

Article

PM_{2.5} Prediction Based on Random Forest, XGBoost, and Deep Learning Using Multisource Remote Sensing Data

Mehdi Zamani Joharestani ^{1,2,†}, Chunxiang Cao ^{1,2,*}, Xiliang Ni ^{1,2,†}, Barjeece Bashir ^{1,2} and Somayeh Talebiesfandarani ^{1,2}

¹ State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100101, China

² University of Chinese Academy of Science, Beijing 100049, China

* Correspondence: caocx@radi.ac.cn; Tel.: +86-139-1161-0226

† These authors contributed equally to this work.

Received: 23 May 2019; Accepted: 2 July 2019; Published: 4 July 2019



Abstract: In recent years, air pollution has become an important public health concern. The high concentration of fine particulate matter with diameter less than 2.5 μm (PM_{2.5}) is known to be associated with lung cancer, cardiovascular disease, respiratory disease, and metabolic disease. Predicting PM_{2.5} concentrations can help governments warn people at high risk, thus mitigating the complications. Although attempts have been made to predict PM_{2.5} concentrations, the factors influencing PM_{2.5} prediction have not been investigated. In this work, we study feature importance for PM_{2.5} prediction in Tehran's urban area, implementing random forest, extreme gradient boosting, and deep learning machine learning (ML) approaches. We use 23 features, including satellite and meteorological data, ground-measured PM_{2.5}, and geographical data, in the modeling. The best model performance obtained was $R^2 = 0.81$ ($R = 0.9$), MAE = 9.93 $\mu\text{g}/\text{m}^3$, and RMSE = 13.58 $\mu\text{g}/\text{m}^3$ using the XGBoost approach, incorporating elimination of unimportant features. However, all three ML methods performed similarly and R^2 varied from 0.63 to 0.67, when Aerosol Optical Depth (AOD) at 3 km resolution was included, and 0.77 to 0.81, when AOD at 3 km resolution was excluded. Contrary to the PM_{2.5} lag data, satellite-derived AODs did not improve model performance.

Keywords: PM_{2.5}; prediction; XGBoost; random forest; deep learning; feature importance

1. Introduction

As a consequence of urbanization and industrialization, air pollution has become one of the most important public health concerns [1–6]. The PM_{2.5} pollutant is defined as fine inhalable particles with diameters less than 2.5 μm [7]. The association of high PM_{2.5} concentration and cancer, cardiovascular disease, respiratory disease, metabolic disease, and obesity has been proven [8–11].

In Tehran, the capital of Iran, the annual PM_{2.5} concentration of $86.8 \pm 33 \mu\text{g m}^{-3}$ (based on 4 years of observations, from 2015 to 2018) significantly exceeds the World Health Organization (WHO) guideline [8]. Gasoline and diesel vehicles, industrial emissions, and dust storms are the main reasons for high PM_{2.5} concentration in Tehran. Taghvaee et al. [12] reported that diesel exhaust and industrial emissions have a greater impact on cancer risks (~70%) than other air pollution sources in Tehran. Tehran is not the city in Iran with the worst air pollution, however, it has received more attention [8,12–18] because of its large population (estimated to be 9 million in 2019 [19]). Dehghan et al. [18] investigated the impact of Tehran's air pollution on the mortality rate related to respiratory diseases. They reported that from 2005 to 2014, high concentrations of O₃, NO₂, PM₁₀, and PM_{2.5} were strongly associated with

34,000 deaths. Additional research has confirmed these results [20–22]. Arhami et al. [14] investigated seasonal trends in the composition and sources of PM_{2.5} and carbonaceous aerosols. They proposed that motor vehicles are the major contributors to air pollution, particularly during winter.

Predicting the PM_{2.5} concentration is necessary for social planning and management, to mitigate the impact of air pollution on public health. In recent years, there have been successes in Aerosol Optical Depth (AOD) estimation using remote sensing technology, and this parameter has become part of PM_{2.5} prediction research [23–28]. Several attempts have been made to predict PM_{2.5} concentration utilizing regression and machine learning techniques, in addition to climatic variables and remote sensing data [29–34]. For instance, Li et al. [35] used the Moderate Resolution Imaging Spectroradiometer (MODIS) derived AOD product at 10 km resolution (AOD10) with meteorological data, in addition to PM_{2.5} historical observations, for PM_{2.5} prediction in China. Xiliang Ni et al. [36] utilized the satellite-derived MODIS AOD at 3 km resolution (AOD03), in addition to meteorological data, to estimate the spatial distribution of PM_{2.5} concentration in the Beijing, Tianjin, and Hebei regions using a backpropagation neural network. There have been other attempts to predict PM_{2.5} using times series modeling, such as a recurrent neural network [16,33,37–39]. Li et al. [35] introduced a geo-intelligent, deep learning method to predict PM_{2.5} over part of China, with a performance of $R^2 = 0.88$.

Although several attempts have been made to predict PM_{2.5} concentration, the relationship between features that influence PM_{2.5} concentration prediction is still not well understood [37]. Only a few studies, of limited extent, have investigated the importance of these features on PM_{2.5} concentration prediction [13,35,40]. Hadei et al. [40] assessed the influence of holidays on air pollution variations. The small number of studies done on the prediction of air pollution in Tehran has not performed well. For example, Shamsoddini et al. [13] used five air pollution stations in Tehran and meteorological data to predict PM_{2.5}, using an artificial neural network and a random forest. They achieved a maximum value of $R^2 = 0.49$ and used a built-in Random Forest (RF) function as an estimation of feature importance. Nabavi et al. [41] also tried to estimate the spatial pattern of PM_{2.5} over Tehran using AOD10 and 1 km MAIAC data. They achieved a maximum value of $R^2 = 0.68$.

We focus on highly distributed air pollution monitoring (APM) sites in Tehran's urban area. Missing values in Tehran's ground-measured APM sites are a severe problem. In the urban area of Tehran, of 42 total APM sites, the rate of missing values in 11 of these for our study period is more than 75%. The difficulty with the missing data problem is also present in satellite-derived AODs, particularly AOD03 (96%) and AOD10 (63%). Nabavi et al. [41] reported that the AOD retrieval algorithm based on a dark target considers the brightness of scenes as an indicator of the existence of aerosols. However, in urban areas, structures such as building roofs and streets act as bright surfaces, leading to miscalculation in AOD retrieval. It has been reported that 80% of AOD data from 2003 to 2017 was discarded because of this issue [41]. In this work, Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and Deep Neural Network (DNN) machine learning methods were used to investigate PM_{2.5} concentration prediction. The performance of the predictions was evaluated using R^2 , root mean square error (RMSE), and mean absolute error (MAE) metrics. A total of 23 features including AOD03, AOD10, meteorological data, geographic information of APM sites (latitude, longitude, and altitude), and other auxiliary features were used to predict PM_{2.5}. Finally, utilizing different methods, the importance of features for PM_{2.5} concentration prediction was evaluated and compared. In addition, the most important features for PM_{2.5} prediction were determined.

2. Experiments

2.1. Study Area

The study area is Tehran, the capital city of Iran, and the study period is from 2015 January 1 to the end of 2018. Tehran is located between 35.50° and 35.88° North and 51.1° and 51.7° East, at the northern center of Iran. Tehran has a cold semi-arid climate, with annual average relative humidity of

46%, annual precipitation of 429 mm, and temperature range from -5 to 38 $^{\circ}\text{C}$. Elevation in Tehran varies significantly from 1117 to 1712.6 m above sea level. Tehran's urban extent and population has been increasing over the last few decades. According to the national census conducted in 2016, the population was 8.69 million, which was 10.8% of the total population of the country (80.28 million). Based on the newest revision of the UN World Urbanization Prospects, the population of Tehran is estimated to be ~ 9 million in 2019 [19]. Although Tehran is not the most polluted city in Iran, more people are exposed to air pollution than in other cities because of its high population density. Tehran is located on the southern slope of Alborz Mountain, which has a significant effect on the region's weather. Air pollution in Tehran mainly originates from three major sources: transportation, industry, and dust storms. Tehran is supported by 42 air pollution monitoring stations that were established by the Department of Environment (23 stations) and the Municipality of Tehran (19 stations). The nearest meteorological station to the APM sites is Mehrabad, located in the urban area. The average distance of APM sites from Mehrabad is approximately 10 km (1 to 20 km for APM stations with missing data less than 75%).

2.2. Data

Ground measured PM_{2.5} concentration, satellite-derived AODs at 3 and 10 km spatial resolution, and meteorological data were utilized (see Table 1). Also, other auxiliary data was used, such as APM site geographic information (longitude, latitude, and altitude) and historical parameters of records such as day of year, day of week, and season. The study period was from 2015 to the end of 2018.

Table 1. List of data and study information.

