



# Air pollution in marmara region before and during the COVID-19 outbreak

Ibrahim Arpaci · Serhat Kilicarslan ·  
Omer Aslan · Ibrahim Ozturk

Received: 29 October 2022 / Accepted: 9 May 2023 / Published online: 30 May 2023  
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2023, corrected publication 2023

**Abstract** The lockdowns and curfews during the COVID-19 pandemic have halted economic and transportation activities across the world. This study aims to investigate air pollution levels in the Marmara region, particularly in Istanbul, before and during the COVID-19 pandemic. The study used real data provided by the General Directorate of Meteorology and applied three machine learning algorithms (ANN, RBFreg, and SMOREg) to analyze air pollution data. In addition, a one-sample t-test was performed to compare air pollution levels before and during the COVID-19 pandemic in the Marmara region and Istanbul. The results of the study showed a significant reduction in the particulate matter (PM) value, which indicates the degree of air pollution, in both the Marmara region and Istanbul during the COVID-19 pandemic. The one-sample t-test results showed that

the reduction in air pollution levels was statistically significant in both areas ( $t=11.45$ ,  $p<.001$  for the Marmara region, and  $t=3.188$ ,  $p<.001$  for Istanbul). These findings have important practical implications for decision-makers planning for a more sustainable environment. Overall, the study provides valuable insights into the impact of the COVID-19 pandemic on air pollution levels in the Marmara region, particularly in Istanbul. The application of machine learning algorithms and statistical analysis provides a rigorous approach to the investigation of this important issue by comparing before and during the COVID-19 outbreak.

**Keywords** Air pollution · COVID-19 · Türkiye · Regression

---

I. Arpaci · S. Kilicarslan (✉) · O. Aslan  
Faculty of Engineering and Natural Sciences, Department  
of Software Engineering, Bandirma Onyedi Eylul  
University, Balikesir, Türkiye  
e-mail: skilicarslan@bandirma.edu.tr

I. Arpaci  
e-mail: iarpaci@bandirma.edu.tr

O. Aslan  
e-mail: oaslan@bandirma.edu.tr

I. Ozturk  
Faculty of Engineering, Department of Electrical  
and Electronics Engineering, Osmaniye Korkut Ata  
University, Osmaniye, Türkiye  
e-mail: ibrahimozturk@osmaniye.edu.tr

## Introduction

These days, air pollution is one of the largest threats to environmental and public health. The levels of  $CO_2$  carbon dioxide,  $NO$  (nitric oxide),  $NO_2$  nitrogen dioxide,  $SO_2$  sulfur dioxide, PM (particulate matter), and  $SO_3$  (ozone) are increasing worldwide, which can be categorized as toxic gases in the air (Adam et al., 2021). Those gases increase the risk of heart attacks, hypertension, potentially fatal respiratory tract infections, and many other diseases. Traffic, social mobility, and industrial and agricultural activities are the major factors affecting air quality (Jephcote et al.,

2021). It is important to know that several measures must be taken to reduce air pollution and increase air quality. The lockdown periods might reduce the main sources of pollutants.

A novel virus type was first reported in December 2019, in Wuhan, China. The coronavirus, called COVID-19, has rapidly spread around China as well as other parts of the world. The World Health Organization (WHO) declared COVID-19 a pandemic on April 11, 2020. On the same day, the first coronavirus case was also confirmed in Türkiye. A nationwide lockdown was imposed in Türkiye on March 16, 2020, while facing the first death and 47th case on March 16, 2020. The quarantine and lockdown periods in response to the COVID-19 pandemic remained in April, May, and June of 2020. In this study, we would like to observe whether the lockdown period statistically reduces air pollution significantly or not (Kılıçarslan et al., 2023; Adem & Kılıçarslan, 2021).

Air pollution data is recorded at fully automated monitoring stations belonging to the Ministry of Environment and Urbanization, established in all 81 provinces in Türkiye. Since 2012, the air quality data have been continuously published in real-time through the National Air Quality Monitoring Network (NAQMN) in a new format (Ministry, 2022). The air monitoring network has 195 stable and 4 mobile air quality monitoring stations in Türkiye.

In this research, the amount of air pollution before and during the COVID-19 lockdown was investigated. As far as we know, this research study is the first to measure air pollution before and during the COVID-19 outbreak. To measure air pollution, we used three machine learning algorithms, namely, ANN, RBFreg, and SMOreg, to classify the collected data. Besides, a t-test was used to find out air pollution levels before and during the COVID-19 pandemic. The experiments were carried out on real data that was provided by the General Directorate of Meteorology in Türkiye. Since the city of Istanbul and the Marmara region are the most crowded and industrialized regions, the study was focused on Istanbul as well as the Marmara region. The experimental test results showed that the level of air pollution decreased sharply during the COVID-19 lockdown.

The study helps to extract data-driven insights into the impact of a pandemic on air pollution levels. It is mainly focused on the Marmara region and the city of Istanbul by the comparison of before and during

the COVID-19 outbreak. The findings of the research using ANN, RBFreg, and SMOreg algorithms show a significant reduction in air pollution levels during the pandemic. The outcome of the data-driven analysis has important practical implications for decision-makers in planning for a more sustainable environment, especially in the Marmara region and Istanbul.

