Journal of Environmental Health and Sustainable Development

Determining Effective Factors Regarding Weather and Some Types of Air Pollutants in Seasonal Changes of PM_{10} Concentration Using Tree-Based Algorithms in Yazd City

Zohre Ebrahimi-Khusfi [1], Mohsen Ebrahimi-Khusfi [2], Ali Reza Nafarzadegan [3]*, Mojtaba Soleimani-Sardo [4]

[1] *Department of Environmental Science and Engineering, Faculty of Natural Resources, University of Jiroft, Jiroft, Iran.*
[2] *Department of Geography, Yazd University, Yazd, Iran.*
[3] *Department of Natural Resources Engineering, University of Hormozgan, Bandar-Abbas, Hormozgan, Iran.*
[4] *Department of Environmental Science and Engineering, Faculty of Natural Resources, University of Jiroft, Jiroft, Iran.*

ARTICLE INFO

ORIGINAL ARTICLE

Article History:
Received: 13 November 2023
Accepted: 20 January 2024

***Corresponding Author:**
Ali Reza Nafarzadegan
Email:
a.r.nafarzadegan@hormozgan.ac.ir
Tel:
+98 917 1163934

Keywords:
Air Pollution,
Particulate Matter,
Dust,
Machine Learning,
Random Forest.

ABSTRACT

Introduction: This study was carried out with the aim of determining weather parameters and air pollutants affecting seasonal changes of particulate matter of less than 10 microns (PM_{10}) in Yazd city using Random Forest (RF) and extreme gradient boosting (Xgboost) models.

Materials and Methods: The required data was obtained from 2018 to 2022. Levene's test was applied to investigate the significant difference in the variance of PM_{10} values in 4 different seasons, and Boruta algorithm was used to select the best predictive variables. RF and Xgboost models were trained using two-thirds of the input data and were tested using the remaining data set. Their performance was evaluated based on R^2, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Nash–Sutcliffe Model Efficiency Coefficient (NSE).

Results: The RF showed a higher performance in predicting PM_{10} in all the study seasons ($R^2 > 0.85$; RMSE < 22). The contribution of dust concentration and relative humidity in spring PM_{10} changes was more than other variables. For summer, wind direction and ozone were identified as the most important variables affecting PM_{10} concentration. In the autumn and winter, air pollutants and dust concentration had the greatest effect on PM_{10}, respectively.

Conclusion: RF model could explain more than 85% of PM_{10} seasonal variability in Yazd city. It is recommended to use the model to predict the changes of this air pollutant in other regions with similar climatic and environmental conditions. The results can also be useful for providing suitable solutions to reduce PM_{10} pollution hazards in Yazd city.

Citation: Ebrahimi-Khusfi Z, Ebrahimi-Khusfi M, Nafarzadegan AR, et al. ***Determining Effective Factors Regarding Weather and Some Types of Air Pollutants in Seasonal Changes of PM_{10} Concentration Using Tree-Based Algorithms in Yazd City*.** J Environ Health Sustain Dev. 2024; 9(1): 2180-94.

Introduction

Particulate matter with a diameter of less than 10 μm (PM_{10}) is composed of a variety of solid and liquid compounds originating from crustal matter and anthropogenic activities [1, 2]. PM_{10}, as one of the most important air pollutants, has adverse effects on public health and the environment [3, 4]. Increasing the incidence of cardiovascular and respiratory diseases and mortality rate are among the adverse consequences of degraded air quality due to the increase of PM_{10} in urban areas [5-8].

Complex interactions between different

[DOI: 10.18502/jehsd.v9i1.15115]

parameters have a great impact on spatial and temporal variations of PM_{10} concentrations. Although anthropogenic activities play an important role in particular matters variations [9, 10], several studies have linked these variations to changes in meteorological conditions and some other air pollutants [11-14]. Therefore, it is important to conduct a comprehensive investigation to recognize the contribution of PM_{10} drivers to the control and management of pollution caused by this air pollutant.

Previous studies focused on exploring the correlations between some climatic parameters and PM_{10} [15, 16], and some other stressed the spatial prediction of PM_{10} based on different models of machine learning (ML) [17-19]. However, predicting the seasonal changes of PM_{10} and determining the share of factors affecting it in different seasons has not been extensively studied; however, it has received more attention in the present study.

According to the previous works, temperature, rainfall, wind speed, wind direction, air pressure, sunshine duration, and relative humidity have been identified as the most important meteorological drivers controlling PM_{10} concentrations which can be used to model and estimate PM_{10} in urban environments [14, 20, 21].

In addition to climatic drivers, dust events also affect changes in particular matters concentration especially in desert environments [22-24]. Therefore, dust-related indicators such as dust events frequency can be used to estimate and model PM_{10}. Dust concentration (DC) changes from one place to another place, especially over cities located in arid regions and adjacent to dust-prone areas [25, 26]. Therefore, it can play an important role in changing the concentration of particular matters and can be considered as another dust-related indicator to predict temporal and spatial variation of PM_{10}; but, so far it has not been used in air pollutants modeling studies. Hence, in this study, along with climatic variables mentioned above, DC and total dust event frequency (TDF) was also used to estimate daily changes in PM_{10} concentration in Yazd city, and to determine its relative importance in estimating PM_{10} in different seasons. The city is located near Yazd-Ardakan plain, which is one of the most sensitive dust-prone areas in Central Iran, and the dust entering it, has endangered the health of the inhabitants of this region [27].

In modeling and predicting PM_{10} concentration, some models of ML have shown better performance than others. The least absolute shrinkage and selection operator (LASSO), support vector regression (SVR), random forest (RF), K-Nearest Neighbour (KNN), extreme gradient boosting (Xgboost), and artificial neural network (ANN) were evaluated for spatial prediction of PM_{10} concentrations in Ankara, and the best performance was obtained from ANN model [17]. Among the RF, Xgboost, and artificial neural networks (ANN), the RF and Xgboost models had the best performance in predicting PMs [28], and between the Gradient boosted regression (GBR) and RF, the second model [29] has been successfully applied to predict PM_{10}. The high capacity of the RF model compared to Naïve Bayes (NB) and KNN for spatial prediction of PM_{10} in Italy has been proven in another study [19]. The RF ,bagged classification, and regression trees (Bagged CART) models showed the same and higher performance compared with the mixture discriminate analysis (MDA) in estimating PM_{10} in Barcelona Province [18]. In general, based on recent works, the RF, Xgboost, ANN, and Bagged CART models are optimal models for estimating PM_{10} concentrations. In this study, tree-based models (RF and Xgboost) were used to estimate daily PM_{10} concentrations across Yazd city.