Data Type	Parameter	Abbreviation	Unit	Period	Source
Climatic	Temperature	T	$^{\circ}\text{C}$		
	Temperature max	T_max	$^{\circ}\text{C}$		
	Temperature min	T_min	$^{\circ}\text{C}$		
	Relative humidity	RH	%		
	Daily rainfall	Rainfall	mm	2015.1–2018.12	IRAN Meteorological Organization
	Visibility	Visibility	km		
	Wind speed	Windsp	m/s		
	Sustained wind speed	ST_windsp	m/s		
	Air pressure	Air_pressure	hPa		
Ground measured	Dew point	Dew point	$^{\circ}\text{C}$		
	PM _{2.5}	PM _{2.5}	$\mu\text{g m}^{-3}$	2015.1–2018.12	airnow.tehran.ir aqms.doe.ir
Satellite products	MODIS AODs from Aqua satellite	AOD03 AOD10	unitless unitless	2015.1–2018.12	NASA Atmosphere Archive & Distribution System (LAADS) Archive

2.2.1. PM_{2.5} Air pollution Data

The National Department of Environment and Municipality of Tehran city have set up 42 APM stations, distributed as illustrated in Figure 1. Air pollution parameters such as PM_{2.5}, PM₁₀, CO, O₃, NO₂, and SO₂ are recorded by APM sites. The daily average of the PM_{2.5} data is accessible through the Tehran's Municipality ICT website [42] and Air Pollution Monitoring System platform of the Department of Environment [43]. In addition, station parameters such as altitude, longitude, and latitude were obtained from the same sources.

2.2.2. Aerosol Optical Depth (AOD) Data

Aerosol Optical Depth is recognized as the accumulated attenuation factor over a perpendicular column of unit cross section [44]. The Moderate Resolution Imaging Spectroradiometer (MODIS) AOD products are well known in air pollution studies. The MODIS instrument is installed in both Aqua and Terra satellites. Recently, the MODIS atmospheric analysis team published a new level-2 collection 6.1 global scale aerosol optical depth product. Products are offered in two spatial resolutions of 10 km and 3 km. These products are labeled MOD04_L2 and MOD04_3K, respectively, for Aqua

satellite-derived AOD. The AOD is calculated based on variations in the Dark Target (DT) and Deep Blue (DB) Aerosol retrievals algorithm, over urban areas. The recent version of the product offered by the MODIS atmosphere algorithm developer's team has improved the estimation of AOD values. In particular, improvements have been made for areas with extremely variable topography, such as Iran. However, there is still a high rate of missing values, and bias in AOD values is observed for our study area.

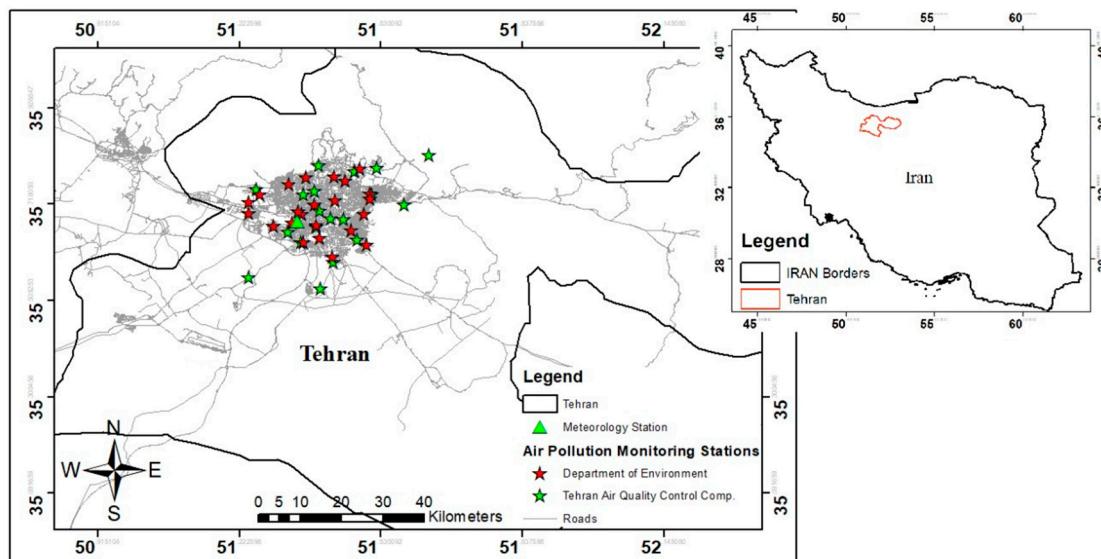


Figure 1. Study area, situated in the northern central part of Iran. There are 42 air pollution monitoring stations installed in the urban area of Tehran by two organizations. The Department of the Environment has 23 monitoring stations (red stars) and the Municipality of Tehran city has 19 stations (green stars). The nearest weather station to the urban area is Mehrabad meteorology station, marked by a green triangle.

The aerosol optical depth products (Both 10 and 3 km spatial resolution) were downloaded for the study period of 2015 to the end of 2018, from the NASA Atmosphere Archive & Distribution System (LAADS) archive portal [45]. Products are in the Hierarchical Data Format (HDF) with several subdatasets. The “AOD_550_Dark_Target_Deep_Blue_Combined” subdataset from the 10 km resolution product (MOD04_L2) and “Optical_Depth_Land_And_Ocean” subdataset from MOD04_3k were extracted from the main datasets.

2.2.3. Meteorological Data

The climatic features utilized in this study include air temperature (T), maximum and minimum air temperature (T_{max}, T_{min}), relative humidity (RH), daily rainfall, visibility, wind speed (Windsp), sustained wind speed (ST_windsp), air pressure, and dew point. Climatic data from Mehrabad weather station was used for all APM stations, because it is the nearest meteorological station to the study area and the APM stations are all within about 10 km of it. Data were downloaded from the Iran Meteorological Organization (IMO) portal [46].

2.3. Methodology

This work involved sampling, data pre-processing, data aggregation, and using three modeling methods for prediction and validation of $PM_{2.5}$ concentration. Out of 42 available sites, 37 APM sites were used in the modeling. The purpose of the current study is to find the best model to predict $PM_{2.5}$ at selected sites, using climatic, satellite, and auxiliary data. We analyzed and introduced the most important features for the study area, based on several methods. The Random Forest and XGBoost algorithms have built-in methods for detection of important features, but for deep learning

we implemented feature permutation to evaluate feature importance. In addition, a recursive feature removing and model training based on XGBoost modeling was carried out, and the mean absolute error was used as a reference for feature removal during modeling.

2.3.1. Data Preprocessing and Matching

Data processing and matching is necessary because the data was obtained from different sources. Daily climatic data was downloaded from the IMO portal. There are a few missing values or in some cases, full day missing records. Therefore, we used interpolation to estimate and fill in the missing data. The PM_{2.5} data was collected from the Tehran Municipality ICT website [42] and the Air Pollution Monitoring System platform of the Iran Environment Department [43]. Also, the altitude, longitude, and latitude of each APM station were recorded from those references, to be used later in modeling and for AOD data sampling reference. The time format of the data offered by the department of the environment was not in Julian format and thus was converted to be compatible with the other data. Missing values for short or long periods are a common problem in air pollution monitoring stations. This happens when there is a critical failure or temporary power cutoff [17,47]. For Tehran's APM sites, there are many missing values that cannot be compensated by interpolation.

Next, we downloaded both the 10 km (MOD04_L2) and 3 km (MOD04_3k) spatial resolution MODIS AOD products of the Aqua satellite, from the NASA Atmosphere Archive & Distribution System (LAADS) portal. Products are in HDF file format and have multiple subdatasets. We used the "AOD_550_Dark_Target_Deep_Blue_Combined" subdataset from the MOD04_L2 and the "Optical_Depth_Land_And_Ocean" subdataset from the MOD04_3k. Considering all 37 APM sites location, AOD values were sampled for the entire study period (four years, equal to 1460 days).

The AOD and PM_{2.5} data, in addition to the altitude, longitude, and latitude of each station, were merged together. The same climatic data were concatenated to this data, based on the sampling date. Day of year (DoY), season, and weekday were also obtained for each day and added to the database. We also used the PM_{2.5} and rainfall with a lag of one and two days, so new columns were added to the database as PM_{2.5_lag1}, PM_{2.5_lag2}, Rainfall_{_lag1}, and Rainfall_{_lag2}. Finally, the PM_{2.5} monitoring organization as a Boolean value (zero = Tehran Municipal, one = DOE) and distance of each station from Mehrabad Weather station, were added to the database. Descriptive statistics over meteorology parameters, PM_{2.5}, and AOD values were calculated to evaluate the datasets. The mean, standard deviation, maximum and minimum values of features, and the 25, 50, and 75% quartile values of each parameter were calculated. This is illustrated in the Supplementary Materials Section S3 and Tables S1 and S2.

2.3.2. Normalization

Data normalization is an important step for many machine-learning estimators, particularly when dealing with deep learning. The preferred range of features for most ML approaches is between -1 to 1. Features with a wider range can cause instability during the model training [48]. Standardization was used to standardize the features by deducting the mean and scaling the data, with the variance of feature (Z_i) calculated as

$$Z_i = \frac{x_i - \bar{x}}{\delta_x} \quad (1)$$

where x_i , \bar{x} , and δ_x are the sample values, mean, and standard deviation, respectively.

After applying standard normalizations, train and test datasets were prepared. Dataset records were shuffled and split to 70% for the train and 30% for the test. As a result of the high rate of missing values in AOD03 (94%), the training was carried out once including the AOD03 and then excluding AOD03. Records for each step were purified of missing values, based on which feature was used for modeling.

2.3.3. Random Forest Modeling

Random forest, introduced by Ho [49], is a supervised ensemble learning method that acts based on the decision tree. It can be used for both classification and regression, and it is very flexible and fast. To conduct RF analysis, it is necessary to adjust a model's hyperparameters. A grid search for model performance optimization was carried out with the 10-fold cross-validation technique based on the R^2 metric. Table 2 shows the RF hyperparameter ranges and the optimized values detected by the grid search.

