

Article

Machine Learning for Determining Interactions between Air Pollutants and Environmental Parameters in Three Cities of Iran

Abdullah Kaviani Rad ¹, Redmond R. Shamshiri ^{2,*} , Armin Naghipour ³, Seraj-Odeen Razmi ⁴ ,
Mohsen Shariati ⁵, Forogh Golkar ⁶ and Siva K. Balasundram ⁷ 

¹ Department of Soil Science, School of Agriculture, Shiraz University, Shiraz 71946-85111, Iran; arad@adaptiveagrotech.com

² Leibniz Institute for Agricultural Engineering and Bioeconomy, 14469 Potsdam, Germany

³ Clinical Research Development Center, Imam Reza Hospital, Kermanshah University of Medical Sciences, Kermanshah 67148-69914, Iran; armin.naghipour@kums.ac.ir

⁴ Department of MBA, Faculty of Management, University of Tehran, Tehran 14179-35840, Iran; serajrazmi@ut.ac.ir

⁵ Department of Environmental Planning, Management, and Education, Faculty of Environment, University of Tehran, Tehran 14179-35840, Iran; mohsen.shariati@ut.ac.ir

⁶ Department of Water Engineering & Oceanic and Atmospheric Research Center, College of Agriculture, Shiraz University, Shiraz 71946-85111, Iran; fgolkar@shirazu.ac.ir

⁷ Department of Agriculture Technology, Faculty of Agriculture, University Putra Malaysia, Serdang 43400, Selangor, Malaysia; siva@upm.edu.my

* Correspondence: rshamshiri@atb-potsdam.de



check for updates

Citation: Rad, A.K.; Shamshiri, R.R.; Naghipour, A.; Razmi, S.-O.; Shariati, M.; Golkar, F.; Balasundram, S.K. Machine Learning for Determining Interactions between Air Pollutants and Environmental Parameters in Three Cities of Iran. *Sustainability* **2022**, *14*, 8027. <https://doi.org/10.3390/su14138027>

Academic Editors: Laura Marziali and Laura Fantozzi

Received: 14 May 2022

Accepted: 28 June 2022

Published: 30 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract: Air pollution, as one of the most significant environmental challenges, has adversely affected the global economy, human health, and ecosystems. Consequently, comprehensive research is being conducted to provide solutions to air quality management. Recently, it has been demonstrated that environmental parameters, including temperature, relative humidity, wind speed, air pressure, and vegetation, interact with air pollutants, such as particulate matter (PM), NO₂, SO₂, O₃, and CO, contributing to frameworks for forecasting air quality. The objective of the present study is to explore these interactions in three Iranian metropolises of Tehran, Tabriz, and Shiraz from 2015 to 2019 and develop a machine learning-based model to predict daily air pollution. Three distinct assessment criteria were used to assess the proposed XGBoost model, including R squared (R²), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). Preliminary results showed that although air pollutants were significantly associated with meteorological factors and vegetation, the formulated model had low accuracy in predicting (R²_{PM_{2.5}} = 0.36, R²_{PM₁₀} = 0.27, R²_{NO₂} = 0.46, R²_{SO₂} = 0.41, R²_{O₃} = 0.52, and R²_{CO} = 0.38). Accordingly, future studies should consider more variables, including emission data from manufactories and traffic, as well as sunlight and wind direction. It is also suggested that strategies be applied to minimize the lack of observational data by considering second- and third-order interactions between parameters, increasing the number of simultaneous air pollution and meteorological monitoring stations, as well as hybrid machine learning models based on proximal and satellite data.

Keywords: air pollution; quality; meteorological factors; vegetation; interaction; modeling; machine learning; XGBoost; AQI; Iran



Copyright: © 2022 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As a result of the increasing demand for energy in the previous 50 years, air pollution has expanded dramatically, with a threatening acceleration that cannot be eliminated [1–4]. Approximately 91% of the population of the world lives in regions with high levels of air pollution [5], which contributes to the deaths of seven million people annually [6]. According to Statista [7], nearly 1.1 million Americans live in zones with high levels of PM_{2.5}, and more than 90% of European citizens are exposed to PM exceeding the

WHO standard [8]. Neurological and psychological disruptions, eye irritation, and the progression of various diseases such as asthma, Alzheimer's, Parkinson's, autism, and low birth weight (LBW) are among the short-term and long-term consequences of air pollution [9]. Annual financial damages of O₃ and PM_{2.5} in healthcare sector are estimated to be \$5.5–12.5 billion and \$48.6–140.7 billion, respectively [10]. Premature death and air pollution-related diseases have been documented to cause financial losses in India totaling \$28.8 billion and \$8 billion, respectively [11]. Air pollution causes \$2.9 trillion in economic losses to the global economy [12]. These economic statistics only refer to the financial losses caused by air pollution in the public health sector.

In addition to human health, air pollution is also a significant threat to ecosystems [13]. According to research by Ito et al. [14], plants exposed to NO₂ had a lower dry weight. Examining the effects of air pollution on lichens in northeastern Norway by Hogda et al. [15] showed that lichen-rich areas fell from 30% in 1973 to 1.5% in 1992. Furthermore, Bignal et al. [16] explored the impacts of air pollutant emissions on vegetation along two highways in the UK and observed that the deforestation rate rose. SO₂, O₃, and NO_x can alter the physiological processes that affect plants' growth patterns by damaging the leaf cuticles and affecting the conductivity of the stomata, thereby having a direct impact on the photosynthesis system, leaf longevity, and carbon allocation [17,18]. Moreover, these pollutants can change the competitive balance between plant species and alter the composition of the plant community; thus, they can reduce crop yield and, subsequently, economic effectiveness in agricultural systems [19]. According to assessments, O₃ is responsible for yield losses of 7–12% in wheat and 3–5% in corn. The economic losses from O₃ on 23 crops in Europe in 2000 were estimated to be around €6.7 billion [20]. Vlachokostas et al. [21] reported that the economic damage to O₃-induced crops in Thessaloniki (Greece) is estimated at roughly €43 million per year. Hence, air pollution imperils both public health and economic development.

Presently, air pollution is a substantial concern for emerging Asian economies [22]. Approximately 70% of air pollution-related deaths occur in Asia and the Pacific [23], and 98% of cities with a population of more than 100,000 in these regions do not follow WHO air quality policies [24]. As an Asian country, Iran is facing various environmental challenges such as arid and semi-arid climate, water crisis, soil salinity, desertification, floods, and air pollution [25–27], which ranked 23rd among 106 countries in air pollution [28]. Extensive use of fossil fuels, an antiquated transportation system, and industrial activities besides natural dust are the leading reasons for air pollution in Iran [29]. It has been estimated that a one percent rise in the production of gasoline in Iran would result in a 0.59 percent increase in the country's carbon emissions [30]. In 2014, Iran ranked eighth out of 27 countries in terms of CO₂ emissions from energy consumption [31]. The financial consequences of CO₂ emissions in some Asian countries from 1970 to 2018 are shown in Figure 1.

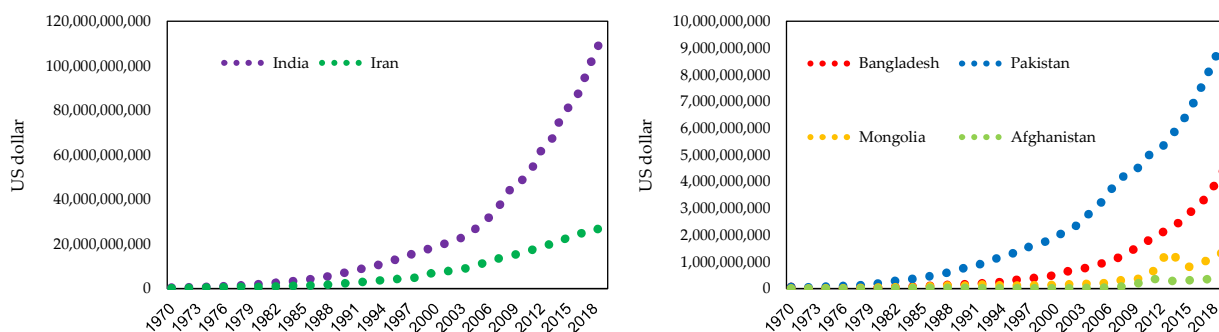


Figure 1. The economic loss caused by CO₂ emissions in some Asian countries, data source: [32].

In order to encounter ecological issues, environmental engineering provides low-cost and practical solutions [33]. Monitoring the concentration of ambient air pollutants to maintain public health and sustainable development is a helpful solution [34]. Recently, various technologies have been employed to monitor air pollution [35]. In Iran, the Department

of Environment (DOE) is monitoring air quality and executing the national strategies to reduce air pollution [29]. Although reducing air pollution by declining industrial activities and traffic is an approach that policymakers and administrators follow, the effectiveness of this strategy in decreasing pollution risks is controversial [36]. Since urbanization and rising energy usage ultimately render ineffective air pollution management policies [37]. More so, lowering emissions does not directly reduce air pollution since different characteristics, such as topography and climatic factors, are also involved in air quality.

Meteorological parameters, directly or indirectly, play a crucial role in ambient air quality by impacting the formation, emission, and deposition of pollutants [38]. In a study by Liu et al. [39], the concentration of air pollutants was depended on geography. Jayamurugan et al. [40] reported that the concentration of pollutants is influenced by wind speed, wind direction, relative humidity, and temperature. Zhang et al. [41] found that the increase in PM_{2.5} concentrations might be due to the relative humidity fluctuation. Yang et al. [42] examined the interaction between PM_{2.5} and meteorological parameters in Chinese cities for 22 months and found a positive correlation between relative humidity and PM_{2.5}. In most zones, wind speed demonstrated a negative relationship with PM_{2.5}. Lou et al. [43] observed that low humidity drives the accumulation of PM_{2.5}. In another study by Zhou et al. [44] in Beijing and Nanjing, the seasonal average of PM_{2.5}, PM₁₀, SO₂, CO, and NO₂ was significantly correlated with wind speed, and relative humidity had a contrasting impact on pollutant accumulation. The Pearson correlation analysis also demonstrated a significant relationship between air pollutants and meteorological parameters in Iran [45]. Therefore, to efficiently control of air quality, it is necessary to explore the interactions between air pollution and meteorological factors based on long-term daily data. In this regard, Fan et al. [46] showed that improving air quality in some Chinese cities was associated with changes in weather conditions. Sunday and Haruna [47] concluded that evaluating the effect of climatic conditions on seasonal changes in pollutant concentrations might help to reduce ambient air pollution.

In addition to meteorological factors, plants also affect air quality [48]. Although plants are victims of air pollution, they have protection mechanisms to absorb air pollutants [49]. Hence, vegetation is one of the major sources of ecosystem services to enhance the quality of urban life [50] by preventing the release of contaminants [51]. In an investigation by Klingberg et al. [52], NO₂ levels were lower in vegetated areas, indicating that leaf area and tree bark can be critical elements in improving air quality [53]. The obtained results by Jeanjean et al. [54] demonstrated that trees trap 7% of air pollutants. In an examination in the United States by Nowak et al. [55], forest trees dragged 17.4 million tons of air pollutants, saved \$6.8 billion in public health costs in 2010. In an analysis by Wu et al. [56] in Shenzhen, China, it was found that the removal of PM_{2.5} by vegetation was nearly 1000 tons in 2015, and the average removal rate was measured at 16 g m⁻² per year. Alonso et al. [57] documented that vacating vegetation increases O₃ levels. Mirsanjari et al. [58] detected that a drop in dense vegetation and the extension of regions with poor vegetation were positively associated with an increase in air pollution in Karaj (Iran). Despite the fact that urban vegetation is pinpointed as an eco-friendly solution, Xing and Brimblecombe [59] found that cities rarely remove more than 1% of air pollutants via plants. Moreover, the deposition of pollutants on the branches and leaves does not appreciably improve air quality. Nemitz et al. [60] declared that urban vegetation in the UK reduced PM_{2.5} by an average of 1%. According to Viippola et al. [61], there is inadequate empirical evidence for ameliorating urban air pollution by forests. According to this conflicting evidence, more investigations are required to focus on the efficiency of vegetation in lowering air pollution [62].

The interaction of air pollutants with meteorological factors and plants is explored in order to forecast the behavior of pollutants and the status of air quality. As much as constructing a forewarning system is essential to protect humans against detrimental consequences of air pollution [63], forecasting air pollution using machine learning, neural networks, and deep learning has recently been addressed by many researchers. Environ-

mental sciences, including weather prediction, soil erosion, waste disposal, dust storms, and air pollution, make extensive use of machine learning techniques [64–66]. Conventional air pollution prediction techniques can be divided into statistical methods, artificial intelligence, and numerical forecasting [67]. Sharma et al. [68] used time-series analysis of 2009–2017 data to predict New Delhi air quality. Kaya and Oguducu [69] developed a 4, 12, and 24-h forecasting model based on deep learning using PM₁₀ hourly data from Istanbul (Turkey) between 2014 and 2018. By applying the classification and regression tree method, Gocheva-llieva et al. [70] presented a model for forecasting daily PM₁₀ concentration with 90% accuracy in Ruse and Pernik (Bulgaria). Madan et al. [71] mentioned that a variety of machine learning methods, including linear regression, decision tree, random forest, neural network, and support vector machine, have been used to predict quality of air. The air quality prediction model developed by Mahalingam et al. [72] using the neural network algorithm and support vector machine proved effective. Pasupuleti et al. [73] found that the random forest method is more accurate in comparison to regression and decision tree for predicting pollutants ($r_{\text{CO}} = 0.79$, $r_{\text{O}_3} = 0.79$, $r_{\text{NO}_2} = 0.70$, $r_{\text{PM}_{2.5}} = 0.86$, and $r_{\text{PM}_{10}} = 0.79$). In another study, Pan [74] demonstrated that the Extreme Gradient Boosting (XGBoost) significantly outperforms random forest, multiple linear regression, decision tree, and support vector machine algorithms for hourly PM_{2.5} concentration forecasts ($r_{\text{PM}_{2.5}} = 0.95$). Furthermore, Ma et al. [75] conducted a study in the northern United States and demonstrated that XGBoost was able to accurately model PM_{2.5} interactions with the environment. Liu et al. [76] revealed that the integration of the ridge regression (RR) model and the XGBoost algorithm had more generalization ability than conventional machine learning techniques for forecasting pollutants. Kumar and Pande [77] recognized that XGBoost had the highest amount of linearity between predicted and real data. Therefore, many air quality models have been developed [78]. However, air pollution is driven by a complex combination of meteorological factors, physical obstacles, and chemical reactions among pollutants [79] that lower the model precision.