It is worth noting that in studies on PM_{10} estimation and prediction, the selection of the most appropriate estimators is very important because the existence of a collinearity effect between them can lead to a decrease in estimation accuracy. This issue has not been considered in most of the recent studies [17, 19, 29], while the authors have tried to choose the best estimators to predict the target variable (PM_{10}) using various techniques such as removing variables with minimum variance and Boruta algorithm [30]. Generally, the main objectives of this study were (i) to identify optimal variables among the air pollutants, meteorological factors,

[DOI: 10.18502/jehsd.v9i1.15115]

TDF, and DC for predicting daily PM_{10} concentration in different seasons, (ii) to evaluate the performance of RF and Xgboost models in modeling and estimating PM_{10}, and (iii) to determine the relative importance of best PM_{10} predictors in different seasons in Yazd city.

Materials and Methods

Study city

The study area of the current research is Yazd city located in the center of Iran. The area of the city is about 1629 square kilometers, and it is located in the path of dusty winds originating from Yazd-Ardakan plain, one of the critical centers of dust production in central Iran. There are three air pollutants measurement stations in Yazd city. After checking the statistics of these stations, it was found that the station located in Sanat Square has many statistical deficiencies. To investigate the objectives of this research, the information related to the other two stations, which have a common statistical period from 2018 to 2022, as well as their geographical location, was used as shown in Figure 1.

The methodology of the present research is shown in Figure 2, and more details are provided in the following sections.

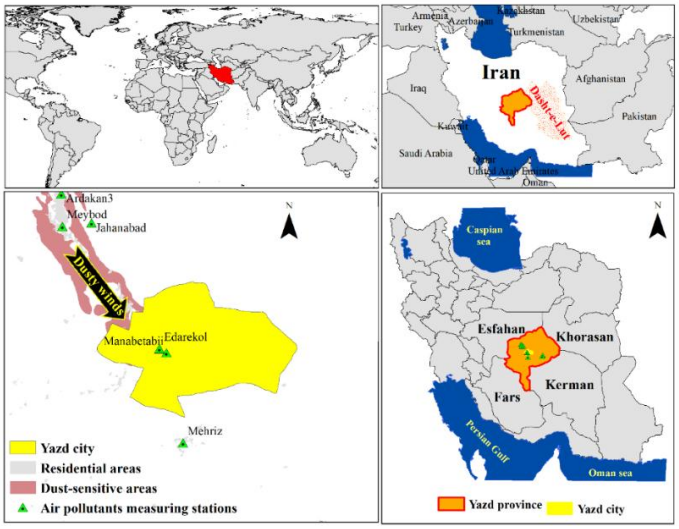


Figure 1: The geographical location of the study area and air pollutants measuring stations in Yazd province

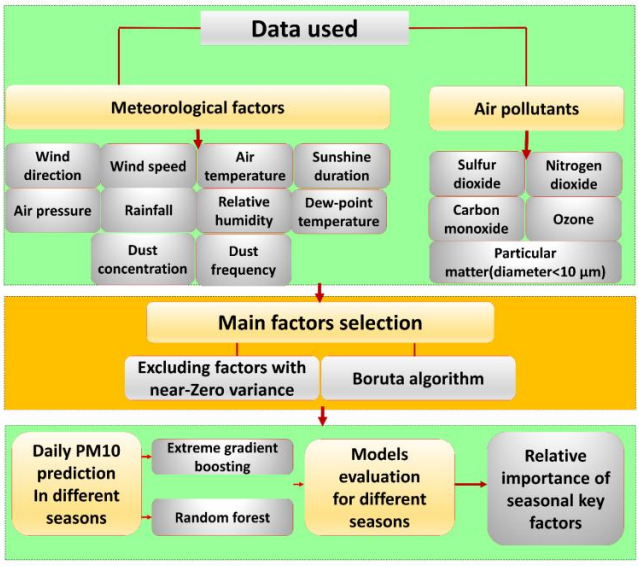


Figure 2: Schematic diagram of estimating process of daily PM_{10} concentration and determining key factors in different seasons

[DOI: 10.18502/jehsd.v9i1.15115]

Data used

In this study, the data related to some air pollutants (SO_2, NO_2, O_3, and CO) as well as weather parameters mentioned in Table 1 were used. Meteorological and air pollutants data were respectively obtained from meteorological and environmental protection organizations of Yazd Province.

Meteorological data including air temperature, rainfall, relative humidity, wind speed, wind direction, sunshine duration, dew-point temperature, and air pressure were collected on a daily scale. Data on horizontal visibility (HV), internal dust-event codes, and external dust-event codes were recorded in synoptic stations. Internal events are recorded with codes 07 to 09, 30 to 35, and 98 while external events are recorded with code 06 in the synoptic stations [31]. In this study, the data recorded at the Yazd synoptic were acquired from the Meteorological Organization of Yazd Province for the same statistical period as air pollutants data. They were utilized to calculate TDF (Equation 1), and DC [32] as expressed by Equations (1-3):

$$TDF = IDF + EDF \quad (1)$$

where, TDF is total dust frequency, and IDF and EDF refer to internal dust-event frequency and external dust-event frequency, respectively.

$$DC(\mu g/m^3) = \frac{3802.29}{visibility^{0.84}} \quad \text{visibility (km)} < 3.5 \quad (2)$$

$$DC(\mu g/m^3) = exp(-0.11 \times visibility + 7.62) \quad \text{visibility (km)} \geq 3.5 \quad (3)$$

DC was calculated in the 3-houlry scales from 2018 to 2022, and daily DC was then calculated by averaging 3-hourly DC for all the dusty days of the study period. For days without dust occurrence, the DC value was considered zero. In general, according to the study background and the researchers' understanding of the factors affecting the PM_{10} concentration, data on sixteen factors were initially collected (Table 1), and based on the techniques described below, the main factors affecting PM_{10} concentrations were then selected to predict daily PM_{10} concentrations during different seasons and analyze the relative importance of the dominant factors affecting them in the Yazd city.

Table 1: Collected variables to predict daily changes of PM_{10} in Yazd city

No.	Description	Abbreviation	Unit
1	Sulfur dioxide	SO_2	ppb
2	Nitrogen dioxide	NO_2	ppb
3	Ozone	O_3	ppb
4	Carbon monoxide	CO	ppm
5	Mean wind speed	WSmean	m/s
6	Air temperature	T_{min}	ºC
7	Air pressure	Pr	hpa
8	Rainfall	Rain	mm
9	Relative humidity	RH	%
10	Sunshine duration	SSD	hours
11	Dew-point temperature	Td	ºC
12	Horizontal visibility	HV	m
13	Wind direction	WD	°
14	Dust concentration	DC	$\mu g/m^3$
15	Total dust frequency	TDF	

Statistical analysis and selection of main predictors

In this stage, at first, the significant difference in the variance of the target variable values (PM_{10}) between the four seasons was investigated using Levene's test.