Table 2. The Random Forest (RF) grid search hyperparameters.

Parameter	Range	Optimum Value
n_estimators	70 to 150	130
max_features	[Auto, SQRT, Log2]	SQRT
min_samples_split	[2,4,8]	2
bootstrap	[True, False]	False

2.3.4. Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) is a successful machine learning library based on a gradient boosting algorithm proposed by Tianqi Chen [50]. It has better control against overfitting by using more regularized model formalization, in comparison to prior algorithms. It has a high rate of success in Kaggle competitions, particularly for the structured features [50]. The XGBoost similar to the random forest is tuned using hyperparameters. A grid search on hyperparameters with 10-fold cross-validation was carried out to find the best model based on R^2 metrics (see Table 3).

Table 3. Extreme Gradient Boosting regression modeling hyperparameters from the grid search.

Parameter	Range	Optimum Value
n_estimators	70 to 1000	200
max_depth	1 to 10	8
gamma	0.1 to 1	0.7
min_child_weight	3 to 10	8

2.3.5. Deep Learning

Deep learning is one of the machine learning methods that is based on its ancestor—the Artificial Neural Network (ANN) [48,51]. Due to significant developments in hardware and algorithms, deeper hidden layers with more neurons per layer can be implemented, performing deep neural network modeling. These developments have allowed deep learning to progress from research to industrial applications. Recently, it has been shown to have comparable performance to human experts [52–54]. For PM_{2.5} prediction with deep learning, there have been some attempts with different structures, including a deep neural network, long and short term memory (LSTM), and a convolutional neural network (CNN) [33,35,37,38,55,56]. One of the most challenging problems in deep learning methods is missing values. The missing values in our study area were very high, and thus CNN and LSTM could not be applied. Therefore, in this study, we used a six-layer deep neural network with an Adam optimizer (see Figure 2). Here, L2 and L1 regularization were applied to the layers to avoid the model over-fitting issue [57]. The structure of the deep neural network used in this study is presented in Table 4.

Table 4. The deep learning layer configuration. A six-layer neural network with a “relu” activation function that is equipped with regularization to avoid overfitting, was used.

Layer	Layer Type	Neurons Count	Regularization Type	Regularization Value	Activation Function
1	Input	270	None	0	relu
2	Hidden	120	L2	0.002	relu
3	Hidden	70	L2	0.002	relu
4	Hidden	50	L2	0.002	relu
5	Hidden	20	L2, L1	0.001, 0.001	relu
6	Output	1	None	0	relu

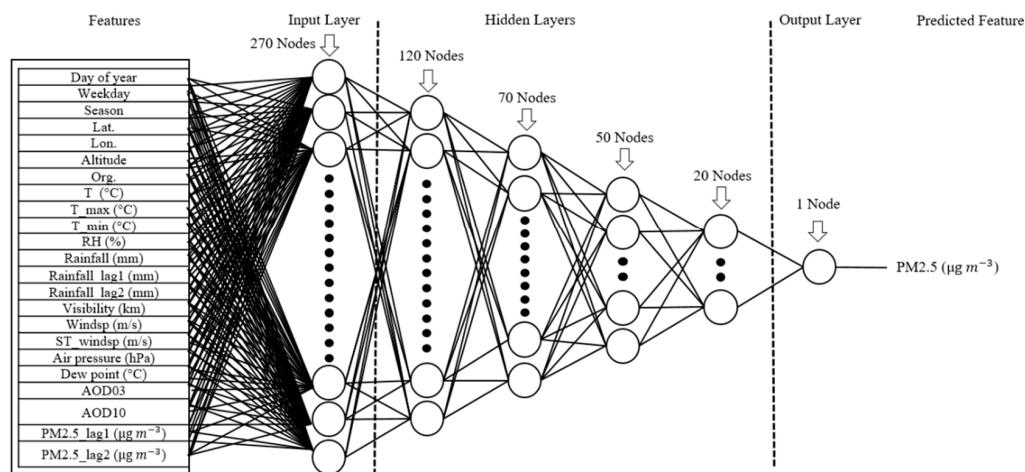


Figure 2. The deep neural network configuration with six layers (4 hidden layers) employed in this study to predict PM_{2.5} concentration value.

2.3.6. Feature Importance Assessment

A Spearman correlation coefficient analysis was carried out, to evaluate the mutual associations of the features (for more details please refer to Supplementary Material Section S2). Moreover, features were analyzed for their importance in PM_{2.5} concentration prediction. This provides a better understanding of the trained model and feature importance. Eliminating the features with negative or neutral effect on the model’s performance can improve the cost and prediction performance.

In this study, we used three methods for feature importance estimation. First, we utilized built-in functions of random forest and XGBoost regression that estimate feature importance, based on the impurity variance of decision tree nodes, a fast but not perfect method. Second, features permutation was implemented. In this step, the performance (R^2 , MAE, and RMSE) of a well-trained deep neural network considering all features was obtained. Performance of the trained model was evaluated by permuting just one of the features in each round. Lower performance can be seen for permuted features with higher importance. This method is fast and reliable and has very low computation cost. Third, XGBoost was used for recursive feature elimination. Here, MAE metrics were used for model performance evaluation. In the first step, a model was trained using all features and the performance of the model was measured. In step 2, model training and performance calculation were repeated by excluding just one of the features at once and including the others. In the third step, the excluded feature that caused the highest prediction performance (lowest MAE) was removed from the features list and this step was repeated, until just three features remained for modeling.

3. Results

Descriptive statistics of meteorological data and APM sites for the entire 4 year study period are illustrated in the Supplementary Materials Section S3 and Tables S1 and S2. The missing value rate for most of the variables is less than 1.3% and for visibility only is 10.19%. The missing value for PM_{2.5} was

approximately 54.11%. This is caused by a critical failure of stations or power outage, maintenance, and so on [17]. The AOD03 data show a high rate of missing values of approximately 94.09%. Thus, although there is an improvement in version 6.1 of the MODIS AOD product, it is not acceptable for this study area. The proportion of missing values in AOD10 is approximately 63.13%, which is better in comparison with AOD03. However, the product's spatial resolution (10 km) is not high enough and multiple APM stations will share the same AOD value, while sampling the AOD10. The minimum and maximum altitude of APM stations are 1023 m and 1758 m above sea level. Five of the 42 APM stations have no instrument for PM_{2.5} concentration measurement. The histograms of features are illustrated in Section S4 Figure S1 of Supplementary Materials. The PM_{2.5}, AOD03, AOD10, windsp, and air_pressure histograms have an almost bell shaped distribution, while the other features show no special uniform distribution.

3.1. Model Performance Validation

In this study, we used AOD derived satellite data, in addition to ground measured climatic data and 37 APM stations, to predict the PM_{2.5} concentration. Three methods of machine learning—RF, XGBoost, and deep learning—were used for predictions. The dataset size has a significant impact on model training performance, particularly for the deep neural network. The AOD03 has a high rate of missing values (94%). Therefore, we conducted three tests, including AOD03 and excluding AOD03 and excluding both AODs, for each training. In addition, records with missing values were excluded from training and test datasets. All of the 1900 (including AOD03) and 11,800 (excluding AOD03) non-missing records out of a 41.2 k data size were used for training the models. The R², MAE, and RMSE metrics were used to evaluate and compare the performance of the three methods.

3.1.1. Random Forest

The optimum configuration of random forest modeling was obtained using the built-in grid search function in Python, with the 10-fold cross-validation technique. The optimum values are shown in Table 2. For records size, the best performance is seen while excluding the AOD03 from the model input features. The prediction metrics show R² = 0.78 (R = 0.88), MAE = 10.8 µg/m³, and RMSE = 14.54 µg/m³. The predicted vs. observed scattered points are distributed around the y = x reference line and the density of points is closer to the reference line. This exhibits a reasonable prediction of PM_{2.5} (see Figure 3a,b). Excluding both AODs shows almost the same performance as a test with AOD10.

3.1.2. XGBoost

The scatter plot of predicted PM_{2.5} versus observed values using the XGBoost method is illustrated in Figure 4. Figure 4a shows the scatterplot of predicted vs. observed PM_{2.5} values considering all parameters. Figure 4b shows the scatterplot of predicted vs. observed PM_{2.5} values excluding the AOD03 variable. The scattered points are distributed around the y = x reference line, demonstrating a reasonable prediction of PM_{2.5} concentration. Excluding both AODs shows the same performance as a test with AOD10.

3.1.3. Deep Learning

Deep neural networks are very sensitive to the input features range and easily become unstable during the training process. In the first step, 1900 (including AOD03) and 11,800 (excluding AOD03) records were selected, based on non-missing records. Moreover, a standard scaler was used to normalize the features to (-1,1). The model was trained based on 70% of selected records that were shuffled in advance. The scatter plots of predicted PM_{2.5} concentration versus observed values, considering all features, is illustrated in Figure 5. The best model performance was achieved by excluding AOD03 from the input features, with R² = 0.77 (R = 0.88), MAE = 10.99 µg/m³, and RMSE = 14.86 µg/m³. Although the R² value obtained by deep learning is lower than for RF and XGBoost, the distribution

of predictions vs. observed points around the $y = x$ reference line still demonstrates acceptable performance. Excluding both AODs shows the same performance as a test with AOD10.

