DOI: 10.5604/01.3001.0015.7027
Volume 110 • Issue 1 • January 2022

Journal of Achievements in Materials and Manufacturing Engineering

International Scientific Journal published monthly by the World Academy of Materials and Manufacturing Engineering

PM 2.5 modelling during paddy stubble burning months using artificial intelligence techniques

V. Sangwan *, S. Deswal
Department of Civil Engineering, National Institute of Technology Kurukshetra, India
* Corresponding e-mail address: vikramaditya_6180078@gmail.com
ORCID identifier: https://orcid.org/0000-0001-7285-1873 (V.S.);
https://orcid.org/0000-0001-7256-142X (S.D.)

ABSTRACT

Purpose: In this study, the artificial intelligence techniques namely Artificial Neural Network, Random Forest, and Support Vector Machine are employed for PM 2.5 modelling. The study is carried out in Rohtak city of India during paddy stubble burning months i.e., October and November. The different models are compared to check their respective efficacies and also sensitivity analysis is performed to know about the most vital parameter in PM 2.5 modelling.

Design/methodology/approach: The air pollution data of October and November months from the year 2016 to 2020 was collected for the study. The months of October and November are chosen as paddy stubble burning and major festivities using fireworks occur during these months. The untoward data entries viz. zero values, blank data, etc. were eliminated from the gathered data set and thereafter 231 observations of each parameter were left for the conduct of the presented study. The different models i.e., ANN, RF, SVM, etc. had PM 2.5 as an output variable while relative humidity, sulfur dioxide, nitrogen dioxide, nitric oxide, carbon monoxide, ozone, temperature, solar radiation, wind direction and wind speed acted as input variables. The prototypes created from the training data set are verified on the testing data set. A sensitivity analysis is also done to quantify impact of various parameters on output variable i.e., PM 2.5.

Findings: The performance of the SVM_RBF based model turned out to be the best with the performance parameters being the coefficient of determination, root mean square error, and mean absolute error. In the sensitivity test, sulphur dioxide (SO_2) was adjudged as the most vital variable.

Research limitations/implications: The quantification capacity of the generated models may go beyond the used data set of observations.

Practical implications: The artificial intelligence techniques provide precise estimation and forecasting of PM 2.5 in the air during paddy stubble burning months of October and November.

Originality/value: Unlike the past research work that focus on modelling of various air pollution parameters, this study in specific focuses on the modelling of most vital air pollutant i.e., PM 2.5 that too specifically during the paddy stubble burning months of October and November when the air pollution is at its peak in northern India.

Keywords: PM 2.5, ANN, Random Forest, SVM, Pollution, Burning, Stubble

Reference to this paper should be given in the following way:
V. Sangwan, S. Deswal, PM 2.5 modelling during paddy stubble burning months using artificial intelligence techniques, Journal of Achievements in Materials and Manufacturing Engineering 110/1 (2022) 16-26. DOI: https://doi.org/10.5604/01.3001.0015.7027

ANALYSIS AND MODELLING

1. Introduction

The rising air pollution has been a matter of grave concern since many years now. The menace of air pollution is usually caused by vehicles, industries, and fires, etc. The various studies conducted at different places across the globe successfully infer that every year a large number of living beings meet death in wake of toxicity in air [1,2]. The air pollutants leave a deadly impact on the health of living organisms is a well-established fact [3]. The major air pollutants include particulate matter of sizes 2.5 and 10 microns, carbon monoxide, sulphur dioxide, nitrogen dioxide, etc. [4,5]. Out of all the deadly air pollutants, one of the most hazardous pollutants is the particulate matter with the size up to 2.5 micrometres. The fatality and vastness of the PM 2.5 invite extensive research and higher degree of attention. The areas of south east Asia are well known for the cultivation of paddy. Due to several reasons the paddy growing farmers tend to burn the paddy stubble in-situ after harvesting. This trend of paddy stubble burning every year results in dangerous ascend of pollution in the air of corresponding regions [6]. And along with the other air pollutants, the PM 2.5 pollutant becomes a great matter of concern during the paddy stubble burning months of October and November. Moreover, the pollution emitted from vehicles, industries, etc. add to the toxicity of the air. In India the paddy harvesting season also witnesses many festivals viz. Dussehra, Diwali, etc., and the celebrations of these festivals are usually marked by the burning of the fire crackers. As the paddy is harvested at the advent of winter season, the low temperature traps the pollution from various sources within the ambit of the earth surface and hence makes it hard to the pollutants to escape into the atmosphere. This phenomenon creates a blanket of toxic smog over the surface of land. Thus, it becomes imperative in view of a safe planet to check and control the levels of air pollutants in the air especially the levels of major pollutant PM 2.5 [7].

There have been myriad studies in the recent years about the estimation and modelling of particulate matter of size 2.5 microns in the air [8]. These days the soft computing approaches have found their successful application in a wide range of fields such as air quality, hydrology, soil characteristics, etc. [9]. Therefore, after rigorous and thoughtful review of the feasibility and utility of the machine learning approaches in a variety of fields, it was decided to carry out the estimation of PM 2.5 concentration in Rohtak (Haryana), India during the months of paddy stubble burning by employing approaches viz. Random Forest (RF), Artificial Neural Network (ANN), and Support Vector Machine (SVM). The primary aim of the present research is to carry out the comparative study towards the utility of the mentioned artificial intelligence approaches i.e., ANN, RF and SVM in estimating the PM 2.5. Furthermore, a sensitivity test was also conducted in order to learn about the input variable to which the estimation of PM 2.5 in Rohtak is most sensitive. The city of Rohtak is counted among the most polluted areas not only in India but also in the world especially during the paddy stubble burning months. The study area is landlocked and has considerable exposure to the vehicular and industrial pollution. The northern parts of India get choked dangerously during the months of October and November, thereby leading to a number of health complications in the habitants of the region [10]. The selected soft computing methods are known for their feasible applications in addressing the irregularities of the environment.