The existence of near-zero variance (NZV) predictors and the strong correlation between predictor variables can lead to increased errors in modeling and decrease its accuracy (Kuhn 2008

[DOI: 10.18502/jehsd.v9i1.15115]

[33]). For this reason, it is necessary to identify such variables before modeling. Predictors with NZV can be detected based on the fraction of unique values of less than 10% [34] which was used in the present work. The Boruta is also a useful tool for determining the multicollinearity of predictors and selecting the main factors affecting the target variables (PM_{10}, in this study). This algorithm has a 7-step process as follows [35]:

i. A copy of all the variables is added to develop system information with at least five shadows.
ii. The correlation of the added variables with the response variable is removed.
iii. A random forest classification is performed on the extended information system, and the calculated Z-scores are collected.
iv. The maximum Z score is explored among the shadow variables.
v. A two-way equality test is performed for variables whose importance is not determined in the previous step.
vi. Variables with very low importance and shadow variables are removed.
vii. This process is repeated until the importance of all the variables is determined, or the algorithm reaches the predetermined range of RFS.

After completing the mentioned steps, some variables are confirmed for modeling, and others are rejected, and the confirmed variables can be used for modeling and forecasting.

Prediction of daily PM_{10} concentration in different seasons

The models used to predict seasonal PM_{10} changes in Yazd city were RF and Xgboost. The RF trains multiple decision trees in parallel with bootstrap data samples and results in an individual prediction (Breiman, 2001). Bootstrapping guarantees that each tree in the RF is unique and plays an effective role in reducing the overall variance of the classification. This method also has the ability to rank predictor variables based on their role in decreasing regression prediction error. In this method, instead of searching for the most important attributes while dividing a node, the best attributes are searched among a random subset of attributes, leading to a wide diversity and a better model. In general, RF is widely used in various sciences for modeling and forecasting purposes [36-38] due to its simplicity, high diversity, and application in both classification and regression [39], which are suitable capacities. In addition, this model provides a good and strong estimate of the importance of variables using sensitivity analysis, and accordingly [40], there is no need to perform sensitivity analysis. The number of trees (ntree) and the number of randomly selected predictors (mtry) are the optimized hyper-parameters of this model.

Xgboost was another model used in this study. Boosting is a progressive method in which a subset of data is randomly selected to reduce the loss function by creating new tree models [41]. The two models of RF and Xgboost differ in how the trees are made. The Xgboost works better than RF if the data between classes is relatively moderated, there is not much noise in the data, and the parameters are fine-tuned [28]. Using the Xgboost model, a forecasting model in the form of a reinforcement set of weakly gradient descending trees generates that optimizes the loss function [42]. In this model, the major tunable hyper-parameters were a booster, max-depth, min-child-weight, colsample-bytree, subsample, and eta.

In the present work, both models were trained based on two-thirds of the daily data related to PM_{10} and the variables confirmed in the previous step, which were tested based on one-third of the remaining data in the spring, summer, autumn, and winter seasons. The technique used to train the models was 10-cross-correlation with 5 repeats.

Schematic diagram of the RF and Xgboost models are presented in Figure 3.

[DOI: 10.18502/jehsd.v9i1.15115]

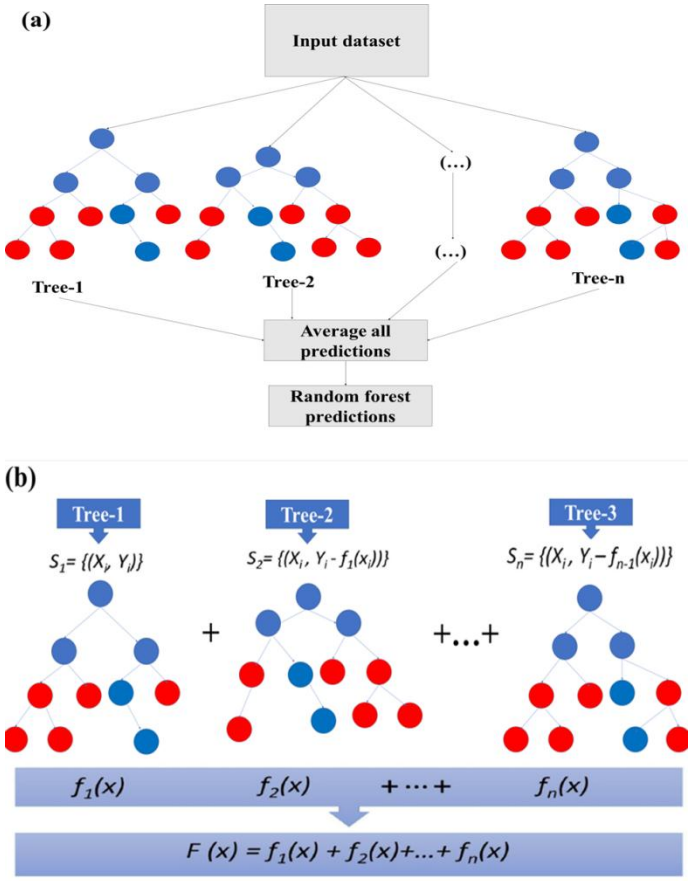


Figure 3: Schematic diagram of RF (a)[43] and Xgboost models (b) [43].

Accuracy assessment of RF and Xgboost models

The performance of the models used to predict seasonal changes in PM_{10} was evaluated using the accuracy metrics presented in Equations 4 to 7:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[PM_{10,obs} - PM_{10,est}\right]^2}, (0 \leq RMSE < +\infty) \quad (4)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|PM_{10,obs} - PM_{10,est}\right|, (0 \leq MAE < +\infty) \quad (5)$$

$$R = \left[\frac{\frac{1}{n}\sum_{i=1}^{n}(PM_{10,obs} - \overline{PM_{10,obs}})(PM_{10,est} - \overline{PM_{10,est}})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(PM_{10,obs} - \overline{PM_{10,obs}})^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(PM_{10,est} - \overline{PM_{10,est}})^2}}\right], (-1 < R \leq +1) \quad (6)$$

$$NSE = 1 - \left[\frac{\sum_{i=1}^{n}\left[PM_{10,obs} - PM_{10,est}\right]^2}{\sum_{i=1}^{n}\left[PPM_{10,obs} - \overline{PM_{10,obs}}\right]^2}\right], (-\infty < NSE \leq 1) \quad (7)$$

where, n is the total number of observations. $(PM_{10,obs})_i$ and PM_{10est} $(PM_{10,est})_i$ refers to the observed and estimated daily PM_{10} , respectively. Also, $\overline{PM_{10,obs}}$ is the averaged observed PM_{10}, and $\overline{PM_{10,est}}$ is the averaged estimated PM_{10}.

