Correlations of Particulate Matter and Associated Weather Factors in the Capital City of the State of Bihar, India: An Analysis and Prediction

Anand Shankar[1,2] , Ashish Kumar[2], Deepak Kumar Singh[2],Uttam Kumar[2], Neha Kumari[2], Saurabh Kumar[2], Kunal Kaushik[2], Dr. Dharma Raju[3], Vivek Sinha[4] & Sandeep Kumar2

[1]Department of Electronics and Communication Engineering
National Institute of Technology, Patna, India
[2] India Meteorological Department, Ministry of Earth Sciences,
Government of India, Patna, India
[3] India Meteorological Department, Ministry of Earth Sciences,
Government of India, Hyderabad, India
[4] India Meteorological Department, Ministry of Earth Sciences,
Government of India, New Delhi, India
Email: anand.shankar@imd.gov.in

ABSTRACT

Fine particulate matter (PM2.5) is the most important environmental risk factor, requiring routine monitoring and analysis for effective management of air quality. Exposure to high levels of PM2.5 can have significant impacts on human health, such as aggravating asthma, causing respiratory problems, and increasing the risk of lung cancer. High levels of PM2.5 can affect visibility and reduce air quality, which can in turn impact weather conditions. For example, during periods of high PM2.5 concentrations, there may be increased haze and smog in the air, which can reduce visibility and make it difficult to see objects at a distance. In some cases, this can also result in lower solar radiation and, as a result, cooler temperatures. This work used a machine-learning approach to predict PM2.5 and examine the association between PM2.5, a variety of contributing factors, trend analysis, and their temporal variations based on air quality data and meteorological data for the metropolitan city of Patna for the period 2016 to 2023. The results show that PM2.5 concentration predictions can be made using the random forest model. In this model, the PM2.5 concentration is significantly affected by the visibility, mean sea level pressure, CO, O3, relative humidity, wind speed, dew point temperature, etc., but there is only a weak link between these parameters. From 2016 to 2023, the data showed a persistence in PM2.5 pollution levels, and the data also revealed substantial variations in PM2.5 concentration and its fluctuations over the different months. The objective of the analysis is to take a close look at the impacts of weather on air pollution in the capital city of Bihar. This type of analysis may be carried out in other cities as well. This research could help air pollution management programmers in Patna, the state capital, as well as all cities lying in the pollution-prone Indo-Gangetic Plains regions.

Keywords: *Air Pollution, PM2.5, Meteorological factors, Random Forest.*

1. Introduction

Air pollution has become a worldwide problem that hurts the environment and makes people sick all over the world. In recent years, as industrialization and urbanization have rapidly progressed, polluting gases from fuel combustion and fugitive dust (Gupta et al. 2022) from traffic and construction have caused frequent occurrences of haze or smog globally under unfavorable climatic circumstances of diffusion. People consider air pollution to be one of the great killers of our time because it is hazardous to their health (Z. Sun and Zhu 2019). Most developing countries, like India, have worsening air quality every year (Swarna Priya and Sathya 2019). Nearly 1,800 people die every day in developing cities because of the dirty air (Autrup 2010; Remoundou and Koundouri 2009). About 90% of deaths from air pollution happen in countries with low or middle incomes. PM2.5 is one of the most significant pollutants in haze-polluted areas (Westervelt et al. 2016). The death rate from air pollution shows that life expectancy drops by nearly three years on average (Taneja et al. 2017). It hinders not only economic growth and has negative effects on people's health (Sharma, Chandra, and Kota 2020), as well as making it more difficult for people to go around. Efforts are currently being made all around the world to better control PM2.5. An additional crucial aim is the efficient control of PM2.5. Analyzing air pollutants and meteorological parameters closely connected with PM2.5 is vital for successful control of the pollutant.

There have been a lot of studies done to try to figure out how to stop and regulate air pollution. Socioeconomic and climatic variables, as well as the presence and quantity of other pollutants, all play a role in determining PM2.5 concentrations. Because to its atmospheric origin, PM2.5 is sensitive to variations in temperature, humidity, and wind speed. PM2.5 is impacted by the same external variables as other anthropogenic pollutants. Because of this, academics have investigated the weather and pollution relationship. (Cifuentes et al. 2021) used statistical models and showed that sun radiation and temperature were the most important factors. Wind and surface turbulence were shown to be particularly sensitive to PM2.5 levels, as discovered by (Park et al. 2021). The dramatic drop in PM2.5 values was driven more by synoptic than local factors, (X. Li et al. 2021) discovered that weather conditions are associated with daily variations in PM2.5 concentration. Seasonal and regional variations in the impact of weather on PM2.5 concentration were observed by (Chen et al. 2018). When compared to other climatic parameters, temperature, humidity, and wind speed had the greatest impact on PM2.5 concentrations. The effects of weather and human activity antecedents on PM2.5 were shown to vary significantly throughout time and space, as discovered by (Jing et al. 2020). According to the work of (Zheng et al. 2019), PM10, SO2, NO2, and CO are the primary factors impacting the concentration of PM2.5, whereas meteorological conditions and O3 are secondary contributors. Using long-term air quality data, (Mingzhi 2017) identified climate, NO2, and O3 as greater causes of PM2.5, while (Licheng Zhang et al. 2020) used a variety of statistical techniques to assess regional and seasonal changes in PM2.5 concentrations. Based on their analysis of the effects of typical severe weather conditions on PM2.5 in Tianjin, (Shao et al. 2021) found that increases in wind speed and decrease in planetary boundary height increases the PM2.5 concentration, with inversion having the greatest impact. Research must therefore incorporate other air pollutants, such as SO2, CO, O3, NO2, and PM10, as well as climatic variables like temperature, wind direction and speed, rainfall, and humidity, in order to provide more precise predictions of PM2.5 concentrations. It is of major scientific importance to investigate an accurate PM2.5 concentration prediction model due to the inherent difficulty in doing so due to the wide variety of factors that might affect PM2.5 concentration. Inaccurate lower boundary conditions, approximation of physical parameters, and a lack of a perfect initial state are just a few of the problems with a PM2.5 concentration forward prediction model based on physical principles, beginning with meteorological elements and pollution circumstances (Cheng et al. 2021). At the same time, as computing power has increased, interest in data-driven statistical approaches has grown. There is a lot of interest in machine learning because of the benefits it offers in automatically refining algorithms via experience (Lei Zhang et al. 2021). Random forests and neural networks, two examples of the more common types of nonlinear machine learning models, have shown promising predictive performance (Delavar et al. 2019). One cannot just apply the integration algorithm as a machine learning algorithm. To accomplish a goal, it constructs and integrates many machine learners. The decision trees in a random forest are all independent yet work together as an ensemble to make predictions (Sadorsky 2021). The neural network approach differs from the standard parametric model approach in that it is a data-driven adaptive strategy that makes no assumptions about the underlying problem model. Neurons can acquire the latent functional correlations between the inputs through training and learning (Lee 2020), even when the underlying rules for issue resolution are unclear. It works well with issues that have sufficient data and observed variables but are difficult to explain using hypotheses and established theories. Machine learning's exceptional learning capabilities has made it increasingly popular for PM2.5 forecasting. Using a deep neural network model, (Wang and Sun 2019) reduced the estimation bias caused by insufficient AOD (aerosol optical depth) by predicting the PM2.5 concentration in the missing AOD (aerosol optical depth) area using data on gaseous pollutants (NO_2, SO_2, CO, and O_3). An effective random forest model for assessing ground PM2.5 was created by (Yang, Xu, and Yu 2020), which took into account reflectance, meteorological field, and land use variables. When discussing PM2.5, certain weather conditions, and land use variables, and also underlined the importance of ground-level issues. (Haiming and Xiaoxiao 2013) chose PM10, sulphur dioxide, nitrogen dioxide, temperature, pressure, humidity, wind direction, and wind speed as potential influencer. Radial basis function (RBF) neural network based models were utilized to make PM2.5 forecasts. The findings demonstrated the model's usefulness. (Zheng et al. 2019) combined gaseous pollution and meteorological parameters for a more all-encompassing forecasting system. To forecast the 24-hour PM2.5 concentration, (Shi, Fang, and Ni 2021) suggested a neural network technique based on the attention mechanism. Based on measures of root-mean-square error (RMSE) and mean absolute error (MAE), he concluded that the model was more accurate in its predictions. In a recent study (Lu et al. 2021) suggests, PM2.5 concentrations are affected by a wide variety of social, economic, meteorological and the interaction between pollutants factors.

To deal with PM2.5 air pollution forecasts with sufficient accuracy, (Du et al. 2021) created a hybrid deep learning architecture combining one-dimensional convolutional neural networks and bidirectional long short-term memory networks. By putting four machine learning models through their paces using standard of analysis and cross-validation, (Czernecki, Marosz, and Jędruszkiewicz 2021) proved the high applicability of machine learning to short-term air quality prediction. The aforementioned research concentrated on improving the current model to improve prediction accuracy and performance without considering the model's interpretability or the many components that contribute to PM2.5.

Only a small number of PM2.5 studies (Kumar et al. 2020) have been conducted in the state of Bihar, India. The state of Bihar, India, experiences the subtropical monsoon, a mild and dry winter, and a hot summer, with annual temperature ranges of 1°C to 49.5°. Intensive agriculture has been the primary focus of development. Winter haze is com-mon due to the geographical location and the widespread practice of burning straw outdoors in the region's rural communities. When it comes to air pollution, the capital city of Patna is indicative of other major cities in this region. The public may quickly and easily assess the present state of PM2.5 pollution in Patna and gain a deeper and more intuitive grasp of the state of the city's air quality. Bihar State Pollution Control Board (BSPCB) decision-making bodies can use this information as a foundation for more precise air pollution control efforts. The formulation of urban development plans and the maintenance of sustainable economic development are of the utmost importance. The PM2.5 trend was also addressed in this research, along with variations of PM2.5 (diurnal, monthly, etc.). India has established a number of public awareness programs and policies to reduce pollution (MoEFCC 2019). The impact of air quality data and meteorological data on PM2.5 concentration changes was investigated, and their respective contributions to these changes were quantified using the random forest model. Using the capital city's air quality and meteorological data from 2016–2023 as predictors and PM2.5 concentration as the outcome, a prediction model was developed. Using the SHAP technique, this model identified the most important elements in determining PM2.5 concentrations and assessed the impact of each factor. Each factor's relationship to PM2.5 was calculated using the Pearson technique. Variations and trends in PM2.5 concentration between and within years were evaluated, and several potential causes of such variation in the capital city of Patna were looked into. This can help with the state’s air pollution control and air quality management by providing both theoretical and data support.