2. Area of study

The study is conducted in Rohtak of Haryana state (India). The area of the study is situated on the latitude and longitude of 28.8909° N and 76.5796° E respectively. Rohtak is situated in the north-western region of India and the place is a part of national capital region, as it is nearby New Delhi. Air pollution is rampant in the study area as it is prone to paddy stubble burning, vehicular and industrial pollution. The satellite imagery of Rohtak is described in Figure 1 and is formulated by employing AWiFS.

Methodology and dataset

The air pollution data of Rohtak that is being utilized towards the research study was gathered on daily scale basis from CPCB (Central Pollution Control Board), India [11]. Only the data of October and November months from the year 2016 to 2020 was collected for the study. The months of October and November are chosen as paddy stubble burning and major festivities using fireworks occur during these months. The advent of cold weather further prevents

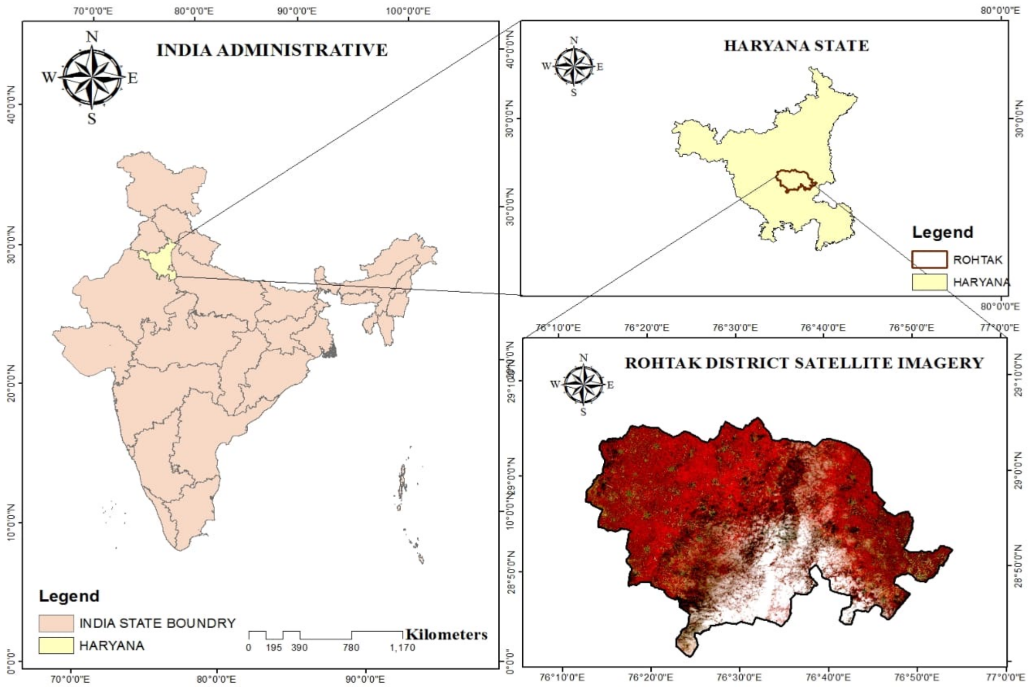

Fig. 1. Area of study (Rohtak, India)

Table 1.
Characteristics of utilized dataset

Parameters	Mean	Standard Error	Median	Standard Deviation	Kurtosis	Skewness	Minimum	Maximum
RH, %	45.2863	1.3973	56.7900	22.0486	-0.3384	-1.1809	0.3900	64.2500
SO_2, µg/m^3	14.0476	0.7171	10.4400	11.3155	11.1660	2.6911	0.0700	81.6300
NO, µg/m^3	12.3459	0.7082	9.2300	11.1750	8.5870	2.5352	0.1100	74.9300
NO_2, µg/m^3	34.8337	1.5560	31.6200	24.5528	0.4384	0.8724	0.3100	122.4100
CO, mg/m^3	0.9741	0.0414	0.8500	0.6539	0.2270	0.9381	0.0200	2.7900
O_3, µg/m^3	30.6248	0.9331	27.8900	14.7244	0.1664	0.4031	0.3400	85.3500
Temp., °C	21.3285	0.3708	20.3800	5.8517	0.7163	0.0095	3.0200	40.4900
SR, W/m^2	109.1401	4.2592	104.3600	67.2095	6.0042	1.8605	1.0400	428.0200
WD, degree	120.2453	2.7652	116.4200	43.6344	-0.2141	-0.1002	10.1000	234.8800
WS, m/s	0.8035	0.0369	0.6300	0.5823	5.6487	2.3166	0.0100	3.4700
PM 2.5, µg/m^3	149.1631	5.5461	138.8900	87.5160	4.0955	1.6061	23.6300	583.3300

the multi-source air pollutants especially the PM 2.5 from escaping into the atmosphere during these months. The untoward data entries viz. zero values, blank data, etc. was removed from the gathered data set and thereafter 231 observations of each parameter were left for the conduct of the presented study. The characteristics of the utilized data with varied units are tabulated in Table 1 in which Relative humidity (RH), sulfur dioxide (SO_2), nitrogen dioxide (NO_2), nitric oxide (NO), carbon monoxide (CO), ozone (O_3), temperature (T), solar radiation (SR), wind direction (WD) and wind speed (WS) are taken as input variables while Particulate matters (PM 2.5) is contemplated as output. In order to arrive at a unique scale for further processing, the data set was subjected to the process of normalization.

$$G = 0.1 + 0.9\frac{(G_i - G_s)}{(G_l - G_s)} \quad (1)$$

G_s and G_l denotes the smallest and largest values of different variables respectively, while the G_i denotes the value on which the normalization is carried out.