This research will first examine these hypotheses: (H1) as urbanization and population grow, air pollution matures annually (Section 3.1); (H2) does vegetation have a significant impact on lowering air pollution? (Section 3.2). Considering few investigations have been conducted on assessing the interactions of air pollutants with the ambient environment in Iran, the current survey's objective is to (i) evaluate relationships between air pollutants, meteorological parameters, and vegetation in three Iranian metropolises between 2015 and 2019 (Section 3.3), and, thus, (ii) develop a XGBoost-based model to predict air quality and assess its performance under real-world conditions (Section 3.4).

2. Materials and Methods

2.1. Case Study

Tehran is the largest metropolis and capital of Iran, and it is located at the geographical position of 35°41' N, 51°26' E [80]. Its altitude is 900 to 1800 m above sea level, and its northern part has cold and dry weather, while its southern part is relatively hot and dry. The yearly temperature ranges from 15° to 18°, though it varies by about 3° in different parts of the city. The area of this city is approximately 730 km⁻², and its population density is estimated at 10,555 people per km⁻². Due to the fact that Tehran, with a population of nearly 13 million, has important governmental, political, economic, and industrial headquarters, there is a significant desire to migrate to it. Tehran's population growth is 4.1% and is expected to grow in the forthcoming years. More than 2 million cars, 500,000 motorcycles, and 5000 industrial units operate in Tehran. Considering that Tehran is the industrial and commercial capital of Iran and uses over 20% of the country's total energy, its air pollution is one of the most prominent environmental issues in Iran [81]. Tehran has two international airports and twelve active air pollution monitoring stations [82].

Tabriz, the capital of East Azerbaijan, is one of Iran's largest and oldest cities, located at the geographical position of 38°4' N, 46°25' E. The area of this city is 324 km⁻², and its altitude is 1350 to 1550 m above sea level. It is the most populous metropolis (1,559,000 people)

in northwestern Iran. Moreover, Tabriz is known as an air pollution hotspot due to its extensive industrial activities. The city has ten municipal districts, an international airport, and eight air pollution sensors.

Shiraz is located in the mountainous region of Zagros at the geographical position of $29^{\circ}36' N$, $52^{\circ}33' E$ and an altitude of 1486 m above sea level. This metropolis has an area of 217 km^{-2} and is divided into eleven municipal districts. In 2016, about 32% of the population of Fars province lived in Shiraz (1,566,000 people) [83], and the population density of this city was 7215 per km^{-2} . Shiraz has an international airport and three active air pollution monitoring stations. Figure 2 shows a schematic of the study zones.

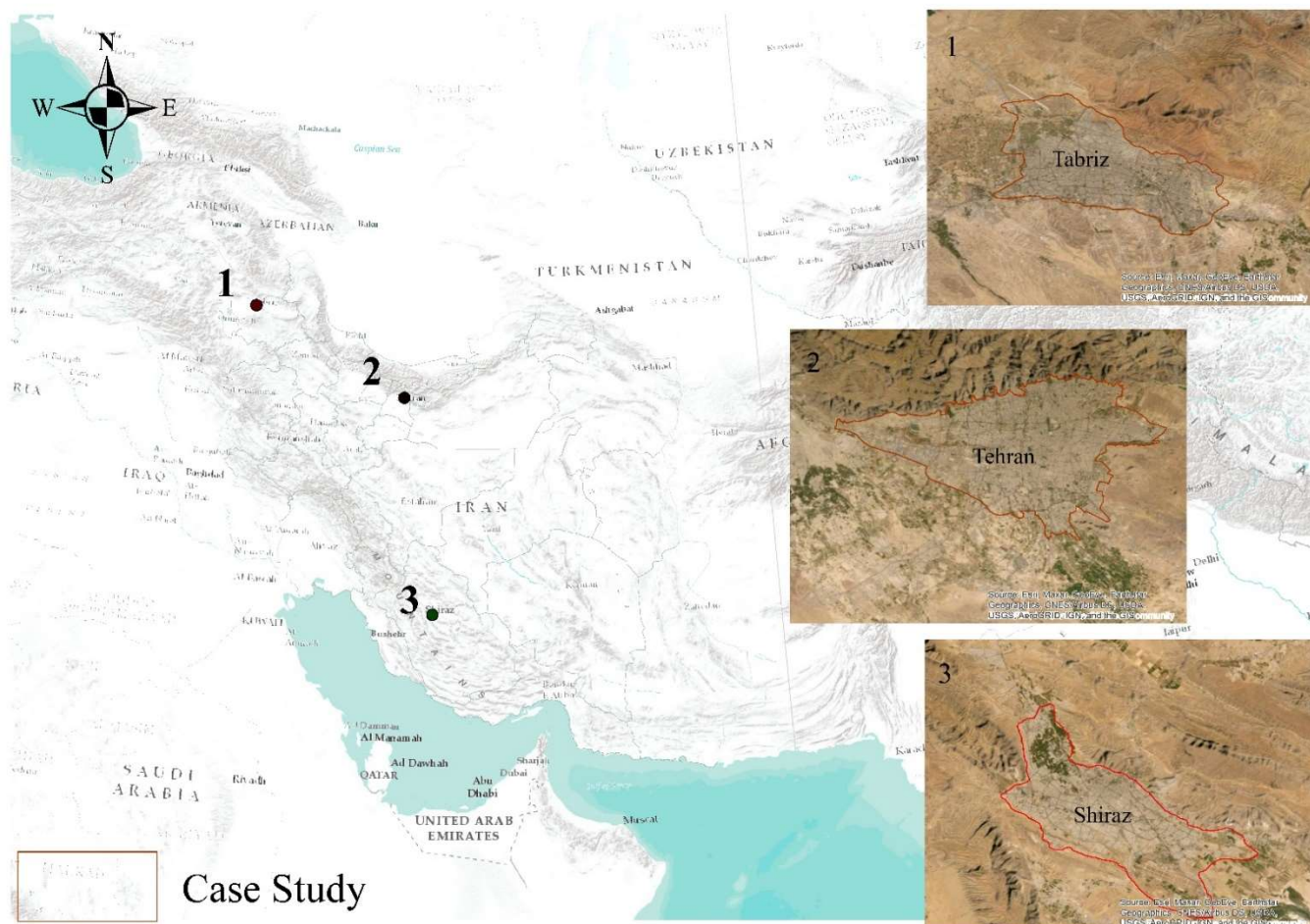


Figure 2. The map of study zones in Iran.

2.2. Data

The air pollution data in this investigation includes the recorded data of each of the parameters CO , O_3 , NO_2 , SO_2 , PM_{10} , $\text{PM}_{2.5}$, and air quality index (AQI), which are registered by the DOE monitoring system. The AQI is equal to the highest amount of pollutant measured per day, and it rises as air pollution worsens. This paper directly acquired the average daily data from January 2015 to December 2019 from the Air Quality Monitoring System (AQMS) (available at <https://aqms.doe.ir>, accessed on 20 June 2021).

The Weather Underground archive (available at <https://www.wunderground.com>, accessed on 20 June 2021) has been used to obtain meteorological data. Approximately 6000 computerized meteorological stations operate at international airports where their data is updated every 1, 3, and 6 h. The meteorological variables in this study include temperature (T, C^0), relative humidity (RH, %), wind speed (WS, mp h^{-1}), and air pressure (AP, mmHg). The international airports connected to this system are Mehrabad Airport

in Tehran (35°41' N, 51°19' E), Shahid Madani Airport in Tabriz (38°7' N, 46°14' E), and Shahid Dastgheib Airport in Shiraz (29°32' N, 52°35' E).

NDVI is a well-known and broadly used indicator for numerically determining vegetation and measuring the health status of plants based on the reflections of light at specific frequencies by plants [84]. In remote sensing studies, data related to the wavelengths of light absorbed and reflected by satellite sensors are used. NDVI demonstrates vegetation in numerals between -1 and 1 , allocates values near to -1 for water, rocky places, sand, and snow at 0.1 or less, and shrubs, grasslands, or old plants between 0.2 and 0.5 . Dense plants, forests, and farms' canopy are between 0.6 and 0.9 [85]. The Sentinel-Hub database (available at <https://apps.sentinel-hub.com/eo-browser>, accessed on 27 June 2021) enables researchers to receive images from different satellites having a variety of indicators for monitoring water, soil, and the atmosphere. NDVI was acquired using the Landsat 8 L1 satellite. It can automatically calculate indices in a selected zone. The data provided by the automatic calculation system includes the maximum, average, and minimum values every ten days and every 20 days. The average data from 2015–2019 for three studied zones was selected as a reference for this study.

2.3. Mapping

The ratio of near-infrared (NIR) and red (R) reflectance is used to calculate NDVI (Equation (1)) [86], and Landsat 8 satellite imagery (OLI_TIRS) was employed for NDVI zoning. The images were downloaded from the USGS database (available at <https://earthexplorer.usgs.gov>, accessed on 27 June 2021) for six days in 2020 and 2021. In preparing this layer, four and five bands of Landsat 8 have been used, and NDVI was implemented on images in the ArcGIS Pro software; thus, a vegetation density layer was obtained.

$$\text{NDVI} = (\text{NIR} - \text{RED}) / (\text{NIR} + \text{RED}) \quad (1)$$

Kriging interpolation was used to zone AQI. Kriging is a robust geostatistical method to evaluate a surface from a scattering of points having z -values [87]. In the present study, after analyzing the distribution pattern of points and the difference between their mean and variance, AQI was used as the value of known z -values and zoned for the study areas. In order to better compare NDVI and AQI, the statistical zoning (Zonal Statistics) in the spatial statistics toolbox in ArcGIS was applied. A Zonal Statistics application computes statistics on raster (value raster) cell values inside the zones defined by another dataset. Its framework produces a raster outcome after calculating only one statistic at a certain time. With the cells corresponding to that zone, this value has become the cell value of the raster output. Since a cell in the output raster may only represent one value, the statistic is generated for only one zone if a zone attribute has overlapped zones [88]. This tool calculates the average value of raster layer cells (NDVI and AQI) by polygons based on the created Thiessen polygons. There is only one spot input attribute for each Thiessen polygon. A Thiessen polygon's corresponding point is relatively closer toward any spot inside it than any other point input feature.

2.4. Statistical Analyzing

SPSS version 18 was utilized for conducting statistical analysis. Indicators (mean and standard deviation) were used in the descriptive statistics processes. The Tukey test and one-way analysis of variance were carried out to compare the means of the variables NDVI, $\text{PM}_{2.5}$, PM_{10} , SO_2 , NO_2 , O_3 , and CO . The Pearson correlation coefficient was employed to evaluate the correlation of the variables T, RH, WS, AP, and NDVI with the variables $\text{PM}_{2.5}$, PM_{10} , SO_2 , NO_2 , O_3 , and CO . The significance level was considered to be <0.05 .

2.5. Modeling

Extreme Gradient Boosting (XGBoost) and Gradient Boosting (GB) are group tree techniques that boost weak learners using the gradient descent architecture. XGBoost, on the other hand, empowers the fundamental GB architecture through algorithmic optimizations.

The XGboost package is part of the Distributed Machine Learning Community. The data is first fitted using a weak regressor. It adopts a weaker regressor to ensure better accuracy of the algorithm without changing the prior regressor, and the procedure is repeated. Each subsequent regressor should incorporate where the preceding regressor failed to perform appropriately. Figure 3 illustrates the flow of the General Boosting algorithm. Initially, it approximates y_1 by assigning numerical values to a decision tree, and then the second tree is adapted based on the previous step's residual, which is $y-y_1$, and so on. By analogy, the algorithm anomaly may be substantially reduced. Table A1 in Appendix A shows the air quality features studied in the modeling process.

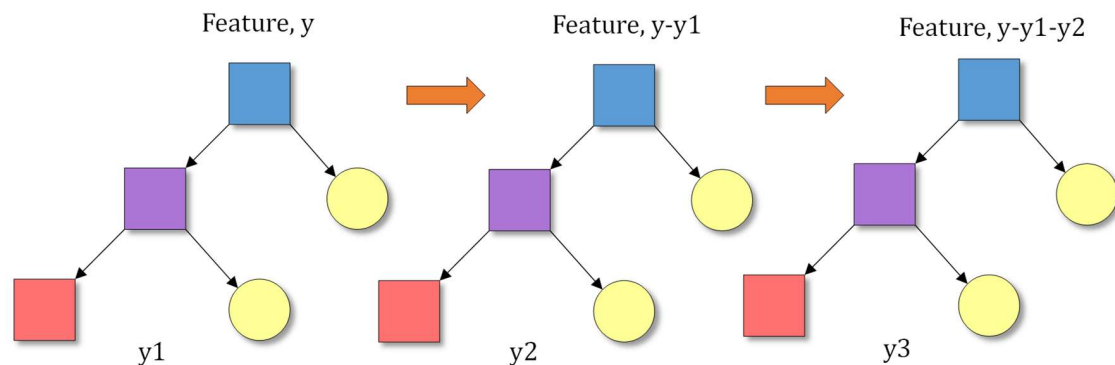


Figure 3. A schematic of GB algorithm, adapted from [89].

Following comprehensively reviewing the related literature of Chen and Guestrin [90], Friedman [91], the GB and XGBoost algorithms are represented as follows:

$$D = x ; y, |D| = n, x \in R^m, y \in R$$

D denotes a dataset, n is the number of samples, m is the number of parameters, and x and y represent the dataset's features and target variable. The ambient air database comprises 1697 samples and five parameters. The prediction results in GB for dataset D are k tree forecasted scores total, which would be determined by a method called the K-additive function, as indicated in Equation (2):

$$\hat{y}_i = \sum_{k=1}^k f_k(x_i), f_k \in F \quad (2)$$

where y_i indicates the forecast of the i -th instance at the k -th boost and x_i is the training dataset's i -th instance sample. The k -th tree's value is $f_k(x_i)$, and the function F represents all decision trees' values. The loss function L_k , as defined in Equation (3), is minimized by GB.

$$L_k = \sum_{i=1}^n L(\hat{y}_i, y_i) \quad (3)$$

Considering GB and XGBoost are both decision tree-based techniques, several tree-related hyper-parameters, such as subsample and max depth, were employed to avoid overfitting and optimize predictive accuracy. Moreover, the learning rate governs the tree weighting that is attached to the model, which is also used to reduce the model's rate of adaption to the training dataset. These hyper-parameters are similarly defined by XGBoost, and their explanations are reported in Table A2 in Appendix A.