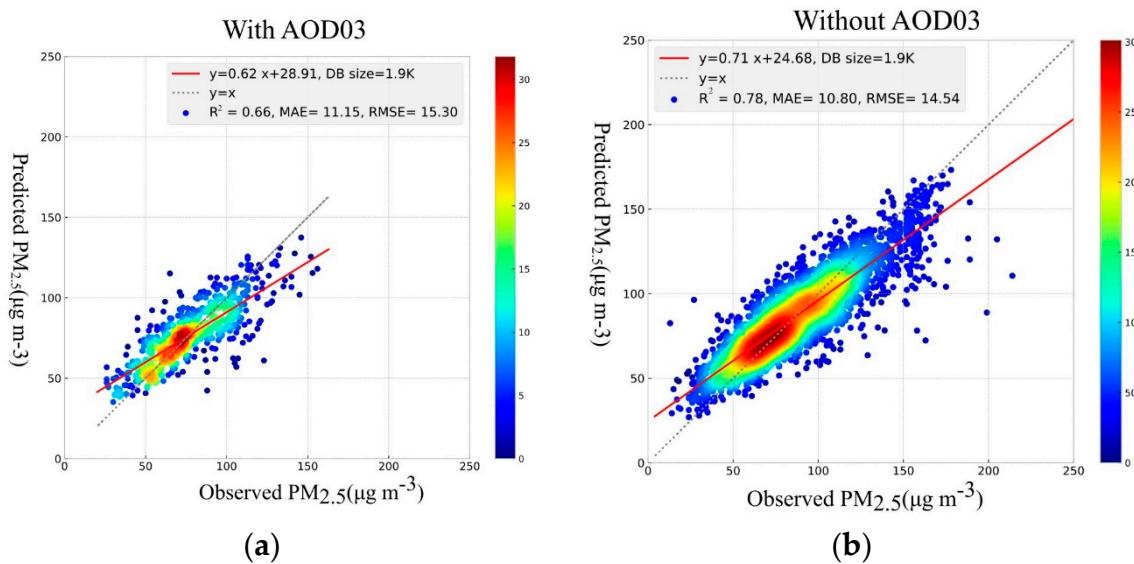


Figure 3. Scatter plot of predicted PM_{2.5} versus observed values using the RF method. The total dataset record was 41.2 k, with 1900 (including AOD03) and 11800 (excluding AOD03) non-missing records available and used for training the model. (a) Scatter plot of predicted vs. observed PM_{2.5} including AOD03. (b) Scatter plot of predicted vs. observed PM_{2.5} excluding the AOD03 variable. The line in red shows a linear regression between observed and predicted PM_{2.5} values for the test dataset.

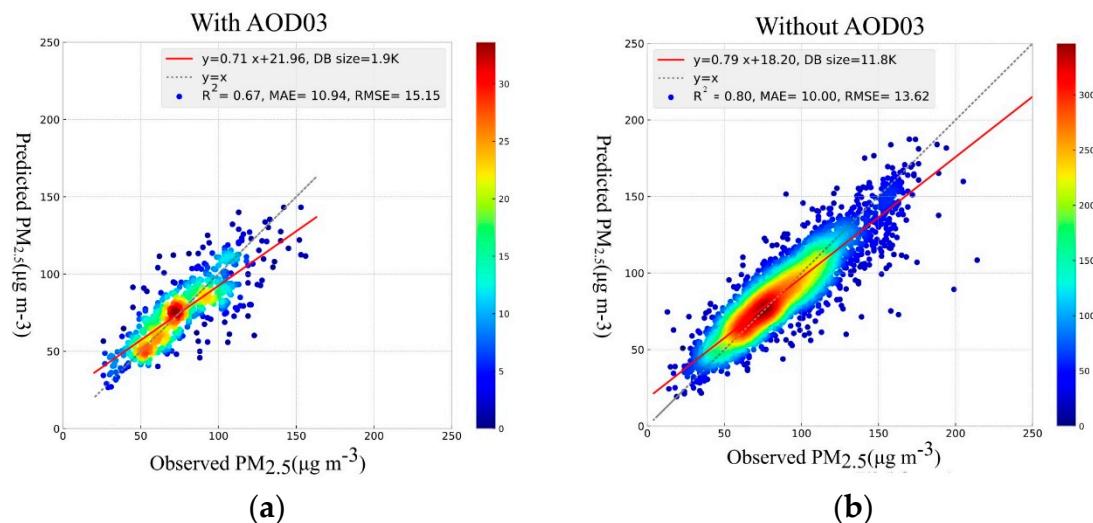


Figure 4. Scatter plot of predicted PM_{2.5} versus observed values using the XGBoost method. (a) Scatter plot of predicted vs. observed PM_{2.5} including AOD03. (b) Scatter plot of predicted vs. observed PM_{2.5} excluding the AOD03 variable. The line in red shows a linear regression between observed and predicted PM_{2.5} values for the test dataset.

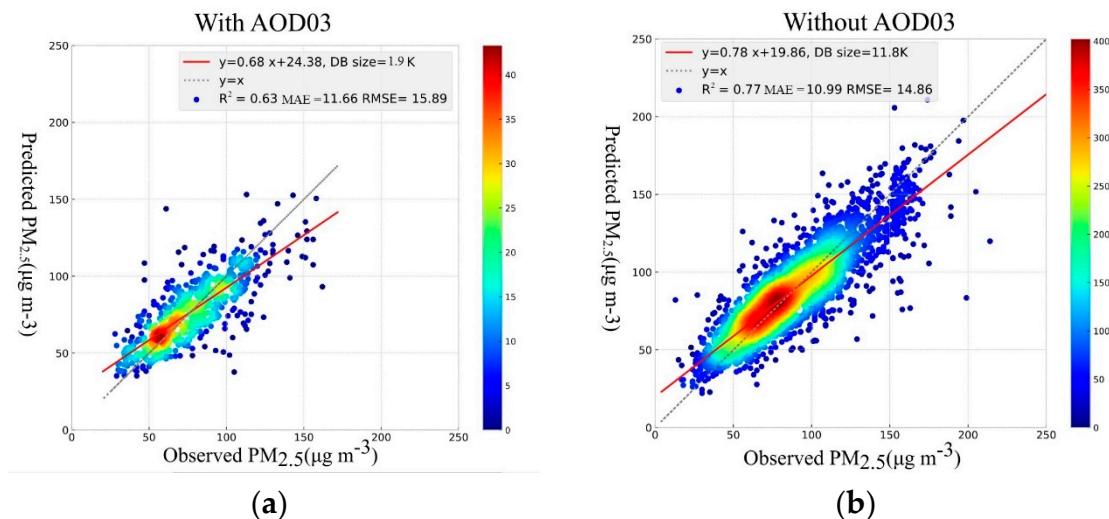


Figure 5. Scatter plot of predicted PM_{2.5} versus observed values using the deep learning method. (a) Scatter plot of predicted vs. observed PM_{2.5} including AOD03. (b) Scatter plot of predicted vs. observed PM_{2.5} excluding the AOD03 variable. The line in red shows a linear regression between observed and predicted PM_{2.5} values for the test dataset.

The results for all three modeling approaches with and without AOD03 and without both AODs, are shown in Table 5. The XGBoost method demonstrates the highest model performance with $R^2 = 0.8$ ($R = 0.894$), MAE = 10.0 $\mu\text{g}/\text{m}^3$, and RMSE = 13.62 $\mu\text{g}/\text{m}^3$, while excluding the AOD03. This shows that AOD03 is not a good feature for PM_{2.5} concentration prediction. In addition, excluding both AODs did not reduce the performance.

Table 5. Three machine learning methods (RF, XGBoost, and deep learning) used to predict PM_{2.5} for 37 air quality monitoring stations. The R^2 , MAE, and RMSE metrics were used to evaluate the prediction accuracy. The study period was from 2015 to the end of 2018.

Method	Include	Record Size	R^2	MAE ($\mu\text{g}/\text{m}^{-3}$)	RMSE ($\mu\text{g}/\text{m}^{-3}$)	Time-Cost (s)
Random Forest	AODs ¹	1900	0.66	11.15	15.30	02
Random Forest	AOD10	11800	0.78	10.80	14.54	17
Random Forest	No AODs	11800	0.78	10.78	14.47	17
XGBoost	AODs	1900	0.67	10.94	15.15	03
XGBoost	AOD10	11800	0.80	10.00	13.62	19
XGBoost	No AODs	11800	0.80	10.00	13.66	19
Deep Learning	AODs	1900	0.63	11.66	15.89	30
Deep Learning	AOD10	11800	0.77	10.88	14.65	87
Deep Learning	No AODs	11800	0.76	11.12	15.11	76

¹ AODs stands for both AOD10 and AOD03.

Therefore, it can be inferred that other features can act as a substitute for AODs. This decreases the importance of AODs on PM_{2.5} prediction. In addition, sample size has significant impact on modeling and prediction performance. Considering the performance metrics, all three ML methods demonstrate almost similar performance. The R^2 values varied from 0.63 to 0.67, excluding AOD03, and 0.77 to 0.80 with AOD10 and without AODs. The best model performance was obtained with the XGBoost ML method, with a very low time cost of 19 s.