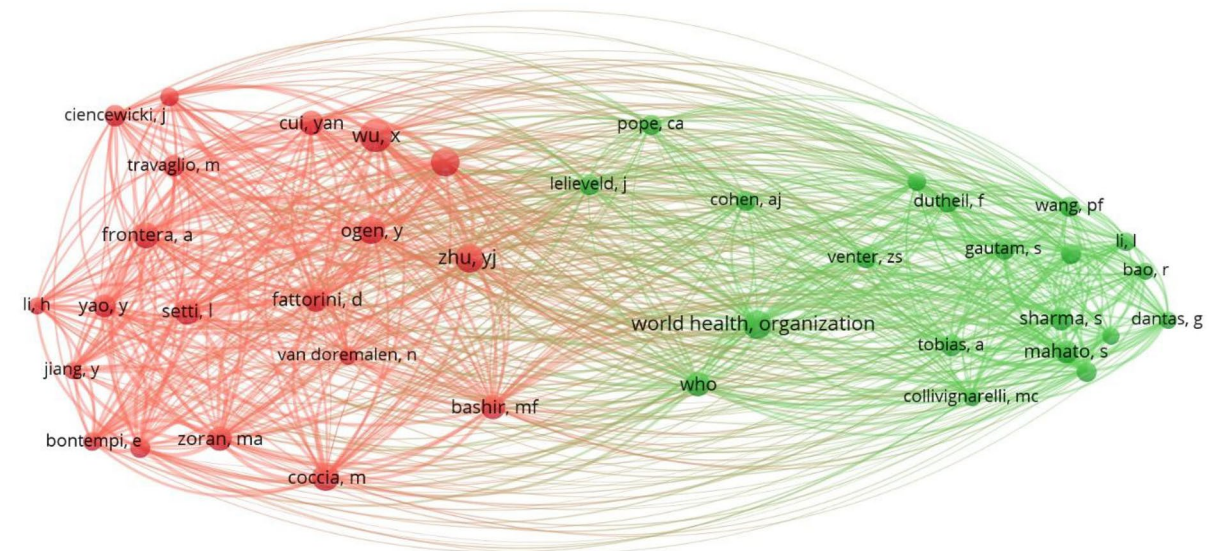
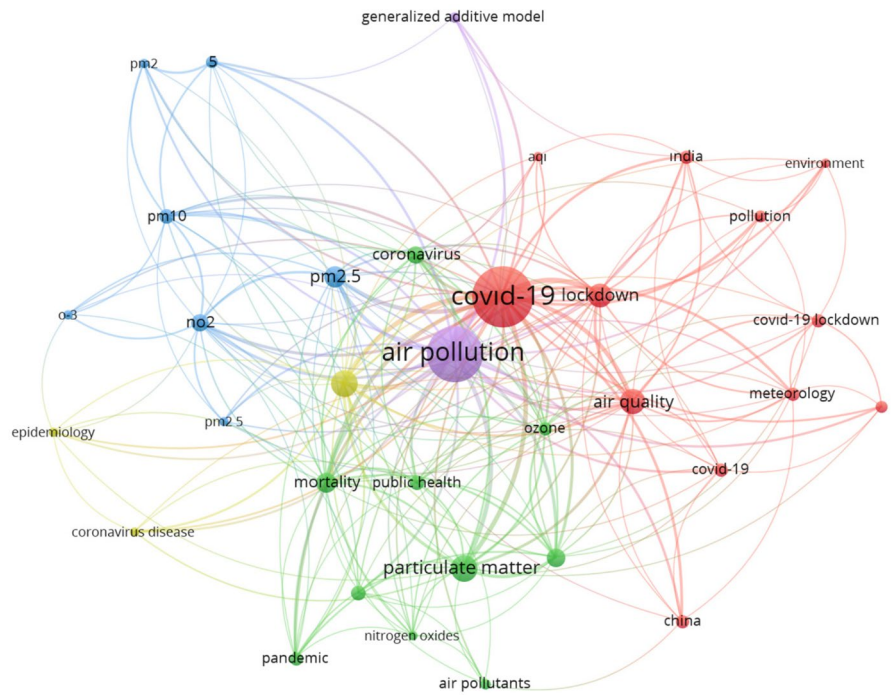
The rest of the paper is organized as follows: In Sect. 2, the bibliometric review is explained. In this section, the studies that were somehow related to COVID-19 were examined. In Sect. 3, a proposed method to analyze the data is presented. In this context, three machine learning classifiers, including ANN, RBFreg, and SMOreg, and a description of the data are given in detail. In Sect. 4, experimental results of machine learning algorithms are listed, and t-test results are compared for air pollution before and during the COVID-19 outbreak. In Sects. 5 and 6, the discussion, key findings, and limitations of the study are listed. Finally, the conclusion is given in Sect. 7.