The XGBoost objective operation includes a regularization mechanism that facilitates the selection of prediction operations and the management of model complexity. The objective function of the XGBoost is obtained by combining the loss function with the regularization term. The loss function controls the model's predictive power, while the regularization term determines the model's complexity. The XGBoost's aim function can be represented as follows in Equation (4):

$$\text{Obj} = \sum_{i=1}^n L(\hat{y}_i, y_i) + \sum_{i=1}^k R(f_i) \quad (4)$$

where L is the loss value that represents the model's compatibility on training dataset, \hat{y}_i is the predicted label, and y_i is the actual label. $R(f)$ is capable of reducing the dynamics of the training tree's functions. It also addresses the overfitting issue. In order to demonstrate the complexity, we should first describe the details of the tree $f(x)$ as Equation (5):

$$f(x) = w_{q(x)}, w \in R^T, q : R^m \rightarrow \{1, 2, \dots, T\} \quad (5)$$

The leaf score vector is represented by w , q is a mapping function that maps data samples to the associated leaf is represented by q , and T is the number of leaves. Equation (6) is based on the equation for penalizing the model's complexity:

$$R(f) = \gamma T + \alpha (\|w\|) + \frac{1}{2} \lambda (\|w\|^2) \quad (6)$$

where γ and α are the hyper-parameters or constant coefficients, α represents each leaf value, and T is the total number of leaves in the tree. $\|w\|^2$ signifies the L2-norm of the leaf weight controlled by the γ term, whereas $\|w\|$ indicates the L1-norm of the leaf weight controlled by the α term. The weights are driven to be modest by L2 regularization (controlled by the reg lambda term), whereas sparsity is encouraged by L1 regularization (controlled by the reg alpha term). The lowest loss reduction is determined by the hyper-parameter γ towards further division.

A hyper-parameter wmc (min child weight) maintains the depth of the tree, similar to alpha, and a substantial wmc could therefore make the system more precise in the splitting process. The objective function of XGBoost is optimized via gradient descent. The model is an additive model, which means that it introduces a tree to the model every time the forecast outcome equals the sum of the previous and new trees. So, among these equations, at the t -th step, Equation (7) calculates the target at each step, and f_t is used to minimize the error, which seems to minimize errors between both the predicted and measured output with loading f_t .

$$\text{Obj}^{(t)} = \sum_{i=1}^n L(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + R(f_t) + \text{constant} \quad (7)$$

To compute the second-order Taylor equivalent as indicated in Equation (8), we do not have a gradient per each optimization process.

$$\text{Obj}^{(t)} = \sum_{i=1}^n \left[L(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + R(f_t) + \text{constant} \quad (8)$$

where g_i is Equation (9) and h_i is Equation (10):

$$g_i = \partial_{\hat{y}_i^{(t-1)}} L(y_i, \hat{y}_i^{(t-1)}) \quad (9)$$

$$h_i = \partial_{\hat{y}_i^{(t-1)}}^2 L(y_i, \hat{y}_i^{(t-1)}) \quad (10)$$

After deleting the constant terms and appending regularization of Equation (6), Equation (11) depicts the objective function at the t -th step.

$$\text{Obj}^{(t)} = \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \gamma T + \alpha \sum_{j=1}^T \omega_j + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2 \quad (11)$$

In comparison to GB, another strategy employed in XGBoost to minimize additional overfitting is column subsampling. The use of column subsampling is shown to be more beneficial than typical row subsampling in avoiding fitting problems [92]. The hyper-

parameter “subsample” is used to subsample the data by row, and so its definition is demonstrated in Table A2 in Appendix A, which includes a definition of the “colsample bytree” hyper-parameter. As it is not practical to compute all possibilities of trees at the same time, the tree framework is made by computing the leaf scores, regularization, and objective functions at each level. The tree structure will be replicated in subsequent rounds, decreasing the computational complexity dramatically.

In addition, during the node splitting process, the gain of every characteristic is determined. Iteratively, it determines the optimal dividing point till it exceeds the maximum depth. The nodes are then pruned in a bottom-up direction, resulting in a negative gain. This is how XGBoost categorizes the data, which arrive deep in the trees. The findings were derived using the hyper-parameters. Default variables are defined by XGBoost if variables are not specified, though parameters would be set.

2.5.1. Hyperparameters Optimization

Hyper-parameter optimization is the process of determining which hyper-parameters for a particular learning algorithm achieve the best possible results whenever tested on validation data. Equation (12) represents hyper-parameter optimization:

$$x^* = \underset{x \in X}{\operatorname{argmin}} f(x) \quad (12)$$

$f(x)$ denotes an objective score to reduce which is assessed in the test dataset; x indicates a collection of hyper-parameters that gives the minimum score value; and x means any value in the X domain. Determining the model hyper-parameters that result in the highest validation set metric score is crucial. Another challenge concerning hyperparameter optimization is that evaluating the objective function to determine the score is exceedingly expensive. Whenever users attempt alternative hyper-parameters, they must train the model on the training sets, predict outcomes on the validation dataset, and afterwards evaluate the validation metric. With multiple hyper-parameters and models, which include combinations of deep neural networks, this operation is impossible to perform manually. The four typical hyper-parameter optimization methods include (i) grid search, (ii) manual search, (iii) random search, and (iv) Bayesian optimization. Grid search is a conventional technique of hyperparameter optimization that executes a comprehensive search across a portion of the training algorithm’s hyperparameter space. There is a need to characterize a border to execute a grid search since the machine learning algorithm parameter area potentially comprises spaces with actual or limitless values for some parameters. Grid search has a high-dimensional space issue, and it can frequently be simply parallelized since the hyperparameter values used by the algorithms are generally independent of one another.

2.5.2. Preprocessing Dataset

The ambient air dataset is divided into training and testing datasets by 80% and 20%, respectively. The training dataset was employed in model training and optimization. When both of the individuals’ characteristics are numerical, the mean of their up and down values is used to fill in the missing values in the dataset. Feature scaling is a method used to standardize the range of features. In this regard, we use normalization to re-scale features in the range [0, 1]. To normalize our data, we can apply the min-max scaling to each feature column, where the new value x_{norm} of a sample x can be computed by Equation (13):

$$x_{\text{norm}} = (x - x_{\text{min}}) / (x_{\text{max}} - x_{\text{min}}) \quad (13)$$

2.5.3. Xgboost Training and Hyper-Parameter Optimization

The XGBoost regressor of the target variable with Grid Search Optimization was used on the training data after preprocessing the training and test datasets. The hyper-parameters are fine-tuned using Grid Search Optimization. `Learning_rate`, `n_estimators`,

min_child weight, max_depth, subsample, gamma, reg_lambda, and booster are the eight parameters that were tuned in this study. The learning rate improves the model's stability and robustness, although the min child weight, max depth, subsample, and gamma control over-fitting. Similarly, the reg_lambda regularization parameter penalizes complex models. The evaluation indicator is the mean squared error value of 6-fold stratified cross-validation of training examples, which varies depending on the objective function considered. The main objective function is the XGBoost algorithm with different varieties of hyper-parameters. Figure 4 illustrates the proposed model.

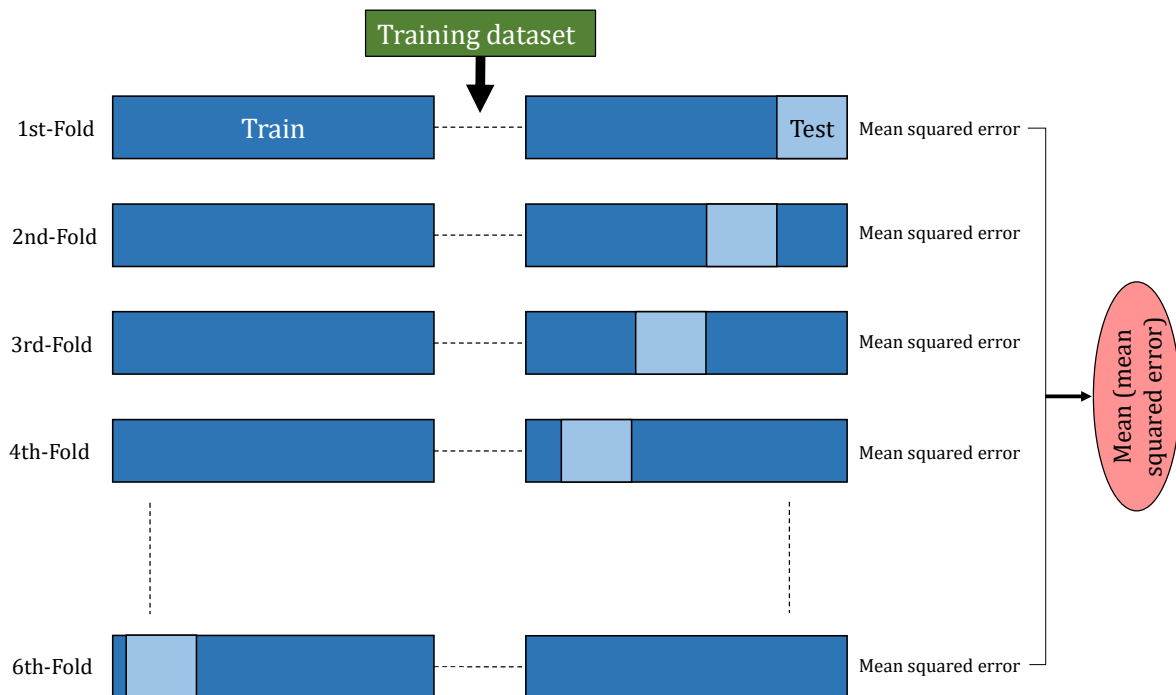


Figure 4. Grid Search Optimization proposes a set of hyper-parameters for six-fold cross-validation of the XGBoost model; adapted from [89].

In Figure 4, cross-validation of a training phase and assessment of mean and mean squared error values for a set of XGBoost hyper-parameters are shown. The Grid Search Optimization attempts to determine and present the highest mean and mean squared error number potential. It determines the model with the maximum mean squared error value for prediction on holdout test data when the specified number of iterations is finalized. On holdout test data, a variety of evaluation metrics were employed to assess the performance of the selected optimized XGBoost model.

2.5.4. Evaluation Metrics

The developed method's performance was assessed utilizing evaluation metrics such as RMSE (root mean squared error), R^2 (R squared), and MAE (mean absolute error), which were determined using formulae. Machine learning has a single number to evaluate a model's performance, whether this is during training, cross-validation, or monitoring after deployment. One of the most frequently utilized measurements is root mean square error. This is a simple scoring method that is also consistent with several of the most basic statistical assumptions.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N \|y(i) - \hat{y}(i)\|^2}{N}}$$

The mean absolute error (MAE) measures the difference in errors between paired observations describing the same occurrence. Comparisons of expected against observed,

further time versus initial time, and one measuring technique versus are instances of Y versus X .

$$\text{MAE} = \frac{\sum_{i=1}^n |y_i - x_i|}{n}$$

The coefficient of determination, sometimes called coefficient, is the fraction of the variation in the dependent variable that is predicted from the independent variable(s), denoted R^2 and pronounced “R squared”. This is a statistic used with statistical models whose primary objective is either to forecast future results or to evaluate hypotheses based on other data. According to the fraction of the overall variance of outputs described by the model, it allows assessment of how well observed results are duplicated by the model. If \bar{y} is the mean of the observed data (y_i) and f_i is model forecasted values:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

$$SS_{\text{res}} = \sum_i (y_i - f_i)^2 = \sum_i e_i^2$$

$$SS_{\text{tot}} = \sum_i (y_i - \bar{y})^2$$

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$

3. Results and Discussion

3.1. Changes of Pollutants Emission

Although air pollution in the metropolitan regions was expected to increase in the 2015–2019 period owing to population and urbanization growth according to the hypothesis H1, preliminary results demonstrated a declining trend of CO and PM_{2.5} in Tehran. PM₁₀, SO₂, T, and WS had a stable trend. Exclusively, NO₂ and RH were slightly increased (Figure 5A). The annual average of PM_{2.5} in Tabriz significantly decreased from 70 $\mu\text{g m}^{-3}$ in 2015 to 50 $\mu\text{g m}^{-3}$ in 2019. There were no significant variations in the mean changes of SO₂ and NO₂, while the levels of O₃ and CO had an increasing trend in 2017–2019. Furthermore, T, RH, WS, and AP were relatively unchanged (Figure 5B). In Shiraz, there was no considerable difference between 2018–2019 and 2015–2017 for the PM_{2.5} level. PM₁₀ and SO₂ levels dropped dramatically at the same time. NO₂ and O₃ experienced a significant increase, whereas CO only decreased in 2019. WS and T had a consistent trend, while RH enhanced (Figure 5C). In addition, AP was unchanged in all cities. Accordingly, since no substantial increase in all or the majority of pollutants was recorded, the hypothesis H1 is rejected. More so, several pollutants in 2015–2019 showed a lowering or an unchanging trend.

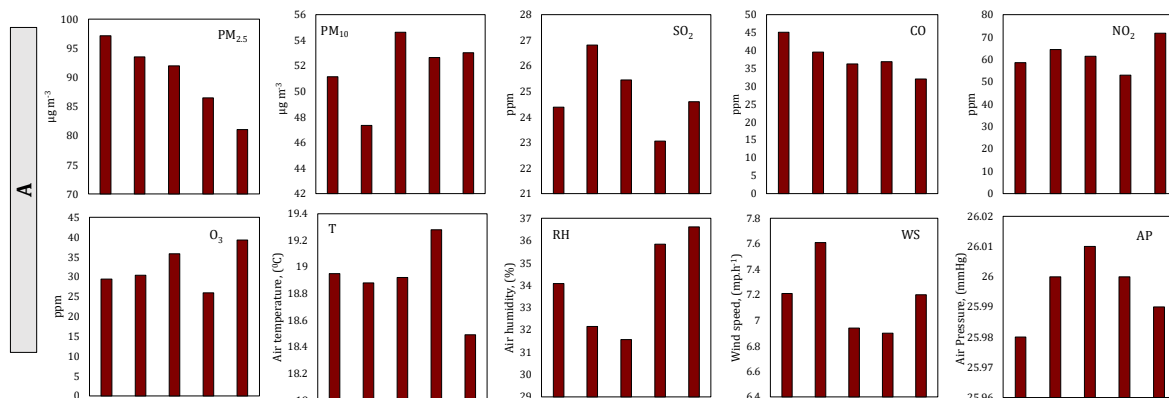


Figure 5. Cont.