Results

Main predictors of daily PM_{10} changes in different seasons

The results obtained from Levene's test indicated that there was a significant difference between the variance of target variable (PM_{10}) seasonal values in the air of Yazd city (Levene Statistic = 2.98 (Sig < 0.05); Table 2). Therefore, the selection of predictive variables and modeling was done separately for each season.

After excluding the variables with NZV, the most important PM_{10} predictors for each season were selected based on Boruta algorithm. The results are presented in Figures 4 and 5, respectively.

[DOI: 10.18502/jehsd.v9i1.15115]

Table 2: Homogeneity of variance for PM_{10} values in all seasons based on Levene's test

Test of homogeneity of variance				
Target variable	**Levene statistic**	**df_1**	**df_2**	**Sig.**
PM_{10}	2.984	3	1172	0.030

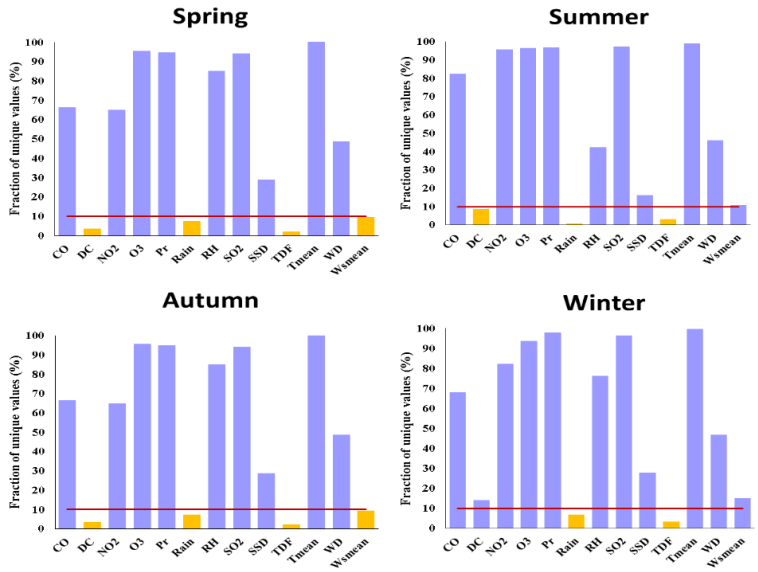


Figure 4: Fraction of unique values related to all the variables used to predict PM_{10} concentrations in different seasons

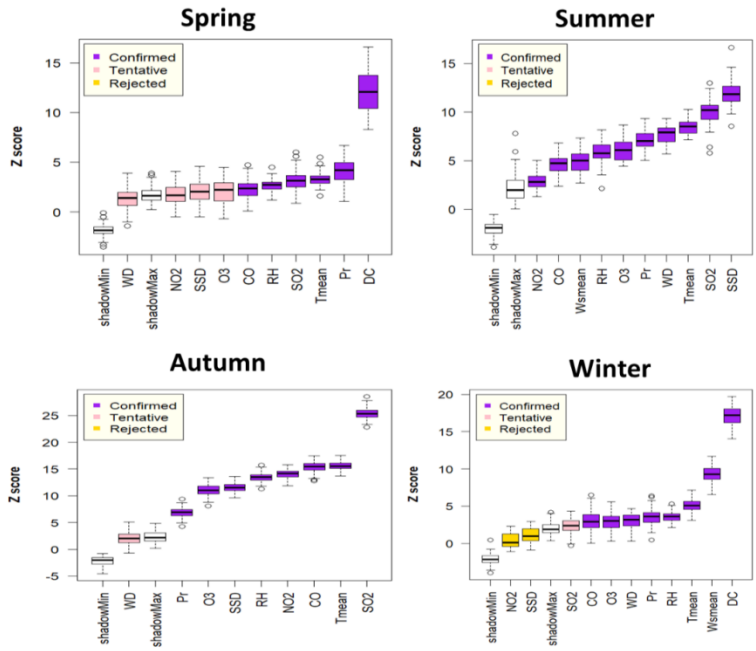


Figure 5: Selected variables by Boruta algorithm for predicting daily PM_{10} concentrations in different seasons

Performance of RF and Xgboost tree models in predicting daily PM_{10} concentrations in different seasons

In this study, in order to predict the concentration of target variable, the data related to the days in which information was recorded for all the confirmed variables in Figure 5 were used. The total number of these days for spring, summer,

[DOI: 10.18502/jehsd.v9i1.15115]

autumn, and winter were 301, 306, 288, and 281 days respectively. The capacity of the two different machine learning models, namely RF and Xgboost, to predict daily PM_{10} concentration in different seasons across the urban environment of Yazd City, was tested using a ten-fold CV with 5 replications. The results are summarized in Table 3. As observed, the values of R^2 and Nash–Sutcliffe Model Efficiency Coefficient (NSE) obtained from RF model for both test and training data sets in all study seasons were higher than the values obtained from the Xgboost model. In contrast, Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) values obtained from RF model were lower than those obtained from the Xgboost model. Accordingly, RF was chosen as the most effective model for predicting daily PM_{10} concentration in all the seasons in Yazd city. The observed and predicted PM_{10} values in testing data set for each season based on the best model (RF) were shown in Figure 6.

Table 3: Performance of RF and Xgboost models for predicting daily PM_{10} concentration in different seasons

RF model								
Evaluation metrics	**R^2**		**RMSE**		**MAE**		**NSE**	
Datasets	Train	Test	Train	Test	Train	Test	Train	Test
Spring	0.92	0.87	29.6	15.1	13.7	10.2	0.84	0.85
Summer	0.91	0.92	19.6	18.4	11	12.4	0.83	0.85
Autumn	0.94	0.94	11.9	12.9	8.5	9.3	0.93	0.92
Winter	0.93	0.88	28.9	21.6	14.3	12.2	0.88	0.87
Xgboost model								
Evaluation metrics	**R^2**		**RMSE**		**MAE**		**NSE**	
Datasets	Train	Test	Train	Test	Train	Test	Train	Test
Spring	0.49	0.47	49.8	28.5	26.1	21.5	0.46	0.45
Summer	0.65	0.64	29.7	31	18.4	21.9	0.63	0.59
Autumn	0.95	0.95	9.3	10.8	7.1	7.8	0.95	0.94
Winter	0.91	0.80	26.9	27	18.7	16.9	0.90	0.80