3.2. Feature Importance Assessment

3.2.1. RF and XGBoost Feature Importance Ranking

Some features do not contribute to the modeling and only increase the complexity of the model. Therefore, we conducted a feature importance assessment, to detect and eliminate useless features. The RF and XGBoost have a built-in function that evaluates the features importance. The feature importance bar graph plot based on RF and XGBoost modeling is shown in Figures 6 and 7. The features are sorted based on their importance. In both RF and XGBoost, PM2.5_lag1 and visibility show significant importance compared to the other features.

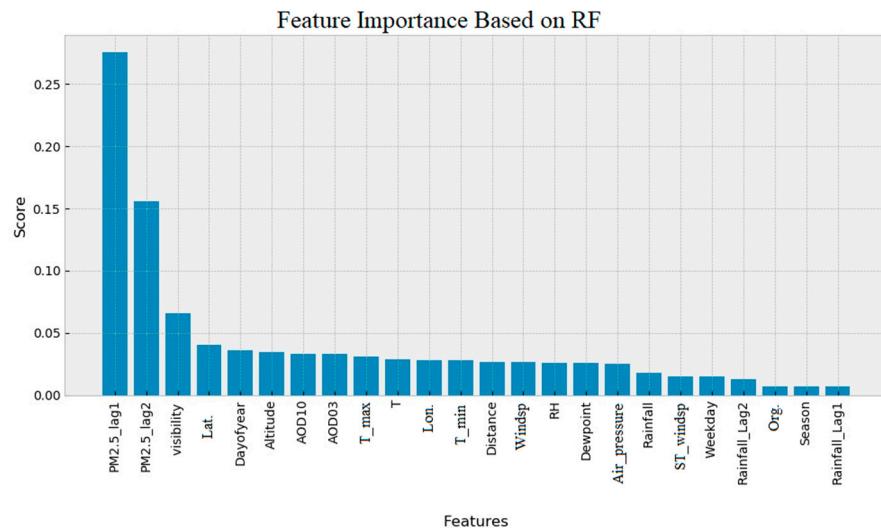


Figure 6. Feature importance bar graph based on random forest modeling.

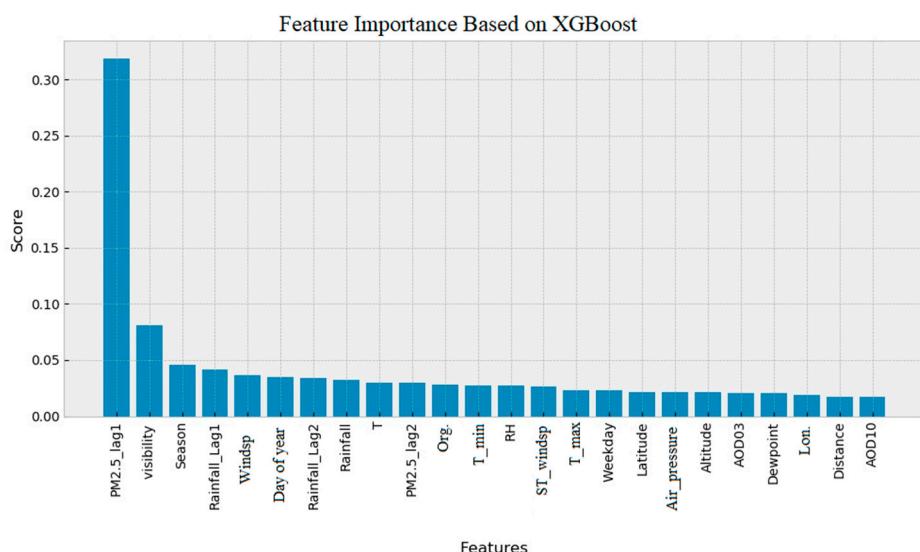


Figure 7. Feature importance bar graph based on the XGBoost feature importance built-in function.

However, there are large differences between RF and XGBoost feature importance ranking results. For example, AOD10 has the lowest rank in XGBoost feature ranking, while AOD10 is ranked seventh in the feature importance ranking by the RF method. Some studies have reported that the feature importance ranking built-in function of RF is biased and unreliable [58] and suggest carrying out the features permutation for feature importance ranking.

3.2.2. Feature Permutation Using Deep Neural Network

The Table 6 shows the feature permutation impact on the prediction performance of a well-trained DNN model. It is reasonable that by permuting an important feature, lower prediction performance be obtained.

Considering the feature importance ranking obtained by features permutation, we repeated DNN training 23 times. In round one, PM2.5_lag1 was used as the input feature and model performance was measured. In the second round of training, features with the rank of 1 and 2 were used as input features. This procedure was repeated to cover all 23 features. In each step, the R^2 value was measured to evaluate model performance. The best model performance during this procedure was obtained using the 15 most important features (from PM2.5_lag1 to dew point), with R^2 of 0.776. In the Table 6 column “ R^2 based on ranking”, the result of this procedure is presented, and less important features are marked with bold font. A negative effect on R^2 value was observed after adding more features. Therefore, the best DNN model performance using useless feature reduction is a R^2 of 0.776.

Table 6. Features permutation of a well-trained deep neural network (DNN) model and features permutation effect on the prediction performance.

Permuted Feature	R^2	MAE ($\mu\text{g m}^{-3}$)	RMSE ($\mu\text{g m}^{-3}$)	Ranking	R^2 Based on Ranking
PM2.5_lag1	0.21	20.63	27.32	1	0.528
Windsp	0.53	15.09	21.06	2	0.564
Visibility	0.54	15.09	20.92	3	0.613
ST_windsp	0.57	14.48	20.26	4	0.620
RH	0.58	14.64	20.08	5	0.704
T_min	0.61	14.62	19.27	6	0.718
Altitude	0.62	14.28	19.03	7	0.737
T	0.64	13.58	18.58	8	0.741
PM2.5_lag2	0.66	13.48	18.02	9	0.740
Day of year	0.68	13.07	17.50	10	0.749
Air_pressure	0.68	12.98	17.37	11	0.752
T_max	0.69	12.91	17.28	12	0.758
Season	0.69	12.80	17.21	13	0.763
Weekday	0.69	12.98	17.20	14	0.774
Dew point	0.71	12.26	16.49	15	0.776
AOD10	0.72	12.15	16.32	16	0.776
Rainfall_Lag2	0.72	11.93	16.25	17	0.771
Distance	0.73	11.97	16.08	18	0.773
Lat.	0.73	11.99	16.07	19	0.765
Rainfall_Lag1	0.74	11.70	15.82	20	0.768
Lon.	0.75	11.70	15.56	21	0.760
Rainfall	0.75	11.33	15.41	22	0.771
Org. ¹	0.75	11.41	15.40	23	0.760
Well Trained Model	0.77	10.88	14.65	-	-

¹ Org. stands for Organization.

3.2.3. MAE Based Feature Elimination Using XGBoost

In addition to the other methods explained above, we conducted a recursive XGBoost training procedure, with feature removal based on MAE metrics. In the first step, a model was trained using all features and the performance of the model measured using MAE metrics. In the second step, the training was repeated 23 times, by removing one of the features at each round and measuring MAE metrics. In the third step, feature with the lowest effect on model performance (lowest MAE) was removed from the total features. The procedure was repeated using 22 features and so on until three features remained. The results are illustrated in Figure 8.

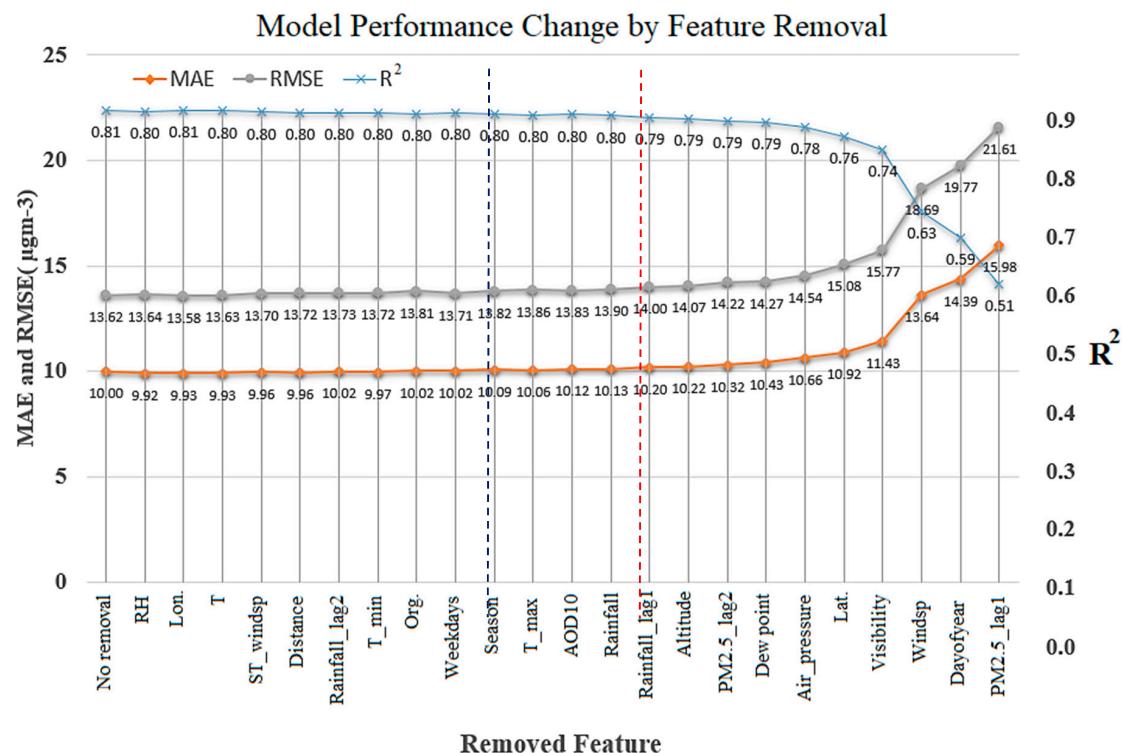


Figure 8. Feature removal using the XGBoost machine learning method, based on MAE metrics. In each step, one feature was removed based on its impact on model performance. From the dotted red line to the right side, features have higher impact on model performance than features on the left side. From the left side of the figure to the blue line, MAE is still below $10 \mu\text{g}/\text{m}^3$.