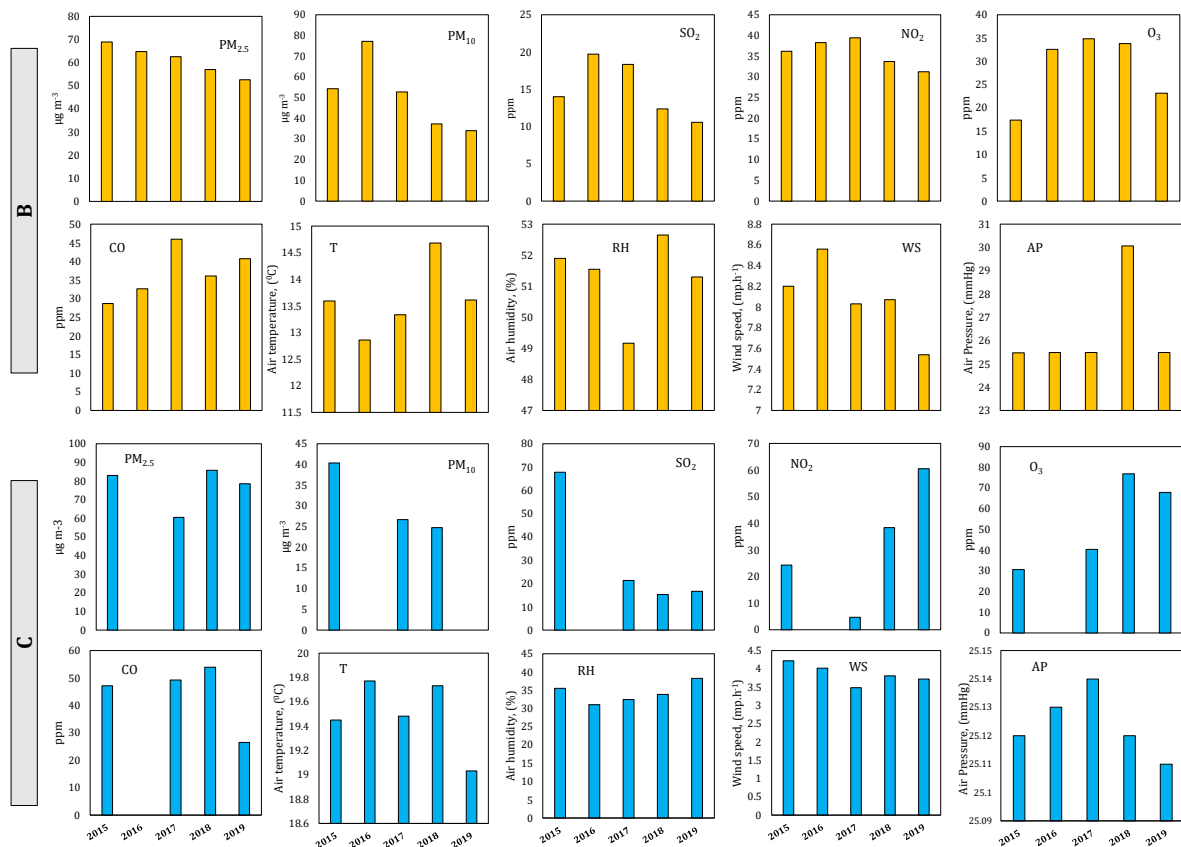


Figure 5. The trend of changes in air pollutants and meteorological parameters in Tehran: (A), Tabriz: (B), and Shiraz: (C) during 2015–2019.

3.2. Interactions between Air Pollutants and Vegetation

The mean difference and standard deviation of NDVI in Tehran, Shiraz, and Tabriz were statistically significant in pairs ($p < 0.05$) (Table A3 in Appendix A). The correlation of $PM_{2.5}$ with NDVI was $r = -0.15$ in Tehran, $r = 0.12$ in Shiraz, and $r = -0.25$ in Tabriz, which was the only non-significant correlation in Shiraz ($p = 0.218$). The correlation between PM_{10} and NDVI was $r = 0.08$ in Tehran, $r = 0.20$ in Shiraz, and $r = 0.08$ in Tabriz, all of which were insignificant. The correlation of SO_2 with the NDVI index was not significant in all cities. The correlation of NO_2 with NDVI was significant in Shiraz and Tabriz. The correlation of O_3 with NDVI was $r = 0.54$ in Tehran, $r = -0.03$ in Shiraz, and $r = 0.34$ in Tabriz, and only Shiraz had an insignificant correlation ($p = 0.801$). The correlation of CO with NDVI in Tehran ($p = 0.183$) and Tabriz ($p = 0.066$) was insignificant. Relationships between NDVI and pollutants in Tehran, Tabriz, and Shiraz are illustrated in Figure 6A–C.

In response to hypothesis 2 (H2), this research reveals that vegetation may have a minor effect in lowering $PM_{2.5}$, SO_2 , and CO emissions in Iranian cities ($r_{PM_{2.5}} = -0.03$, $r_{SO_2} = -0.08$, and $r_{CO} = -0.17$) (Table A4 in Appendix A). This finding is consistent with the results of prior studies that found plants to be inefficient in lowering pollution levels. According to Yli-Pelkonen et al. [93], the influence of urban vegetation on enhancing air quality and lowering pollutants in Helsinki, Finland, was low.

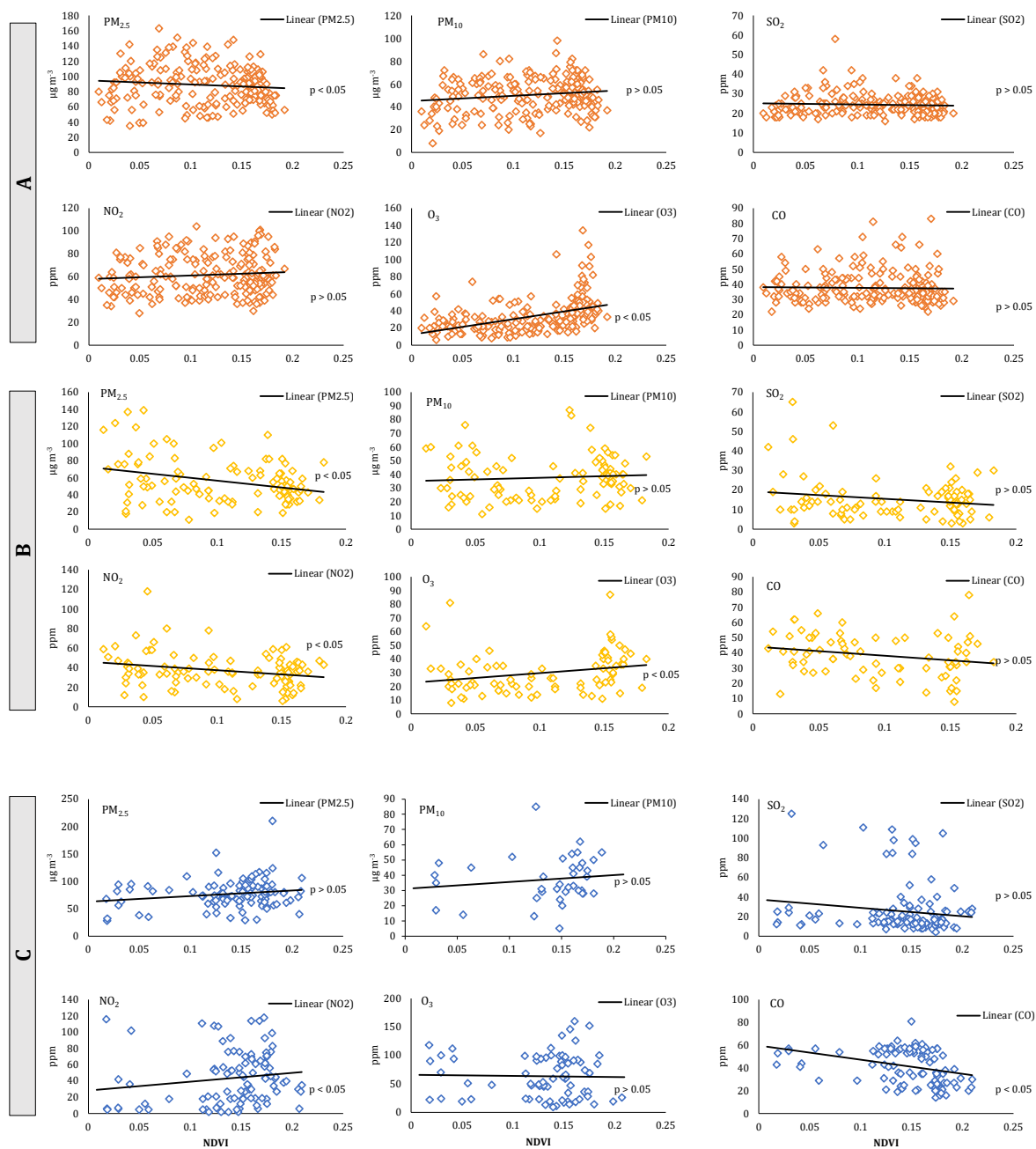


Figure 6. Relationships between air pollutants and NDVI in Tehran (A), Tabriz (B), and Shiraz (C) from 2015 to 2019. Value of the NDVI is between 0 and 1.

Numerous investigations have been conducted to demonstrate visually and spatially the relationship between vegetation and air pollution. In this context, Zhou et al. [94] used the Pearson correlation coefficient to examine the association between NDVI and pollutants in Chinese cities. They recognized that regions with higher NDVI had lower AQI, and there was a negative relationship between NDVI and AQI, so that increasing 0.1 NDVI units reduced AQI by 3.75 units (95% confidence interval). Zheng et al. [95] also evaluated the connection between air pollution and land use in Hangzhou, China, and identified that areas with low NDVI and high surface temperature had high concentrations of PM, NO₂, SO₂, and CO. Prakasam et al. [96] examined satellite images during 2001–2021 and identified that decreasing vegetation was clearly connected with poor air quality in Himachal Pradesh (India). Sun et al. [97] reported that concentrations of PM_{2.5}, PM₁₀, CO, NO₂, and SO₂ were negatively correlated with NDVI levels. Figures 7–9 reveal that

regions with lower vegetation have higher pollutant emissions, leading to an increase in AQI. Although the figures demonstrated that AQI is higher in regions with low NDVI, the contribution of vegetation in decreasing pollution cannot be deemed effective.

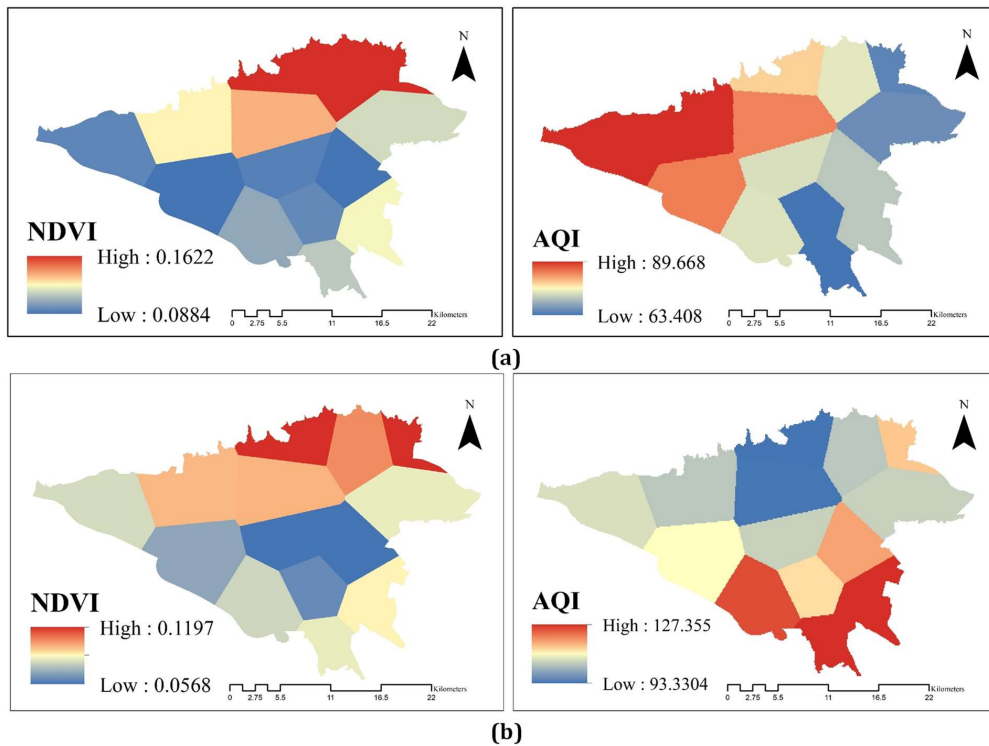


Figure 7. Relationships between NDVI and AQI in Tehran; AQI is collected from air pollution sensors disturbed in the city, and NDVI is obtained from Landsat 8; (a) 6 November 2020; and (b) 2 June 2021.

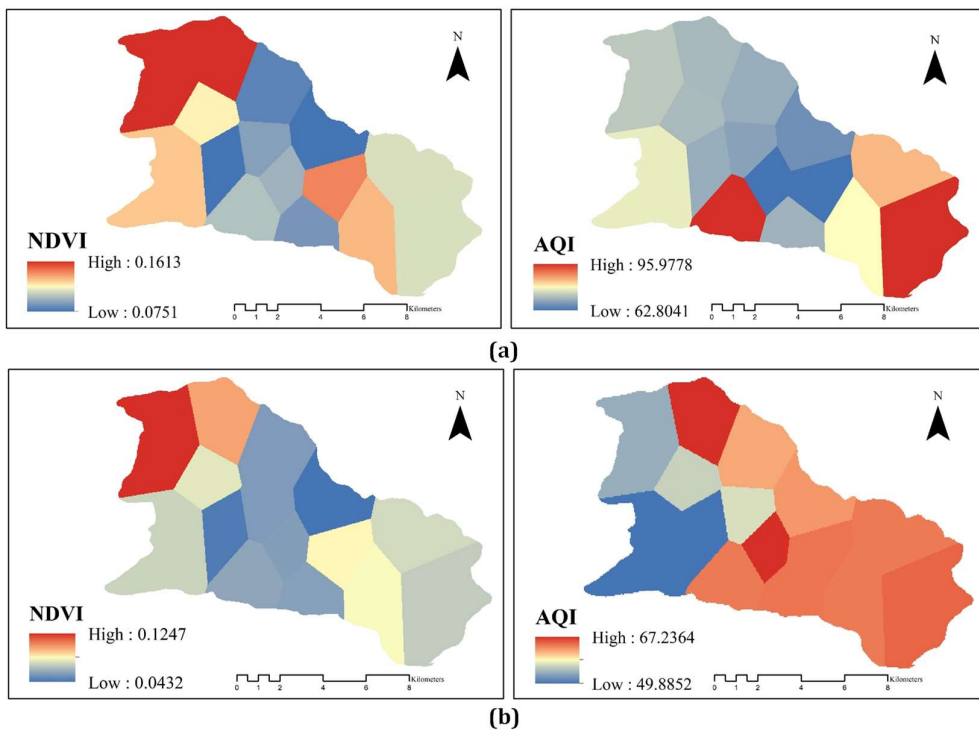


Figure 8. Relationships between NDVI and AQI in Tabriz; AQI is collected from air pollution sensors disturbed in the city, and NDVI is obtained from Landsat 8; (a) 1 November 2020; and (b) 29 May 2021.