## Bibliometric Review

Scientific studies on the links between COVID-19 and air pollution have begun to appear since 2020. There are 338 articles that include the keywords “air pollution” and “COVID-19” within the title. A bibliometric mapping analysis using VOSviewer was conducted, including the studies indexed in the Web of Science (WOS) database. The results revealed that the most commonly used keywords in these studies are “COVID-19” ( $f=182$ ), “air pollution” ( $f=148$ ), and “particulate matter” ( $f=36$ ). Figure 1 illustrates the most commonly used author keywords in these studies.

The results indicated that Zhu et al. (2020) (461 WOS citations), Wang et al. (2020) (328 citations), Le et al. (2020) (283 citations), Fattorini and Regoli (2020) (277 citations), and Venter et al. (2020) (255 citations) were the most cited authors. Furthermore, “Science of the Total Environment” (24 documents, 1536 citations), “Environmental Pollution” (21 documents, 1097 citations), and “Environmental Research” (24 documents, 388 citations) were the most cited journals. Figure 2 indicates that there are 40 co-cited authors with a minimum of 30 citations. Wu et al. (2020) (110 citations), the World Health

**Fig. 1** The most used keywords



**Fig. 2** The most co-cited authors

Organization (85 citations), and Zhu et al. (2020) (84 citations) were the most co-cited sources.

The bibliometric mapping analysis results indicated that several studies focused on air pollution during the pandemic. However, a limited

number of studies focused on the comparison of air pollution before and during the COVID-19 outbreak. Therefore, the current study aimed to investigate air pollution before and during the pandemic.

Applying three machine learning algorithms and a one-sample t-test to real data provided by the General Directorate of Meteorology, the study employs a novel methodology to investigate air pollution levels in the Marmara region and Istanbul before and during the COVID-19 pandemic. This makes it distinct from previous studies. This strategy offers an in-depth, data-driven investigation of the pandemic's effects on air pollution levels. The study contributes to our knowledge of the impact of the COVID-19 pandemic on air pollution levels in the Marmara region and the city of Istanbul. The findings show a significant reduction in air pollution levels during the pandemic, which has important practical implications for decision-makers in planning for a more sustainable environment. The study provides important practical implications for decision-makers in the Marmara region and Istanbul in terms of planning for a more sustainable environment. The findings suggest that the reduction in air pollution levels during the pandemic could be maintained through sustainable policies and practices in the future. The study highlights the significance of air pollution for public health in the Marmara region and Istanbul. The findings suggest that reducing air pollution levels could have positive effects on public health, particularly during pandemics such as COVID-19.

## Methods and Material

### Machine Learning Methods

Developing machine learning models involves several steps, and performance indices are typically used to evaluate the effectiveness of each step. Data preprocessing is the first step that involves preparing the data for use in the machine learning model (Al-Sharafi et al., 2022). Feature selection, as a second step, involves selecting the most relevant features from the dataset to use in the machine learning model. In the study, six features, including carbon dioxide, nitric oxide, nitrogen dioxide, sulfur dioxide, particulate matter (PM), and ozone concentration levels, were used. Model selection involves selecting the appropriate machine learning algorithm to use for the given problem. In the study, we have used three machine learning algorithms; Artificial Neural Network (ANN), Radial Basis Functions Regression (RBFreg),

and Sequential Minimal Optimization Regression (SMOreg). Performance indices for this step include accuracy, recall, F1 score, precision, and area under the ROC curve.

### Artificial Neural Network

The idea of an artificial neural network (ANN) is a machine learning algorithm based on the development of mathematical models that reference the functionality of the human brain, and can be imitated using silicon and wires as living neurons and dendrites (Schalkoff, 1997). The ANN has the ability to make quick decisions on samples. In other words, ANN has been developed to produce accurate and fast solutions to complex problems (Kaastra & Boyd, 1996; Zhang et al., 1998). The ANN model consists of three layers; input, hidden, and output (Yöntem et al., 2019). In the ANN architecture, the best net value is obtained with the help of the activation functions used in the hidden layer. After obtaining the best net value, the output layer is presented with the output values produced by the network. In the ANN method, the optimum number of neurons in the hidden layer was found by the trial-and-error method. Neuron numbers in the hidden layer were 10. The hyperbolic tangent sigmoid transfer function (tansig) and linear transfer function (purelin) were used for the activation of the hidden and output layers of the network, respectively.

### Radial Basis Functions Regression (RBFreg)

Radial Basis Functions Regression (RBFreg) was developed by Witten et al. (2016) to provide solutions to estimation problems. The RBFreg model appears to be a curve-fitting approach suitable for multi-dimensional datasets during the training phase. Thus, it provides the most ideal curve for the prediction result, close to the true value of the datasets. The RBFreg model consists of input, hidden, and output layers. RBFreg can learn to estimate nonlinear problems and the underlying trend using many Gaussian curves (Pereira & Aires, 2018). The training of the RBFreg algorithm works by finding the cell centers and optimizing the weights in the output layer. We have implemented the RBFreg method with standard parameters, which were as follows: batch size=100, numDecimalPlaces=2, numFunctions=2, numTree=2,

numThreads = 1, poolSize = 1, ridge = 0.01, seed = 1, and tolerance 1.0E-6 with one hidden layer.