The remainder of this paper is organized as follows: In Section 2, the study areas, the observed dataset, and the proposed methodology are presented. The extensive assessment and the discussion of the results are provided in Section 3. Finally, we

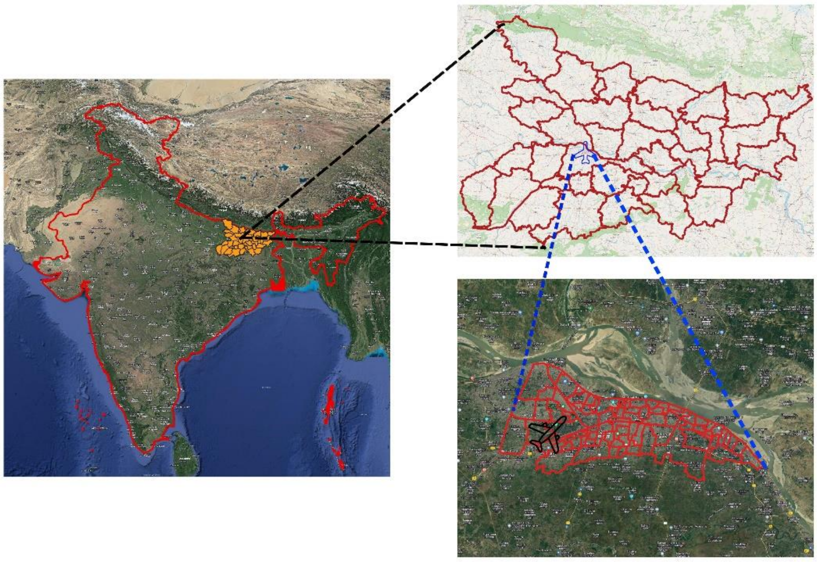

Figure 1: The geographical location of the Capital City of Patna (a) India (b) State of Bihar, India (c) Capital Cities of Patna.

Table 1. The presentation of the data used in the analysis and the prediction of PM 2.5.

Type	Name	Unit	Value Range	Source
Air Quality Data	PM2.5	$\mu g/m^3$	3.1-1049	Bihar State Pollution Contro Board,Govt. of Bihar
	NO_2	$\mu g/m^3$	1-328.2	
	CO	mg/m^3	0-26.8	
	SO2	$\mu g/m^3$	1-1568	
	O3	$\mu g/m^3$	1-778.3	
Meteorological Data	Dry Bulb Temperature	°C	4.6- 44.4	India Meteorological Department
	Dew Point Temperature	°C	0.1- 39.4	
	Relative Humidity	%	7.7-100	
	Wind Speed	knots	0-30	
	Present Weather	coded	0-99	
	Past Weather	coded	--	
	Station Level Pressure	hPa	--	

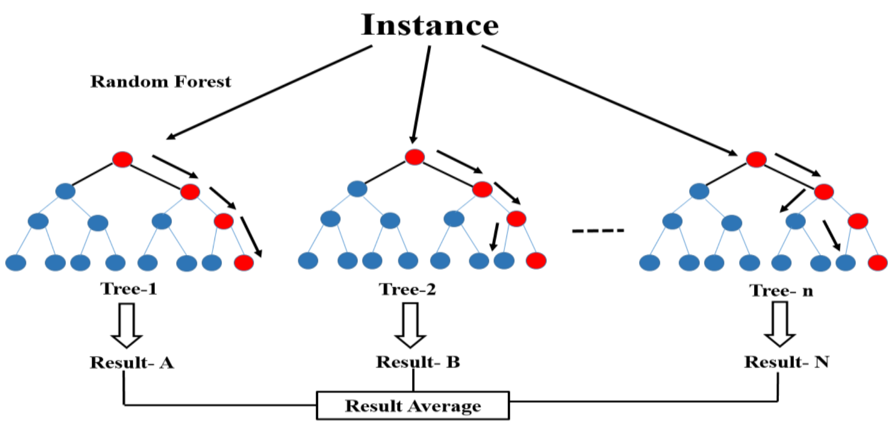

Figure 2: Schematic diagram of the random forest principle.

provide conclusions in Section 4. At the end, a list of acronyms is provided in Table A1 to make it easier to read articles.

2. Materials and Methods

2.1 Data

Air quality and meteorological data were analyzed for this investigation. The hourly air quality data shared by the Bihar State Pollution Control Board and corresponding weather data collected from the state's India Meteorological Department were used for the analysis and prediction of PM 2.5. Air quality and corresponding meteorological data were measured for the period 2016–2023 (CO, mg/m3; other pollutants, PM2.5, PM10, NO2, CO, SO2, and O3 μg/m3) and corresponding hourly meteorological parameters from Patna Airport (Dry Bulb Temperatures (°C), Dew Point Temperature (°C), Past and Present Weather(Code), Relative Humidity (RH%), Pressure (hPa), Average Wind Speed (Knots) etc.) were collected for the studied area presented in Fig. 1. Table 1 displays the details of the data information that was used in this analysis.

2.2 Random Forest Prediction Model

Bagging and RF (Breiman 2001) are two representative parallelization approaches among ensemble-learning algorithms in which individual learners do not have a substantial dependence on each other, they can be made at the same time. Bagging works by first employing the bootstrap approach to select a subset of training samples from a larger dataset, then using those examples to train a relatively inexperienced learner, before finally combining the trained learners into a single one. Both the classification and regression tasks contribute to the final result by voting on the output of the prediction. RF is a more extensive form of bagging. The basic algorithmic concept is depicted in Fig.2.

RF employs a decision tree trained with the classification and regression tree (CART) algorithm as a weak learner and includes a random selection of characteristics in the training process. The typical decision tree uses a node's best feature (out of N possible characteristics) to split the tree into left and right branches. However, RF picks a feature to divide the decision tree's left and right branches at random from among Nsub (Nsub < N) sample features on the node. The model's applicability is thus expanded even further. In each iteration of bagging's random sampling process, about 36.8% of the training data is left out of the kth tree's creation. We refer to these as kth tree out-of-bag samples. These additional data are not part of the modelling process but can be used to check the accuracy of the model.

In conclusion, RF constructs a single regression decision subtree via the bootstrap method and a random selection of F-characteristics for node splitting. The aforementioned steps are repeated numerous times to build T regression decision subtrees, and then each tree in the resulting random forest is allowed to develop naturally without being trimmed. The final forecast is the average of all the sample-training decision trees. Fig.3 is a flowchart depicting the algorithm for the random forest. Due to its ability to handle high-dimensional data and immunity to over fitting, the random forest technique has become increasingly popular. And it gets good results for default value problems while still providing an objective estimate of the significance of each attribute. Training with this method can be performed in a very parallel fashion. It's fast for training huge samples, quite flexible across datasets, and accurate in its predictions.

In our research, we performed a grid search to identify the best model parameters for achieving the best prediction. In order to find the optimal combination of settings for a given problem, the grid search approach iteratively cycles through all of the available parameters. Table 2 shows the explanations and settings of several of the most important random forest parameters utilized in this investigation; the remaining parameters were left at their default levels.

2.3 Data Analysis Method

The degree to which people are able to comprehend the rationale behind their choices is referred to as explain ability. The foundation of machine learning is an algorithm that, given data, seeks out potential patterns and relationships, and ultimately, creates

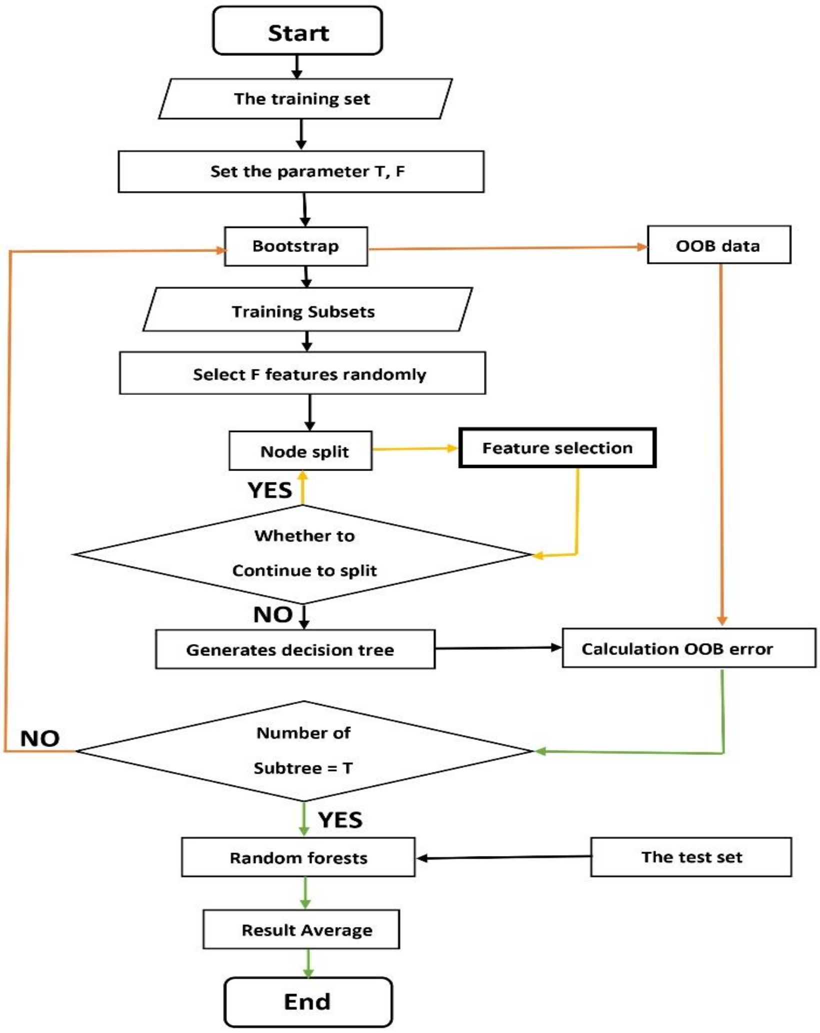

Figure 3: Flowchart of the proposed random forest regression models

Table 2. Best parameters achieved during the grid search of random forest regression models

Name	**Meaning**	**Values**
N_estimators	Number of tress in the forest	200
max_features	Number of features to consider when looking for the best split	sqrt
Max_depth	Maximum depth of the tree	10
bootstrap	Bootstrap samples are used when building trees.	True
criterion	Measure the quality of a split	mse
Oob_score	Whether or not the generalisation score should be estimated using out-of-bag samples.	True
Random_state	Adjusts how many replicates are used for bootstrapping the samples used to construct trees and how many features are considered when determining the optimal split at each node.	20

judgments or predictions based on those findings. People will have an easier time comprehending the reasoning behind particular choices or forecasts to the extent that the phenomenon in question is explicable. Not only are people pleased with the results of the model, but they are also thinking more about the factors that contribute to those results. This kind of thinking assists in the optimization of the model and its characteristics, and it also has the potential to assist in better comprehending the model itself and improving the model's overall quality.