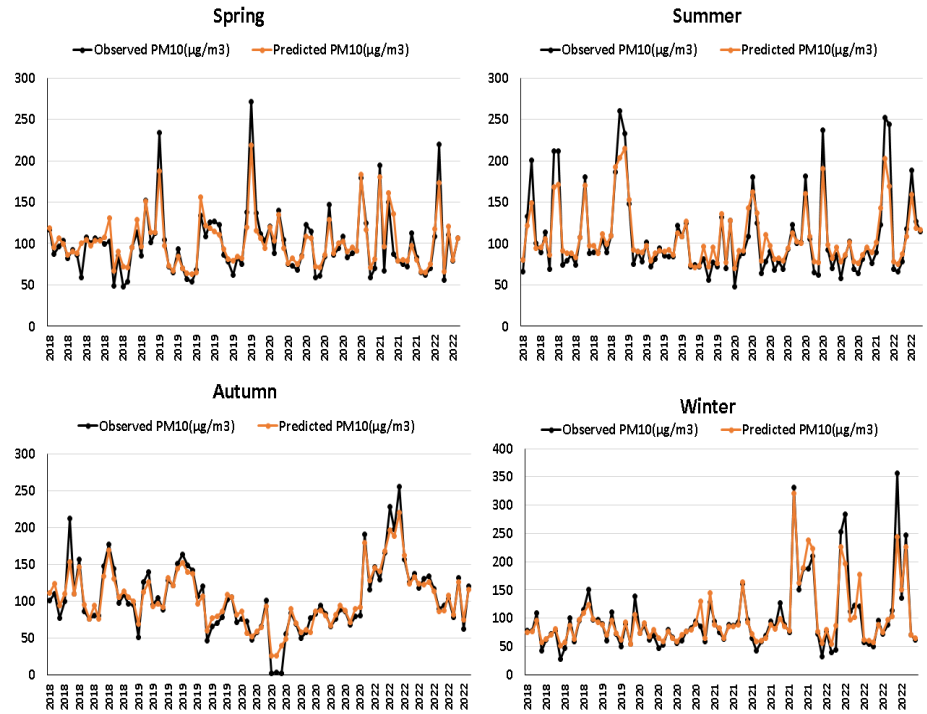
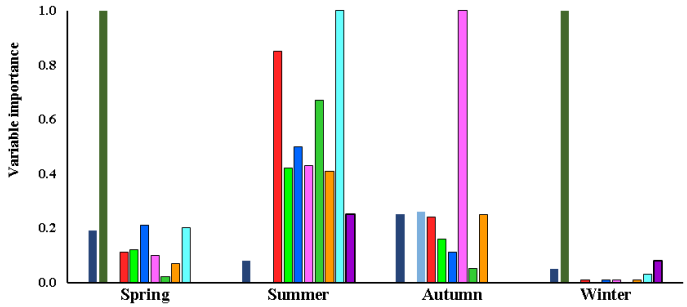


Figure 6: Observed and predicted PM_{10} concentrations based on the best model in testing data sets for different seasons

[DOI: 10.18502/jehsd.v9i1.15115]

The relative importance of factors affecting daily PM_{10} concentrations in different seasons

Relative importance of factors influencing PM_{10} in different seasons was reported using RF model because it was selected as a more efficient model in the previous step. The values were normalized between 0-1 and were presented in Figure 7.



	Spring	Summer	Autumn	Winter
CO	0.19	0.08	0.25	0.05
DC	1			1.0
NO2			0.26	
O3	0.11	0.85	0.24	0.01
Pr	0.12	0.42	0.16	0
RH	0.21	0.5	0.11	0.01
SO2	0.1	0.43	1	0.01
SSD	0.02	0.67	0.05	
Tmean	0.07	0.41	0.25	0.01
WD	0.2	1		0.03
Wsmean		0.25		0.08

Figure 7: Relative importance of main factors affecting daily PM_{10} concentration in different seasons in Yazd city

The results showed that for predicting daily PM_{10} concentration in the spring, DC, RH, WD, CO, and O_3 were the most important factors, respectively. The WD followed by O_3, SSD, and RH had a greater role in predicting daily PM_{10} in the summer. The most important predictor variables of daily PM_{10} in the autumn were SO_2, NO_2, T, CO, and O_3, respectively. DC had played the most important part in predicting daily PM_{10} in the winter, while other variables made a much smaller contribution. In general, the variable of DC in spring and winter, SO_2 in autumn, and WD in summer had a greater impact in predicting the daily PM_{10}, compared to other parameters.

Discussion

Particulate matter ≤ 10 µm (PM_{10}), as one of the most important air pollutants, have adverse effects on public health and the environment Therefore, it is of great important to predict and identify the factors affecting their concentration changes in different regions. In this study, after removing the variables with low variance, Boruta algorithm was used to select the best predictive variables of PM_{10} seasonal changes in weather. According to the results of this study, in spring and autumn, the fraction of unique values was less than 10% for 4 variables of DC, Rain, TDF, and Wsmean, and a higher percentage for other variables. In summer, the lowest values of variance were observed in DC, Rain, and TDF variables, and in the winter season, it was only observed for the last two variables. Therefore, they were excluded, and other variables were used to introduce Boruta algorithm in each season separately. According to Boruta algorithm results, some variables were not confirmed to model and predict daily PM_{10} concentrations in the study seasons. Unconfirmed and tentative variables for predicting PM_{10} in spring were WD, NO_2, SSD, and O_3 while in autumn; the only variable was NO_2 variable. For winter, NO_2, SSD, and SO_2 were not confirmed for predicting daily PM_{10}, while for summer, all the variables were confirmed. Finally, the remaining parameters (marked in blue in Figure 4) were confirmed and used to predict the target variable in different seasons. In one study, T_{max}, T_{min}, WS_{max}, WD_{max}, evaporation, and rainfall were selected as independent variables for PM_{10} prediction in all the seasons in Isfahan [44]. In addition to these parameters, maximum and

[DOI: 10.18502/jehsd.v9i1.15115]

minimum values of relative humidity and sunshine hours were chosen to predict seasonal PM_{10} in Ahvaz city [45]. Most of the climate variables selected in the previous studies were confirmed in the current study and were largely consistent with the findings of this stage in the present study. Kliengchuay W, et al., found that some air pollutants such as NO_2 and CO also had an effect on the daily concentration of PM_{10}, and in the context of choosing these variables, it could be concluded that the findings of this study were consistent with the reports of these researchers [14]. One of the reasons regarding the similarity of the selected parameters for predicting the PM_{10} in Isfahan and Ahvaz in all the seasons was that all the climatic variables entered the modeling process without performing collinear analysis. This was while this important issue was taken into account in the present study. For this purpose, Boruta algorithm was used separately to select predictor variables in each season.