Table 2.
Correlation between PM 2.5 and input variables

Parameters	RH, %	SO_2, µg/m³	NO, µg/m³	NO_2, µg/m³	CO, mg/m³	Ozone, µg/m³	Temp., °C	SR, W/m²	WD, degree	WS, m/s	PM 2.5, µg/m³
RH, %	1.0000										
SO_2, µg/m³	0.1034	1.0000									
NO, µg/m³	0.2264	0.2714	1.0000								
NO_2, µg/m³	0.3601	0.2193	0.7171	1.0000							
CO, mg/m³	0.2345	-0.2624	-0.1518	-0.0895	1.0000						
O_3, µg/m³	0.0260	0.2709	0.3364	0.4026	-0.1504	1.0000					
Temp., °C	-0.1282	0.0822	-0.0168	-0.0970	-0.0095	-0.0129	1.0000				
SR, W/m²	0.6352	0.1066	0.2714	0.3418	0.0206	0.1978	0.0381	1.0000			
WD, degree	0.2953	0.2199	0.1404	0.0407	0.0843	-0.0356	0.1529	0.1558	1.0000		
WS, m/s	0.0163	-0.0124	0.1509	0.1899	-0.0652	0.1499	0.1053	0.2197	-0.0276	1.0000	
PM 2.5, µg/m³	-0.1506	0.0022	0.2067	0.1177	-0.1757	0.2171	-0.0234	0.0779	0.0043	0.0246	1.0000

Table 2 shows a bivariate correlation between PM 2.5 and the input variables using the entire dataset on a unique scale. It is important to note that a correlation between two variables does not imply causation.

A random separation of the normalized data was done in the portions of 70% and 30% for the purpose of training and testing respectively. This separation was done so as to use the larger data set (70%) for model development and smaller data set (30%) for model validation.

So as to quantify the reliability of the models created by employing artificial intelligence approaches, three different performance parameters namely Coefficient of Determination (R^2), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) were quantified. The expressions for these performance parameters are as follows:

$$R^2 = \frac{Q\Sigma NF-(\Sigma N)(\Sigma N)}{\sqrt{Q(\Sigma N^2)-(\Sigma N)^2}\sqrt{Q(\Sigma F^2)-(\Sigma F)^2}} \tag{2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{Q}(F_i-N_i)^2}{Q}} \tag{3}$$

$$MAE = \frac{1}{Q}\sum_{i=1}^{Q}|F_i - N_i| \tag{4}$$

In the above equations, F denotes the forecasted values; N denotes the noted values and Q denotes the quantity or number of values.

3. Artificial intelligence techniques

3.1. Random Forest

The random forest approach is an organized assembly of trees. In this machine learning algorithm, the tree predictors are generated from input vectors by employing random vector samples. Randomly selected input variables are employed to generate a tree. At each node of the tree, the distinct variables are organized [12]. In order to generate specific trees, a training data set is carved out of the arbitrarily chosen parameters. In contrast to the output the adulteration in the parameters is quantified with the help of Gini index [13]. The two user-defined variables that are vital for the random forest analysis include: In forest the trees that are to be grown and the input variables that are to be employed at nodes so as to generate a tree. On the basis of the arbitrary selection, the input variables are sorted out of the variables set through the best split [14]. The random forest approach forms trees to a great extent without trimming. The technique forms independent classes of trees by utilizing arbitrary specimens of data. This is termed as bagging. Thereafter, the most favourable class of trees is discovered through the process of voting. The regression technique performs precisely and efficiently when subjected to complex and vast dataset. The arbitrary sampling and collective approach are the main reasons behind the accurate and extensive predictions [15,16]. The RF approach is swift and effective in application but at times the running time performance may get compromised a bit.

3.2. Artificial Neural Network

The ANN approach is considered as a hugely popular and broadly practised machine learning approaches. It is usually employed for the statistical estimation and its performance have its basis in the mechanism of human nervous system [17]. In addition to the one or more hidden layers, there are output and input layers which are single and the complex link between these layers of output and input is modelled by employing the artificial neural network technique [18]. The input variables are in consistency with the quantity of nodes in the layers of input. The input film hardly assists the processing, though it disseminates the network data. The hidden layers may assist in data processing but usually the ultimate processing occurs in context of the output layer [19]. The veiled layers, neurons, coefficients of momentum, etc. are some of the parameters that administer the networks [20,21]. At times it becomes a very time demanding phenomenon to look out for factors viz. veiled layers, neurons, etc. in the network through trial-and-error method so as to optimally utilize the artificial neural network. The back propagation algorithm is employed in the current study.

