$ estimation results can achieve better results if the training parameters such as number of hidden layers, neurons, learning rate, validations cycles etc. are well tuned and optimized. In this study, Back Propagation Artificial Neural Network (BPANN) is developed to predict the $PM_{2.5}$ concentrations. The overall dataset of the two years (2017-2018) was subcategorized randomly into three categories with 75% data for training, and 15% each for the validation and testing of the model. The neural network model used in the study is prepared by using neural toolbox of MATLAB developed by the MathWorks (Natick, Massachusetts). Once the training get completed, the input parameters for the year 2019 are fed to the network to forecast the $PM_{2.5}$ concentrations. The performance function used in the training of neural network was mean square error. The number of validation checks was limited to six and the most fitting well trained function was selected for the prediction of $PM_{2.5}$ concentrations. The Mean Square Error is calculated by using the Equation -1.

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(o_i - t_i)^2 \qquad (1)$$

Where N represents the number of training samples, *o* represents the current output and *t* represents the target (actual or true) value.

The schematic representation of the artificial neural network is shown in Fig -2. The essential components of any neural network are input layer, hidden layers, hidden neurons and an output layer. The fine tuning of neural networks depends on hit and trial method. The optimum number of hidden layers, hidden neurons, training function and learning rate are decided based on the outcomes of the initial model runs.

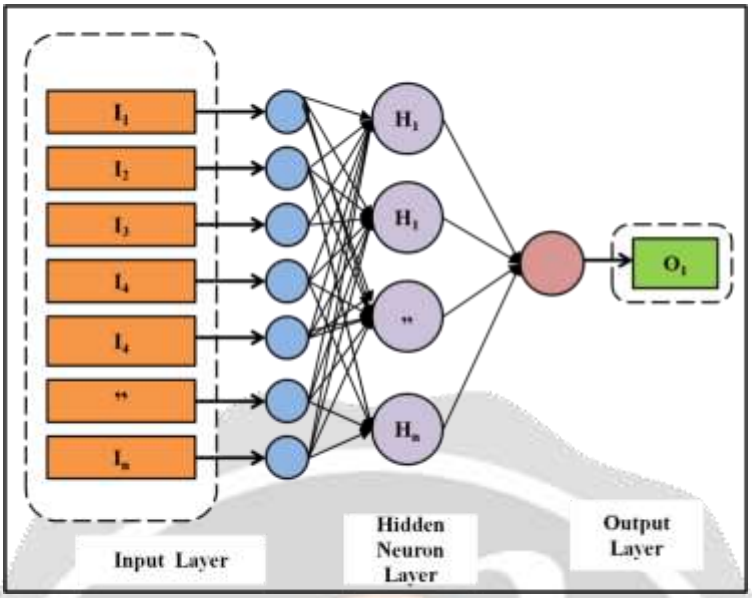

Fig -2: Schematic representation of Artificial Neural Network Model

3. RESULTS AND DISCUSSIONS

3.1 Performance Evaluation of Trained ANN models

Each artificial neural network model developed had an input layer containing seven nodes for the input parameters and hidden layer with six neurons and an output layer with single output node for $PM_{2.5}$ concentration. The neural networks developed were trained using the sigmoid transfer function based on error back propagation algorithm. The networks are trained using the two years (2017-2018) dataset of AOD, meteorological parameters and $PM_{2.5}$ concentrations. The dataset was first refined before being fed to the neural networks so as to make sure that the temporal resolution of every parameter was same. The matrix was prepared using data points of particularly selected days for which we had the all the available input and output variables available. The training results of the neural networks developed are summarized in Table -2.

Table -2: Regression analysis and performance evaluation of the ANN models

Station Name	AOD	No. of Samples	Correlation Coefficients				MSE
			Train	Validate	Test	Overall	
NSIT Dwaka, Delhi	Aqua	511	0.80	0.65	0.76	0.77	2525
	Terra	513	0.67	0.73	0.71	0.68	1668
	Mean	552	0.79	0.73	0.71	0.76	2905
	INSAT	536	0.80	0.75	0.78	0.79	2570
Central School, Lucknow	Aqua	372	0.76	0.73	0.77	0.76	1674
	Terra	403	0.82	0.80	0.74	0.80	1289
	Mean	429	0.81	0.70	0.62	0.76	942
	INSAT	424	0.81	0.78	0.60	0.77	1267
IGSC Planetarium Complex, Patna	Aqua	321	0.87	0.89	0.86	0.87	1709
	Terra	320	0.92	0.84	0.87	0.90	1589
	Mean	348	0.87	0.73	0.88	0.85	1376
	INSAT	374	0.89	0.87	0.84	0.88	1579

The sample size denote the available number days out of the two years for which the common dataset points are available. Since the daily AOD data is not available for all around the year the total number of sample pairs is less than 730 (365 × 2). The correlation coefficients obtained during the training of neural networks are tabulated in the Table 2. The high correlation values (~0.8) obtained denotes that the models are well trained and finely tuned for the estimation of $PM_{2.5}$ concentrations. The Mean Square Error started decreasing as the training moved forward and the minimum obtained value are mentioned. The ANN networks developed for the Patna location had the best performance index (R = 0.88) values while the Delhi's ANN models performance (R = 0.75) were slightly low as compared to the other two sites.