The Shapley value (Lundberg and Lee 2017) served as inspiration for the additive explanation model known as SHAP. The purpose of this approach is to compute the contribution of each feature to the prediction of an instance X in order to explain the prediction of that instance. SHAP assigns the output values to the Shapely values that are associated with each feature, and the feature values of a data instance serve as "contributors" in this context. Measures of the contributions that each feature makes to a machine-learning model are referred to as shapely values. The following is how the Shapley value for feature Xj in the model should be interpreted (Ziqi Li 2022):

$$Shapely(X_j) = \sum_{S \subseteq N \setminus \{j\}} \frac{k!(p-k-1)!}{p!} (f(S \cup \{j\} - f(s)) \qquad (1)$$

where p is the number of features, N\{j} is the set of all possible combinations of the features except X_j, S is the set of features in N\{j},f(S) is the model prediction using features from S, and f (S∪{j}) is the model prediction using features from S and Xj. Shapley value of a feature is its marginal contribution to the model prediction averaged over all possible models with different permutations of features, as indicated by the interpretation of Equation (1). (Lundberg and Lee 2017) developed SHAP because they recognised that the complexity of computing Shapley values was a major barrier to their widespread use. To quantify the impact of features on the final output value, it computes the Shapley value of each feature value and provides the following justification:

$$g(z') = \emptyset_0 + \sum_{j=1}^{M} \emptyset_j \, Z_j' \qquad (2)$$

where g is the explanatory model, M is the number of input features is the typical mean of the target variable across all samples $z' \in \{0,1\}^M$s the simplified features and indicates whether the corresponding feature exists (1 or 0), $\Phi_j \in$ R is the feature attribution for feature j, the Shapely values, and Φ_0 is the feature attribution for feature j. For example, with the X we're discussing, every single feature value is "present" (1 for each of the simplified features). The preceding formula can now be written more simply as:

$$g(z') = \emptyset_0 + \sum_{j-1}^{M} \emptyset_j \qquad (3)$$

That which shifts the expected result from the mean to the predicted result, namely g(z'), can be thought of as the contribution of the total of the Shapley values of each feature. The benefit of SHAP is that it makes it evident if a given attribute aids or hinders the prediction. The SHAP package in Python was used for the study presented in this paper. The current analysis's random forest prediction model is explained by using this library.

3. Results and Discussion

The investigation results of PM2.5 over the capital cities of Patna and the results of predicted models are presented in this section.

3.1 Time Series Analysis of PM2.5

Here, we detail the results of our study into the dynamics between meteorological and pollution variables and PM2.5 concentration across time. Diurnal, monthly, and annual variations are used to describe them. The variation in the PM2.5 levels for a location is a complex interplay of emissions, environmental factors, such as geography and meteorological factors (Alimissis et al. 2018; Ganguly et al. 2019; Nair et al. 2007).

3.1.1. Diurnal Variation of PM2.5

Fig. 4 displays the PM2.5 monthly diurnal variation in the capital cities of Patna averaged over the more than seven years (2016 to march 2023),the figure displays the diurnal mean PM2.5 readings .The primary peak of PM2.5 concentrations is observed around 0230 hours in the night, and a secondary peak is observed at 1200 hours in the afternoon. Season and location can cause a change of up to two hours in the morning peak hours. Late winter sunrises and the start of human activity push cities' winter peaks later in the day than their summer counterparts. Unlike in developed countries, the diurnal variation in PM2.5

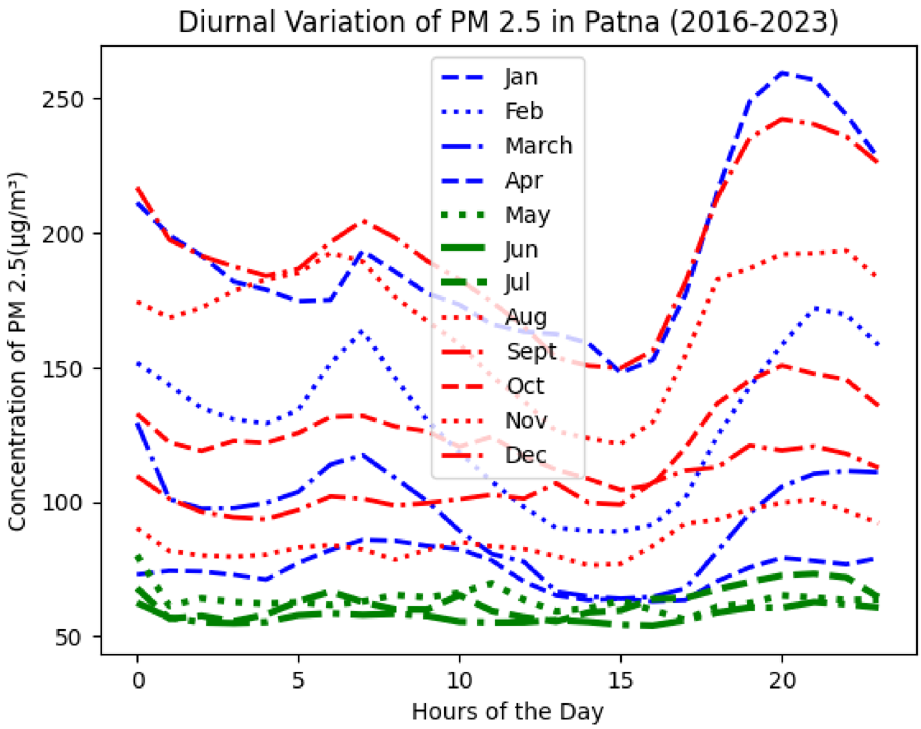

Figure 4: Diurnal variation of PM2.5.

concentrations in our region is not solely driven by transportation or industrial activities; instead, it is the atmospheric conditions, particularly the PBL layer and low-level wind that determine the concentration of particulate matter. The same diurnal pattern is observed in all four months, i.e., November, December, January, and February, suggesting that the same mechanism governs the concentration, transportation, and dispersal of PM2.5 during these months. In the absence of strong convective heating during the winter months, PBL height starts falling early in the afternoon, and the lowest level is attained sometime in mid-night around 0230 hrs. This results in the confinement of available particulate matter to a smaller area, thus increasing its concentration. In terms of intra-seasonal variation, the PBL height can come down to less than 500 meters during peak winter, compared to 3–4 kilometers during peak summer afternoon; thus, even when the absolute amount of PM2.5 remains the same, there will be a very large variation in concentration value between summer and winter months.

3.1.2. Monthly Variation of PM2.5

PM2.5 follows the same general pattern as other polluting emissions like NO2, CO, and O3. Since baseline PM2.5 values remain high during the winter due to persistent atmospheric conditions (Sreekanth, Niranjan, and Madhavan 2007; Tiwari et al. 2013; Tyagi et al. 2017) and increased emissions (Guo et al. 2017, 2019; Schnell et al. 2018), the maximum values are reached during this time of year. It is usually very high during the November to February. But the corresponding rainfall had a significant impact on the monthly concentration of PM2.5 (Shown in Fig.5). It is usually very high during the post-monsoon and winter seasons, which last from November to February. But the correlated rainfall had a significant impact on the monthly concentration of PM2.5. As of the post-monsoon season (October to December) and winter season (January to February), both have low rainfall. So, the highest concentration of PM2.5 was reported in these months. Due to wet scavenging and washout by rain during the South-West monsoon, PM2.5 levels are lowest during the monsoon months (JJAS) (Singh, Singh, and Biswal 2021). During later parts of the monsoon season, if rainfall diminishes, PM2.5 greatly increases, as in September 2021 and 2020. Straw burning during the harvesting season of Kharif and winter cooling may be to blame for this increase in PM2.5. In addition, the low wind speed,

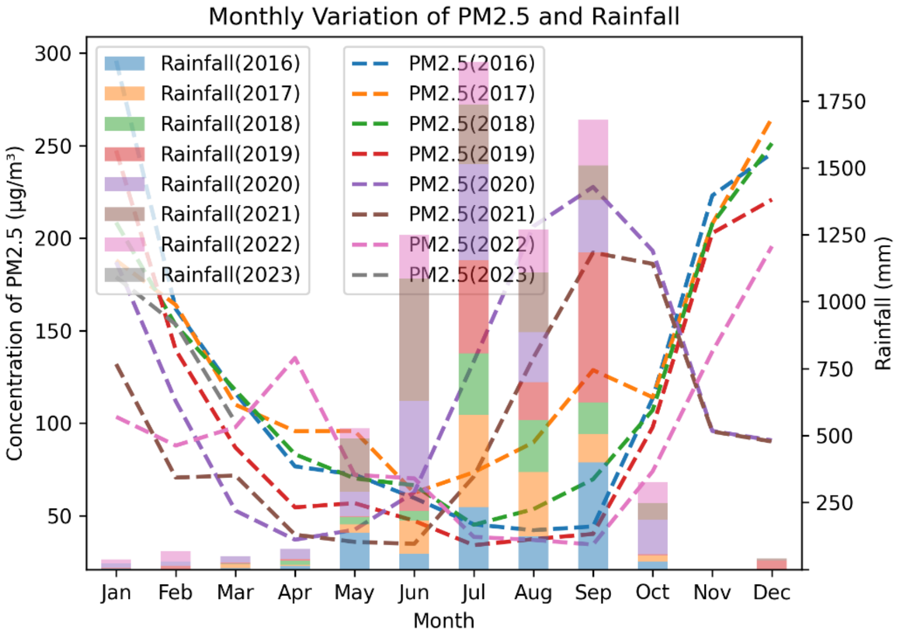

Figure 5: Plots of Monthly Variation of PM2.5 and Corresponding Rainfall.