3.3. Support Vector Machine

The SVM approach is derived from numeral theory of learning. It utilizes the vectors in a machine so as to categorize the sets of data into the scope of two proportions. The classification of the sets of data is usually executed by the machines, which use one portion of the dataset for training while other portion for the testing purpose [22]. It is very convenient for the vector machine to categorize the sets of data in two classes in an environment space of two dimensions by employing a number of lines and a specific super line [23]. The super line of the most appropriate character is at the largest distance from the border lines. In the given study, SVM is applied using two kernel functions i.e., radial basis function (SVM_RBF) and polynomial function (SVM_poly).

4. Results and discussion

4.1. Parameters set for modelling techniques

The phenomenon of development of models has its ground in the hit and trial method. In this study, modelling methodology involves two stages: training, and testing. The models were developed with training data and an unseen set of testing dataset is employed to judge the execution quality of developed prototypes. A plethora of trials were executed in order to search the optimum worth of user-based variables. Optimum values of user-defined parameters are listed in Table 3 below. The suitability of the modelling approaches is measured by the statistical measures for all the modelling stages. Figure 2 illustrates the trends of the input variables against PM 2.5 on a time-series plot at a unique scale.

Table 3.
Model-specific parameters for SVM (RBF & poly), RF and ANN

Modelling Techniques	Model-Specific Parameters
SVM_RBF	C = 0.3, Gamma = 10
SVM_poly	C = 0.3, Exponent = 3
RF	Features = 2, Trees = 90
ANN	Rate of learning = 0.3, Momentum = 0.2, Hidden film neuron = 10, Iterations = 200

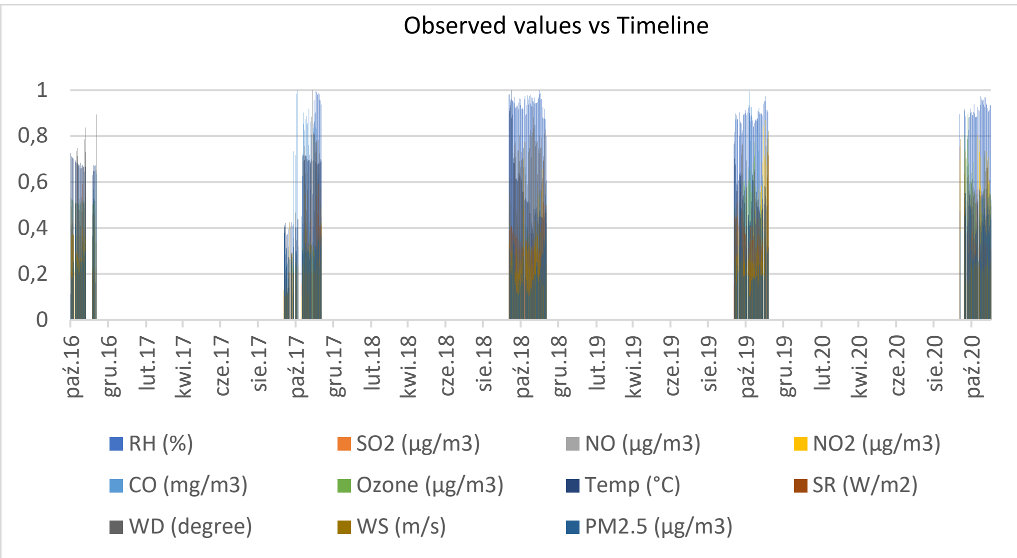

Fig. 2. PM 2.5 and input variables on time-series plot

4.2. ANN results

The model calibrated with the model-specific parameters for ANN is depicted in Table 3. The main parameters involve adjustment of rate of learning, momentum, quantity of hidden film neurons, and iterations. The observed output data of PM 2.5 is plotted against the predicted data of PM 2.5 for the training and testing phases in Figure 3 and Figure 4, respectively. The reliability of created models is quantified in the light of three statistical parameters: R^2, RMSE, and MAE. The fitting of the predicted data with the line of perfect prediction is good in training (Fig. 3) as compared to testing (Fig. 4). The attained values of R^2, RMSE and MAE during training are 0.6595, 0.0668, and 0.0858, respectively; while during testing are 0.3861, 0.0752, and 0.1019, respectively. The model is working fine with training dataset as compared to the testing dataset.