3.2 Evaluation of Estimated $PM_{2.5}$ Concentrations

The trained neural networks were provided with never seen dataset for the year 2019 to estimate the PM2.5 concentrations. The dataset was reconfigured before being fed to the trained network to make sure that the data matrix contained only common data points. Out of 365 days in the year 2019, the number of samples denote the days for which the predicting model provided the estimated values. The estimated values were linearly regressed with the ground truths and the performances including correlation coefficient, root mean square error and absolute percentage errors are tabulated in Table -3.

Table -3: Correlations, RMSE and absolute percentage error for estimated and observed $PM_{2.5}$ concentrations

Station Name	AOD	No of Samples	Correlation Coefficient	RMSE (µg m^{-3})	Absolute % Error
NSIT Dwaka, Delhi	Aqua	238	0.70	45.4	31
	Terra	243	0.70	45.6	27
	Mean	275	0.69	49.9	30
	INSAT	258	0.67	51.9	31
Central School, Lucknow	Aqua	219	0.65	49.0	41
	Terra	229	0.66	46.1	43
	Mean	253	0.66	45.6	38
	INSAT	246	0.56	57.7	65
IGSC Planetarium Complex, Patna	Aqua	215	0.87	44.2	26
	Terra	236	0.85	50.1	53
	Mean	249	0.83	53.9	35
	INSAT	236	0.80	61.3	40

For Delhi, the MODIS Terra AOD based ANN network provided best performance. The Pearson's correlation coefficient was 0.70 and had the lowest RSME value of 45.6 µg m^{-3} and absolute percentage error of 27%. In case of Lucknow, in spite of well trained model, the estimated values showed variation from observed values. The MODIS Mean AOD based ANN network provided a maximum correlation value of 0.66 with RMSE of 45.6 µg m^{-3} and absolute percentage error of 38%. For Patna, the ANN model provided promising results. The highest correlation value of 0.87 was provided by MODIS Aqua AOD based ANN with the lowest RMSE value of 44.2 µg m^{-3} and absolute percentage error of 26%

3.3 Statistical Analysis of the Estimated $PM_{2.5}$ Concentrations

The average $PM_{2.5}$ concentrations were calculated for the observed ground truths and the estimated values obtained from the developed neural network models. The results showed that most of the time ANN model overestimated against the lower ground level values. The average monthly estimated values were close to the observed ground-level values but the overall estimation suggests that the model based values were up scaled as shown in Fig. The statistical summary suggests that the inter-quartile range of the estimated and the observed $PM_{2.5}$ is same up to some extent for the ANN models developed using MODIS AOD, but for the INSAT-3D AOD the inter-quartile range is more and the values are more dispersed over greater range. This possible explanation for this may include the different operating wavelengths of the sensors (MODIS – 550 µm and INSAT – 660 µm). The results for both satellite AOD are in close proximity and the use of INSAT-3D can be suggested for the studies based on Indian subcontinent. The average monthly variation shows that the estimated values were quite close to the ground values around the year except during the monsoon season (July-September) possible due to the poor retrieval of AOD values because of the washout of tropospheric atmospheric particles.

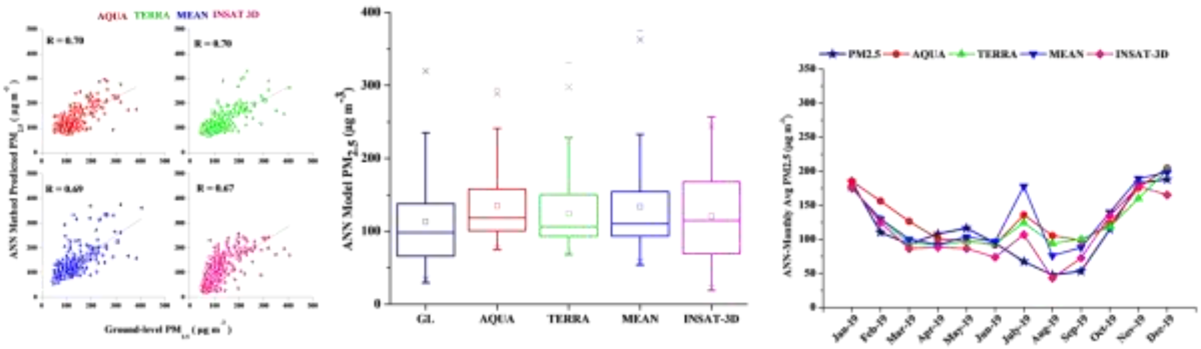

Fig -3: Correlation Analysis, Satistical Summary and Average Monthly Variation of Estimated and Observed $PM_{2.5}$ concentrations for Delhi