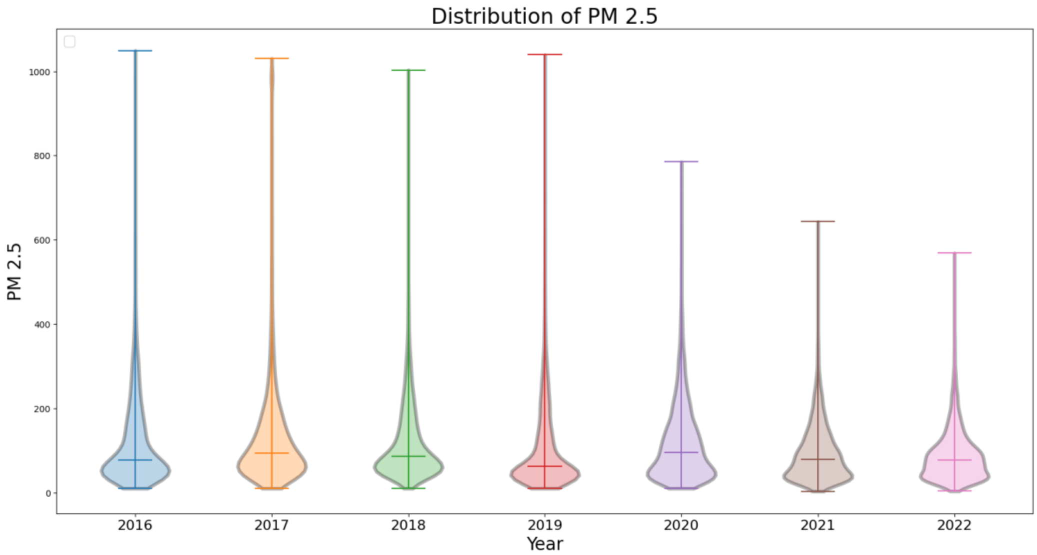

Figure 6: Distribution of PM2.5 across the years.

low temperature, short length of sunshine, high pressure, and high relative humidity that prevailed at the time all contributed to the buildup of PM2.5. Starting in March, PM2.5 levels hovered at about 100 µg/m3, where they stayed until the month's end. During this time, temperatures rose dramatically, and urban heating and pre-monsoon showers diminished PM2.5.

3.1.3 Inter-Annual variation of PM2.5

In this analysis, we looked at how PM2.5 levels in the capital city of Patna have changed over time (presented in Fig. 6). It shows that 2017 and 2020 were the years with the highest median pollution concentration. But the highest upper margin values for PM2.5 were in 2016, even though the average indicative of the worst pollution dropped. The number of extreme readings went up in 2019 in spite of a low median for pollution. In 2022, the violin plot broadened around the median, suggesting that this year the number of days with low PM2.5 concentration are more than usual. But from 2020 on, there will be a minor decline in extreme PM2.5 and outlier concentrations. But the

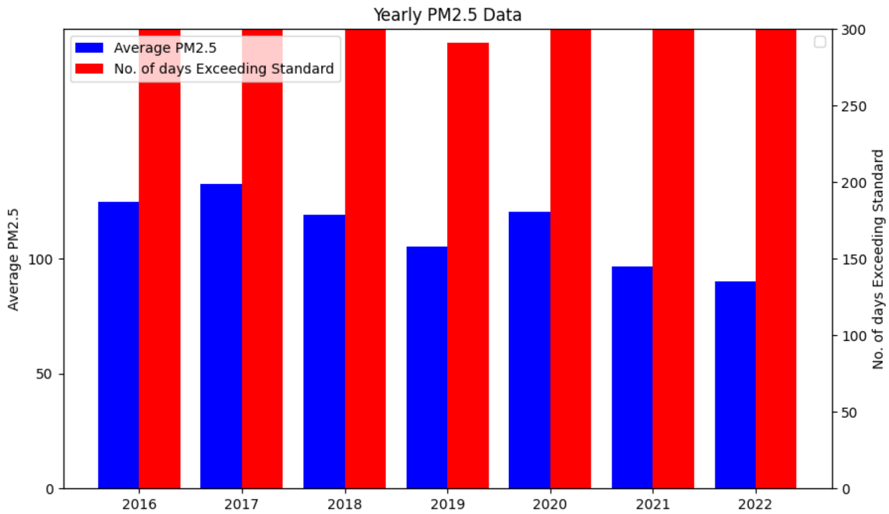

Figure 7: Yearly average concentration of PM2.5 and associated number of days exceeding the standard level of PM2.5.

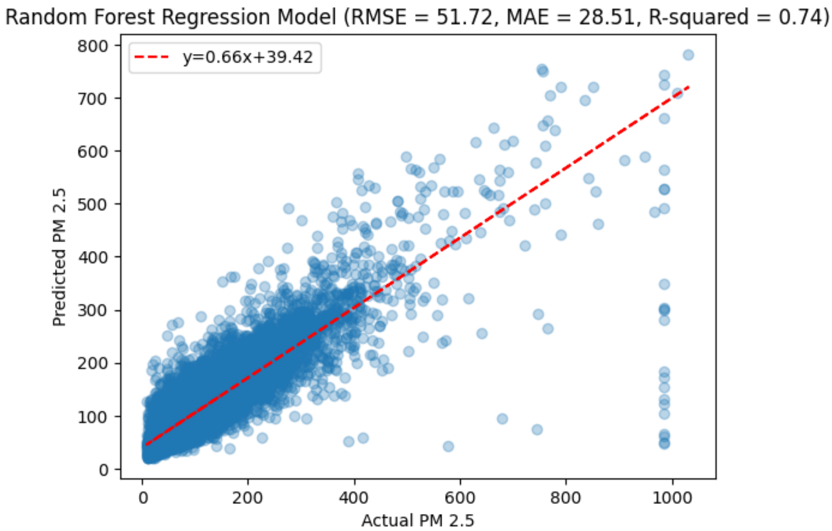

Figure 8: Prediction Outcome of PM2.5 through Random Forest Regression Models.

frequency of low pollution concentrations increases over time. The capital city of Patna's reduction in air pollution may have contributed to this positive outcome.

The annual average concentration limit for PM2.5 is set at 40 μg/m3 and the 24-hour average concentration limit is set at 60μg/m3 by the Ambient Air Quality Standard in India. Fig.7 shows that between 2016 to 2022, both the number of days in which PM2.5 concentrations were over the threshold and the annual average concentration of PM2.5.

3.2 Model Evaluation and Prediction

Hourly Atmospheric pollutants (NO2, CO, SO2, and O3), meteorological conditions (Wind Direction, Wind Speed, Air Temperature, Dew Point Temperature, Relative Humidity, Cloud Amount, Present Weather, Past Weather ,Visibility ,MSLP,SLP) were incorporated into a random forest prediction model for Patna from 2016 to 2023. Eighty percent of the dataset were used for training and the remaining were used for testing. The model's result was the hourly concentration of PM2.5. Fig.8 displays the results of fitting the test samples. The density scatter plot clearly demonstrates the model's high predictive quality. The sample points cluster together and are roughly dispersed on either side of the straight line. The prediction model's estimates of PM2.5 concentrations were remarkably close to the measured values. Equation: Y = 0.66X + 39.42 was