*Sequential Minimal Optimization Regression*

The Sequential Minimal Optimization Regression (SMOreg) algorithm proposed by Smola and Schölkopf (2004) is a machine learning algorithm based on statistical learning theory. SMOreg is an improved support vector machine (SVM). SMOreg was developed to predict complex regression problems (Shevade et al., 2000; Smola and Schölkopf, 2004). In the SMOreg algorithm, there is an error term ( $\epsilon$ ), a regulation factor (C), and kernel parameters that can directly affect the performance. The error term controls noise on the performance of the model and the regulation factor is used to minimize error in the model. Also, the kernel parameter affects the performance of the model in the regression process (Shevade et al., 2000; Smola and Schölkopf, 2004). While this method was used with the Weka software in the study, the necessary parameters were determined as complexity parameter = 0.1001, regressor optimizer = RegSMOImproved by the trial and error method.

**Material**

On the collected dataset, we performed ANN, RBFreg, and SMOreg algorithms to identify air pollution before and during the COVID-19 pandemic. We measured air pollution based on six different gas concentration levels. Missing data points were removed from the dataset. In order to classify the data points correctly, the RMSE, MAE, MAPE, and  $R^2$  errors were measured for performing the ANN, RBFreg, and SMOreg algorithms. Furthermore, a t-test was performed to compare the air pollution before and during the COVID-19 periods for a given dataset. The details of the material given are as follows:

*Study Area*

The study focuses on the Marmara region, which is the most developed geographical region among the seven main geographical regions of Türkiye. Although it has the second-smallest area, it has the largest population in the country. The Marmara region comprises 10 provinces (or il) including Balıkesir, Bilecik, Bursa, Edirne, İstanbul, Kırklareli, Kocaeli, Sakarya, Tekirdağ, and Yalova. Hence, air pollutants are also recorded at stations in the Marmara Region. The parameters monitored in each province are given in Table 1. National Air Quality Monitoring Network Data Set on the portal for meteorological data ([https://sim.csb.gov.tr/STN/STN\\_Report/StationDataDownloadNew](https://sim.csb.gov.tr/STN/STN_Report/StationDataDownloadNew) visit date: 24.04.2023).

*Parameters*

A number of different criteria for pollutants are used on the air quality index (AQI). However, there are six available and common gaseous air pollutants illustrated in Table 1: Carbon monoxide (CO), ozone  $O_3$ , sulfur dioxide  $SO_2$ , nitrogen dioxide  $NO_2$ , and particulate matter, which is currently split into PM10 and PM2.5 size fractions. PM10 (particulate monitors) is particles with a diameter of 10  $\mu m$  or less; PM2.5 are particles with a diameter of 2.5  $\mu m$  or less.

*Data*

In order to determine the local air quality, the data generated by the air quality stations belonging to the Ministry of Environment and Urbanization in provinces of the Marmara region were examined in Table 1. The data was captured in daily periods through the stations between 11 March 2015, and 22 March 2022 to measure the impact of multiple pollutants before, during, and after the COVID-19

**Table 1** Summary of monitored air pollutants at each location discussed in this paper

| Pollutant Symbol  | Pollutant Name     | Standard          | Definition                                       |
|-------------------|--------------------|-------------------|--|
| PM <sub>10</sub>  | Particulate Matter | Mg/m <sup>3</sup> | Particles with a diameter of 10 $\mu m$ or less  |
| PM <sub>2.5</sub> | Particulate Matter | Mg/m <sup>3</sup> | Particles with a diameter of 2.5 $\mu m$ or less |
| SO <sub>2</sub>   | Sulfur Dioxide     | Mg/m <sup>3</sup> | Gaseous air pollutant                            |
| CO                | Carbon Monoxide    | Mg/m <sup>3</sup> | Gaseous air pollutant                            |
| NO <sub>2</sub>   | Nitrogen Dioxide   | Mg/m <sup>3</sup> | Gaseous air pollutant                            |
| O <sub>3</sub>    | Ozone              | Mg/m <sup>3</sup> | Gaseous air pollutant                            |

pandemic period. In this study, three different techniques have been tested to obtain the best model to predict the 2015–2021 data set of the studied air pollutants, namely ANN, RBFreg, and SMOReg. The best model obtained will be used to estimate 2022 air quality levels. The data set is divided into K groups by random method or other criteria, such as grouped by yearly data. One group is chosen for the testing data set and the remaining K-1 groups are used as the training data set. The process is repeated K times and calculates the average model performance. The K value is different depending on the researchers. Some research articles applied CV with only the training data set, and fixed testing data set with the previous year's data (Moustris et al., 2013; Sayegh et al., 2014; S. Yadav et al., 2015).

## Results

### Machine Learning Results

In the study, the predictive performances of reference Particulate Matter (PM10) values by using the climate parameters measured with machine learning techniques, including SMOReg, RBFreg, and ANN models, were investigated. In the study, experimental evaluations of the dataset were carried out using the k-fold=10 cross-validation method. The performances of the predictive algorithms were compared according to MAE, RMSE, MAPE, and  $R^2$  criteria. The results obtained in the experimental evaluation are presented in Table 2. The flow chart of the study is given in Fig. 3.