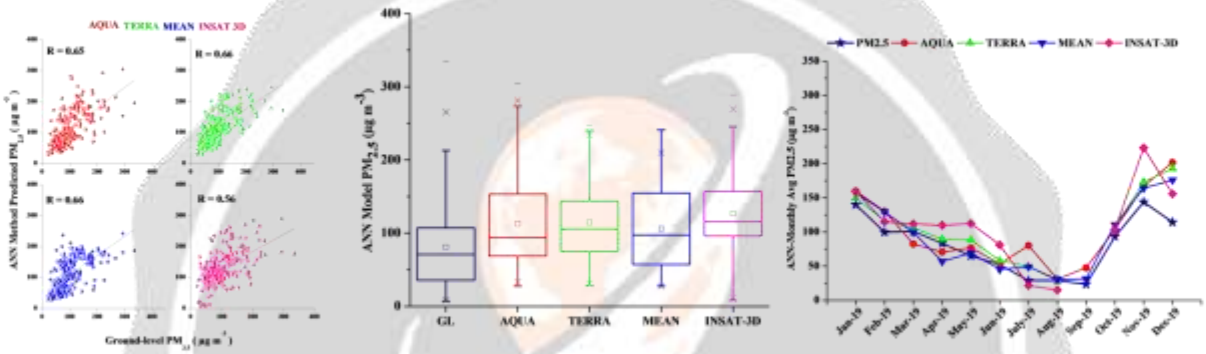

Fig -4: Correlation Analysis, Satistical Summary and Average Monthly Variation of Estimated and Observed $PM_{2.5}$ concentrations for Lucknow

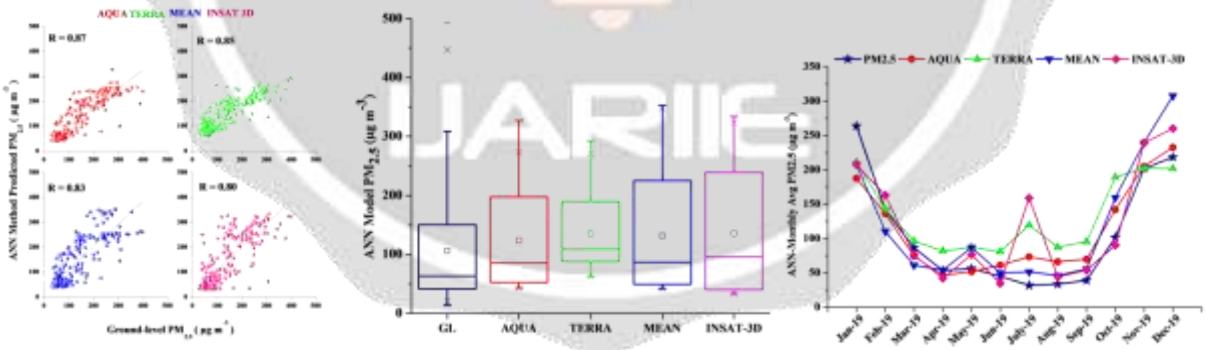

Fig -5: Correlation Analysis, Satistical Summary and Average Monthly Variation of Estimated and Observed $PM_{2.5}$ concentrations for Patna

The annual mean and the standard deviations are calculated for every estimating model provided values and are shown in Table -4. On closer analysis, it can be deduced that the minimum values for each estimating model is greater than the actual ground-observed values, but INSAT-3D values were much close to the actual minimum values for every location chosen. The annual average mean values of the predicted models showed a slight variation from the ground-level annual mean values. For Delhi, the variation lied between 6.2–19.2%, for Lucknow the variation increased from 31–56.2 % and for Patna the variation lied between 16.7–28%. However such variations are expected because there are so many uncertainties involved in the process including AOD retrievals and human error involved in the measurements of meteorological parameters and the $PM_{2.5}$ concentrations. The overall estimation shows that ANN is a powerful tool and can be used for the future prediction of ground level $PM_{2.5}$ concentrations.