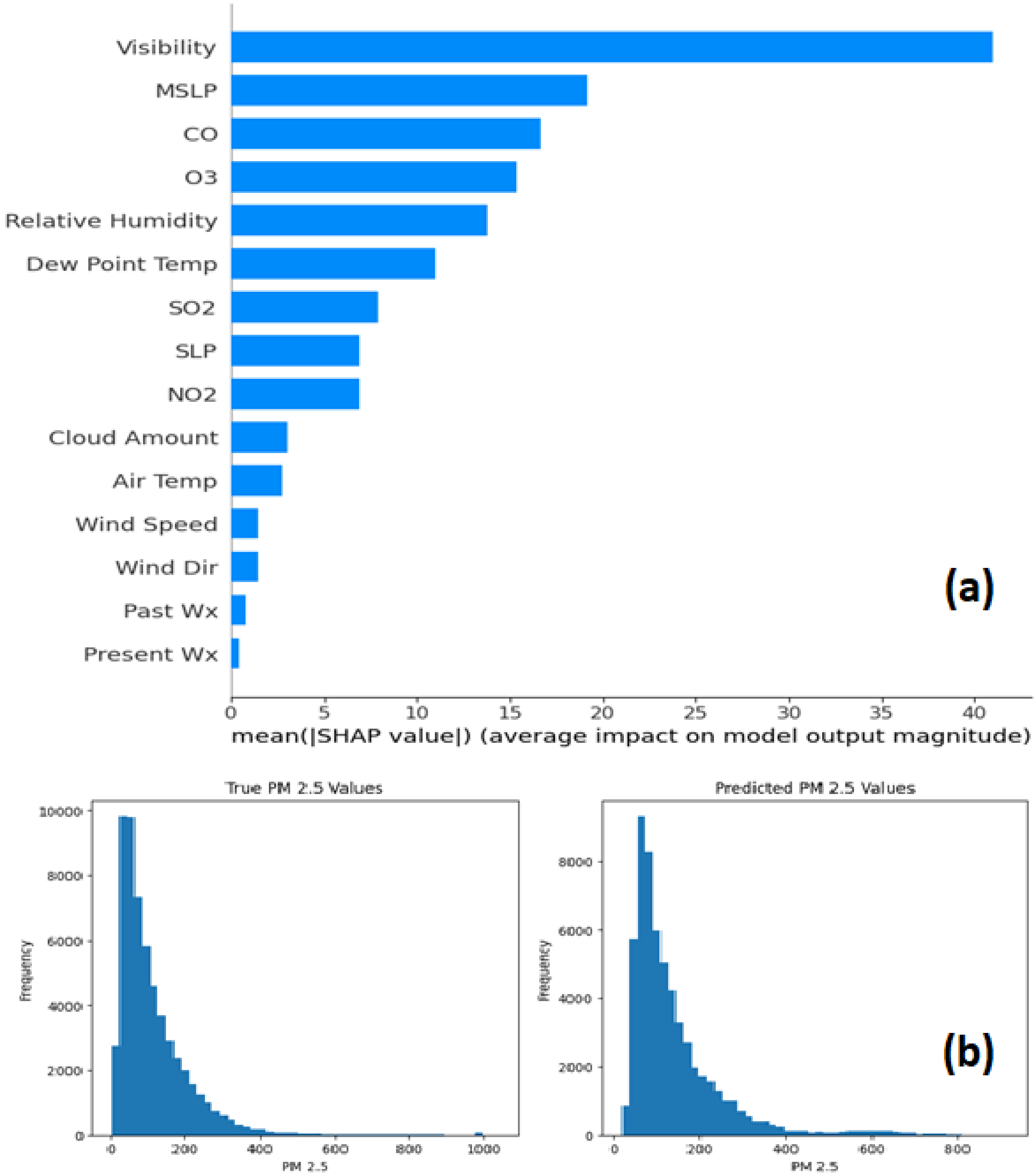

Figure 9: Contribution of features to predicted values (a) Ranking of feature contributions (b) Variation of True and Predicted PM2.5.

found to be the best fit. The results were satisfactory, with an R2 as high as 0.74 and RMSE and MAE values of 52.97 and 28.51 µg/m3, respectively. The aforementioned findings show that PM2.5 concentrations might be predicted with the chosen method. In a similar vein, a random forest model was employed to forecast PM2.5 (X. Gao et al. 2022; Zhiyuan Li et al. 2021). This allows for dissection of the impact of a variety of variables.

3.3 Predictive Factors' Impact on PM2.5 Levels

Our model results and the roles of the influencing elements were explained using the SHAP technique, presented in Fig. 9. As can be seen in Fig. 9, visibility has the highest impact on the concentration of PM2.5 with a shap value greater than 40. Hence, we can determine that PM2.5 is a major contributor to the low visibility. The influencing factors of PM2.5 are in the order of

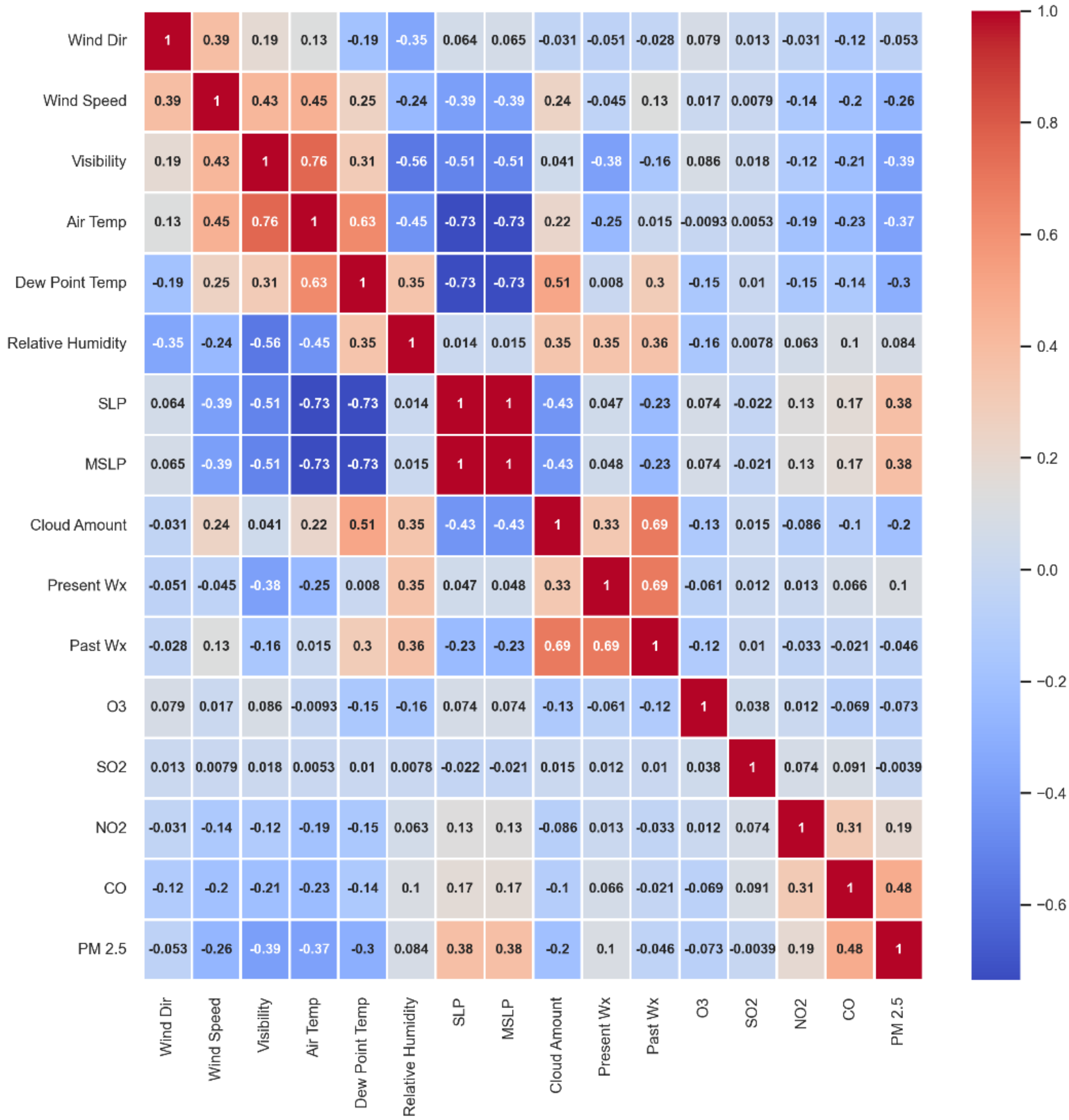

Figure 10: Heat-map of the correlation between PM2.5 and its predictors.

visibility > MSLP>CO>O3>Relative Humidity>SLP>NO2>Cloud Amount>Air Temperature>Wind Speed with a shap value greater than 2.5. The greatest PM2.5 concentrations were found to occur between 45 and 70% relative humidity (RH) (Lou et al. 2017). Also, the plots of predicted and actual PM 2.5 are presented in Fig. 9(b), which clearly signifies that both have similar values. Hence, the prediction is precise and accurate.

3.4 Correlation between PM2.5 and Predictors

The relationship between PM2.5 and the investigated factors was determined using the Pearson correlation technique. The outcome is presented in Fig. 10. Among the major atmospheric pollutants, CO was highly correlated (0.48) with PM2.5, and O3 concentration was negatively but weakly correlated (-0.073) with PM2.5. The main reason is that an increase in the PM2.5 concentration can increase the scattering of solar radiation in the visible and near-infrared bands, thus reducing the photochemical rate and, finally, leading to a decrease in the O3 concentration. Significant positive correlations were observed between NO2 and PM2.5, with coefficients of 0.19. Thus the order of correlation with the pollution parameters is CO > NO2 > O3 > SO2. In fact, CO

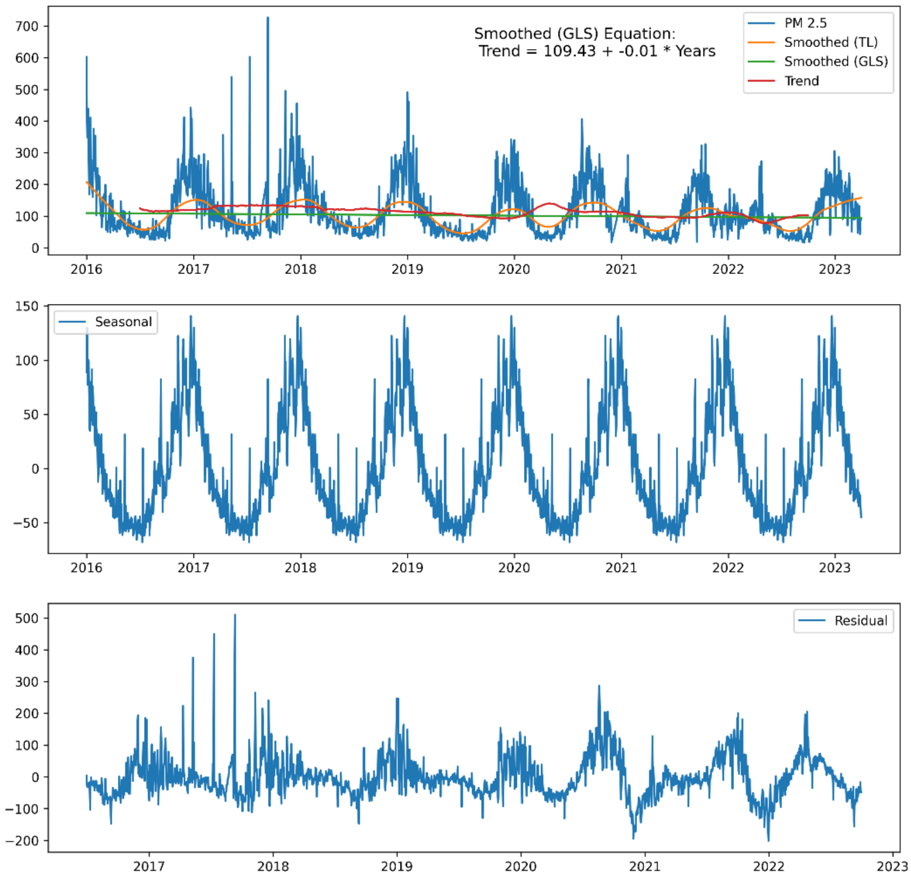

Figure 11: STL decomposition and GLS regression of daily mean PM2.5 for the capital cities of Patna over a period of seven years (2016–2023) Trends presented in (a), seasonal components presented in (b), and residuals shown in (c) The red line predicted the trend lines.

and NO2 concentrations were positively correlated, indicating that the emission sources of these two pollutants were similar; for example, they may have been the burning of straw and coal (M. Li et al. 2017). In addition, the photochemical reactions of NO2, CO, and SO2 can generate nitrate and carbonate, which can lead to an increase in PM2.5(S. Zhang et al. 2021).