These results demonstrate that removing RH in the first step improved the model. Model performance did not decrease when RH, longitude, T, sustained wind speed, distance, rainfall_lag2, T_min, org., and weekdays (located on the left side of the dashed blue line in Figure 8) were removed. In addition, removing season, T_max, AOD10, and rainfall did not change the R^2 value and had a small effect on MAE and RMSE. Feature dependency may be the reason for the low changes in model performance. The best model performance obtained by this method was $R^2 = 0.81$.

4. Discussion

Using different methods for feature importance evaluation, we achieved slightly different results. However, in most of the methods, historical observations of PM_{2.5}, wind speed, visibility, day of the year, altitude, and temperature were very important in modeling. The features importance ranking based on different modeling approaches is presented in Table 7. The rankings median value for each feature is presented in Table 7. Features are sorted from top down, based on the median of feature importance rankings. Features such as latitude are important for RF and XGBoost, but rank lower in the deep learning features permutation method. This is because some features are dependent and can be replaced by other features.

The features used in this study can be divided into three categories. The first category is features that directly carry spatial information, such as latitude, longitude, and altitude. The second category is features that indirectly carry APM station spatial information, such as AOD10, AOD03, PM2.5_lag1, and PM2.5_lag2. The third category is the parameters that are shared for all stations, such as meteorological data and day of year, day of week, and sampling season.

Table 7. Features importance ranking based on different modeling approaches. The median value of rankings for each feature is calculated and shown as a median ranking column. Features from top down are sorted based on median ranking.

Features	Ranking					R^2 Based on Median of Rankings Using XGBoost
	Permuted Features DNN	RF Built in	XGBoost Built in	XGB Feature Removal	Median of Rankings	
PM2.5_lag1	1	1	1	1	1	0.509
Visibility	3	3	2	4	3	0.597
Windsp	2	13	5	3	4	0.699
Day of year	10	5	6	2	5.5	0.761
Altitude	7	6	19	9	8	0.776
PM2.5_lag2	9	2	10	8	8.5	0.776
T	8	9	9	21	9	0.783
Lat.	19	4	17	5	11	0.784
T_min	6	11	12	17	11.5	0.785
T_max	12	8	15	13	12.5	0.792
RH	5	14	13	23	13.5	0.794
Air_pressure	11	16	18	6	13.5	0.799
Season	13	22	3	14	13.5	0.797
AOD10	16	7	23	12	14	0.800
Rainfall	22	17	8	11	14	0.798
Dew point	15	15	20	7	15	0.800
Rainfall_Lag1	20	23	4	10	15	0.799
Weekday	14	19	16	15	15.5	0.800
ST_windsp	4	18	14	20	16	0.804
Rainfall_Lag2	17	20	7	18	17.5	0.803
Distance	18	12	22	19	18.5	0.803
Org.	23	21	11	16	18.5	0.805
Lon.	21	10	21	22	21	0.805

The Spearman's correlation coefficient heat map of features is shown in Figure 9. The PM_{2.5} historical observation values have the highest correlation to PM_{2.5}. The air pressure, AOD03, and AOD10 have positive correlation with the PM_{2.5}. Visibility, wind speed, rainfall, altitude, and latitude show a negative correlation with the PM_{2.5}. Also, high correlation between other features reveals features dependency on each other. Dependent features can be predicted using other features, and subsequently can be eliminated from modeling to reduce the model complexity and cost of prediction.

Air temperature and pressure, dew point, and RH are dependent on altitude. We did not use the exact meteorological parameters for each station because of a lack of data; however, meteorological parameters can be modeled and predicted for each station based on altitude and other available features. Based on the Spearman correlation heat map, RH and air pressure are highly correlated with temperature. Considering the cost of prediction, they can be used as substitutes for each other. Some features such as temperature, RH, and pressure have a seasonal trend, and thus the feature "day of year" can facilitate modeling and improve the performance. Its ranking varies from 2 to 10 with a median value of 5.5.

In this study, three-machine learning techniques (RF, Deep learning, and XGBoost) were used to predict PM_{2.5} concentration. The XGBoost technique demonstrated the highest performance and an acceptable time of training. To detect features importance, permutation and recursive feature removal, in addition to RF and XGBoost built-in functions, were used. Some of differences in features importance ranking could be a result of features dependency. However, overall, the features rankings obtained in this paper are logical and beneficial for future studies.

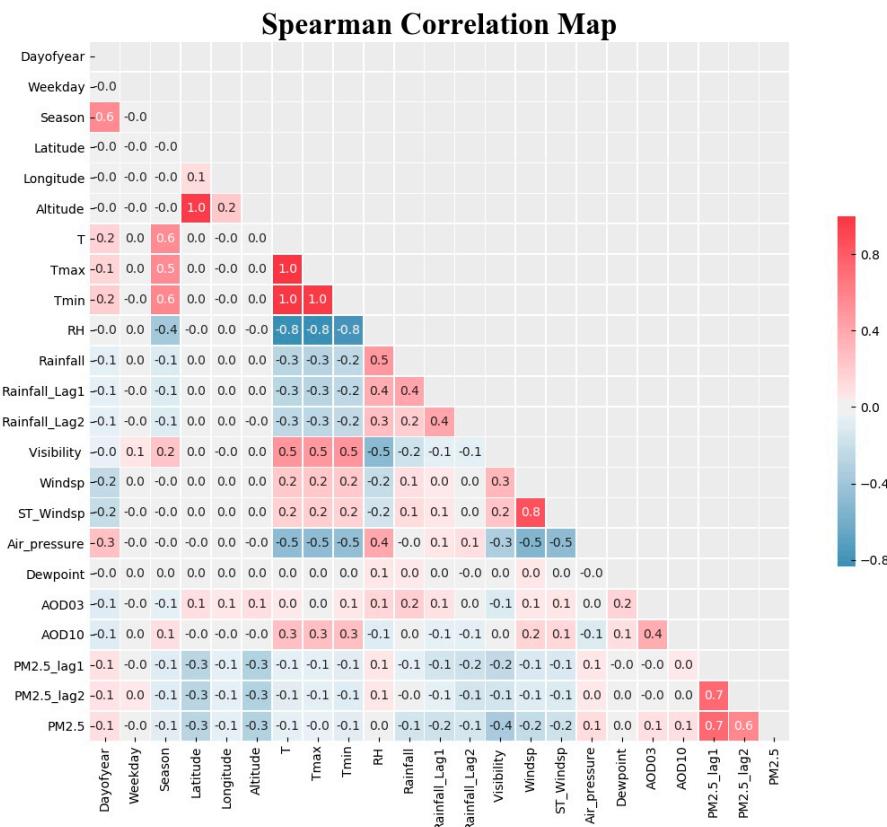


Figure 9. Spearman's correlation coefficient heat map for the study variables shown above. Positive correlations are marked by red while negative correlations are marked by blue.

5. Conclusions

In this study, we utilized Random forest, XGBoost, and Deep learning machine learning techniques to predict PM_{2.5} concentration in Tehran's urban area. Widely distributed ground measured PM_{2.5} data, meteorological features, and remote sensing AOD data were used. In previous research, different methods and features were employed for PM_{2.5} concentration prediction. However, few studies dealt with the limitations of our study area, including the high rate of PM_{2.5} and AOD missing values. In addition, the air pollution monitoring sites in our study area were densely distributed, with just one available weather station. We also utilized 3 and 10 km MODIS AOD products, and geographic properties of the monitoring stations such as latitude, longitude, and topography. Also included were historical observation values of PM_{2.5} and rainfall, in addition to day of year, day of week, and season. Features importance and correlation were evaluated using the Spearman correlation method, permutation, recursive feature removal, and default built-in functions of the XGBoost and RFF techniques.