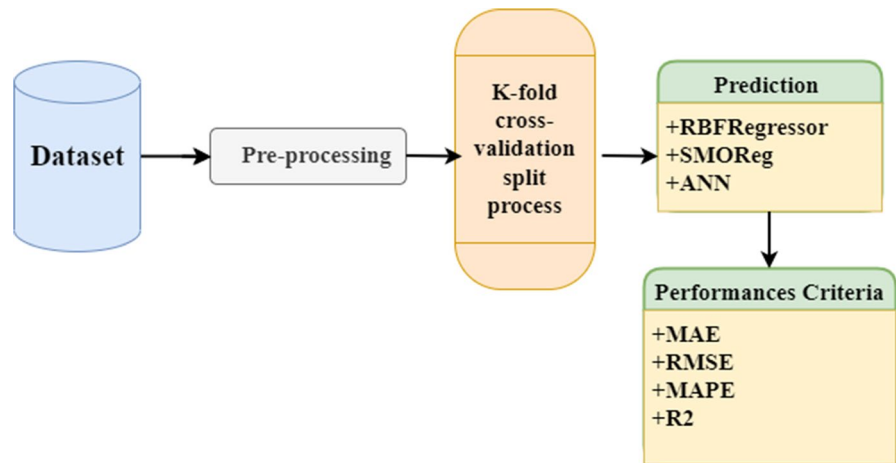
Based on stations, the success of estimating PM-10 of the three models in terms of the  $R^2$  performance measure varied between 0.50 and 0.958. However, using only the  $R^2$  evaluation criterion among estimation studies is an erroneous approach. Therefore, it is ensured that the estimation process is made more realistic by using different evaluation criteria. For the Balikesir station, the RBFRegressor model represented the relevant data proportionally better than SMOReg, and ANN models by 0.3%, and 0.1% respectively. In Table 2, it is seen that the best performance is obtained at Balikesir station with MAE 2.6805, RMSE 4.0741, MAPE 11.9786, and  $R^2$  0.9582. In Table 4, it is seen that poor performance is obtained at Istanbul station with MAE 12.6994,

**Table 2** Performance measures of the PM10 prediction models for datasets

| Cities     | Methods      | MAE     | RMSE    | MAPE %  | $R^2$  |
|------------|--------------|---------|---------|---------|--------|
| Balikesir  | RBFRegressor | 2.6805  | 4.0741  | 11.9786 | 0.9582 |
|            | SMOReg       | 4.2193  | 5.9329  | 21.2874 | 0.9139 |
|            | ANN          | 2.6355  | 4.1365  | 11.1369 | 0.9574 |
| Bilecik    | RBFRegressor | 2.6266  | 4.0628  | 28.3049 | 0.8345 |
|            | SMOReg       | 2.7302  | 4.2538  | 28.1909 | 0.8202 |
|            | ANN          | 2.9558  | 4.4574  | 33.8358 | 0.7988 |
| Bursa      | RBFRegressor | 4.7805  | 7.0976  | 21.2061 | 0.9439 |
|            | SMOReg       | 5.5694  | 7.9204  | 26.2775 | 0.9301 |
|            | ANN          | 8.9520  | 12.2371 | 45.2808 | 0.8357 |
| Edirne     | RBFRegressor | 19.5026 | 24.8331 | 30.2844 | 0.7190 |
|            | SMOReg       | 19.795  | 25.626  | 44.6604 | 0.7030 |
|            | ANN          | 8.4817  | 10.6792 | 58.2998 | 0.6662 |
| Istanbul   | RBFRegressor | 10.4587 | 14.7064 | 31.1158 | 0.6009 |
|            | SMOReg       | 10.8154 | 15.1340 | 32.8404 | 0.5686 |
|            | ANN          | 12.6994 | 16.9937 | 39.6141 | 0.5076 |
| Kırklareli | RBFRegressor | 1.1100  | 1.8198  | 8.1883  | 0.9518 |
|            | SMOReg       | 1.3778  | 2.5063  | 9.8638  | 0.9288 |
|            | ANN          | 1.1297  | 1.7253  | 8.6919  | 0.9573 |
| Kocaeli    | RBFRegressor | 5.5910  | 7.7896  | 15.5996 | 0.8679 |
|            | SMOReg       | 5.5108  | 7.7718  | 15.4546 | 0.8686 |
|            | ANN          | 5.9742  | 7.9633  | 16.8022 | 0.8665 |
| Sakarya    | RBFRegressor | 8.6276  | 11.8294 | 25.8207 | 0.8561 |
|            | SMOReg       | 8.9272  | 12.8861 | 26.1590 | 0.8299 |
|            | ANN          | 9.4141  | 12.8193 | 28.2537 | 0.8351 |
| Tekirdağ   | RBFRegressor | 3.5257  | 4.8476  | 30.1486 | 0.8280 |
|            | SMOReg       | 3.8892  | 5.5682  | 31.5041 | 0.7757 |
|            | ANN          | 3.8054  | 5.1506  | 33.6802 | 0.8084 |
| Yalova     | RBFRegressor | 7.2405  | 9.5000  | 18.5724 | 0.8626 |
|            | SMOReg       | 7.5838  | 9.8780  | 20.1693 | 0.8529 |
|            | ANN          | 7.7717  | 10.6544 | 19.9938 | 0.8273 |