In terms of the meteorological parameters, a positive correlation was found between PM2.5 and station level pressure and relative humidity, which were adversely linked with air temperature, dew point temperature, wind direction, and wind speed. This is due to the fact that seasonal pollution is more severe. High pressure and low wind speed support a stable state for the near-surface atmosphere during cold periods, such as winter, which strengthens the thermal inversion layer and reduces the diffusion of pollutants (Ma et al. 2021). There is less PM2.5 since the wind is blowing faster and diluting the particles. More precipitation means more moisture is removed, which is a key factor in the decrease in PM2.5 concentration (B. Gao et al. 2019; Y. Sun et al. 2019). The RH correlation coefficients with PM2.5 concentrations were less than 0.1, making them statistically insignificant. A recent study found that PM2.5 concentrations increased with increasing relative humidity (RH = 45–70%) (Lou et al. 2017). Low correlation may explain why the RH was constantly different from 45% to 60% and the PM2.5 levels were usually high.

3.5 Trends Analysis of PM2.5

In order to determine the slope of the trend component, GLS was used to decompose the daily mean PM2.5 time series for the capital cities of Patna into trend, seasonal, and residual components. Fig. 11 displays the STL decomposition of the daily mean PM2.5 and the GLS-fitted models. The corresponding equation, which displays the GLS linear regression slope with a 95% confidence interval, reveals no trends of PM2.5 across Patna's capital cities.

Thus, this research confirms that there is an urgent need to mitigate PM2.5. The PM2.5 levels in Patna may have been lowered by the combination of public awareness programmes (MoEFCC 2019) and pollution mitigation projects.

4. Conclusion

Using the hourly air quality and associated meteorological datasets from 2016 to 2023 of the capital cities of Patna, India, this study built a random forest model to estimate PM2.5 concentrations in the coming years. Following an assessment of the model's predictive ability, the SHAP technique was used to dissect the relative importance of each component in determining the final outcome. An in-depth analysis of PM2.5, including temporal patterns and trends as well as their associations with other factors, was determined. This led to the following inferences: The random forest model performed exceptionally well in predicting the PM2.5 level. The model was explained using the SHAP technique. It was discovered that the PM2.5 concentration is significantly affected by Visibility, MSLP, CO, and O3. There was a positive correlation between NO2, CO, MSLP, and SLP, while there was a negative correlation between PM2.5 and the air temperature, dew point temperature, wind speed, etc. During the study period, there were no significant changes in PM2.5, and it showed significant seasonal variability. However, the winter and monsoon months show the highest and lowest concentrations in Patna, India. Higher emissions and lower PBLH are connected to the winter maximums. Wet deposition and higher soil moisture during the monsoon months result in less dust re-suspension, making those months the cleanest. In 2019, heating had a major impact on PM2.5 concentration, and along with weather variations, this caused large variations in PM2.5 levels during heating and non-heating seasons. Despite the efforts of the local governments, we have not found any studies that indicate a major worsening in air quality in the capital cities of Patna, India. To the best of our knowledge, this is the first study that shows a multi-disciplinary approach for the analysis and prediction of PM2.5 in the capital city of Patna. New policies and regulations may be enacted as tools to reduce air pollution to curb vehicle emissions, road dust re-suspension and other fugitive emissions, cleaner fuel, biomass, and municipal solid waste (MSW) burning, industrial pollution, construction and demolition activities, and so on. These are all addressed by implementing source-sector-specific measures aimed at reducing air pollution. Despite economic development, enforcing the NCAP laws more strictly can hasten the decline in pollution and lessen its effects on people's health throughout India. This study's findings will be used to bolster India's National Clean Air Programme.

Acknowledgments

The authors would like to express their appreciation to the Bihar State Pollution Control Board, Government of Bihar for their constant motivation and for providing the timely pollution datasets. The authors also expressed gratitude to the Ministry of Earth Sciences, Government of India, for collaborating these specific types of study for the benefit of policymakers and stakeholders.

Appendix A. Acronyms

Acronyms	Full Name
SLP	Station Level Pressure
MSLP	Mean Sea Level Pressure
RF	Random Forest
GLS	Generalized Least Square
PBLH	Planetary Boundary Layer Height
TL	Trend Lines
STL	Seasonal and Trend decomposition using Loess
LOESS	Locally estimated scatterplot smoothing
MSW	Municipal Solid Waste

References

Alimissis, A., K. Philippopoulos, C. G. Tzanis, and D. Deligiorgi. 2018. “Spatial Estimation of Urban Air Pollution with the Use of Artificial Neural Network Models.” Atmospheric Environment 191: 205–13. https://doi.org/10.1016/j.atmosenv.2018.07.058.

Autrup, Herman. 2010. “Ambient Air Pollution and Adverse Health Effects.” Procedia - Social and Behavioral Sciences 2(5): 7333–38.

Breiman, L. 2001. “Random Forests.” Mach. Learn. 45: 5–32.

Chen, Ziyue et al. 2018. “Understanding Meteorological Influences on PM2.5 Concentrations across China: A Temporal and Spatial Perspective.” Atmospheric Chemistry and Physics 18(8): 5343–58.

Cheng, Fang Yi et al. 2021. “Evaluation of Real-Time PM2.5 Forecasts with the WRF-CMAQ Modeling System and Weather-Pattern-Dependent Bias-Adjusted PM2.5 Forecasts in Taiwan.” Atmospheric Environment 244(August 2020): 117909. https://doi.org/10.1016/j.atmosenv.2020.117909.

Cifuentes, Felipe et al. 2021. “Hourly Ozone and Pm2.5 Prediction Using Meteorological Data – Alternatives for Cities with Limited Pollutant Information.” Aerosol and Air Quality Research 21(9).

Czernecki, Bartosz, Michał Marosz, and Joanna Jędruszkiewicz. 2021. “Assessment of Machine Learning Algorithms in Short-Term Forecasting of Pm10 and Pm2.5 Concentrations in Selected Polish Agglomerations.” Aerosol and Air Quality Research 21(7): 1–18.

Delavar, Mahmoud Reza et al. 2019. “A Novel Method for Improving Air Pollution Prediction Based on Machine Learning Approaches: A Case Study Applied to the Capital City of Tehran.” ISPRS International Journal of Geo-Information 8(2).

Du, Shengdong, Tianrui Li, Yan Yang, and Shi Jinn Horng. 2021. “Deep Air Quality Forecasting Using Hybrid Deep Learning Framework.” IEEE Transactions on Knowledge and Data Engineering 33(6): 2412–24.

Ganguly, Nandita D., Chris G. Tzanis, Kostas Philippopoulos, and Despina Deligiorgi. 2019. “Analysis of a Severe Air Pollution Episode in India during Diwali Festival - A Nationwide Approach.” Atmosfera 32(3): 225–36.

Gao, Bing et al. 2019. “Interactions between Rainfall and Fine Particulate Matter Investigated by Simultaneous Chemical Composition Measurements in Downtown Beijing.” Atmospheric Environment 218(April): 1–9.

Gao, Xiuyan et al. 2022. “Analysis of Atmospheric Pollutants and Meteorological Factors on PM2.5 Concentration and Temporal Variations in Harbin.” Atmosphere 13(9).

Guo, Hao et al. 2017. “Source Apportionment of PM2.5 in North India Using Source-Oriented Air Quality Models.” Environmental Pollution 231: 426–36. http://dx.doi.org/10.1016/j.envpol.2017.08.016.

Guo, Hao, Sri Harsha Kota, Shovan Kumar Sahu, and Hongliang Zhang. 2019. “Contributions of Local and Regional Sources to PM2.5 and Its Health Effects in North India.” Atmospheric Environment 214(July): 116867. https://doi.org/10.1016/j.atmosenv.2019.116867.

Gupta, Vidhu, Lalita Bisht, Akash Deep, and Sneha Gautam. 2022. “Spatial Distribution, Pollution Levels, and Risk Assessment of Potentially Toxic Metals in Road Dust from Major Tourist City, Dehradun, Uttarakhand India.” Stochastic Environmental Research and Risk Assessment 36(10): 3517–33. https://doi.org/10.1007/s00477-022-02207-0.

Haiming, Zheng, and Shang Xiaoxiao. 2013. “Study on Prediction of Atmospheric PM2.5 Based on RBF Neural Network.” Proceedings - 2013 4th International Conference on Digital Manufacturing and Automation, ICDMA 2013: 1287–89.

Jing, Ziyun et al. 2020. "Effects of Meteorological Factors and Anthropogenic Precursors on PM2.5 Concentrations in Cities in China." Sustainability (Switzerland) 12(9): 1–13.

Kumar, Amrendra, Ishwar Chandra Yadav, Anuradha Shukla, and Ningombam Linthoingambi Devi. 2020. "Seasonal Variation of PM2.5 in the Central Indo-Gangetic Plain (Patna) of India: Chemical Characterization and Source Assessment." SN Applied Sciences 2(8): 1–13. https://doi.org/10.1007/s42452-020-3160-y.

Lee, In Ho. 2020. "Topological Invariant Prediction via Deep Learning." Journal of the Korean Physical Society 76(5): 401–5.

Li, Meng et al. 2017. "MIX: A Mosaic Asian Anthropogenic Emission Inventory under the International Collaboration Framework of the MICS-Asia and HTAP." Atmospheric Chemistry and Physics 17(2): 935–63.

Li, Xiaolan et al. 2021. "Impacts of Synoptic Forcing and Topography on Aerosol Pollution during Winter in Shenyang, Northeast China." Atmospheric Research 262(July): 105764. https://doi.org/10.1016/j.atmosres.2021.105764.