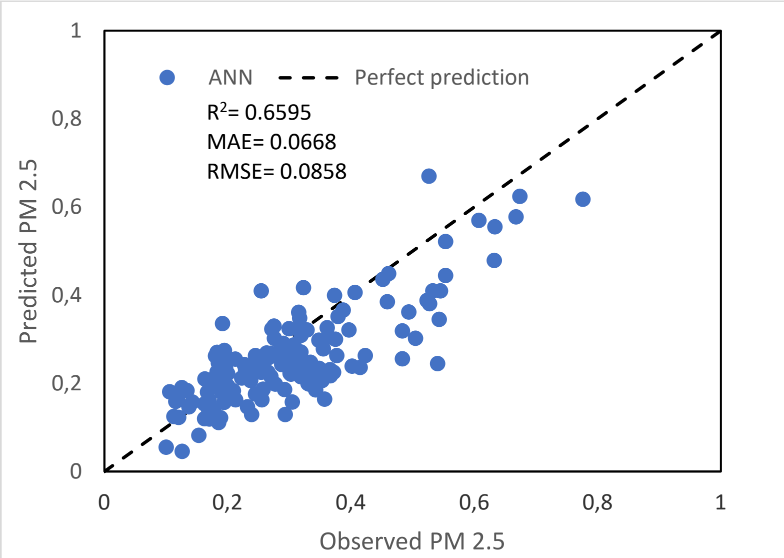

Fig. 3. Predicted vs Observed PM 2.5 training data employing ANN

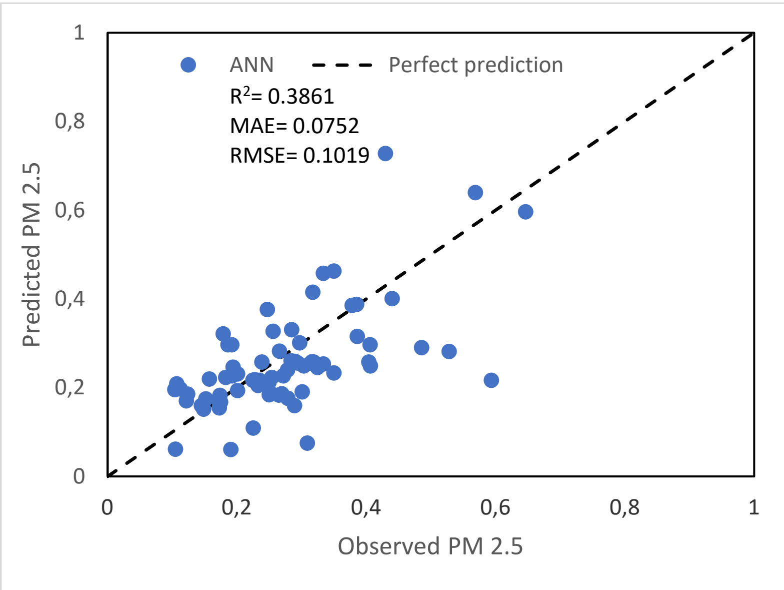

Fig. 4. Predicted vs Observed PM 2.5 testing data employing ANN

4.3. RF results

The tuning parameters for RF prototype are quantity of trees, and total characteristics which need to be adjusted by judging the performances of the trained models on the testing data sets. The adjusted parameters for the model are given in Table 3. The graphs of training (Fig. 5) and testing data (Fig. 6) sets are plotted for the predicted values of PM2.5 against the observed values of PM 2.5. The predicted training data shows good agreement with the real PM 2.5 data and resides closer to the line of perfect prediction (Fig. 5), while on the other hand, the predicted data does not fit perfectly with the observed data (Fig. 6) and scatters more from the line of perfect prediction. After judging the performance in accordance with the statistical variables, the training results (R^2=0.9677, RMSE=0.0372 and MAE=0.0270) are far better than the testing results R^2=0.5378, RMSE=0.0818 and MAE=0.0647) of the RF model.

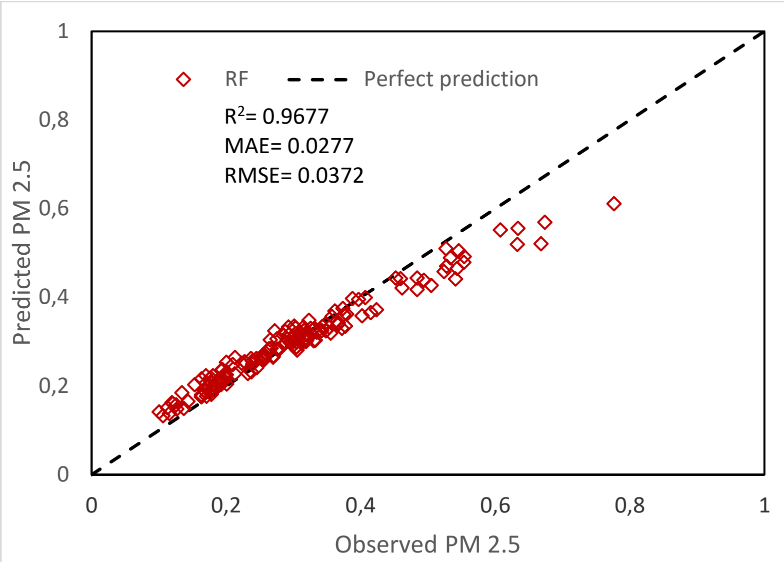

Fig. 5. Predicted vs Observed PM 2.5 training data employing RF

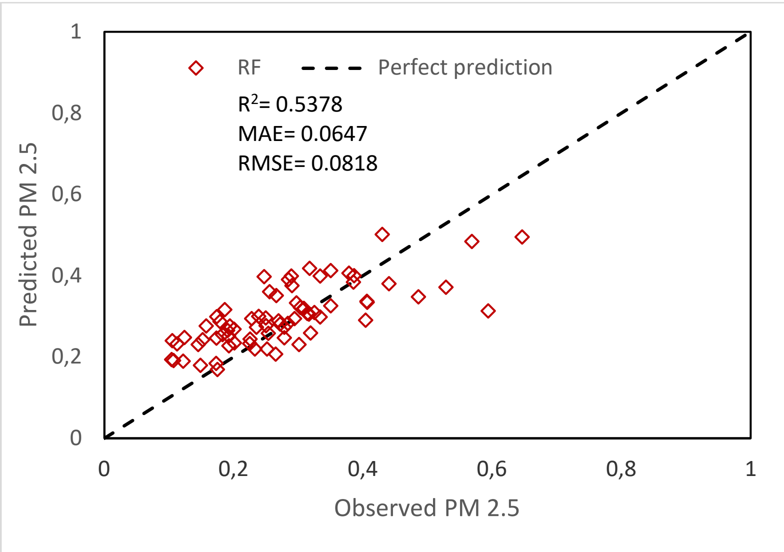

Fig. 6. Predicted vs Observed PM 2.5 testing data employing RF

4.4. SVM results

Two kernel functions employed with the SVM are Polynomial (poly) kernel and Radial Basis Function (RBF) kernel. The training and testing results are given Figure 7, and Figure 8, respectively. The performance evaluation parameters for the trained and tested models are provided in Table 4 and 5, respectively. The accomplishment of the RBF function trained model is comparatively better than the poly function-based model when implemented on the testing data sets (Tab. 5). From the evaluation of the statistical parameters (Tab. 5) during testing, SVM_RBF model (R^2=0.6135, RMSE=0.0733 and MAE=0.0559) outperforms the SVM_poly model (R^2=0.3130, RMSE=0.0974, and MAE=0.0732) in predicting the observed PM2.5 values.