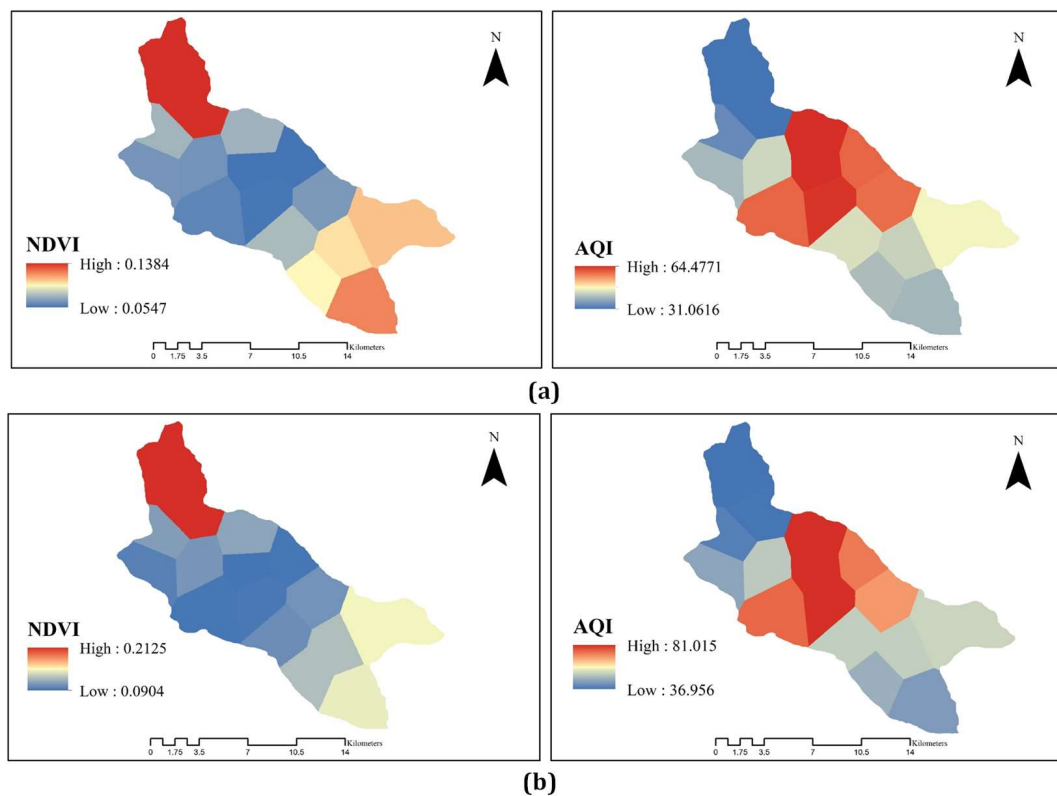


Figure 9. Relationships between NDVI and AQI in Shiraz; AQI is collected from air pollution sensors disturbed in the city, and NDVI is obtained from Landsat 8; (a) 15 November 2020; and (b) 26 May 2021.

3.3. Interactions between Air Pollutants and Meteorological Factors

The mean of $PM_{2.5}$ in Tehran (25.8 ± 90.2) was statistically different from Shiraz and Tabriz. The mean of PM_{10} in Tehran (51.7 ± 16.2) had a significant difference with Shiraz and a slight difference with Tabriz. The mean of SO_2 in Tehran (24.9 ± 5.8) was insignificantly different from Shiraz and considerably different from Tabriz. The mean of NO_2 in Tehran (62.3 ± 16.3) was significantly different from Shiraz and Tabriz. Tehran's mean O_3 level (32.5 ± 19.4) clearly varied with Shiraz and Tabriz. The mean of CO in Tehran (38 ± 10.7) showed a significant difference with Shiraz and Tabriz.

In analyzing the average daily data of meteorological parameters, T in Tehran (66 ± 17.9) recorded a slight difference with Shiraz and a significant difference with Tabriz. The mean of RH in Tehran (34 ± 17.6) showed an insignificant difference with Shiraz and a significant difference in Tabriz. The mean of WS in Tehran (7.2 ± 3) had a significant difference from Shiraz and Tabriz. The difference between the mean AP in Tehran (26 ± 0.1) and Shiraz and Tabriz was significant (Table A5 in Appendix A).

In Shiraz and Tabriz, all correlations of meteorological parameters with $PM_{2.5}$ were significant. Tehran recorded the highest negative correlation of $PM_{2.5}$ with WS ($r = -0.38$) and Tabriz recorded the highest positive correlation of $PM_{2.5}$ with RH ($r = 0.24$). The most negative relationship of PM_{10} with RH was recorded in Shiraz ($r = -0.26$) and the most positive correlation with T was recorded in Shiraz ($r = 0.24$). The most negative correlation of SO_2 with WS was recorded in Tehran ($r = -0.28$) and the most positive relationship with AP was recorded in Tehran ($r = 0.24$). The most negative correlation of NO_2 with WS was observed in Tehran ($r = -0.31$) and the most positive relationship with T was observed in Shiraz ($r = 0.14$). The most negative correlation of O_3 with RH ($r = -0.42$) was observed in Tehran, and the most positive relationship with T ($r = 0.50$) was also observed in Tehran. The most negative correlation of CO with WS was recorded in Tehran ($r = -0.28$) and the most positive relationship with AP in Shiraz ($r = 0.28$) (Table A6 in Appendix A).

The findings of this study are correlated with several previous studies. According to the obtained results by Qiao et al. [98], RH, WS, and T are the main factors affecting air quality in China. In addition, Jayamurugan et al. [40] observed a significant negative correlation between RH and PM, and this correlation is similar to the outcomes of the PM relationship analysis in Shiraz. In a study by Zhou et al. [44], the mean of O₃ had the highest positive correlation with T, which was also recorded in Tehran. In addition, in an investigation by Kayes et al. [99], most pollutants had a negative relationship with T and RH, and this outcome was also observed in Tabriz. In a study by Sezer Turalioğlu et al. [100] in Erzurum, Turkey, higher SO₂ concentrations were associated with lower T, lower WS, and higher RH. Moreover, the results of linear and nonlinear regression analyses of SO₂ with meteorological parameters showed a moderate and weak connection between this pollutant and meteorological parameters in Elazig [101]. In research by Ilten and Selici [102] in Balikesir, higher concentrations of total daily particulate matter and SO₂ were associated with lower T, lower WS, higher AP, and higher RH. In an analysis by Kliengchuay et al. [103] in Mae Hong Son Province, Thailand, PM₁₀ concentrations were significantly associated with RH ($r = -0.37$). The results of Spearman analysis in research by Jassim et al. [104] in Bahrain revealed that the correlation coefficient between RH and the concentrations of PM₁₀ and PM_{2.5} was $r = -0.595$ and $r = -0.526$, respectively, which was a remarkable negative relationship. There was a considerable positive correlation between temperature and PM₁₀ ($r = 0.42$) and PM_{2.5} ($r = 0.48$). The correlations between PM₁₀ and RH and T in Tehran and the correlations between PM_{2.5} and PM₁₀ with RH and T in Shiraz were similar to the obtained outcomes by Jassim et al. [104].

3.4. Model Evaluation

The Grid Search optimization approach is used in the phase of Hyper-parameter optimization to apply various combinations of XGBoost parameters and try to optimize the mean squared error on 6-fold stratified cross-validation on each of the models. It was hypothesized that the Grid Search optimization algorithm iterations could provide an ideal set of parameters. Table A7 in Appendix A expresses the model assessment results and demonstrates that the model performance (R² test) in the daily forecast of PM_{2.5}, PM₁₀, NO₂, SO₂, O₃, and CO emissions was 0.36, 0.27, 0.46, 0.41, 0.52, and 0.38, respectively, which indicates better performance in predicting gaseous pollutant emissions. However, this accuracy is insufficient to predict air pollution on an urban scale.

It was expected that the forecasting model output should be closer to the sensors near airports since the meteorological and air pollution stations were distinct. Therefore, the data generated by the model and the actual data for November 2021 were compared with each other, and it was found that the distance factor is not related to the performance of the model. Since there is no linear and significant relationship between the performance of the model and real data reported by sensors in Tehran (Figure 10A), Tabriz (Figure 10B), and Shiraz (Figure 10C), the model's performance in predicting air pollution varies region by region, which means it is not usable in real conditions. It appears that various algorithms should have been employed in addition to XGBoost. However, it is more necessary to explore the modeling challenges than to have a more efficient model. Numerous prior articles made use of the neural network model, although data training for this model requires large and long-term data [105], which was not obtainable for this research.

Considering that air pollution prediction can be effective for controlling urban operations, the air pollution status depends on various environmental factors that make it difficult to predict the concentration of pollutants [106]. It has been reported that modeling dynamic real-world phenomena like air pollution is a significant challenge owing to their non-linearity and high dimensional sample space [107]. In this study, it appears that several factors were effective in reducing the accuracy of the model, including the unavailability of information regarding pollutant emissions from sources such as factories and traffic, the high dynamics of environmental parameters, and the lack of data due to sensor errors. Kang et al. [108] reported that sensor flaws or incomplete data make it exceedingly difficult

to forecast air quality through modeling. Additionally, since linear regression techniques are not efficient for predicting time-dependent data [109], it is challenging to predict air pollution despite invalid and missing inputs [110]. Liao et al. [111] documented that shallow statistical methods and flawed sensors restrict the air quality prediction process. On the basis of the findings, the following summary of obstacles and solutions corresponding to reliable modeling to forecast air quality is presented:

- The use of deep learning techniques to improve prediction [111,112];
- This survey did not consider second- and third-order interactions between parameters. Researchers should, therefore, address these interactions in the modeling process;
- It is suggested that in machine learning-based investigations, correlations across weather stations and nearby air quality stations should be explored to improve prediction accuracy [113]. In addition, it is necessary to develop dynamic and integrated air quality models employing hybrid machine learning algorithms [108];
- Modeling the emission from sources, chemical reactions of pollutants, and urban activities is required to improve forecasting accuracy [114], which was not considered in the present investigation. Eventually, clean air may only be restored whenever governments shift their approach toward sustainable environmental strategies [115].

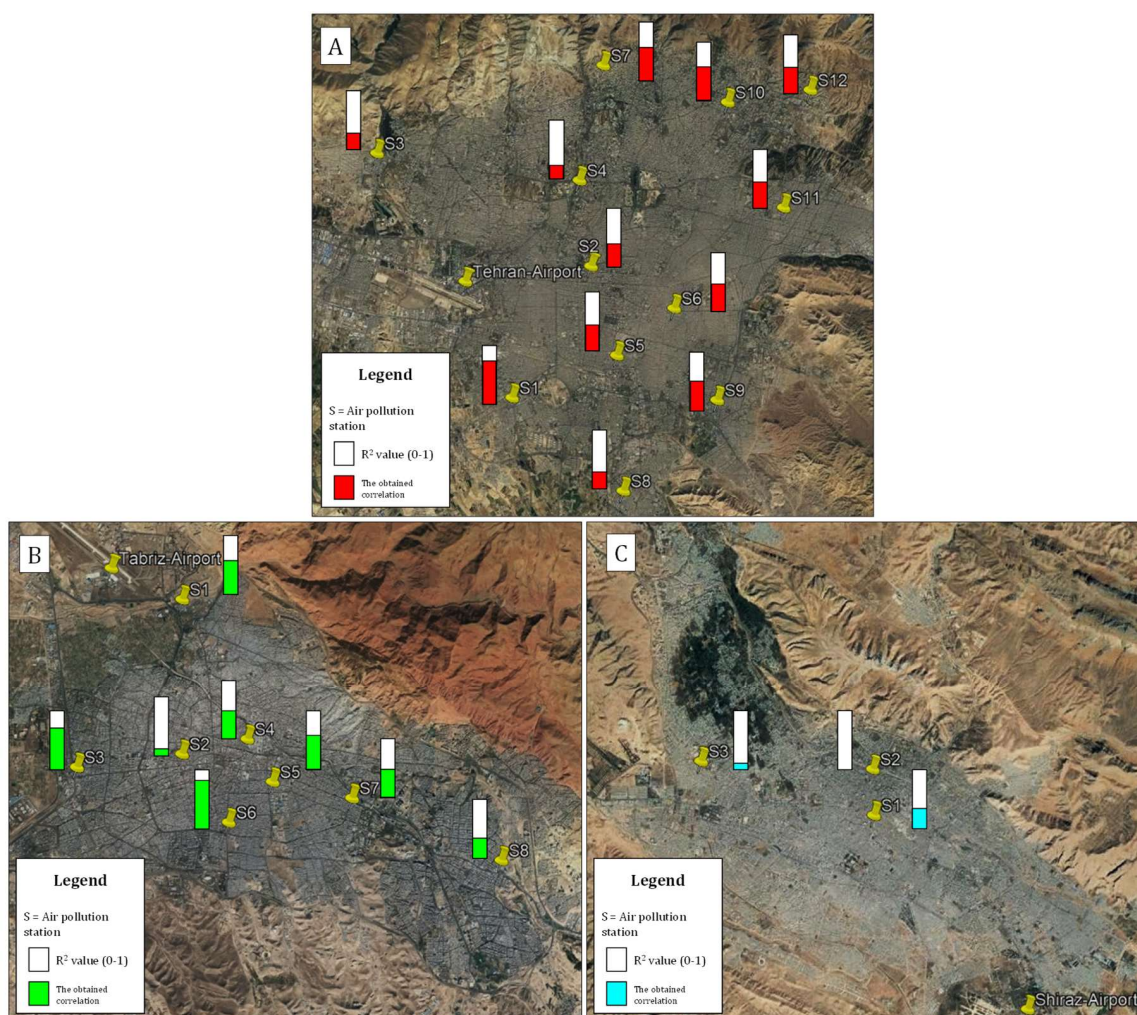


Figure 10. Correlations of the AQI predicted by the model with the AQI reported by sensors in Tehran (A), Tabriz (B), and Shiraz: (C) in November 2021. The stations are listed in order of their distance from the airport.