Li, Zhiyuan et al. 2021. "A Practical Framework for Predicting Residential Indoor PM2.5 Concentration Using Land-Use Regression and Machine Learning Methods." Chemosphere 265: 129140. https://doi.org/10.1016/j.chemosphere.2020.129140.

Li, Ziqi. 2022. "Extracting Spatial Effects from Machine Learning Model Using Local Interpretation Method: An Example of SHAP and XGBoost." Computers, Environment and Urban Systems 96(November 2021): 101845. https://doi.org/10.1016/j.compenvurbsys.2022.101845.

Lou, Cairong et al. 2017. "Relationships of Relative Humidity with PM2.5 and PM10 in the Yangtze River Delta, China." Environmental Monitoring and Assessment 189(11).

Lu, Xiaoman et al. 2021. "Estimating Hourly PM2.5 Concentrations Using Himawari-8 AOD and a DBSCAN-Modified Deep Learning Model over the YRDUA, China." Atmospheric Pollution Research 12(2): 183–92. https://doi.org/10.1016/j.apr.2020.10.020.

Lundberg, Scott M., and Su In Lee. 2017. "A Unified Approach to Interpreting Model Predictions." In Advances in Neural Information Processing Systems, Red Hook, NY, USA, 4766–75. https://dl.acm.org/doi/10.5555/3295222.3295230.

Ma, Jinghui et al. 2021. "PM2.5 Concentration Distribution Patterns and Influencing Meteorological Factors in the Central and Eastern China during 1980–2018." Journal of Cleaner Production 311(May): 127565. https://doi.org/10.1016/j.jclepro.2021.127565.

Mingzhi, Luo. 2017. "Analysis of Influence Factors of PM2.5 in Chengdu Based on VAR Model." IOP Conference Series: Earth and Environmental Science 63(1): 1–7.

MoEFCC. 2019. "NCAP, MoEFCC, 2019. Ministry of Environment, Forest & Climate Change NCAP-National Clean Air Programme (2019)." : 1–106. http://moef.gov.in/wp-content/%0Auploads/2019/05/NCAP_Report.pdf.

Nair, Vijayakumar S. et al. 2007. "Wintertime Aerosol Characteristics over the Indo-Gangetic Plain (IGP): Impacts of Local Boundary Layer Processes and Long-Range Transport." Journal of Geophysical Research Atmospheres 112(13): 1–15.

Park, Il Soo et al. 2021. "Meteorological Characteristics during Periods of Greatly Reduced PM2.5 Concentrations in March 2020 in Seoul." Aerosol and Air Quality Research 21(9).

Remoundou, Kyriaki, and Phoebe Koundouri. 2009. "Environmental Effects on Public Health: An Economic Perspective." International Journal of Environmental Research and Public Health 6(8): 2160–78.

Sadorsky, Perry. 2021. "A Random Forests Approach to Predicting Clean Energy Stock Prices." Journal of Risk and Financial Management 14(2): 48.

Schnell, Jordan L. et al. 2018. “Exploring the Relationship between Surface PM2.5 and Meteorology in Northern India.” Atmospheric Chemistry and Physics 18(14): 10157–75.

Shao, Min et al. 2021. “Responses in PM2.5 and Its Chemical Components to Typical Unfavorable Meteorological Events in the Suburban Area of Tianjin, China.” Science of the Total Environment 788: 147814. https://doi.org/10.1016/j.scitotenv.2021.147814.

Sharma, Shubham, Mina Chandra, and Sri Harsha Kota. 2020. “Health Effects Associated with PM2.5: A Systematic Review.” Current Pollution Reports 6(4): 345–67.

Shi, Pengfei, Xiaolong Fang, and Jianjun Ni. 2021. “Applied Sciences An Improved Attention-Based Integrated Deep Neural Network.”

Singh, Vikas, Shweta Singh, and Akash Biswal. 2021. “Exceedances and Trends of Particulate Matter (PM2.5) in Five Indian Megacities.” Science of the Total Environment 750: 141461. https://doi.org/10.1016/j.scitotenv.2020.141461.

Sreekanth, V., Kandula Niranjan, and B. L. Madhavan. 2007. “Radiative Forcing of Black Carbon over Eastern India.” Geophysical Research Letters 34(17): 1–5.

Sun, Yue et al. 2019. “Distinct Impacts of Light and Heavy Precipitation on PM2.5 Mass Concentration in Beijing.” Earth and Space Science 6(10): 1915–25.

Sun, Zhuanlan, and Demi Zhu. 2019. “Exposure to Outdoor Air Pollution and Its Human-Related Health Outcomes: An Evidence Gap Map.” BMJ Open 9(12): 1–18.

Swarna Priya, R. M., and P. Sathya. 2019. “Statistical Analysis of Air Pollutants in Ambient Air, Reality of Sensors and Corrective Measures in India.” 2019 Innovations in Power and Advanced Computing Technologies, i-PACT 2019: 1–6.

Taneja, Shweta, Nidhi Sharma, Kettun Oberoi, and Yash Navoria. 2017. “Predicting Trends in Air Pollution in Delhi Using Data Mining.” India International Conference on Information Processing, IICIP 2016 - Proceedings: 1–6.

Tiwari, S. et al. 2013. “Diurnal and Seasonal Variations of Black Carbon and PM2.5 over New Delhi, India: Influence of Meteorology.” Atmospheric Research 125–126: 50–62. http://dx.doi.org/10.1016/j.atmosres.2013.01.011.

Tyagi, S. et al. 2017. “Characteristics of Absorbing Aerosols during Winter Foggy Period over the National Capital Region of Delhi: Impact of Planetary Boundary Layer Dynamics and Solar Radiation Flux.” Atmospheric Research 188: 1–10. http://dx.doi.org/10.1016/j.atmosres.2017.01.001.

Wang, Xinpeng, and Wenbin Sun. 2019. “Meteorological Parameters and Gaseous Pollutant Concentrations as Predictors of Daily Continuous PM2.5 Concentrations Using Deep Neural Network in Beijing–Tianjin–Hebei, China.” Atmospheric Environment 211: 128–37. https://doi.org/10.1016/j.atmosenv.2019.05.004.

Westervelt, D. M. et al. 2016. “Quantifying PM2.5-Meteorology Sensitivities in a Global Climate Model.” Atmospheric Environment 142: 43–56. http://dx.doi.org/10.1016/j.atmosenv.2016.07.040.

Yang, Lijuan, Hanqiu Xu, and Shaode Yu. 2020. “Estimating PM2.5 Concentrations in Yangtze River Delta Region of China Using Random Forest Model and the Top-of-Atmosphere Reflectance.” Journal of Environmental Management 272(July): 111061. https://doi.org/10.1016/j.jenvman.2020.111061.

Zhang, Lei et al. 2021. “Spatiotemporal Causal Convolutional Network for Forecasting Hourly PM2.5 Concentrations in Beijing, China.” Computers and Geosciences 155(October 2020): 104869. https://doi.org/10.1016/j.cageo.2021.104869.

Zhang, Licheng et al. 2020. “Spatiotemporal Variations and Influencing Factors of PM2.5 Concentrations in Beijing, China.” Environmental Pollution 262.

Zhang, Si et al. 2021. “Rapid Sulfate Formation from Synergetic Oxidation of SO2 by O3 and NO2 under Ammonia-Rich Conditions: Implications for the Explosive Growth of Atmospheric PM2.5 during Haze Events in China." Science of the Total Environment 772(2): 144897. https://doi.org/10.1016/j.scitotenv.2020.144897.

Zheng, Yu et al. 2019. "Analysis of PM2.5 Concentrations in Heilongjiang Province Associated with Forest Cover and Other Factors." Journal of Forestry Research 30(1): 269–76. https://doi.org/10.1007/s11676-018-0640-7.

Alimissis, A., K. Philippopoulos, C. G. Tzanis, and D. Deligiorgi. 2018. "Spatial Estimation of Urban Air Pollution with the Use of Artificial Neural Network Models." Atmospheric Environment 191: 205–13. https://doi.org/10.1016/j.atmosenv.2018.07.058.

Autrup, Herman. 2010. "Ambient Air Pollution and Adverse Health Effects." Procedia - Social and Behavioral Sciences 2(5): 7333–38.

Breiman, L. 2001. "Random Forests." Mach. Learn. 45: 5–32.

Chen, Ziyue et al. 2018. "Understanding Meteorological Influences on PM2.5 Concentrations across China: A Temporal and Spatial Perspective." Atmospheric Chemistry and Physics 18(8): 5343–58.

Cheng, Fang Yi et al. 2021. "Evaluation of Real-Time PM2.5 Forecasts with the WRF-CMAQ Modeling System and Weather-Pattern-Dependent Bias-Adjusted PM2.5 Forecasts in Taiwan." Atmospheric Environment 244(August 2020): 117909. https://doi.org/10.1016/j.atmosenv.2020.117909.

Cifuentes, Felipe et al. 2021. "Hourly Ozone and Pm2.5 Prediction Using Meteorological Data – Alternatives for Cities with Limited Pollutant Information." Aerosol and Air Quality Research 21(9).

Czernecki, Bartosz, Michał Marosz, and Joanna Jędruszkiewicz. 2021. "Assessment of Machine Learning Algorithms in Short-Term Forecasting of Pm10 and Pm2.5 Concentrations in Selected Polish Agglomerations." Aerosol and Air Quality Research 21(7): 1–18.

Delavar, Mahmoud Reza et al. 2019. "A Novel Method for Improving Air Pollution Prediction Based on Machine Learning Approaches: A Case Study Applied to the Capital City of Tehran." ISPRS International Journal of Geo-Information 8(2).

Du, Shengdong, Tianrui Li, Yan Yang, and Shi Jinn Horng. 2021. "Deep Air Quality Forecasting Using Hybrid Deep Learning Framework." IEEE Transactions on Knowledge and Data Engineering 33(6): 2412–24.

Ganguly, Nandita D., Chris G. Tzanis, Kostas Philippopoulos, and Despina Deligiorgi. 2019. "Analysis of a Severe Air Pollution Episode in India during Diwali Festival - A Nationwide Approach." Atmosfera 32(3): 225–36.