RMSE 16.9937, MAPE 39.6141, and  $R^2$  0.5076 with the ANN model. When the  $R^2$  performance measure of the RBFRegressor model was compared with those of the other models, at Bilecik, Bursa, Kırklareli, Kocaeli Sakarya, Tekirdag, and Yalova stations,

**Fig. 3** The flow chart of study



showed the best  $R^2$  value. The RBFRegressor and ANN models were compared with those of the other models, at Kırklareli station, which showed the best MAE, RMSE, and MAPE values. Table 2 shows the differences between the experimental evaluation performance results. These differences between performance criteria are due to too much missing data in the datasets.  $R^2$  represents the rate of variance explained by the independent variables of the dependent variable (Fisher, 1922). MAE represents the average absolute differences between observed and predicted values (Willmott & Matsuura, 2005). RMSE is a quadratic metric that measures the size of the error and is often used to find the distance between the predicted values and the observed values of the estimator (Willmott & Matsuura, 2005). MAPE is expressed as the mean absolute percent error to measure estimation accuracy in regression and time series models. Lewis (1982) states that models with a MAPE value of less than 10% are “very good”, models between 10–20% are “good”, models between 20–50% are “acceptable”, and models above 50% are “wrong and faulty”. Since the MAPE value in our study is between 20–50% for all stations, the results obtained are expressed as acceptable. The performances of the Istanbul station with the lowest estimation performance result and the Kırklareli station with the highest performance for all models are visually compared in Figs. 4 and 5. In Kırklareli station, the regression lines of all models almost coincide with the  $x=y$  line. In addition, it is observed that the Istanbul station deviates more from the regression line  $x=y$  of all models.

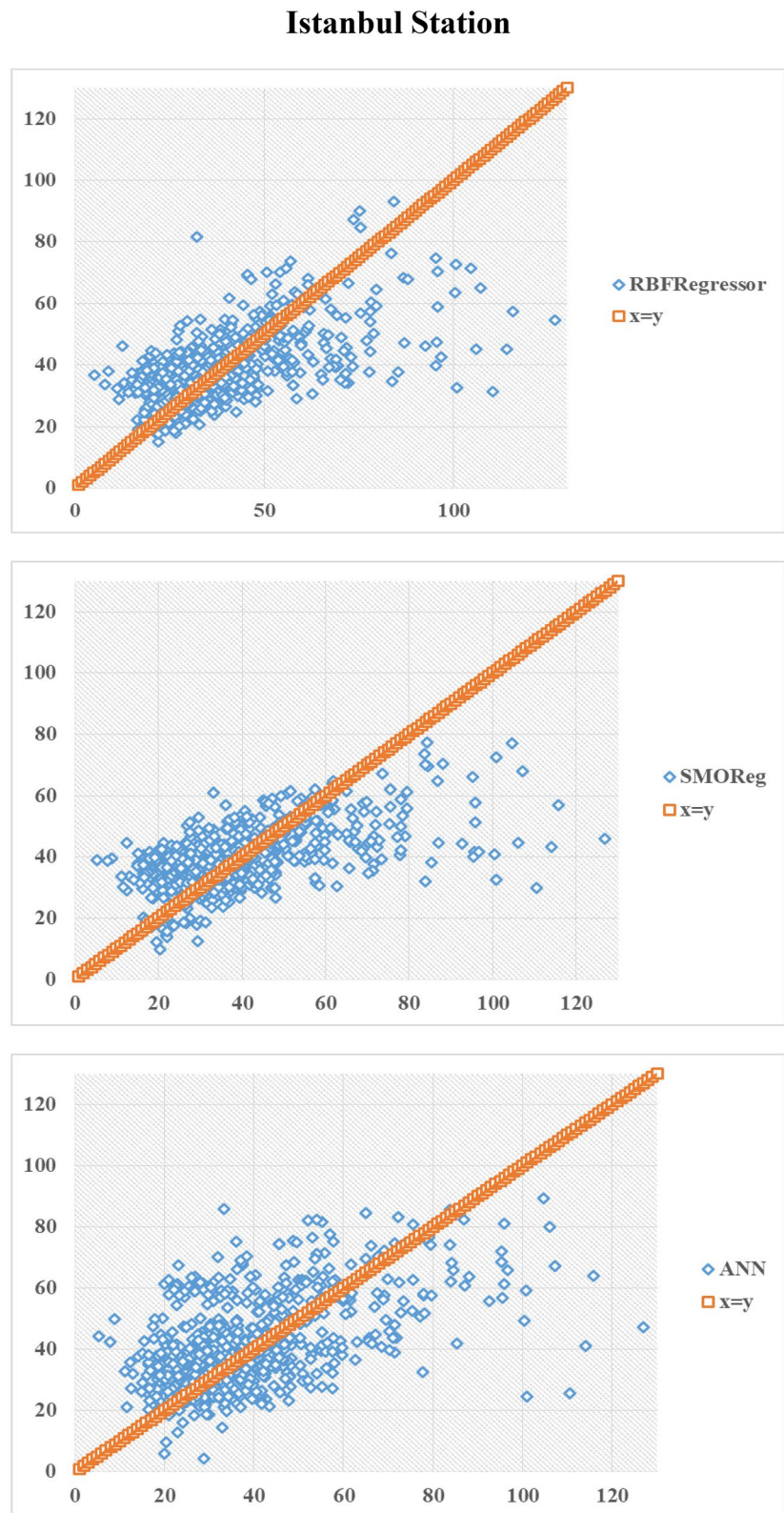
Comparison of Air Pollution Before and During the COVID-19 Outbreak