4. Conclusions

Air pollution is an inevitable phenomenon caused by the development of industry and urbanization in recent decades, which has adversely affected human and ecosystem health. Although some actions have been taken to reduce it, they have not been significantly efficient. Given the fact that air pollutants, including PM, NO₂, SO₂, O₃, and CO, interact with their surrounding environment, many researchers use the interaction of the pollutants with vegetation and meteorological parameters, such as temperature, relative humidity, wind speed, and air pressure, to create and develop air quality forecasting models. The present research attempted to explore the relationships between air pollutants and the ambient environment from a statistical perspective in Tehran, Tabriz, and Shiraz, Iran, then create a model for predicting air pollution using the machine learning method. In the case of regression, the improved XGBoost algorithm was applied in the suggested strategy for the model. Three distinct assessment criteria were used to assess the proposed technique, including R², Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). The Grid Search optimization was employed as a hyper-parameter optimization method in modeling, and it was shown to be a beneficial approach for obtaining the optimum hyper-parameters. According to the obtained results of experiments, it may be inferred that the proposed forecasting model could enhance the level of decision-making during air quality prediction.

Although in all three cities, there were evident connections between pollutants and meteorological factors and vegetation, it was not adequate to allow modeling to accurately predict daily air pollution (R²_{PM_{2.5}} = 0.36, R²_{PM₁₀} = 0.27, R²_{NO₂} = 0.46, R²_{SO₂} = 0.41, R²_{O₃} = 0.52, and R²_{CO} = 0.38). It has been found that in addition to meteorological factors, other factors are also involved in the diffusion of air pollutants in the atmosphere, such as sunlight, wind direction, and chemical reactions of pollutants. It appears that factors such as lack of data caused by sensor errors, lack of data regarding polluting sources such as factories and traffic, and the high dynamics of environmental conditions have driven the reduction in the accuracy of the model. Thus, it is concluded that for modeling and predicting air pollution, examining only the interaction of pollutants with meteorological and vegetation parameters is not sufficient. Furthermore, the spatial diversity of pollution monitors and meteorological stations made it difficult to develop a model for predicting air pollution region by region. The following strategies can be effective for future studies: (1) the number of air pollution and meteorological monitoring stations should be equal; (2) using small and low-cost sensors to develop the pollution monitoring network; (3) the problem of data loss due to sensor errors must be solved by deep learning methods; and (4) integration of satellite observations with proximal data.

Author Contributions: Conceptualization, A.K.R. and R.R.S.; writing—original draft preparation, A.K.R., A.N., M.S. and S.-O.R.; review and editing, A.K.R., R.R.S. and F.G.; resources, R.R.S. and S.K.B.; supervision, R.R.S., F.G. and S.K.B.; project administration, A.K.R. and R.R.S.; analyzing, A.N. and S.-O.R.; modeling, S.-O.R.; mapping, M.S.; funding acquisition, R.R.S. and S.K.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Acknowledgments: The authors would like to acknowledge the editorial support from the AdaptiveAgroTech Consultancy Network.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Features of air quality.

Feature No.	Feature Description	Type
1	Relative Humidity	Numerical
2	Air Pressure	Numerical
3	Temperature	Numerical
4	NDVI	Numerical
5	Wind Speed	Numerical

Table A2. XGBoost regressor hyperparameters, adapted from [101].

PM _{2.5}			PM ₁₀		
Parameter	Value	Description	Parameter	Value	Description
Learning rate	0.02	Shrink the weights on each step	Learning rate	0.0095	Shrink the weights on each step
n_estimators	350	Number of trees to fit	n_estimators	500	Number of trees to fit
Reg_lambda	0.25	L2 regularization term on weights	Reg_lambda	0	L2 regularization term on weights
Booster	gbtree	Select the model for each iteration	Booster	gbtree	Select the model for each iteration
min_chid_weigth	1	Minimum sum of weights	min_chid_weigth	5	Minimum sum of weights
max_depth	6	Maximum depth of a tree	max_depth	4	Maximum depth of a tree
gamma	0	The minimum loss reduction needed for splitting	gamma	0	The minimum loss reduction needed for splitting
subsample	0.82	Control the sample's proportion	subsample	0.83	Control the sample's proportion
NO ₂			SO ₂		
Parameter	Value	Description	Parameter	Value	Description
Learning rate	0.1	Shrink the weights on each step	Learning rate	0.04	Shrink the weights on each step
n_estimators	300	Number of trees to fit	n_estimators	300	Number of trees to fit
Reg_lambda	0.2	L2 regularization term on weights	Reg_lambda	6	L2 regularization term on weights
Booster	gbtree	Select the model for each iteration	Booster	gbtree	Select the model for each iteration
min_chid_weigth	3	Minimum sum of weights	min_chid_weigth	4	Minimum sum of weights
max_depth	7	Maximum depth of a tree	max_depth	6	Maximum depth of a tree
gamma	0	The minimum loss reduction needed for splitting	gamma	0	The minimum loss reduction needed for splitting
subsample	0.92	Control the sample's proportion	subsample	0.91	Control the sample's proportion
O ₃			CO		
Parameter	Value	Description	Parameter	Value	Description
Learning rate	0.1	Shrink the weights on each step	Learning rate	0.1	Shrink the weights on each step
n_estimators	300	Number of trees to fit	n_estimators	200	Number of trees to fit
Reg_lambda	0.5	L2 regularization term on weights	Reg_lambda	2	L2 regularization term on weights
Booster	gbtree	Select the model for each iteration	Booster	gbtree	Select the model for each iteration
min_chid_weigth	5	Minimum sum of weights	min_chid_weigth	5	Minimum sum of weights
max_depth	6	Maximum depth of a tree	max_depth	4	Maximum depth of a tree
gamma	0	The minimum loss reduction needed for splitting	gamma	0	The minimum loss reduction needed for splitting
subsample	0.92	Control the sample's proportion	subsample	0.97	Control the sample's proportion

Table A3. The difference of NDVI in the cities.

Variable	Total Mean ± SD	City 1	Mean ± SD	City 2	Mean ± SD	p
NDVI	0.1 ± 0.1	Tehran	0.1 ± 0.1	Shiraz	0.2 ± 0.1	0.001
		Shiraz	0.2 ± 0.1	Tabriz	0.1 ± 0.2	0.031
				Tabriz	0.1 ± 0.2	0.001

One-way ANOVA; Tukey test for post hoc test; significance level was set at 0.05. Significant quantities are shown in bold format.

Table A4. The correlation between air pollutants and NDVI.

Variable	Total		Tehran		Shiraz		Tabriz	
	r	p	r	p	r	p	r	p
PM _{2.5}	−0.03	0.565	−0.15	0.024	0.12	0.218	−0.25	0.010
PM ₁₀	0.11	0.047	0.08	0.232	0.20	0.222	0.08	0.466
SO ₂	−0.08	0.135	−0.07	0.274	−0.12	0.176	−0.07	0.503
NO ₂	0.05	0.287	0.06	0.358	0.23	0.022	−0.20	0.044
O ₃	0.40	0.001	0.54	0.001	−0.03	0.801	0.34	0.002
CO	−0.17	0.001	−0.09	0.183	−0.42	0.001	−0.22	0.066

r: The Pearson correlation test, the significance level was considered 0.05. Significant quantities are shown in bold format.

Table A5. The difference between average air pollution and meteorological data.

Variable	Total	City 1	Mean ± SD	City 2	Mean ± SD	p
	Mean ± SD					
PM _{2.5}	76.3 ± 32.7	Tehran	90.2 ± 25.8	Shiraz	79.4 ± 31.7	0.001
		Shiraz	79.4 ± 31.7	Tabriz	61.2 ± 32.8	0.001
		Tabriz	61.2 ± 32.8	Shiraz	61.2 ± 32.8	0.001
PM ₁₀	49.7 ± 45.9	Tehran	51.7 ± 16.2	Shiraz	35.3 ± 14.7	0.001
		Shiraz	35.3 ± 14.7	Tabriz	50.6 ± 66.7	0.786
		Tabriz	50.6 ± 66.7	Shiraz	50.6 ± 66.7	0.001
SO ₂	21.6 ± 15.4	Tehran	24.9 ± 5.8	Shiraz	26.0 ± 27.3	0.130
		Shiraz	26.0 ± 27.3	Tabriz	14.9 ± 8.8	0.001
		Tabriz	14.9 ± 8.8	Shiraz	14.9 ± 8.8	0.001
NO ₂	47.7 ± 23.5	Tehran	62.3 ± 16.3	Shiraz	41.4 ± 29.6	0.001
		Shiraz	41.4 ± 29.6	Tabriz	41.4 ± 29.6	0.001
		Tabriz	41.4 ± 29.6	Shiraz	35.6 ± 18.4	0.001
O ₃	36.6 ± 25.1	Tehran	32.5 ± 19.4	Shiraz	61.5 ± 36.5	0.001
		Shiraz	61.5 ± 36.5	Tabriz	30.2 ± 16.8	0.015
		Tabriz	30.2 ± 16.8	Shiraz	30.2 ± 16.8	0.001
CO	38.3 ± 13.1	Tehran	38.0 ± 10.7	Shiraz	40.9 ± 15.6	0.001
		Shiraz	40.9 ± 15.6	Tabriz	36.9 ± 14.1	0.044
		Tabriz	36.9 ± 14.1	Shiraz	36.9 ± 14.1	0.001
T	63.2 ± 18.3	Tehran	66.0 ± 17.9	Shiraz	67.1 ± 16.1	0.165
		Shiraz	67.1 ± 16.1	Tabriz	56.5 ± 19.0	0.001
		Tabriz	56.5 ± 19.0	Shiraz	56.5 ± 19.0	0.001
RH	39.9 ± 20.6	Tehran	34.0 ± 17.6	Shiraz	34.2 ± 20.0	0.957
		Shiraz	34.2 ± 20.0	Tabriz	51.3 ± 19.2	0.001
		Tabriz	51.3 ± 19.2	Shiraz	51.3 ± 19.2	0.001
WS	6.4 ± 3.3	Tehran	7.2 ± 3.0	Shiraz	3.8 ± 1.8	0.001
		Shiraz	3.8 ± 1.8	Tabriz	8.1 ± 3.2	0.001
		Tabriz	8.1 ± 3.2	Shiraz	8.1 ± 3.2	0.001
AP	25.8 ± 4.1	Tehran	26.0 ± 0.1	Shiraz	25.1 ± 6.9	0.001
		Shiraz	25.1 ± 6.9	Tabriz	26.4 ± 6.9	0.006
		Tabriz	26.4 ± 6.9	Shiraz	26.4 ± 6.9	0.001

SD: Standard deviation, one-way analysis of variance and Tukey test, significance level was considered to be 0.05. Significant quantities are shown in bold format.

Table A6. The obtained correlations between air pollutants and meteorological parameters.

Variable 1	Variable 2	Total		Tehran		Shiraz		Tabriz	
		r	p	r	p	r	p	r	p
PM _{2.5}	T	0.01	0.912	−0.08	0.001	0.21	0.001	−0.24	0.001
	RH	−0.09	0.001	0.03	0.186	−0.18	0.001	0.24	0.001
	WS	−0.26	0.001	−0.38	0.001	0.09	0.013	−0.21	0.001
	AP	−0.07	0.001	0.17	0.001	−0.14	0.001	−0.08	0.001

Table A6. Cont.

Variable 1	Variable 2	Total		Tehran		Shiraz		Tabriz	
		r	p	r	p	r	p	r	p
PM ₁₀	T	0.04	0.021	0.21	0.001	0.24	0.001	0.01	0.907
	RH	−0.03	0.038	−0.21	0.001	−0.26	0.001	0.01	0.877
	WS	0.02	0.208	−0.22	0.001	0.19	0.001	0.04	0.154
	AP	−0.01	0.693	0.01	0.573	−0.22	0.001	−0.01	0.602
SO ₂	T	−0.06	0.001	−0.22	0.001	−0.16	0.001	−0.18	0.001
	RH	−0.05	0.002	0.07	0.002	0.10	0.001	0.06	0.025
	WS	−0.17	0.001	−0.28	0.001	0.03	0.358	−0.12	0.001
	AP	−0.02	0.175	0.24	0.001	0.10	0.002	0.02	0.474
NO ₂	T	0.08	0.001	−0.03	0.228	0.14	0.001	−0.16	0.001
	RH	−0.19	0.001	−0.05	0.060	−0.07	0.046	0.06	0.010
	WS	−0.16	0.001	−0.31	0.001	−0.19	0.001	−0.18	0.001
	AP	−0.04	0.002	0.13	0.001	−0.01	0.746	−0.07	0.003
O ₃	T	0.29	0.001	0.50	0.001	−0.01	0.708	0.43	0.001
	RH	−0.24	0.001	−0.42	0.001	−0.02	0.605	−0.39	0.001
	WS	−0.09	0.001	0.14	0.001	−0.04	0.313	0.29	0.001
	AP	0.05	0.002	−0.33	0.001	0.05	0.207	0.24	0.001
CO	T	−0.09	0.001	−0.09	0.001	−0.20	0.001	−0.14	0.001
	RH	0.04	0.013	−0.03	0.281	0.10	0.003	0.17	0.001
	WS	−0.21	0.001	−0.28	0.001	−0.07	0.045	−0.14	0.001
	AP	−0.15	0.001	0.11	0.001	0.28	0.001	−0.22	0.001

r: The Pearson correlation, and the significance level was set at 0.05. Significant quantities are shown in bold format.

Table A7. The obtained results by model evaluation.