4.5. Comparing models

The agreement plot of the estimated data of PM2.5 using SVM, RF and ANN is depicted in Figure 9. SVM_RBF is working better as compared to SVM_poly, so RBF kernel function in SVM is plotted against the RF and ANN models.

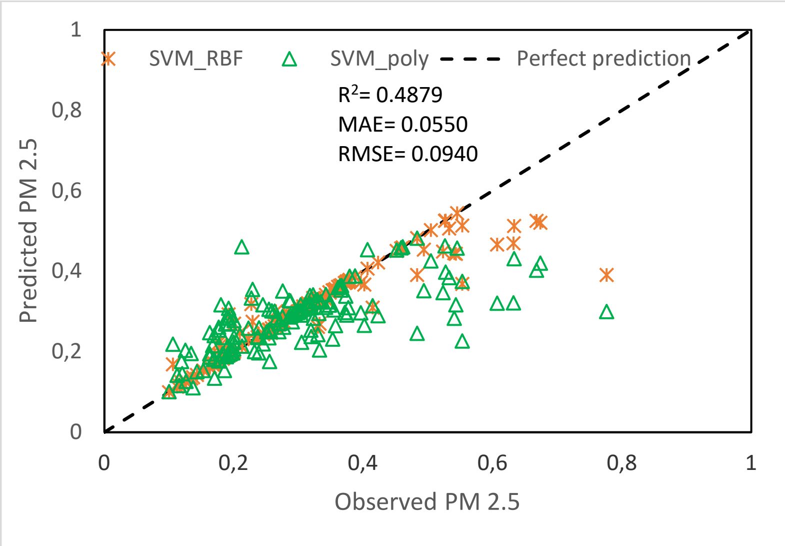

Fig. 7. Predicted vs observed PM 2.5 training data employing SVM_RBF and SVM_poly

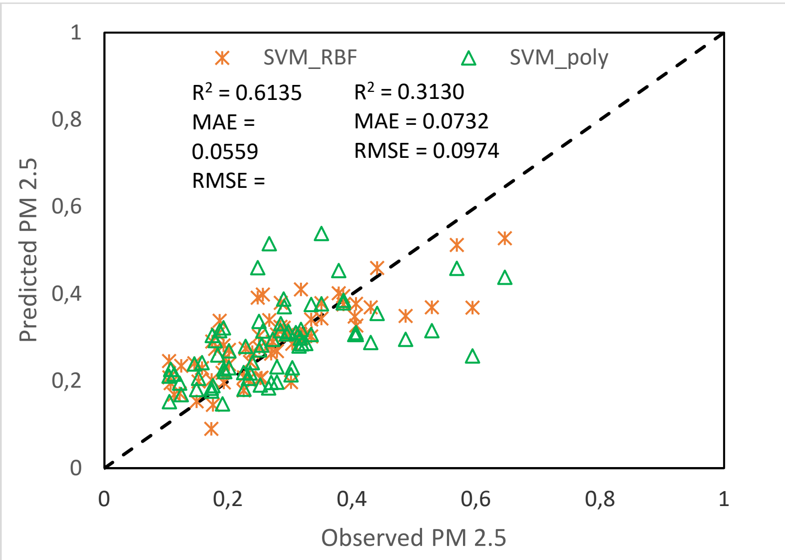

Fig. 8. Predicted vs observed PM 2.5 testing data employing SVM_RBF and SVM_poly

Table 4.
Assessment of performance for SVM (RBF & poly), RF and ANN models (Training)

Model (train)	SVM_RBF	SVM_poly	RF	ANN
R^2	0.8797	0.4879	0.9677	0.6595
MAE	0.0174	0.0550	0.0270	0.0668
RMSE	0.0490	0.0944	0.0372	0.0858

Table 5.
Assessment of performance for SVM (RBF & poly), RF and ANN models (Testing)

Model (test)	SVM_RBF	SVM_poly	RF	ANN
R^2	0.6135	0.3130	0.5378	0.3861
MAE	0.0559	0.0732	0.0647	0.0752
RMSE	0.0733	0.0974	0.0818	0.1019