In this study, the performance of RF and Xgboost models was evaluated based on R^2, RMSE, MAE, and NSE. Compared to Xgboost model, RF showed a higher performance in predicting PM_{10} in all the seasons and was selected as the best predictive model. The adjusted R^2 value in this model was almost 0.9 for all the seasons, indicating that confirmed variables using Boruta algorithm could explain almost 90 % variability of daily PM_{10} in different seasons. The successful application of the RF model for predicting PM_{10} was also proved by Mallet [29] and Tella Balogun [19], which confirmed the findings obtained from this study. The higher accuracy of RF for PM_{10} prediction could be due to the following strengths. This model was highly iterative in nature, which made bootstrap data points for robust and stable forecasts. There were user-friendly OpenSource R libraries, some of which were designed to handle large numbers of input variables. Furthermore, the model had an inherent high capacity to handle nonlinear and high-order interactions between predictors [46]. In addition, RF approach had less restrictive assumptions and was more flexible [47].

In general, some variables such as CO, O_3, RH, SO_2, and T_{mean} had been selected to predict seasonal PM_{10} changes in Yazd city. This suggested that the changes in all the mentioned variables had an impact on pollution changes caused by PM_{10} in the weather during all the seasons. In addition to these parameters, the changes of some variables such as spring and winter DC, summer and winter WS mean, and autumn NO_2 had influenced the changes of PM_{10} in the weather of Yazd city. These results showed the complex relationship between climatic parameters and PM_{10} in different seasons. Such behaviors had also been reported in some of the previous studies regarding PM_{10} and other air pollutants in different parts of the world. The production and destruction of O_3 in Baghdad city was affected by different levels of solar radiation in different seasons[48], and seasonal changes of air dust had different effects on surface solar radiation in Taklamakan desert [49, 50]. Soil moisture followed by air pollution hours, soil heat flux, air pressure, vegetation cover, wind direction, and evaporation were identified as the most important environmental factors affecting PM_{10} changes in Isfahan [51]. Although Isfahan is located in the vicinity of Yazd city, the relative importance of environmental factors affecting PM_{10} changes in these cities was different. Difference in the study period selected for Isfahan (2013 to 2019) and Yazd (2018 to 2022), as well as the study scale selected, which was annual for Isfahan and seasonal for Yazd, were the main reasons for the difference in the selection of some predictor variables of PM_{10} in these cities.

In general, the high contribution of DC in predicting PM_{10} concentrations in the spring and winter indicated the local origin of dust particles in these seasons in Yazd. One of these sources was the Yazd-Ardakan plain, where a the large amount of soil particles is separated from the soil surface every year due to the phenomenon of wind erosion [52], and because the direction of the prevailing winds is from the side of this plain towards the city of Yazd [53], a large part these particles are transferred to this city. Considering that the increase in DC was caused by the increase in the frequency of dust events and dusty days, and the fact that there was a significant

[DOI: 10.18502/jehsd.v9i1.15115]

relationship between seasonal changes in PM_{10} and dusty days in Birjand city [54], with relatively similar climatic conditions to Yazd, the findings of another researcher confirmed the findings of the present research. Probably, the cold weather and increase in the traffic of automobiles, followed by the increase in the consumption of fossil fuels in the autumn, were the reasons for the increase in sulfur dioxide gas and the greater contribution of this factor in predicting the PM_{10} in this season. The significant contribution of WD in predicting PM_{10} concentrations in summer indicated that regional sources had a great impact on PM concentration in the dusty city of Yazd in this season. Guerra and Lane [55] pointed out a significant relationship between the wind direction and the concentration of PMs, which to some extent confirmed findings of this study. Although the mentioned variables were influential in predicting the concentration of PM_{10} in the study area, the influence of other climatic parameters and air pollutants should not be ignored. For example, in winter, one of the reasons for increasing air pollution was air temperature inversions [56], and this might be the reason why this variable appeared in the list of the main variables predicting winter PM_{10} concentration. The effect of temperature fluctuations on PM_{10} concentration variations had been proven in a study conducted by Plocoste and Calif [15]. Wind speed affected PM_{10} concentration by affecting the horizontal dissemination [57]. Moreover, on warm days, the wind can increase the concentration of particular matters by removing soil particles and road dust [58]. In several studies, sunshine duration had been identified as a key factor affecting the concentration of PM_{10}, which was consistent with the findings of this research. Solar radiation was also identified as another major factor affecting the change in PM_{10} concentration in different seasons because its changes were a function of changes in the concentration of airborne particles, cloud covers, and anthropogenic pollutions in the urban areas [59, 60]. In agreement with the findings of this study, the important role of some drivers such as wind speed, wind direction, air pressure, relative humidity, air temperature, rainfall, and O_3 in PMs changes had also been reported by Duarte and Schneider [61] for urban environments. Most of these variables were successfully used to predict PM_{10} concentration in previous studies [18, 29] and confirmed outcomes of this study.

Conclusion

Analysis of daily changes in air pollutants concentrations and determining the factors influencing these changes can help to better manage air quality in urban environments. In this work, the efficiency of RF and Xgboost models for predicting daily PM_{10} concentrations in different seasons was evaluated, and the most important factors affecting it were identified by the best fitted model. The major findings of this study are as follows:

- According to the accuracy metrics, the RF model is better than the Xgboost model for predicting the daily PM_{10} concentration in all seasons.
- DC has the greatest relative importance in predicting daily PM_{10} concentrations in spring and winter.
- WD and SO_2 are identified as the most important factors affecting the daily changes of PM_{10} concentration in summer and autumn, respectively.

One of the limitations of this study was the lack of access to information on air pollutants in all the stations of Yazd city in a similar and long-term period. Furthermore, the incompleteness of PM_{10} information in other air pollution monitoring stations in Yazd province was one of the limitations of this study regarding the analysis of spatial variations of PM_{10} at a larger scale. Having such information provides a broader insight into areas with a higher risk of pollution in the province. Therefore, it is suggested that if PM_{10} information is recorded in the coming years for all the air pollution measuring stations in Yazd province, they should be used for spatial analysis of PM_{10} changes at the provincial scale.

Acknowledgments

This research was supported by the University of Jiroft under grant NO: 4812-01-1.The authors would like to thank the vice-chancellor of

[DOI: 10.18502/jehsd.v9i1.15115]

education and research in the university for his support and the department of environment and the meteorological organization of Yazd province for providing the required information on air pollutants and meteorological parameters.

Conflict of Interest

The authors declared no conflict of interest.

Funding

This research was supported by the University of Jiroft under grant NO: 4812-01-1.

Ethical considerations

The authors fully addressed ethical issues including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publishing and/or submission, redundancy, etc.

Code of ethics

This study was carried out with the aim of determining weather parameters and air pollutants affecting seasonal changes of particulate matter of less than 10 microns using meteorological data and air pollutants measures. The ethics code is not applicable for this research.

Authors' contributions

Zohre Ebrahimi-Khusfi and Ali Reza Nafarzadegan contributed to the study's conception and design. Data collection, statistical analysis and data processing were performed by Zohre Ebrahimi-Khusfi, Mohsen Ebrahimi-Khusfi, and Ali Reza Nafarzadegan. The first draft of the manuscript was written by Zohre Ebrahimi-Khusfi and Mojtaba Soleimani Sardoo. Zohre Ebrahimi-Khusfi and Ali Reza Nafarzadegan were responsible for reviewing and revising the manuscript and supervising the project. All the authors read and approved the final manuscript.