In comparison to RF and Deep learning methods, XGBoost achieved the best performance of approximately $R^2 = 0.81$ ($R = 0.9$), MAE = 09.92 $\mu\text{g}/\text{m}^3$, and RMSE = 13.58 $\mu\text{g}/\text{m}^3$, with very low cost of time (19 s). Although a DNN model was used for modeling and prediction, XGBoost with its simple structure, performed better. However, all three ML methods performed similarly and R^2 varied from 0.63 to 0.67, when Aerosol Optical Depth (AOD) at 3 km resolution was included, and 0.77 to 0.81, when AOD at 3 km resolution was excluded. Based on feature importance ranking, we found that there are features with high dependency on other features. Therefore, some features can be ranked differently based on machine learning structure. We investigated 23 features and determined that by using eight to 12 features, we can achieve acceptable PM_{2.5} prediction performance. For example, with MAE based XGBoost feature removal, by using only nine of the most important features, such as PM2.5_lag1, day of year, wind speed, visibility, latitude, air pressure, dew point, PM2.5_lag2, and

altitude (see Figure 8), an acceptable performance of $R^2 = 0.79$ ($R = 0.888$), $MAE = 10.20 \mu\text{g}/\text{m}^3$, and $RMSE = 14 \mu\text{g}/\text{m}^3$ was obtained.

Most notably, this is the first study, to our knowledge, to investigate the importance of features for $\text{PM}_{2.5}$ concentration prediction. New features such as latitude, longitude, altitude, and dew point, in addition to day of year, day of week, and season were utilized in a way that has not been done in previous work. However, some limitations are worth noting. Although we have achieved reasonable $\text{PM}_{2.5}$ prediction performance, satellite-derived AODs did not have a significant impact on predictions. Yet, historical values of $\text{PM}_{2.5}$ are necessary for reasonable $\text{PM}_{2.5}$ prediction. In particular, AOD03 has a very high rate of missing values. Thus, it is not useful for our study area. Spatial distribution pattern prediction of $\text{PM}_{2.5}$ is limited without historical values of $\text{PM}_{2.5}$. Future work will focus on images with high spatial resolution, based on the important features introduced in this research.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-4433/10/7/373/s1>, Figure S1: The histogram bar plot of features, Table S1: Descriptive statistics of climatic parameters, $\text{PM}_{2.5}$, and AODs, Table S2: Descriptive statistics of $\text{PM}_{2.5}$ at APM stations of Tehran. The list is sorted based on the rate of missing values for $\text{PM}_{2.5}$ parameter.

Author Contributions: Conceptualization, M.Z.J. and X.N.; Formal analysis, M.Z.J.; Investigation, M.Z.J.; Methodology, M.Z.J. and X.N.; Software, M.Z.J.; Supervision, C.C.; Validation, M.Z.J.; Visualization, M.Z.J.; Writing—original draft, M.Z.J.; Writing—review & editing, M.Z.J., X.N., B.B., and S.T.

Funding: The study was supported by the project of the National Key R&D Program of China “Research of Key Technologies for Monitoring Forest Plantation Resources” (2017YFD0600900) and the National Natural Science Foundation of China “Research of Remote Sensing Inversion Algorithm for Forest Biomass Based on Allometric Scale and Resource Limited Model” (grant no. 41701408).

Acknowledgments: The authors would like to thank the anonymous reviewers whose comments significantly improved this manuscript. Three authors, Mehdi Zamani Joharestani, Barjeece Bashir, and Somayeh Talebiesfandarani acknowledge the University of Chinese Academy of Sciences (UCAS), the Chinese Academy of Sciences (CAS), and the World Academy of Sciences (TWAS) for awarding the CAS-TWAS President’s Fellowship and support to carry out this research. The authors would like to acknowledge the Iran Meteorological Organization for climatic data, Tehran Municipality and Department of Environment for air pollution data, and NASA LAADS DAAC for aerosol optical depth data.

Conflicts of Interest: The authors declare no conflicts of interest

References

1. Riojas-Rodríguez, H.; Romieu, I.; Hernández-Ávila, M. Air pollution. In *Occupational and Environmental Health*; Oxford University Press: Oxford, UK, 2017; pp. 345–364, ISBN 9780190662677.
2. Brunekreef, B.; Holgate, S.T. Air pollution and health. *Lancet* **2002**, *360*, 1233–1242. [[CrossRef](#)]
3. Guarnieri, M.; Balmes, J.R. Outdoor air pollution and asthma. *Lancet* **2014**, *383*, 1581–1592. [[CrossRef](#)]
4. Akimoto, H. Global Air Quality and Pollution. *Science* **2003**, *302*, 1716–1719. [[CrossRef](#)] [[PubMed](#)]
5. Wang, Z. Energy and Air Pollution. In *Comprehensive Energy Systems*; Elsevier: Amsterdam, Netherlands, 2018; Volume 1–5, pp. 909–949, ISBN 9780128095973.
6. Nowak, D.J.; Crane, D.E.; Stevens, J.C. Air pollution removal by urban trees and shrubs in the United States. *Urban For. Urban Green.* **2006**, *4*, 115–123. [[CrossRef](#)]
7. Shen, H.; Li, T.; Yuan, Q.; Zhang, L. Estimating Regional Ground-Level $\text{PM}_{2.5}$ Directly From Satellite Top-Of-Atmosphere Reflectance Using Deep Belief Networks. *J. Geophys. Res. Atmos.* **2018**, *123*, 13875–13886. [[CrossRef](#)]
8. Al Hanai, A.H.; Antkiewicz, D.S.; Hemming, J.D.C.; Shafer, M.M.; Lai, A.M.; Arhami, M.; Hosseini, V.; Schauer, J.J. Seasonal variations in the oxidative stress and inflammatory potential of $\text{PM}_{2.5}$ in Tehran using an alveolar macrophage model; The role of chemical composition and sources. *Environ. Int.* **2019**, *417*–427. [[CrossRef](#)]
9. Laden, F.; Schwartz, J.; Speizer, F.E.; Dockery, D.W. Reduction in fine particulate air pollution and mortality: Extended follow-up of the Harvard Six Cities Study. *Am. J. Respir. Crit. Care Med.* **2006**, *173*, 667–672. [[CrossRef](#)] [[PubMed](#)]

10. Evans, J.; van Donkelaar, A.; Martin, R.V.; Burnett, R.; Rainham, D.G.; Birkett, N.J.; Krewski, D. Estimates of global mortality attributable to particulate air pollution using satellite imagery. *Environ. Res.* **2013**, *120*, 33–42. [[CrossRef](#)]
11. Rojas-Rueda, D.; de Nazelle, A.; Teixidó, O.; Nieuwenhuijsen, M.J. Health impact assessment of increasing public transport and cycling use in Barcelona: A morbidity and burden of disease approach. *Prev. Med. (Baltim.)* **2013**, *57*, 573–579. [[CrossRef](#)]
12. Taghvaei, S.; Sowlat, M.H.; Hassanvand, M.S.; Yunesian, M.; Naddafi, K.; Sioutas, C. Source-specific lung cancer risk assessment of ambient PM_{2.5}-bound polycyclic aromatic hydrocarbons (PAHs) in central Tehran. *Environ. Int.* **2018**, *120*, 321–332. [[CrossRef](#)]
13. Shamsoddini, A.; Aboodi, M.R.; Karami, J. Tehran air pollutants prediction based on Random Forest feature selection method. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. ISPRS Arch.* **2017**, *42*, 483–488. [[CrossRef](#)]
14. Arhami, M.; Shahne, M.Z.; Hosseini, V.; Roufigar Haghighat, N.; Lai, A.M.; Schauer, J.J. Seasonal trends in the composition and sources of PM_{2.5} and carbonaceous aerosol in Tehran, Iran. *Environ. Pollut.* **2018**, *239*, 69–81. [[CrossRef](#)] [[PubMed](#)]
15. Arhami, M.; Hosseini, V.; Zare Shahne, M.; Bigdeli, M.; Lai, A.; Schauer, J.J. Seasonal trends, chemical speciation and source apportionment of fine PM in Tehran. *Atmos. Environ.* **2017**, *153*, 70–82. [[CrossRef](#)]
16. Qi, Y.; Li, Q.; Karimian, H.; Liu, D. A hybrid model for spatiotemporal forecasting of PM_{2.5} based on graph convolutional neural network and long short-term memory. *Sci. Total Environ.* **2019**, *664*, 1–10. [[CrossRef](#)] [[PubMed](#)]
17. Shahbazi, H.; Karimi, S.; Hosseini, V.; Yazgi, D.; Torbatian, S. A novel regression imputation framework for Tehran air pollution monitoring network using outputs from WRF and CAMx models. *Atmos. Environ.* **2018**, *187*, 24–33. [[CrossRef](#)]
18. Dehghan, A.; Khanjani, N.; Bahrampour, A.; Goudarzi, G.; Yunesian, M. The relation between air pollution and respiratory deaths in Tehran, Iran—using generalized additive models. *BMC Pulm. Med.* **2018**, *18*. [[CrossRef](#)] [[PubMed](#)]
19. UN-DESA World Urbanization Prospects: The 2018 Revision. *Dep. Econ. Soc. Aff.* **2018**, *2*.
20. Ansari, M.; Ehrampoush, M.H. Meteorological correlates and AirQ + health risk assessment of ambient fine particulate matter in Tehran, Iran. *Environ. Res.* **2019**, *141*–150. [[CrossRef](#)]
21. Faridi, S.; Shamsipour, M.; Krzyzanowski, M.; Künzli, N.; Amini, H.; Azimi, F.; Malkawi, M.; Momeniha, F.; Gholampour, A.; Hassanvand, M.S.; et al. Long-term trends and health impact of PM_{2.5} and O₃ in Tehran, Iran, 2006–2015. *Environ. Int.* **2018**, *114*, 37–49. [[CrossRef](#)]
22. Hadei, M.; Hopke, P.K.; Nazari, S.S.H.; Yarahmadi, M.; Shahsavani, A.; Alipour, M.R. Estimation of mortality and hospital admissions attributed to criteria air pollutants in Tehran metropolis, Iran (2013–2016). *Aerosol Air Qual. Res.* **2017**, *17*, 2474–2481. [[CrossRef](#)]
23. Wang, Z.; Chen, L.; Tao, J.; Zhang, Y.; Su, L. Satellite-based estimation of regional particulate matter (PM) in Beijing using vertical-and-RH correcting method. *Remote Sens. Environ.* **2010**, *114*, 50–63. [[CrossRef](#)]
24. Gupta, P.; Christopher, S.A.; Wang, J.; Gehrig, R.; Lee, Y.; Kumar, N. Satellite remote sensing of particulate matter and air quality assessment over global cities. *Atmos. Environ.* **2006**, *40*, 5880–5892. [[CrossRef](#)]
25. Engel-Cox, J.A.; Holloman, C.H.; Coutant, B.W.; Hoff, R.M. Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality. *Atmos. Environ.* **2004**, *38*, 2495–2509. [[CrossRef](#)]
26. van Donkelaar, A.; Martin, R.V.; Brauer, M.; Kahn, R.; Levy, R.; Verdúzco, C.; Villeneuve, P.J. Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: Development and application. *Environ. Health Perspect.* **2010**, *118*, 847–855. [[CrossRef](#)] [[PubMed](#)]
27. Ma, Z.; Hu, X.; Huang, L.; Bi, J.; Liu, Y. Estimating ground-level PM_{2.5} in China using satellite remote sensing. *Environ. Sci. Technol.* **2014**, *48*, 7436–7444. [[CrossRef](#)] [[PubMed](#)]
28. Geng, G.; Zhang, Q.; Martin, R.V.; van Donkelaar, A.; Huo, H.; Che, H.; Lin, J.; He, K. Estimating long-term PM_{2.5} concentrations in China using satellite-based aerosol optical depth and a chemical transport model. *Remote Sens. Environ.* **2015**, *166*, 262–270. [[CrossRef](#)]
29. Shang, Z.; Deng, T.; He, J.; Duan, X. A novel model for hourly PM_{2.5} concentration prediction based on CART and EELM. *Sci. Total Environ.* **2019**, *651*, 3043–3052. [[CrossRef](#)]
30. Wen, C.; Liu, S.; Yao, X.; Peng, L.; Li, X.; Hu, Y.; Chi, T. A novel spatiotemporal convolutional long short-term neural network for air pollution prediction. *Sci. Total Environ.* **2019**, *654*, 1091–1099. [[CrossRef](#)]