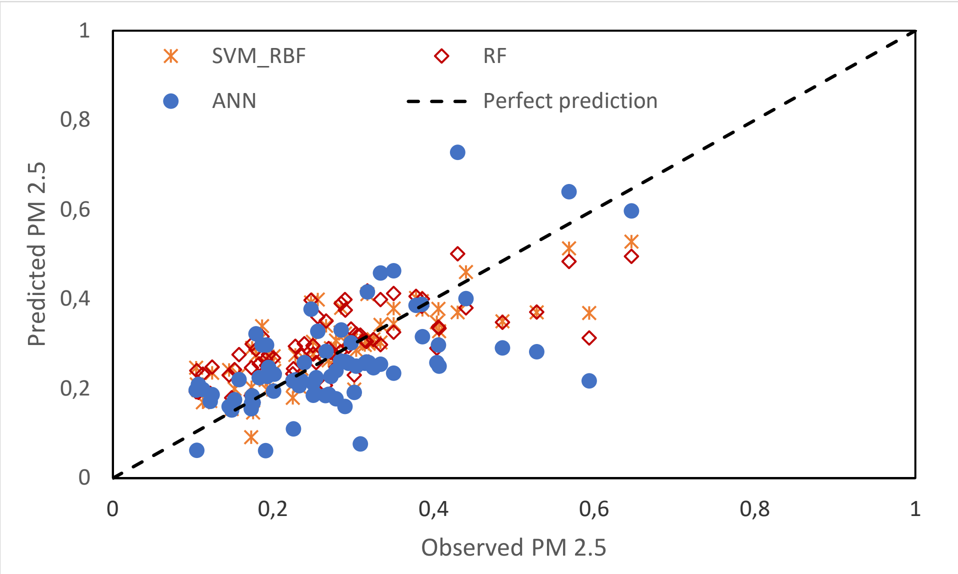

Fig. 9. Comparison plot of PM 2.5 data (Testing) employing SVM_RBF, RF and ANN

The plot conveys information regarding the scattering of the predicted data with reference to the agreement line using different models. From the agreement plot of the testing data, the modelling efficacy of SVM using RBF as kernel function is superior in contrast to RF and ANN based models. The scattering of the predicted data using SVM models is relatively less as compared to the other models. Table 5 confirms this statement as the error values (RMSE and MAE) are relatively less than the RF and ANN models. Hence SVM models using RBF kernel function outperforms RF and ANN models. The performance of the modelling tools (SVM or any other modelling tool) is dependent on the dataset. The RBF based SVM works well for the dataset of this study. The model efficacy is nowhere connected with the level of pollution, PM 2.5, etc. The dataset in entirety rather than the variations in pollution play a role in model efficacy [4].

4.6. Sensitivity analysis

The sensitivity analysis results clearly state that there is no great fluctuation in the models after omitting specific input parameters on the prediction of the observed output data. The sensitivity analysis of PM 2.5 is quantified taking into account the behaviour of each parameter whilst employing the blend of different input parameters towards the prediction of PM 2.5 during the paddy stubble burning season. This analysis is conducted by employing the best approach of this study i.e., SVM_RBF. The impact of each variable on the sensitivity of PM2.5 is discovered by observing the corresponding values of Coefficient of Determination (R^2) and Root Mean Square Error (RMSE) as depicted in Table 6. It may be noted that the analysis implies that sulphur dioxide (SO_2) [R^2=0.48] is the most important parameter. The lowest value of coefficient of determination (R^2) authenticates SO_2 as the most important parameter in the sensitivity analysis. The other sulphur oxides (SO_x) are formed as a result of high concentrations of SO_2 in the atmosphere. The reactions of these sulphur oxides with different compounds in the atmosphere leads to the formation of particulate matter (PM) pollutants. The weather conditions impact the particulate matter (PM 2.5, PM 10, etc.) concentration, which in turn impacts the entire dynamics of the atmosphere [15].

Table 6.
PM 2.5 sensitivity test (Input Combination (SVM_RBF)

Parameter removed	Training		Testing	
	R^2	RMSE	R^2	RMSE
-	0.88	0.0490	0.78	0.0733
RH, %	0.85	0.0527	0.58	0.0753
SO_2, μg/m^3	0.82	0.0594	0.48	0.0839
NO, μg/m^3	0.86	0.0535	0.59	0.0745
NO_2, μg/m^3	0.84	0.0560	0.55	0.0800
CO, mg/m^3	0.84	0.0550	0.65	0.0693
Ozone, μg/ m^3	0.82	0.0581	0.53	0.0795
Temp., °C	0.86	0.0520	0.61	0.0733
SR, W/m^2	0.80	0.0611	0.57	0.0762
WD, degree	0.84	0.0538	0.53	0.0794
WS, m/s	0.88	0.0484	0.58	0.0761

5. Conclusions

The presented study had the target of estimating PM 2.5 during the paddy stubble burning months i.e., October and November. In order to meet the objective of precise forecasting of PM 2.5, varied models were created by employing artificial intelligence techniques namely ANN, RF and SVM (SVM_RBF and SVM_poly). The study concludes that out of all the developed models, the reliability, efficiency and performance of SVM_RBF driven model is better than the rest of the created models. The SVM_RBF model for PM 2.5 estimation has the values of the performance parameters as R^2=0.61, RMSE=0.073 and MAE=0.055. The sensitivity analysis further indicated that sulphur dioxide (SO_2) is the most vital parameter that influences the estimation PM 2.5. The suggested model can be significantly used in dealing with the threat of air pollution especially PM 2.5 during paddy stubble burning months i.e., October and November.