Pollutant	MAE Train	RMSE Train	R ² Train	MAE Test	RMSE Test	R ² Test
PM _{2.5}	12.4012	16.932	0.432	14.42	19.92	0.36
PM ₁₀	9.2278	12.5375	0.324	10.73	14.75	0.27
NO ₂	8.0582	9.9875	0.552	9.37	11.75	0.46
SO ₂	3.0444	4.182	0.492	3.54	4.92	0.41
O ₃	8.17	12.886	0.624	9.5	15.16	0.52
CO	4.6956	5.9925	0.456	5.46	7.05	0.38

References

- Munsif, R.; Zubair, M.; Aziz, A.; Zafar, M.N. Industrial air emission pollution: Potential sources and sustainable mitigation. In *Environmental Emissions*; IntechOpen: London, UK, 2021. [CrossRef]
- Fenger, J. Air pollution in the last 50 years—From local to global. *Atmos. Environ.* **2009**, *43*, 13–22. [CrossRef]
- Manisalidis, I.; Stavropoulou, E.; Stavropoulos, A.; Bezirtzoglou, E. Environmental and Health Impacts of Air Pollution: A Review. *Front. Public Health* **2020**, *8*, 14. [CrossRef] [PubMed]
- WHO. Air Pollution in the South-East Asia Region. Available online: <https://www.who.int/southeastasia/health-topics/air-pollution> (accessed on 13 February 2022).
- WHO. Air Pollution. Available online: https://www.who.int/health-topics/air-pollution#tab=tab_2 (accessed on 13 February 2022).
- WHO. Air Pollution: Overview. Available online: https://www.who.int/health-topics/air-pollution#tab=tab_1 (accessed on 3 January 2022).
- Statista. The Most Polluted Cities in America. Available online: <https://www.statista.com/chart/24695/us-cities-by-year-round-pm-pollution/> (accessed on 1 January 2022).
- UNECE. Air Pollution and Health. Available online: <https://unece.org/air-pollution-and-health> (accessed on 1 January 2022).
- Ghorani-Azam, A.; Riahi-Zanjani, B.; Balali-Mood, M. Effects of air pollution on human health and practical measures for prevention in Iran. *J. Res. Med. Sci.* **2016**, *21*, 65. [CrossRef] [PubMed]
- Zhang, Y.; Yang, P.; Gao, Y.; Leung, R.L.; Bell, M.L. Health and economic impacts of air pollution induced by weather extremes over the continental U.S. *Environ. Int.* **2020**, *143*, 105921. [CrossRef] [PubMed]
- Pandey, A.; Brauer, M.; Cropper, M.L.; Balakrishnan, K.; Mathur, P.; Dey, S.; Turkoglu, B.; Kumar, G.A.; Khare, M.; Beig, G.; et al. Health and economic impact of air pollution in the states of India: The Global Burden of Disease Study 2019. *Lancet Planet. Health* **2021**, *5*, e25–e38. [CrossRef]

12. IQAir. World's Most Polluted Countries 2020 (PM_{2.5}). Available online: <https://www.iqair.com/world-most-polluted-countries> (accessed on 16 January 2022).
13. Rad, A.K.; Naghipour, A. Impacts of subway development on air pollution and vegetation in Tabriz and Shiraz, Iran. *J. Air Pollut. Health* **2022**, *7*, 121–130. [[CrossRef](#)]
14. Ito, O.; Okano, K.; Totsuka, T. Effects of NO₂ and O₃ Exposure Alone or in Combination on Kidney Bean Plants: Amino Acid Content and Composition. *Soil Sci. Plant Nutr.* **1986**, *32*, 351–363. [[CrossRef](#)]
15. Hogda, K.A.; Tommervik, H.; Solheim, I.; Lauknes, I. Mapping of Air Pollution Effects on the Vegetation Cover in the Kirkenes-Nikel Area Using Remote Sensing. In Proceedings of the 1995 International Geoscience and Remote Sensing Symposium, IGARSS'95. Quantitative Remote Sensing for Science and Applications, Firenze, Italy, 10–14 July 1995; Volume 2, pp. 1249–1251. [[CrossRef](#)]
16. Bignal, K.L.; Ashmore, M.R.; Headley, A.D.; Stewart, K.; Weigert, K. Ecological impacts of air pollution from road transport on local vegetation. *Appl. Geochem.* **2007**, *22*, 1265–1271. [[CrossRef](#)]
17. Winner, W.E.; Atkinson, C.J. Absorption of air pollution by plants, and consequences for growth. *Trends Ecol. Evol.* **1986**, *1*, 15–18. [[CrossRef](#)]
18. Gostin, I. Air Pollution Stress and Plant Response. In *Plant Responses to Air Pollution*; Kulshrestha, U., Saxena, P., Eds.; Springer: Singapore, 2016; pp. 99–117. [[CrossRef](#)]
19. Weber, J.D.; Tingey, D.; Andersen, C. Plant Response to Air Pollution. U.S. Environmental Protection Agency, Washington, DC, EPA/600/A-93/050 (NTIS PB93167260). Available online: https://cfpub.epa.gov/si/si_public_record_Report.cfm?Lab=NHEERL&dirEntryId=50437 (accessed on 3 October 2021).
20. UNECE. Air Pollution and Food Production. Available online: <https://unece.org/air-pollution-and-food-production> (accessed on 10 January 2022).
21. Vlachokostas, C.; Nastis, S.A.; Achillas, C.; Kalogeropoulos, K.; Karmiris, I.; Moussiopoulos, N.; Chourdakis, E.; Baniyas, G.; Limperi, N. Economic damages of ozone air pollution to crops using combined air quality and GIS modelling. *Atmos. Environ.* **2010**, *44*, 3352–3361. [[CrossRef](#)]
22. Narita, D.; Oanh, N.; Sato, K.; Huo, M.; Permadi, D.; Chi, N.; Ratanajaratroj, T.; Pawarmart, I. Pollution Characteristics and Policy Actions on Fine Particulate Matter in a Growing Asian Economy: The Case of Bangkok Metropolitan Region. *Atmosphere* **2019**, *10*, 227. [[CrossRef](#)]
23. United Nations Environment Programme. Restoring Clean Air. Available online: <https://www.unep.org/regions/asia-and-pacific/regional-initiatives/restoring-clean-air> (accessed on 10 January 2022).
24. United Nations Environment Programme. Why Does Air Matter? Available online: <https://www.unep.org/explore-topics/air/why-does-air-matter> (accessed on 10 January 2022).
25. United Nations. UN Iran Country Results Report 2019. Available online: <https://iran.un.org/en/97918-un-iran-country-results-report-2019> (accessed on 28 October 2020).
26. United Nations Development Programme. About Iran. Available online: <https://www.ir.undp.org/content/iran/en/home/countryinfo.html> (accessed on 10 January 2022).
27. Rad, A.K.; Shamshiri, R.R.; Azarm, H.; Balasundram, S.K.; Sultan, M. Effects of the COVID-19 Pandemic on Food Security and Agriculture in Iran: A Survey. *Sustainability* **2021**, *13*, 10103. [[CrossRef](#)]
28. IQAir. Air Quality in Iran. Available online: <https://www.iqair.com/iran> (accessed on 13 February 2022).
29. Hosseini, V.; Shahbazi, H. Urban Air Pollution in Iran. *Iran. Stud.* **2016**, *49*, 1029–1046. [[CrossRef](#)]
30. Mousavi, S.; Mozaffari, Z.; Motamed, M. The effect of higher fuel price on pollutants emission in Iran. *Casp. J. Environ. Sci.* **2018**, *16*, 1–11. [[CrossRef](#)]
31. Economy. Iran—Economic Indicators. Available online: <https://www.economy.com/iran/indicators#ECONOMY> (accessed on 21 January 2022).
32. World Bank. Adjusted Savings: Carbon Dioxide Damage (Current US\$)—Iran, Islamic Rep. Available online: <https://data.worldbank.org/indicator/NY.ADJ.DCO2.CD?end=2019&locations=IR&start=1970&view=chart> (accessed on 21 January 2022).
33. Weiner, R.; Matthews, R.; Vesilind, P.A. *Environmental Engineering*; Butterworth-Heinemann: Oxford, UK, 2003. [[CrossRef](#)]
34. Alvarez-Mendoza, C.I.; Teodoro, A.C.; Torres, N.; Vivanco, V. Assessment of Remote Sensing Data to Model PM₁₀ Estimation in Cities with a Low Number of Air Quality Stations: A Case of Study in Quito, Ecuador. *Environments* **2019**, *6*, 85. [[CrossRef](#)]
35. Vallero, D. Air Pollution Monitoring Changes to Accompany the Transition from a Control to a Systems Focus. *Sustainability* **2016**, *8*, 1216. [[CrossRef](#)]
36. Shih, H.C.; Chen, L.H.; Shih, X.H.; Ma, H.W. Twice the effort: Ineffectiveness of selecting air pollution control targets with emission quantity for risk reduction. *Environ. Int.* **2019**, *125*, 489–496. [[CrossRef](#)]
37. Li, Y.; Chen, K. A Review of Air Pollution Control Policy Development and Effectiveness in China. In *Energy Management for Sustainable Development*; IntechOpen: London, UK, 2018. [[CrossRef](#)]
38. Zhang, H.; Wang, Y.; Hu, J.; Ying, Q.; Hu, X.M. Relationships between meteorological parameters and criteria air pollutants in three megacities in China. *Environ. Res.* **2015**, *140*, 242–254. [[CrossRef](#)]
39. Liu, Y.; Zhou, Y.; Lu, J. Exploring the relationship between air pollution and meteorological conditions in China under environmental governance. *Sci. Rep.* **2020**, *10*, 14518. [[CrossRef](#)]

40. Jayamurugan, R.; Kumaravel, B.; Palanivelraja, S.; Chockalingam, M.P. Influence of Temperature, Relative Humidity and Seasonal Variability on Ambient Air Quality in a Coastal Urban Area. *Int. J. Atmos. Sci.* **2013**, *2013*, 264046. [CrossRef]
41. Zhang, L.; Cheng, Y.; Zhang, Y.; He, Y.; Gu, Z.; Yu, C. Impact of Air Humidity Fluctuation on the Rise of PM Mass Concentration Based on the High-Resolution Monitoring Data. *Aerosol Air Qual. Res.* **2017**, *17*, 543–552. [CrossRef]
42. Yang, Q.; Yuan, Q.; Li, T.; Shen, H.; Zhang, L. The Relationships between PM_{2.5} and Meteorological Factors in China: Seasonal and Regional Variations. *Int. J. Environ. Res Public Health* **2017**, *14*, 1510. [CrossRef] [PubMed]
43. Lou, C.; Liu, H.; Li, Y.; Peng, Y.; Wang, J.; Dai, L. Relationships of relative humidity with PM_{2.5} and PM₁₀ in the Yangtze River Delta, China. *Environ. Monit. Assess.* **2017**, *189*, 582. [CrossRef] [PubMed]
44. Zhou, H.; Yu, Y.; Gu, X.; Wu, Y.; Wang, M.; Yue, H.; Gao, J.; Lei, R.; Ge, X. Characteristics of Air Pollution and Their Relationship with Meteorological Parameters: Northern Versus Southern Cities of China. *Atmosphere* **2020**, *11*, 253. [CrossRef]
45. Ahmadi, H.; Ahmadi, T.; Shahmoradi, B.; Mohammadi, S.; Kohzadi, S. The effect of climatic parameters on air pollution in Sanandaj, Iran. *J. Adv. Environ. Health Res.* **2015**, *3*, 49–61. [CrossRef]
46. Fan, H.; Zhao, C.; Yang, Y. A comprehensive analysis of the spatio-temporal variation of urban air pollution in China during 2014–2018. *Atmos. Environ.* **2020**, *220*, 117066. [CrossRef]
47. Sunday, O.; Haruna, A. Correlation between air pollutants concentration and meteorological factors on seasonal air quality variation. *J. Air Pollut. Health* **2020**, *5*, 11–32. [CrossRef]
48. Brilli, F.; Fares, S.; Ghirardo, A.; de Visser, P.; Calatayud, V.; Munoz, A.; Annesi-Maesano, I.; Sebastiani, F.; Alivernini, A.; Varriale, V.; et al. Plants for Sustainable Improvement of Indoor Air Quality. *Trends Plant Sci.* **2018**, *23*, 507–512. [CrossRef]
49. Gawronski, S.W.; Gawronska, H.; Lomnicki, S.; Sæbo, A.; Vangronsveld, J. Chapter Eight—Plants in Air Phytoremediation. In *Advances in Botanical Research*; Cuyppers, A., Vangronsveld, J., Eds.; Academic Press: Cambridge, MA, USA, 2017; Volume 83, pp. 319–346. [CrossRef]
50. De Carvalho, R.M.; Szlafsztein, C.F. Urban vegetation loss and ecosystem services: The influence on climate regulation and noise and air pollution. *Environ. Pollut.* **2019**, *245*, 844–852. [CrossRef]
51. Barwise, Y.; Kumar, P. Designing vegetation barriers for urban air pollution abatement: A practical review for appropriate plant species selection. *Npj Clim. Atmos. Sci.* **2020**, *3*, 12. [CrossRef]
52. Klingberg, J.; Broberg, M.; Strandberg, B.; Thorsson, P.; Pleijel, H. Influence of urban vegetation on air pollution and noise exposure—A case study in Gothenburg, Sweden. *Sci. Total Environ.* **2017**, *599–600*, 1728–1739. [CrossRef] [PubMed]
53. Setälä, H.; Viippola, V.; Rantalainen, A.L.; Pennanen, A.; Yli-Pelkonen, V. Does urban vegetation mitigate air pollution in northern conditions? *Environ. Pollut.* **2013**, *183*, 104–112. [CrossRef] [PubMed]
54. Jeanjean, A.P.R.; Buccolieri, R.; Eddy, J.; Monks, P.S.; Leigh, R.J. Air quality affected by trees in real street canyons: The case of Marylebone neighbourhood in central London. *Urban For. Urban Green.* **2017**, *22*, 41–53. [CrossRef]
55. Nowak, D.J.; Hirabayashi, S.; Bodine, A.; Greenfield, E. Tree and forest effects on air quality and human health in the United States. *Environ. Pollut.* **2014**, *193*, 119–129. [CrossRef]
56. Wu, J.; Wang, Y.; Qiu, S.; Peng, J. Using the modified i-Tree Eco model to quantify air pollution removal by urban vegetation. *Sci. Total Environ.* **2019**, *688*, 673–683. [CrossRef]
57. Alonso, R.; Vivanco, M.G.; Gonzalez-Fernandez, I.; Bermejo, V.; Palomino, I.; Garrido, J.L.; Elvira, S.; Salvador, P.; Artinano, B. Modelling the influence of peri-urban trees in the air quality of Madrid region (Spain). *Environ. Pollut.* **2011**, *159*, 2138–2147. [CrossRef]
58. Mirsanjari, M.M.; Zarandian, A.; Mohammadyari, F.; Visockiene, J.S. Investigation of the impacts of urban vegetation loss on the ecosystem service of air pollution mitigation in Karaj metropolis, Iran. *Environ. Monit. Assess.* **2020**, *192*, 501. [CrossRef]
59. Xing, Y.; Brimblecombe, P. Role of vegetation in deposition and dispersion of air pollution in urban parks. *Atmos. Environ.* **2019**, *201*, 73–83. [CrossRef]
60. Nemitz, E.; Vieno, M.; Carnell, E.; Fitch, A.; Steadman, C.; Cryle, P.; Holland, M.; Morton, R.D.; Hall, J.; Mills, G.; et al. Potential and limitation of air pollution mitigation by vegetation and uncertainties of deposition-based evaluations. *Philos. Trans. A Math. Phys. Eng. Sci.* **2020**, *378*, 20190320. [CrossRef]
61. Viippola, V.; Whitlow, T.H.; Zhao, W.; Yli-Pelkonen, V.; Mikola, J.; Pouyat, R.; Setälä, H. The effects of trees on air pollutant levels in peri-urban near-road environments. *Urban For. Urban Green.* **2018**, *30*, 62–71. [CrossRef]
62. Wang, J.; Bai, L.; Wang, S.; Wang, C. Research and application of the hybrid forecasting model based on secondary denoising and multi-objective optimization for air pollution early warning system. *J. Clean. Prod.* **2019**, *234*, 54–70. [CrossRef]
63. Chang, Y.-S.; Chiao, H.-T.; Abimannan, S.; Huang, Y.-P.; Tsai, Y.-T.; Lin, K.-M. An LSTM-based aggregated model for air pollution forecasting. *Atmos. Pollut. Res.* **2020**, *11*, 1451–1463. [CrossRef]
64. Sultanbekov, I.R.; Myshkina, I.Y.; Gruditsyna, L.Y. Development of an application for creation and learning of neural networks to utilize in environmental sciences. *Casp. J. Environ. Sci.* **2020**, *18*, 595–601. [CrossRef]
65. Karami, M.; Ahmadi, H.; Karami, K. Environmental impacts assessment of construction and utilization phases of tourism projects in Karun Dam IV, Iran. *Casp. J. Environ. Sci.* **2016**, *14*, 165–175. Available online: https://cjes.guilan.ac.ir/article_1772.html (accessed on 13 May 2021).
66. Kavyanifar, B.; Tavakoli, B.; Torkaman, J.; Mohammad Taheri, A.; Ahmadi Orkomi, A. Coastal solid waste prediction by applying machine learning approaches (Case study: Noor, Mazandaran Province, Iran). *Casp. J. Environ. Sci.* **2020**, *18*, 227–236. [CrossRef]
67. Bai, L.; Wang, J.; Ma, X.; Lu, H. Air Pollution Forecasts: An Overview. *Int. J. Environ. Res. Public Health* **2018**, *15*, 780. [CrossRef]