31. Liu, W.; Guo, G.; Chen, F.; Chen, Y. Meteorological pattern analysis assisted daily PM_{2.5} grades prediction using SVM optimized by PSO algorithm. *Atmos. Pollut. Res.* **2019**. [[CrossRef](#)]
32. Delavar, M.; Gholami, A.; Shiran, G.; Rashidi, Y.; Nakhaeizadeh, G.; Fedra, K.; Hatefi Afshar, S. A Novel Method for Improving Air Pollution Prediction Based on Machine Learning Approaches: A Case Study Applied to the Capital City of Tehran. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 99. [[CrossRef](#)]
33. Qin, D.; Yu, J.; Zou, G.; Yong, R.; Zhao, Q.; Zhang, B. A Novel Combined Prediction Scheme Based on CNN and LSTM for Urban PM_{2.5} Concentration. *IEEE Access* **2019**, *7*, 20050–20059. [[CrossRef](#)]
34. Wang, Q.; Zeng, Q.; Tao, J.; Sun, L.; Zhang, L.; Gu, T.; Wang, Z.; Chen, L. Estimating PM_{2.5} concentrations based on MODIS AOD and NAQPMS data over beijing–tianjin–hebei. *Sensors* **2019**, *19*.
35. Li, T.; Shen, H.; Yuan, Q.; Zhang, X.; Zhang, L. Estimating Ground-Level PM_{2.5} by Fusing Satellite and Station Observations: A Geo-Intelligent Deep Learning Approach. *Geophys. Res. Lett.* **2017**, *44*, 11985–11993. [[CrossRef](#)]
36. Ni, X.; Cao, C.; Zhou, Y.; Cui, X.; Singh, R.P. Spatio-temporal pattern estimation of PM_{2.5} in Beijing-Tianjin-Hebei Region based on MODIS AOD and meteorological data using the back propagation neural network. *Atmosphere* **2018**, *9*, 105.
37. Tong, W.; Li, L.; Zhou, X.; Hamilton, A.; Zhang, K. Deep learning PM_{2.5} concentrations with bidirectional LSTM RNN. *Air Qual. Atmos. Health* **2019**, *12*, 411–423. [[CrossRef](#)]
38. Huang, C.J.; Kuo, P.H. A deep cnn-lstm model for particulate matter (PM_{2.5}) forecasting in smart cities. *Sensors* **2018**, *18*, 2220. [[CrossRef](#)] [[PubMed](#)]
39. Zhou, Y.; Chang, F.J.; Chang, L.C.; Kao, I.F.; Wang, Y.S. Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts. *J. Clean. Prod.* **2019**, *209*, 134–145. [[CrossRef](#)]
40. Hadei, M.; Yarahmadi, M.; Jafari, A.J.; Farhadi, M.; Nazari, S.S.H.; Emam, B.; Namvar, Z.; Shahsavani, A. Effects of meteorological variables and holidays on the concentrations of PM₁₀, PM_{2.5}, O₃, NO₂, SO₂, and CO in Tehran (2014–2018). *J. Air Pollut. Health* **2019**. [[CrossRef](#)]
41. Nabavi, S.O.; Haimberger, L.; Abbasi, E. Assessing PM_{2.5} concentrations in Tehran, Iran, from space using MAIAC, deep blue, and dark target AOD and machine learning algorithms. *Atmos. Pollut. Res.* **2019**, *10*, 889–903. [[CrossRef](#)]
42. Tehran's Municipality ICT Website. Available online: airnow.tehran.ir (accessed on 12 May 2019).
43. Air Pollution Monitoring System platform of the Department of Environment. Available online: aqms.doe.ir (accessed on 12 May 2019).
44. Guleria, R.P.; Kuniyal, J.C.; Rawat, P.S.; Thakur, H.K.; Sharma, M.; Sharma, N.L.; Dhyani, P.P.; Singh, M. Validation of MODIS retrieval aerosol optical depth and an investigation of aerosol transport over Mohal in north western Indian Himalaya. *Int. J. Remote Sens.* **2012**, *33*, 5379–5401. [[CrossRef](#)]
45. Portal, NASA Atmosphere Archive & Distribution System (LAADS) Archive. Available online: <https://ladsweb.modaps.eosdis.nasa.gov> (accessed on 12 May 2019).
46. Iran Meteorological Organization. Available online: <http://www.irimo.ir/far> (accessed on 12 May 2019).
47. Junninen, H.; Niska, H.; Tuppurainen, K.; Ruuskanen, J.; Kolehmainen, M. Methods for imputation of missing values in air quality data sets. *Atmos. Environ.* **2004**, *38*, 2895–2907. [[CrossRef](#)]
48. Mousavi, S.S.; Schukat, M.; Howley, E. Deep Reinforcement Learning: An Overview. In *Lecture Notes in Networks and Systems*; Springer: Cham, Switzerland, 2018; Volume 16, pp. 426–440.
49. Ho, T.K. Random decision forests. In Proceedings of the International Conference on Document Analysis and Recognition, ICDAR, Montreal, QC, Canada, 14–15 August 1995; pp. 278–282.
50. Chen, T.; Guestrin, C. XGBoost. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining—KDD '16, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
51. Schmidhuber, J. Deep Learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [[CrossRef](#)] [[PubMed](#)]
52. Kalash, M.; Rochan, M.; Mohammed, N.; Bruce, N.D.B.; Wang, Y.; Iqbal, F. Malware Classification with Deep Convolutional Neural Networks. In Proceedings of the 2018 9th IFIP International Conference on New Technologies, Mobility and Security, NTMS 2018—Proceedings, Paris, France, 26–28 February 2018; pp. 1–5.
53. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; Volume 2016, pp. 770–778.

54. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems 25 (NIPS 2012), neural information processing systems: University of Toronto. Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
55. Li, T.; Shen, H.; Yuan, Q.; Zhang, L. Deep learning for ground-level PM_{2.5} prediction from satellite remote sensing data. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Valencia, Spain, 22–27 July 2018; Volume 2018, pp. 7581–7584.
56. Xie, J. Deep neural network for PM_{2.5} pollution forecasting based on manifold learning. In Proceedings of the 2017 International Conference on Sensing, Diagnostics, Prognostics, and Control, SDPC 2017, Shanghai, China, 16–18 August 2017; Volume 2017, pp. 236–240.
57. Bengio, Y.; Boulanger-Lewandowski, N.; Pascanu, R. Advances in optimizing recurrent networks. In Proceedings of the ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing—Proceedings, Vancouver, BC, Canada, 26–30 May 2013; 2013; pp. 8624–8628.
58. Strobl, C.; Boulesteix, A.-L.; Zeileis, A.; Hothorn, T. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinform.* **2007**, *8*, 25. [CrossRef] [PubMed]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).