References

[1] J. Yang, Q. Yu, P. Gong, Quantifying air pollution removal by green roofs in Chicago, Atmospheric Environment 42/31 (2008) 7266-7273. DOI: https://doi.org/10.1016/j.atmosenv.2008.07.003

[2] Y.-F. Xing, Y.-H. Xu, M.-H. Shi, Y.-X. Lian, The impact of PM 2. 5 on the human respiratory system, Journal of Thoracic Disease 8/1 (2016) E69-E74. DOI: https://doi.org/10.3978/j.issn.2072-1439.2016.01.19

[3] A.G. Barnett, G.M. Williams, J. Schwartz, A.H. Neller, T.L. Best, A.L. Petroeschevsky, R.W. Simpson, Air pollution and child respiratory health: a case-crossover study in Australia and New Zealand, American Journal of Respiratory and Critical Care Medicine 171/11 (2005) 1272-1278.
DOI: https://doi.org/10.1164/rccm.200411-1586OC

[4] P.A. Aguilera, A. Fernández, R. Fernández, R. Rumí, A. Salmerón, Bayesian networks in environmental modelling, Environmental Modelling and Software 26/12 (2011) 1376-1388.
DOI: https://doi.org/10.1016/j.envsoft.2011.06.004

[5] Minidisk Infiltrometer User's Manual, Decagon Devices, Inc., Pullman, WA, USA, 2014.

[6] C.E. Rasmussen, C.K. Williams, Gaussian processes for machine learning, Vol. 1, MIT Press, Cambridge, Massachusetts, 2006.

[7] H. Wati, P. Lestari, A. Sofyan, A Contribution black carbon (BC) in PM2. 5 from rice straw open burning in district Cianjur, West Java, Indonesia, Proceeding of the 4th Asian Academic Society International Conference "AASIC", Nakhon Pathom, 2016, 575-583.

[8] A. Parsaie, H. Yonesi, S. Najafian, Prediction of flow discharge in compound open channels using adaptive neuro fuzzy inference system method, Flow Measurement and Instrumentation 54 (2017) 288-297. DOI: https://doi.org/10.1016/j.flowmeasinst.2016.08.013

[9] A.H. Zaji, H. Bonakdari, B. Gharabaghi, Reservoir water level forecasting using group method of data handling, Acta Geophysica 66/4 (2018) 717-730. DOI: https://doi.org/10.1007/s11600-018-0168-4

[10] U. Schlink, O. Herbarth, M. Richter, S. Dorling, G. Nunnari, G. Cawley, E. Pelikan, Statistical models to assess the health effects and to forecast ground-level ozone, Environmental Modelling and Software 21/4 (2006) 547-558.
DOI: https://doi.org/10.1016/j.envsoft.2004.12.002

[11] Central Pollution Control Board: central control room for air quality management – al India. Available online: https://app.cpcbccr.com/ccr/#/caaqm-dashboard-all/caaqm-landing/data (accessed in: 22 March 2021)

[12] H. Emami, M. Sorafa, M.R. Neyshabouri, Evaluation of hydraulic conductivity at inflection point of soil moisture characteristic curve as a matching point for some soil unsaturated hydraulic conductivity models, Journal of Science and Technology of Agriculture and Natural Resources 15/59 (2012) 169-182.

[13] L. Breiman, Random Forests – Random Features, Technical Report 567, University of California Berkeley, 1999, 1-28.

[14] L. Breiman, J. Friedman, C.J. Stone, R.A. Olshen, Classification and regression trees, CRC Press, Boca Raton, Florida, 1984.

[15] R.D. Peng, F. Dominici, R. Pastor-Barriuso, S.L. Zeger, J.M. Samet, Seasonal analyses of air pollution and mortality in 100 US cities, American Journal of Epidemiology 161/6 (2005) 585-594. DOI: https://doi.org/10.1093/aje/kwi075

[16] G.D. Rolph, R.R. Draxler, A.F. Stein, A. Taylor, M.G. Ruminski, S. Kondragunta, P.M. Davidson, Descrition and verification of the NOAA smoke forecasting system: the 2007 fire season, Weather and Forecasting 24/2 (2009) 361-378. DOI: https://doi.org/10.1175/2008WAF2222165.1

[17] R. Gupta, Causes of emissions from agricultural residue burning in north-west India: evaluation of a technology policy response, South Asian Network For Development And Environmental Economics, 2012.

[18] A.F. Mashaly, A.A. Alazba, Assessing the accuracy of ANN, ANFIS, and MR techniques in forecasting productivity of an inclined passive solar still in a hot, arid environment, Water SA 45/2 (2019) 239-250. DOI: https://doi.org/10.4314/wsa.v45i2.11

[19] R.S. Govindaraju, Task committee on application of artificial neural networks in hydrology, artificial neural networks in hydrology. ii: hydrologic application, Journal of Hydrologic Engineering 5/2 (2000) 124-136. DOI: https://doi.org/10.1061/(ASCE)1084-0699(2000)5:2(124)

[20] A.J. Smola, B. Schölkopf, A tutorial on support vector regression. Statistics and Computing 14/3 (2004) 199-222. DOI: https://doi.org/10.1023/B:STCO.0000035301.49549.88

[21] D.P. Solomatine, Y. Xue, M5 model trees and neural networks: application to flood forecasting in the upper reach of the Huai River in China, Journal of Hydrologic Engineering 9/6 (2004) 491-501. DOI: https://doi.org/10.1061/(ASCE)1084-0699(2004)9:6(491)

[22] C. Cortes, V. Vapnik, Support vector networks, Machine Learning 20 (1995) 273-297. DOI: https://doi.org/10.1007/BF00994018

[23] S. Heddam, Modelling hourly dissolved oxygen concentration (DO) using dynamic evolving neural-fuzzy inference system (DENFIS)-based approach: case study of Klamath River at Miller Island Boat Ramp, OR, USA, Environmental Science and Pollution Research 21/15 (2014) 9212-9227. DOI: https://doi.org/10.1007/s11356-014-2842-7