References

1.Taheri Shahraiyni H, Sodoudi S. Statistical modeling approaches for PM_{10} prediction in urban areas; A review of 21st-century studies. Atmosphere. 2016;7(2):15.

2.Kchih H, Perrino C, Cherif S. Investigation of desert dust contribution to source apportionment of PM_{10} and $PM_{2.5}$ from a southern Mediterranean coast. Aerosol Air Qual Res. 2015;15(2):454-64.

3.Kermani M, Arfaeinia H, Masroor K, et al. Health impacts and burden of disease attributed to long-term exposure to atmospheric PM_{10}/$PM_{2.5}$ in Karaj, Iran: effect of meteorological factors. Int J Environ Anal Chem. 2020;102(18):6134-50.

4.Lee S, Hong YC, Park H, et al. Combined effects of multiple prenatal exposure to pollutants on birth weight: The Mothers and Children's Environmental Health (MOCEH) study. Environ Res. 2020;181:108832.

5.Tzima K, Analitis A, Katsouyanni K, et al. Has the risk of mortality related to short-term exposure to particles changed over the past years in Athens, Greece?. Environment International. 2018;113:306-12.

6.Sicard P, Khaniabadi YO, Perez S, et al. Effect of O^3, PM_{10} and $PM_{2.5}$ on cardiovascular and respiratory diseases in cities of France, Iran and Italy. Environ Sci Pollut Res. 2019;26(31): 32645-65.

7.Karimi B, Samadi S. Mortality and hospitalizations due to cardiovascular and respiratory diseases associated with air pollution in Iran: A systematic review and meta-analysis. Atmos Environ. 2019;198:438-47.

8.Tahery N, Geravandi S, Goudarzi G, et al. Estimation of PM_{10} pollutant and its effect on total mortality (TM), hospitalizations due to cardiovascular diseases (HACD), and respiratory disease (HARD) outcome. Environ Sci Pollut Res. 2021;28(17):22123-30.

9.Kumar S, Dwivedi S. Impact on particulate matters in India's most polluted cities due to long-term restriction on anthropogenic activities. Environ Res. 2021;200:111754.

10. Millán-Martínez M, Sánchez-Rodas D, de la Campa AMS, et al. Contribution of

[DOI: 10.18502/jehsd.v9i1.15115]

anthropogenic and natural sources in PM_{10} during North African dust events in Southern Europe. Environ Pollut. 2021;290:118065.
11. Alifa M, Bolster D, Mead M, et al. The influence of meteorology and emissions on the spatio-temporal variability of PM_{10} in Malaysia. Atmos Res. 2020;246:105107.
12. Hassan H, Latif MT, Juneng L, et al. Interaction of PM_{10} concentrations with local and synoptic meteorological conditions at different temporal scales. Atmos Res. 2020;241:104975.
13. Arregocés HA, Rojano R, Restrepo G. Meteorological factors contributing to organic and elemental carbon concentrations in PM_{10} near an open-pit coal mine. Environ Sci Pollut Res. 2022;29(19):28854-65.
14. Kliengchuay W, Worakhunpiset S, Limpanont Y, et al. Influence of the meteorological conditions and some pollutants on PM_{10} concentrations in Lamphun, Thailand. J Environ Health Sci Eng. 2021;19(1):237-49.
15. Plocoste T, Calif R. Is there a causal relationship between Particulate Matter (PM_{10}) and air Temperature data? An analysis based on the Liang–Kleeman information transfer theory. Atmos Pollut Res. 2021;12(10):101177.
16. Chuchro M, Sarlej W, Grzegorczyk M, et al. Application of Photo texture analysis and weather data in assessment of air quality in terms of airborne PM_{10} and $PM_{2.5}$ Particulate Matter. Sensors. 2021;21(16):5483.
17. Bozdağ A, Dokuz Y, Gökçek ÖB. Spatial prediction of PM_{10} concentration using machine learning algorithms in Ankara, Turkey. Environ Pollut. 2020;263:114635.
18. Choubin B, Abdolshahnejad M, Moradi E, et al. Spatial hazard assessment of the PM_{10} using machine learning models in Barcelona, Spain. Sci Total Environ. 2020;701:134474.
19. Tella A, Balogun AL, Adebisi N, et al. Spatial assessment of PM_{10} hotspots using Random Forest, K-Nearest Neighbour and Naïve Bayes. Atmos Pollut Res. 2021;12(10):101202.
20. Park H, Kim T, Yang M. The effect of meteorological factors on PM_{10} depletion in the atmosphere and evaluation of rainwater quality. Korean Journal of Remote Sensing. 2020;36(63):1733-41.
21. Zalewska T, Biernacik D, Marosz M. Correlations between 7Be, 210Pb, dust and PM_{10} concentrations in relation to meteorological conditions in northern Poland in 1998–2018. J Environ Radioact. 2021;228: 106526.
22. Draxler RR, Gillette DA, Kirkpatrick JS, et al. Estimating PM_{10} air concentrations from dust storms in Iraq, Kuwait and Saudi Arabia. Atmos Environ. 2001;35(25):4315-30.
23. Hussein T, Li X, Al-Dulaimi Q, et al. Particulate matter concentrations in a middle eastern city–an insight to sand and dust storm episodes. Aerosol Air Qual Res. 2020;20:2780-92.
24. Guan Q, Luo H, Pan N, et al. Contribution of dust in northern China to PM_{10} concentrations over the Hexi corridor. Sci Total Environ. 2019;660:947-58.
25. Rezazadeh M, Irannejad P, Shao Y. Climatology of the middle east dust events. Aeolian Res. 2013;10:103-9.
26. Ebrahimi-Khusfi Z, Mirakbari M, Soleimani-Sardo M. Aridity index variations and dust events in Iran from 1990 to 2018. Ann Am Assoc Geogr. 2022;112(1):123-40.
27. Jalili M, Ehrampoush MH, Mokhtari M, et al. Ambient air pollution and cardiovascular disease rate an ANN modeling: Yazd-Central of Iran. Scientific Reports. 2021;11(1):16937.
28. Czernecki B, Marosz M, Jędruszkiewicz J. Assessment of machine learning algorithms in short-term forecasting of PM_{10} and $PM_{2.5}$ concentrations in selected Polish agglomerations. Aerosol Air Qual Res. 2021;21:200586.
29. Mallet MD. Meteorological normalisation of PM_{10} using machine learning reveals distinct increases of nearby source emissions in the Australian mining town of Moranbah. Atmos Pollut Res. 2021;12(1):23-35.
30. Kursa MB, Rudnicki WR. Feature selection with the Boruta package. J Stat Softw. 2010;36(11):1-13.
31. O'Loingsigh T, McTainsh G, Tews E, et al. The Dust Storm Index (DSI): a method for