The first case in Türkiye was recorded on 11 March 2020, and the first death due to COVID-19 in the country occurred on 15 March 2020 (BBC, 2021). On 4 May 2020, the “Coronavirus Scientific Advisory Board” stated the need for public health interventions such as mandatory mask use and lockdowns (Arpaci et al., 2020). With the highest infection rate in Europe, Türkiye entered its first nationwide lockdown on April 29, 2021 (Arpaci et al., 2021; BBC, 2021).

The present study investigated air pollution before and during the COVID-19 pandemic. Therefore, the date between March 11, 2018, and March 10, 2020 was defined as the pre-COVID-19 pandemic period, and the date between March 11, 2020, and March 11, 2022 was defined as the COVID-19 pandemic period. Accordingly, the first data set included records between the dates of March 11, 2018, and March 10, 2020, and the second data set included records between the dates of March 11, 2020, and March 11, 2022.

The population of Türkiye in 2021 was reported as 84 million 680 thousand 273 people (TUIK, 2021). The Marmara region has the highest population density in Türkiye and more than 30% of the total population lives there. The Marmara region is the most trafficked and industrialized region of Türkiye and one-third of the country’s industrial activities take place in this region. There are 10 cities in the Marmara region, namely Balıkesir, Bilecik, Bursa, Edirne, Istanbul, Kırklareli,

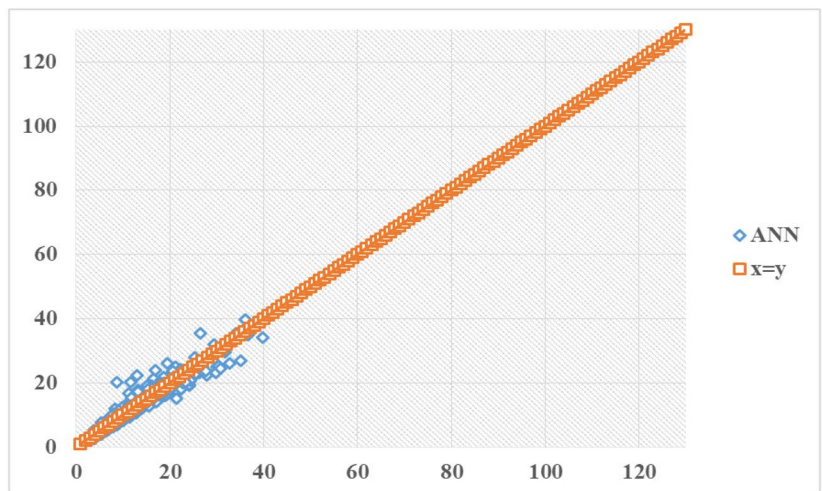
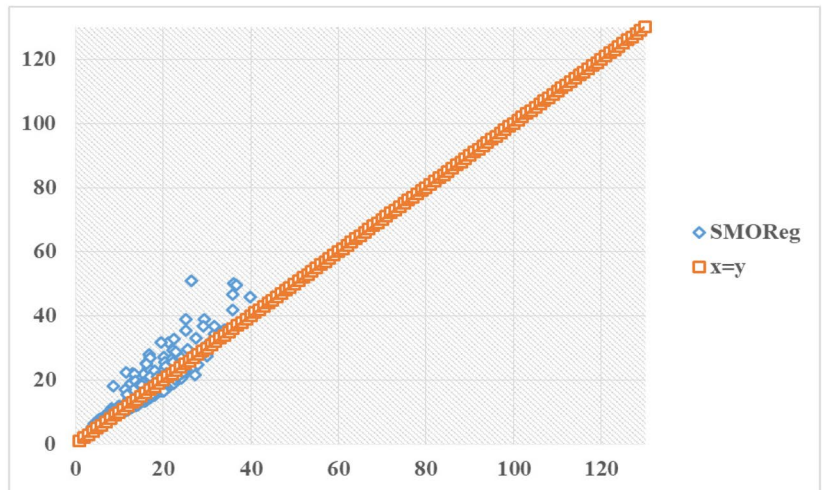
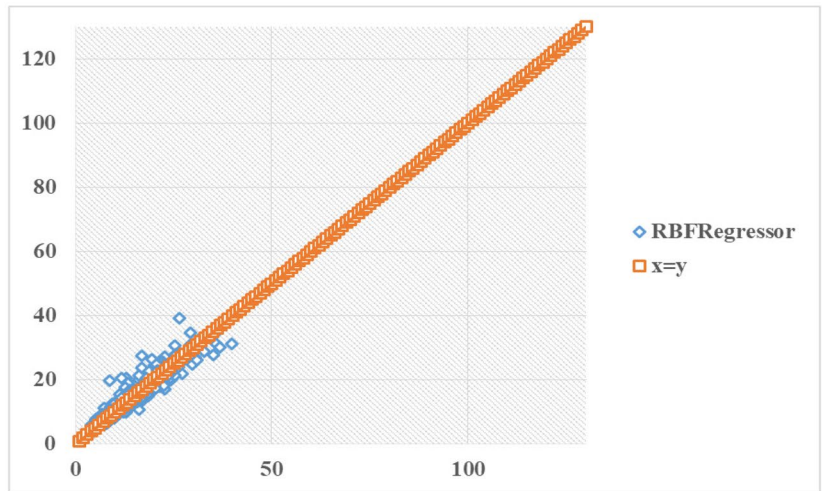
**Fig. 4** The regression plots for values of PM-10 and predicted models (RBFRegressor, SMOReg, ANN) for Istanbul station