68. Sharma, N.; Taneja, S.; Sagar, V.; Bhatt, A. Forecasting air pollution load in Delhi using data analysis tools. *Procedia Comput. Sci.* **2018**, *132*, 1077–1085. [[CrossRef](#)]
69. Kaya, K.; Gunduz Oguducu, S. Deep Flexible Sequential (DFS) Model for Air Pollution Forecasting. *Sci. Rep.* **2020**, *10*, 3346. [[CrossRef](#)]
70. Gocheva-Ilieva, S.G.; Voynikova, D.S.; Stoimenova, M.P.; Ivanov, A.V.; Iliev, I.P. Regression trees modeling of time series for air pollution analysis and forecasting. *Neural Comput. Appl.* **2019**, *31*, 9023–9039. [[CrossRef](#)]
71. Madan, T.; Sagar, S.; Virmani, D. Air Quality Prediction using Machine Learning Algorithms—A Review. In Proceedings of the 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida, India, 18–19 December 2020; pp. 140–145. [[CrossRef](#)]
72. Mahalingam, U.; Elangovan, K.; Dobhal, H.; Valliappa, C.; Shrestha, S.; Kedam, G. A machine learning model for air quality prediction for smart cities. In Proceedings of the 2019 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET), Chennai, India, 21–23 March 2019; pp. 452–457. [[CrossRef](#)]
73. Pasupuleti, V.R.; Kalyan, P.; Reddy, H.K. Air quality prediction of data log by machine learning. In Proceedings of the 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 6–7 March 2020; pp. 1395–1399. [[CrossRef](#)]
74. Pan, B. Application of XGBoost algorithm in hourly PM_{2.5} concentration prediction. In *IOP Conference Series: Earth and Environmental Science*; IOP publishing: Bristol, UK, 2018; p. 012127. [[CrossRef](#)]
75. Ma, J.; Cheng, J.C.P.; Xu, Z.; Chen, K.; Lin, C.; Jiang, F. Identification of the most influential areas for air pollution control using XGBoost and Grid Importance Rank. *J. Clean. Prod.* **2020**, *274*, 64–71. [[CrossRef](#)]
76. Liu, B.; Tan, X.; Jin, Y.; Yu, W.; Li, C. Application of RR-XGBoost combined model in data calibration of micro air quality detector. *Sci. Rep.* **2021**, *11*, 15662. [[CrossRef](#)] [[PubMed](#)]
77. Kumar, K.; Pande, B.P. Air pollution prediction with machine learning: A case study of Indian cities. *Int. J. Environ. Sci. Technol.* **2022**, 1–16. [[CrossRef](#)] [[PubMed](#)]
78. Oliveri Conti, G.; Heibati, B.; Kloog, I.; Fiore, M.; Ferrante, M. A review of AirQ Models and their applications for forecasting the air pollution health outcomes. *Environ. Sci. Pollut. Res. Int.* **2017**, *24*, 6426–6445. [[CrossRef](#)]
79. Russo, A.; Soares, A.O. Hybrid Model for Urban Air Pollution Forecasting: A Stochastic Spatio-Temporal Approach. *Math. Geosci.* **2013**, *46*, 75–93. [[CrossRef](#)]
80. Madanipour, A. “Tehrān”. Encyclopedia Britannica. Available online: <https://www.britannica.com/place/Tehran> (accessed on 13 February 2022).
81. Rad, A.K.; Shariati, M.; Naghipour, A. Analyzing relationships between air pollutants and COVID-19 cases during lockdowns in Iran using Sentinel-5 data. *J. Air Pollut. Health* **2022**, *6*, 209–224. [[CrossRef](#)]
82. Rad, A.K.; Shariati, M.; Zarei, M. The impact of COVID-19 on air pollution in Iran in the first and second waves with emphasis on the city of Tehran. *J. Air Pollut. Health* **2021**, *5*, 181–192. [[CrossRef](#)]
83. World Data. Iran. Available online: <https://www.worlddata.info/asia/iran/index.php> (accessed on 21 January 2022).
84. Carreño-Conde, F.; Sipols, A.E.; de Blas, C.S.; Mostaza-Colado, D. A Forecast Model Applied to Monitor Crops Dynamics Using Vegetation Indices (NDVI). *Appl. Sci.* **2021**, *11*, 1859. [[CrossRef](#)]
85. EOS. NDVI. Available online: <https://eos.com/make-an-analysis/ndvi/> (accessed on 21 January 2022).
86. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [[CrossRef](#)]
87. Oliver, M.A.; Webster, R. Kriging: A method of interpolation for geographical information systems. *Int. J. Geogr. Inf. Syst.* **1990**, *4*, 313–332. [[CrossRef](#)]
88. Hyndman, R.J.; Fan, Y. Sample Quantiles in Statistical Packages. *Am. Stat.* **1996**, *50*, 361–365. [[CrossRef](#)]
89. Budholiya, K.; Shrivastava, S.K.; Sharma, V. An optimized XGBoost based diagnostic system for effective prediction of heart disease. *J. King Saud Univ.—Comput. Inf. Sci.* **2020**, *34*, 4514–4523. [[CrossRef](#)]
90. Chen, T.; Guestrin, C. XGBoost. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [[CrossRef](#)]
91. Friedman, H.F. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [[CrossRef](#)]
92. Bergstra, J.; Bengio, Y. Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* **2012**, *13*, 281–305. Available online: <https://www.jmlr.org/papers/v13/bergstra12a.html> (accessed on 13 May 2021).
93. Yli-Pelkonen, V.; Setälä, H.; Viippola, V. Urban forests near roads do not reduce gaseous air pollutant concentrations but have an impact on particles levels. *Landsc. Urban Plan.* **2017**, *158*, 39–47. [[CrossRef](#)]
94. Zhou, M.; Huang, Y.; Li, G. Changes in the concentration of air pollutants before and after the COVID-19 blockade period and their correlation with vegetation coverage. *Environ. Sci. Pollut. Res. Int.* **2021**, *28*, 23405–23419. [[CrossRef](#)] [[PubMed](#)]
95. Zheng, S.; Zhou, X.; Singh, R.; Wu, Y.; Ye, Y.; Wu, C. The Spatiotemporal Distribution of Air Pollutants and Their Relationship with Land-Use Patterns in Hangzhou City, China. *Atmosphere* **2017**, *8*, 110. [[CrossRef](#)]
96. Prakasam, C.; Aravinth, R.; Nagarajan, B. Estimating NDVI and LAI as a precursor for monitoring air pollution along the BBN industrial corridor of Himachal Pradesh, India. *Mater. Today Proc.* **2022**, *61*, 593–603. [[CrossRef](#)]
97. Sun, S.; Li, L.-J.; Zhao, W.-J.; Qi, M.-X.; Tian, X.; Li, S.-S. Variation in Pollutant Concentrations and Correlation Analysis with the Vegetation Index in Beijing-Tianjin-Hebei. *Huan Jing Ke Xue Huanjing Kexue* **2019**, *40*, 1585–1593. [[CrossRef](#)]

98. Qiao, Z.; Wu, F.; Xu, X.; Yang, J.; Liu, L. Mechanism of Spatiotemporal Air Quality Response to Meteorological Parameters: A National-Scale Analysis in China. *Sustainability* **2019**, *11*, 3957. [[CrossRef](#)]
99. Kayes, I.; Shahriar, S.A.; Hasan, K.; Akhter, M.; Kabir, M.M.; Salam, M.A. The relationships between meteorological parameters and air pollutants in an urban environment. *Glob. J. Environ. Sci. Manag.* **2019**, *5*, 265–278. [[CrossRef](#)]
100. Sezer Turalioglu, F.; Nuhoglu, A.; Bayraktar, H. Impacts of some meteorological parameters on SO₂ and TSP concentrations in Erzurum, Turkey. *Chemosphere* **2005**, *59*, 1633–1642. [[CrossRef](#)] [[PubMed](#)]
101. Akpınar, S.; Oztop, H.F.; Kavak Akpınar, E. Evaluation of relationship between meteorological parameters and air pollutant concentrations during winter season in Elazığ, Turkey. *Environ. Monit. Assess.* **2008**, *146*, 211–224. [[CrossRef](#)] [[PubMed](#)]
102. Ilten, N.; Selici, A.T. Investigating the impacts of some meteorological parameters on air pollution in Balıkesir, Turkey. *Environ. Monit. Assess.* **2008**, *140*, 267–277. [[CrossRef](#)]
103. Kliengchuay, W.; Cooper Meeyai, A.; Worakhunpiset, S.; Tantrakarnapa, K. Relationships between Meteorological Parameters and Particulate Matter in Mae Hong Son Province, Thailand. *Int. J. Environ. Res. Public Health* **2018**, *15*, 2801. [[CrossRef](#)]
104. Jassim, M.S.; Coskuner, G.; Munir, S. Temporal analysis of air pollution and its relationship with meteorological parameters in Bahrain, 2006–2012. *Arab. J. Geosci.* **2018**, *11*, 62. [[CrossRef](#)]
105. Iskandaryan, D.; Ramos, F.; Trilles, S. Air Quality Prediction in Smart Cities Using Machine Learning Technologies based on Sensor Data: A Review. *Appl. Sci.* **2020**, *10*, 2401. [[CrossRef](#)]
106. Tao, Q.; Liu, F.; Li, Y.; Sidorov, D. Air Pollution Forecasting Using a Deep Learning Model Based on 1D Convnets and Bidirectional GRU. *IEEE Access* **2019**, *7*, 76690–76698. [[CrossRef](#)]
107. Niska, H.; Hiltunen, T.; Karppinen, A.; Ruuskanen, J.; Kolehmainen, M. Evolving the neural network model for forecasting air pollution time series. *Eng. Appl. Artif. Intell.* **2004**, *17*, 159–167. [[CrossRef](#)]
108. Kang, G.K.; Gao, J.Z.; Chiao, S.; Lu, S.; Xie, G. Air Quality Prediction: Big Data and Machine Learning Approaches. *Int. J. Environ. Sci. Dev.* **2018**, *9*, 8–16. [[CrossRef](#)]
109. Samal, K.K.R.; Babu, K.S.; Das, S.K.; Acharaya, A. Time Series based Air Pollution Forecasting using SARIMA and Prophet Model. In Proceedings of the 2019 International Conference on Information Technology and Computer Communications—ITCC 2019, Singapore, 16–18 August 2019; pp. 80–85. [[CrossRef](#)]
110. Dua, R.D.; Madaan, D.M.; Mukherjee, P.M.; Lall, B.L. Real Time Attention Based Bidirectional Long Short-Term Memory Networks for Air Pollution Forecasting. In Proceedings of the 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService), Newark, CA, USA, 4–9 April 2019; pp. 151–158. [[CrossRef](#)]
111. Liao, Q.; Zhu, M.; Wu, L.; Pan, X.; Tang, X.; Wang, Z. Deep Learning for Air Quality Forecasts: A Review. *Curr. Pollut. Rep.* **2020**, *6*, 399–409. [[CrossRef](#)]
112. Bellinger, C.; Mohamed Jabbar, M.S.; Zaiane, O.; Osornio-Vargas, A. A systematic review of data mining and machine learning for air pollution epidemiology. *BMC Public Health* **2017**, *17*, 907. [[CrossRef](#)] [[PubMed](#)]
113. Zhu, D.; Cai, C.; Yang, T.; Zhou, X. A Machine Learning Approach for Air Quality Prediction: Model Regularization and Optimization. *Big Data Cogn. Comput.* **2018**, *2*, 5. [[CrossRef](#)]
114. Baklanov, A.; Hänninen, O.; Slørdal, L.H.; Kukkonen, J.; Bjergene, N.; Fay, B.; Finardi, S.; Hoe, S.C.; Jantunen, M.; Karppinen, A.; et al. Integrated systems for forecasting urban meteorology, air pollution and population exposure. *Atmos. Chem. Phys.* **2007**, *7*, 855–874. [[CrossRef](#)]
115. Rad, A.K.; Zarei, M.; Pourghasemi, H.R.; Tiefenbacher, J.P. Chapter 27—The COVID-19 crisis and its consequences for global warming and climate change. In *Computers in Earth and Environmental Sciences*; Pourghasemi, H.R., Ed.; Elsevier: Amsterdam, The Netherlands, 2022; pp. 377–385. [[CrossRef](#)]