Gao, Bing et al. 2019. "Interactions between Rainfall and Fine Particulate Matter Investigated by Simultaneous Chemical Composition Measurements in Downtown Beijing." Atmospheric Environment 218(April): 1–9.

Gao, Xiuyan et al. 2022. "Analysis of Atmospheric Pollutants and Meteorological Factors on PM2.5 Concentration and Temporal Variations in Harbin." Atmosphere 13(9).

Guo, Hao et al. 2017. "Source Apportionment of PM2.5 in North India Using Source-Oriented Air Quality Models." Environmental Pollution 231: 426–36. http://dx.doi.org/10.1016/j.envpol.2017.08.016.

Guo, Hao, Sri Harsha Kota, Shovan Kumar Sahu, and Hongliang Zhang. 2019. "Contributions of Local and Regional Sources to PM2.5 and Its Health Effects in North India." Atmospheric Environment 214(July): 116867. https://doi.org/10.1016/j.atmosenv.2019.116867.

Gupta, Vidhu, Lalita Bisht, Akash Deep, and Sneha Gautam. 2022. "Spatial Distribution, Pollution Levels, and Risk Assessment of Potentially Toxic Metals in Road Dust from Major Tourist City, Dehradun, Uttarakhand India." Stochastic

Environmental Research and Risk Assessment 36(10): 3517–33. https://doi.org/10.1007/s00477-022-02207-0.

Haiming, Zheng, and Shang Xiaoxiao. 2013. “Study on Prediction of Atmospheric PM2.5 Based on RBF Neural Network.” Proceedings - 2013 4th International Conference on Digital Manufacturing and Automation, ICDMA 2013: 1287–89.

Jing, Ziyun et al. 2020. “Effects of Meteorological Factors and Anthropogenic Precursors on PM2.5 Concentrations in Cities in China.” Sustainability (Switzerland) 12(9): 1–13.

Lee, In Ho. 2020. “Topological Invariant Prediction via Deep Learning.” Journal of the Korean Physical Society 76(5): 401–5.

Li, Meng et al. 2017. “MIX: A Mosaic Asian Anthropogenic Emission Inventory under the International Collaboration Framework of the MICS-Asia and HTAP.” Atmospheric Chemistry and Physics 17(2): 935–63.

Li, Xiaolan et al. 2021. “Impacts of Synoptic Forcing and Topography on Aerosol Pollution during Winter in Shenyang, Northeast China.” Atmospheric Research 262(July): 105764. https://doi.org/10.1016/j.atmosres.2021.105764.

Li, Zhiyuan et al. 2021. “A Practical Framework for Predicting Residential Indoor PM2.5 Concentration Using Land-Use Regression and Machine Learning Methods.” Chemosphere 265: 129140. https://doi.org/10.1016/j.chemosphere.2020.129140.

Li, Ziqi. 2022. “Extracting Spatial Effects from Machine Learning Model Using Local Interpretation Method: An Example of SHAP and XGBoost.” Computers, Environment and Urban Systems 96(November 2021): 101845. https://doi.org/10.1016/j.compenvurbsys.2022.101845.

Lou, Cairong et al. 2017. “Relationships of Relative Humidity with PM2.5 and PM10 in the Yangtze River Delta, China.” Environmental Monitoring and Assessment 189(11).

Lu, Xiaoman et al. 2021. “Estimating Hourly PM2.5 Concentrations Using Himawari-8 AOD and a DBSCAN-Modified Deep Learning Model over the YRDUA, China.” Atmospheric Pollution Research 12(2): 183–92. https://doi.org/10.1016/j.apr.2020.10.020.

Lundberg, S.; Lee, S. A. “Unified Approach to Interpreting Model Predictions, Edited. In Proceedings of the 31st International Conference on Neural Information Processing Systems.” In NIPS’17, Red Hook, NY, USA.

Ma, Jinghui et al. 2021. “PM2.5 Concentration Distribution Patterns and Influencing Meteorological Factors in the Central and Eastern China during 1980–2018.” Journal of Cleaner Production 311(May): 127565. https://doi.org/10.1016/j.jclepro.2021.127565.

Mingzhi, Luo. 2017. “Analysis of Influence Factors of PM2.5 in Chengdu Based on VAR Model.” IOP Conference Series: Earth and Environmental Science 63(1): 1–7.

MoEFCC. 2019. “NCAP, MoEFCC, 2019. Ministry of Environment, Forest & Climate Change NCAP-National Clean Air Programme (2019).” : 1–106. http://moef.gov.in/wp-content/%0Auploads/2019/05/NCAP_Report.pdf.

Nair, Vijayakumar S. et al. 2007. “Wintertime Aerosol Characteristics over the Indo-Gangetic Plain (IGP): Impacts of Local Boundary Layer Processes and Long-Range Transport.” Journal of Geophysical Research Atmospheres 112(13): 1–15.

Park, Il Soo et al. 2021. “Meteorological Characteristics during Periods of Greatly Reduced PM2.5 Concentrations in March 2020 in Seoul.” Aerosol and Air Quality Research 21(9).

Remoundou, Kyriaki, and Phoebe Koundouri. 2009. “Environmental Effects on Public Health: An Economic Perspective.” International Journal of Environmental Research and Public Health 6(8): 2160–78.

Sadorsky, Perry. 2021. “A Random Forests Approach to Predicting Clean Energy Stock

Prices." Journal of Risk and Financial Management 14(2): 48.

Schnell, Jordan L. et al. 2018. "Exploring the Relationship between Surface PM2.5 and Meteorology in Northern India." Atmospheric Chemistry and Physics 18(14): 10157–75.

Shao, Min et al. 2021. "Responses in PM2.5 and Its Chemical Components to Typical Unfavorable Meteorological Events in the Suburban Area of Tianjin, China." Science of the Total Environment 788: 147814. https://doi.org/10.1016/j.scitotenv.2021.147814.

Sharma, Shubham, Mina Chandra, and Sri Harsha Kota. 2020. "Health Effects Associated with PM2.5: A Systematic Review." Current Pollution Reports 6(4): 345–67.

Shi, Pengfei, Xiaolong Fang, and Jianjun Ni. 2021. "Applied Sciences An Improved Attention-Based Integrated Deep Neural Network."

Singh, Vikas, Shweta Singh, and Akash Biswal. 2021. "Exceedances and Trends of Particulate Matter (PM2.5) in Five Indian Megacities." Science of the Total Environment 750: 141461. https://doi.org/10.1016/j.scitotenv.2020.141461.

Sreekanth, V., Kandula Niranjan, and B. L. Madhavan. 2007. "Radiative Forcing of Black Carbon over Eastern India." Geophysical Research Letters 34(17): 1–5.

Sun, Yue et al. 2019. "Distinct Impacts of Light and Heavy Precipitation on PM2.5 Mass Concentration in Beijing." Earth and Space Science 6(10): 1915–25.

Sun, Zhuanlan, and Demi Zhu. 2019. "Exposure to Outdoor Air Pollution and Its Human-Related Health Outcomes: An Evidence Gap Map." BMJ Open 9(12): 1–18.

Swarna Priya, R. M., and P. Sathya. 2019. "Statistical Analysis of Air Pollutants in Ambient Air, Reality of Sensors and Corrective Measures in India." 2019 Innovations in Power and Advanced Computing Technologies, i-PACT 2019: 1–6.

Taneja, Shweta, Nidhi Sharma, Kettun Oberoi, and Yash Navoria. 2017. "Predicting Trends in Air Pollution in Delhi Using Data Mining." India International Conference on Information Processing, IICIP 2016 - Proceedings: 1–6.

Tiwari, S. et al. 2013. "Diurnal and Seasonal Variations of Black Carbon and PM2.5 over New Delhi, India: Influence of Meteorology." Atmospheric Research 125–126: 50–62. http://dx.doi.org/10.1016/j.atmosres.2013.01.011.

Tyagi, S. et al. 2017. "Characteristics of Absorbing Aerosols during Winter Foggy Period over the National Capital Region of Delhi: Impact of Planetary Boundary Layer Dynamics and Solar Radiation Flux." Atmospheric Research 188: 1–10. http://dx.doi.org/10.1016/j.atmosres.2017.01.001.

Wang, Chien. Exploiting Deep Learning in Forecasting the Occurrence of Severe Haze in Southeast Asia. https://github.com/keras-team/keras.

Wang, Xinpeng, and Wenbin Sun. 2019. "Meteorological Parameters and Gaseous Pollutant Concentrations as Predictors of Daily Continuous PM2.5 Concentrations Using Deep Neural Network in Beijing–Tianjin–Hebei, China." Atmospheric Environment 211: 128–37. https://doi.org/10.1016/j.atmosenv.2019.05.004.

Westervelt, D. M. et al. 2016. "Quantifying PM2.5-Meteorology Sensitivities in a Global Climate Model." Atmospheric Environment 142: 43–56. http://dx.doi.org/10.1016/j.atmosenv.2016.07.040.

Yang, Lijuan, Hanqiu Xu, and Shaode Yu. 2020. "Estimating PM2.5 Concentrations in Yangtze River Delta Region of China Using Random Forest Model and the Top-of-Atmosphere Reflectance." Journal of Environmental Management 272(July): 111061. https://doi.org/10.1016/j.jenvman.2020.111061.

Zhang, Lei et al. 2021. "Spatiotemporal Causal Convolutional Network for Forecasting Hourly PM2.5 Concentrations in Beijing, China." Computers and Geosciences 155(October 2020):

104869. https://doi.org/10.1016/j.cageo.2021.104869.

Zhang, Licheng et al. 2020. "Spatiotemporal Variations and Influencing Factors of PM2.5 Concentrations in Beijing, China." Environmental Pollution 262.

Zhang, Si et al. 2021. "Rapid Sulfate Formation from Synergetic Oxidation of SO2 by O3 and NO2 under Ammonia-Rich Conditions: Implications for the Explosive Growth of Atmospheric PM2.5 during Haze Events in China." Science of the Total Environment 772(2): 144897. https://doi.org/10.1016/j.scitotenv.2020.144897.

Zheng, Yu et al. 2019. "Analysis of PM2.5 Concentrations in Heilongjiang Province Associated with Forest Cover and Other Factors." Journal of Forestry Research 30(1): 269–76. https://doi.org/10.1007/s11676-018-0640-7.