Table -4: The statistical summary of the observed and estimated $PM_{2.5}$ concentrations

Station Name	$PM_{2.5}$	Minimum ($\mu g\ m^{-3}$)	Annual Average ($\mu g\ m^{-3}$)	Maximum ($\mu g\ m^{-3}$)
NSIT Dwaka, Delhi	Ground-level	29.5	113.3 ± 66	404.1
	Aqua	75.2	135.1 ± 46.2	295.8
	Terra	68.3	124.3 ± 47.2	330.4
	Mean	53.9	134 ± 61.4	374.7
	INSAT	19.1	120.4 ± 60	256.6
Central School, Lucknow	Ground-level	7.2	81 ± 56.7	334.7
	Aqua	27.2	112.6 ± 57.4	304.6
	Terra	28.8	114.4 ± 49.6	247.1
	Mean	27.7	106.1 ± 52.7	241.1
	INSAT	8.2	126.5 ± 52.8	289.2
IGSC Planetarium Complex, Patna	Ground-level	13.3	106.1 ± 91.7	397.7
	Aqua	42.8	123.9 ± 76.8	327.5
	Terra	61.8	135.9 ± 60.1	292.5
	Mean	41.7	131.7 ± 95	352.7
	INSAT	32.8	135.6 ± 100.7	334.3

4. CONCLUSIONS

We estimated the ground-level concentrations of fine particulate matter ($PM_{2.5}$) from Jan-19 to Dec-19 for three major cities viz., Delhi, Lucknow and Patna located on the most polluted region of Indo-Gangetic Plain using the satellite measurements from MODIS , INSAT-3D and meteorological parameters. The spatial and temporal variation of the estimated $PM_{2.5}$ concentrations showed a close agreement with the ground truths. Higher correlation values were obtained for those ANN models which were using MODIS data against INSAT-3D data. The average correlation coefficients obtained after regression analysis of estimated and observed values for Delhi, Lucknow and Patna were 0.69, 0.63 and 0.84 respectively. ANN showed its huge potential in handling complex data and establishing a unique fit function for the estimation of ground-level $PM_{2.5}$ concentrations.

5. REFERENCES

[1]. Bartell, S.M., Longhurst, J., Tjoa, T., Sioutas, C. and Delfino, R.J., 2013. Particulate air pollution, ambulatory heart rate variability, and cardiac arrhythmia in retirement community residents with coronary artery disease. Environmental health perspectives, 121(10), pp.1135-1141.

[2]. Chen, Z.H., Cheng, S.Y., Li, J.B., Guo, X.R., Wang, W.H. and Chen, D.S., 2008. Relationship between atmospheric pollution processes and synoptic pressure patterns in northern China. Atmospheric Environment, 42(24), pp.6078-6087.

[3]. Yao, L., Lu, N. and Jiang, S., 2012, May. Artificial neural network (ANN) for multi-source $PM_{2.5}$ estimation using surface, MODIS, and meteorological data. In 2012 International Conference on Biomedical Engineering and Biotechnology (pp. 1228-1231). IEEE

[4]. Ahmad, M., Alam, K., Tariq, S., Anwar, S., Nasir, J. and Mansha, M., 2019. Estimating fine particulate concentration using a combined approach of linear regression and artificial neural network. *Atmospheric Environment*, *219*, p.117050.

[5]. Gupta, P. and Christopher, S.A., 2009. Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: Multiple regression approach. Journal of Geophysical Research: Atmospheres, 114(D14).

[6]. Gupta, P. and Christopher, S.A., 2009. Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: 2. A neural network approach. Journal of Geophysical Research: Atmospheres, 114(D20).

[7]. Pawar, H., Garg, S., Kumar, V., Sachan, H., Arya, R., Sarkar, C., Chandra, B.P. and Sinha, B., 2015. Quantifying the contribution of long-range transport to Particulate Matter (PM) mass loadings at a suburban site in the North-Western Indo Gangetic Plain (IGP). Atmospheric Chemistry & Physics Discussions, 15(8).

[8]. Justice, C.O., Vermote, E., Townshend, J.R., Defries, R., Roy, D.P., Hall, D.K., Salomonson, V.V., Privette, J.L., Riggs, G., Strahler, A. and Lucht, W., 1998. The Moderate Resolution Imaging Spectroradiometer (MODIS): Land remote sensing for global change research. *IEEE transactions on geoscience and remote sensing*, *36*(4), pp.1228-1249.

[9]. Kumar, S., George, J.P., Sreevathsa, M.R. and Rani, S.I., 2016, May. Comparison of INSAT-3D AOD over Indian region with satellite-and ground-based measurements: a data assimilation perspective. In Remote Sensing of the Atmosphere, Clouds, and Precipitation VI (Vol. 9876, p. 98762A). International Society for Optics and Photonics.

[10]. Paola, J.D. and Schowengerdt, R.A., 1995. A review and analysis of backpropagation neural networks for classification of remotely-sensed multi-spectral imagery. International Journal of remote sensing, 16(16), pp.3033-3058