[DOI: 10.18502/jehsd.v9i1.15115]

monitoring broadscale wind erosion using meteorological records. Aeolian Res. 2014;12:29-40.
32. Shao Y, Yang Y, Wang J, et al. Northeast Asian dust storms: Real-time numerical prediction and validation. Journal of Geophysical Research. 2003;108(D22):4691.
33. Zhou T, Geng Y, Ji C, et al. Prediction of soil organic carbon and the C:N ratio on a national scale using machine learning and satellite data: A comparison between Sentinel-2, Sentinel-3 and Landsat-8 images. Sci Total Environ. 2021;755:142661.
34. Bradley B, Brandon G. Hands-On Machine Learning with R. CRC Press, USA. 2020.
35. Ebrahimi-Khusfi Z, Dargahian F, Nafarzadegan AR. Predicting the dust events frequency around a degraded ecosystem and determining the contribution of their controlling factors using gradient boosting-based approaches and game theory. Environ Sci Pollut Res. 2022;29(24):36655-73.
36. Chen X, Li X. Estimating aerosol optical extinction across eastern China in winter during 2014–2019 using the random forest approach. Atmos Environ. 2022;269:118864.
37. Ruiz-AÍvarez M, Gomariz-Castillo F, Alonso-Sarría F. Evapotranspiration response to climate change in semi-arid areas: using random forest as multi-model ensemble method. Water 2021;13(2):222.
38. Pouyan S, Rahmanian S, Amindin A, et al. Spatial and seasonal modeling of the land surface temperature using random forest. Computers in Earth and Environmental Sciences. 2022:221-34.
39. Lagomarsino D, Tofani V, Segoni S, et al. A tool for classification and regression using random forest methodology: Applications to landslide susceptibility mapping and soil thickness modeling. Environ Model Assess. 2017;22(3):201-14.
40. Polewko-Klim A, Lesiński W, Golińska AK, et al. Sensitivity analysis based on the random forest machine learning algorithm identifies candidate genes for regulation of innate and adaptive immune response of chicken. Poult Sci. 2020;99(12):6341-54.
41. Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. J Anim Ecol. 2008;77(4):802-13.
42. Cui Y, Cai M, Stanley HE. Comparative analysis and classification of cassette exons and constitutive exons. Biomed Res Int. 2017;2017:1-9.
43. Sahour H, Gholami V, Torkaman J, et al. Random forest and extreme gradient boosting algorithms for streamflow modeling using vessel features and tree-rings. Environ Earth Sci. 2021;80:1-14.
44. Masoudi M, Gerami S. Assessment of PM_{10} concentration and its prediction using meteorological parameters in the air of Isfahan, Iran. The Jordan Journal of Earth and Environmental Sciences. 2018;9(2):75-80.
45. Masoudi M, Asadifard E, Rastegar M. Status of PM_{10} as an air pollutant and its prediction using meteorological parameters in Ahvaz, Iran. Eur Respir Rev. 2018;6(2):163-74.
46. Stafoggia M, Bellander T, Bucci S, et al. Estimation of daily PM_{10} and $PM_{2.5}$ concentrations in Italy, 2013–2015, using a spatiotemporal land-use random-forest model. Environment International. 2019;124:170-9.
47. Breiman L. Random forests. Machine Learning. 2001;45:5-32.
48. Nassif WG, Al-Taai OT, Abbood ZM. The influence of solar radiation on ozone column weight over Baghdad city. IOP Conf Ser Mater Sci Eng. 2020;928(7):072089.
49. Fei Y, Xia X, Che H. Dust aerosol drives upward trend of surface solar radiation during 1980–2009 in the Taklimakan Desert. Atmos Sci Lett. 2014;15(4):282-7.
50. Gehlot S, Minnett PJ, Stammer D. Impact of sahara dust on solar radiation at cape verde islands derived from MODIS and surface measurements. Remote Sens Environ. 2015;166:154-62.
51. Ebrahimi-Khusfi Z, Taghizadeh-Mehrjardi R, Kazemi M, et al. Predicting the ground-level pollutants concentrations and identifying the influencing factors using machine learning,

[DOI: 10.18502/jehsd.v9i1.15115]

wavelet transformation, and remote sensing techniques. Atmos Pollut Res. 2021;12(5): 101064.
52. Ekhtesasi M, Sepehr A. Investigation of wind erosion process for estimation, prevention, and control of DSS in Yazd–Ardakan plain. Environ Monit Assess. 2009;159:267-80.
53. Ebrahimi Khusfi Z, Roustaei F, Ebrahimi Khusfi M, et al. Investigation of the relationship between dust storm index, climatic parameters, and normalized difference vegetation index using the ridge regression method in arid regions of Central Iran. Arid Land Res Manag. 2020;34(3):239-63.
54. Ebrahimi A, Ahmadizadeh SR, Rashki A. Variation of PM_{10} and its relationship with dust and climate in Birjand, Iran. Desert. 2022;27(1):97-114.
55. Guerra SA, Lane DD, Marotz GA, et al. Effects of wind direction on coarse and fine particulate matter concentrations in southeast Kansas. J Air Waste Manag **Assoc**. 2006;56(11): 1525-31.
56. Li X, Miao Y, Ma Y, et al. Impacts of synoptic forcing and topography on aerosol pollution during winter in Shenyang, Northeast China. Atmos Res. 2021;262:105764.
57. Galindo N, Varea M, Gil-Moltó J, et al. The influence of meteorology on particulate matter concentrations at an urban Mediterranean location. Water Air Soil Pollut. 2011;215(1):365-72.
58. Kukkonen J, Pohjola M, Sokhi RS, et al. Analysis and evaluation of selected local-scale PM_{10} air pollution episodes in four European cities: Helsinki, London, Milan and Oslo. Atmos Environ. 2005;39(15):2759-73.
59. Jin K, Qin P, Liu C, et al. Impact of urbanization on sunshine duration from 1987 to 2016 in Hangzhou City, China. Atmosphere. 2021;12(2):211.
60. Chakchak J, Cetin NS. Investigating the impact of weather parameters selection on the prediction of solar radiation under different genera of cloud cover: A case-study in a subtropical location. Measurement. 2021;176: 109159.
61. Duarte AL, Schneider IL, Artaxo P, et al. Spatiotemporal assessment of particulate matter (PM_{10} and $PM_{2.5}$) and ozone in a Caribbean urban coastal city. Geosci Front. 2022;13(1): 101168.

[DOI: 10.18502/jehsd.v9i1.15115]