**Fig. 5** The regression plots for values of PM-10 and predicted models (RBFRegressor, SMOReg, ANN) for Kırklareli station

### Kırklareli Station



Kocaeli, Sakarya, Tekirdag, and Yalova. The population of Istanbul, which is the most crowded city in Türkiye, is 15 million 840 thousand 900, and its ratio to the country's population is 18.71%. Therefore, the main analysis is focused on the Marmara region and the city of Istanbul.

A one sample t-test was performed to compare air pollution before and during COVID-19 in the Marmara region. Table 3 shows the comparison results of air pollution before and during COVID-19 in the Marmara region. The mean value of PM10 pollution ( $M=52.64$ ,  $SD=32.33$ ) before the pandemic was significantly lower than the PM10 pollution ( $M=47.07$ ,  $SD=28.97$ ) during the pandemic;  $t$  ( $df=15,896$ ) = 11.45,  $p < 0.001$ . This indicated that air pollution was reduced significantly in the Marmara region due to the lockdowns and curfews during the pandemic.

Further, a one sample t-test was performed to compare air pollution before and during COVID-19 in Istanbul. Table 4 shows the comparison results of the air pollution before and during COVID-19 in Istanbul. The mean value of PM10 pollution ( $M=50.49$ ,  $SD=28.66$ ) before the pandemic was significantly lower than the PM10 pollution ( $M=45.98$ ,  $SD=28.68$ ) during the pandemic;  $t$  ( $df=1643$ ) = 3.188,  $p < 0.001$ . This indicated that air pollution was reduced significantly in Istanbul due to the lockdowns and curfews during the pandemic.

## Discussion

Our results indicated that air pollution has decreased considerably in Istanbul city as well as the Marmara region during the COVID-19 pandemic lockdown. The concentrations of six gases that are indicators

**Table 3** Comparison of the air pollution before and during COVID-19 in Marmara region

|                 |        | N    | Mean   | SD     | <i>t</i> | df     | <i>p</i> |
|-----------------|--------|------|--------|--------|----------|--------|----------|
| PM10            | Before | 7897 | 52.64  | 32.33  | 11.45    | 15,896 | 0.000    |
|                 | During | 8001 | 47.07  | 28.97  |          |        |          |
| PM2.5           | Before | 5228 | 21.05  | 15.81  | 14.88    | 12,450 | 0.000    |
|                 | During | 7224 | 17.148 | 13.36  |          |        |          |
| SO <sub>2</sub> | Before | 7570 | 7.012  | 9.61   | -27.25   | 15,411 | 0.000    |
|                 | During | 7843 | 14.53  | 22.07  |          |        |          |
| NO <sub>2</sub> | Before | 5467 | 96.29  | 206.78 | -3.90    | 13,598 | 0.000    |
|                 | During | 8133 | 113.77 | 284.23 |          |        |          |
| NO <sub>X</sub> | Before | 5423 | 59.68  | 49.19  | 17.22    | 13,403 | 0.000    |
|                 | During | 7982 | 46.50  | 39.16  |          |        |          |
| O <sub>3</sub>  | Before | 5554 | 73.76  | 88.94  | 23.18    | 13,011 | 0.000    |
|                 | During | 7459 | 47.89  | 32.01  |          |        |          |

**Table 4** Comparison of the air pollution before and during COVID-19 in Istanbul

|                 |        | N   | Mean    | SD      | <i>t</i> | df   | <i>p</i> |
|-----------------|--------|-----|---------|---------|----------|------|----------|
| PM10            | Before | 806 | 50.492  | 28.663  | 3.188    | 1643 | 0.001    |
|                 | During | 839 | 45.983  | 28.685  |          |      |          |
| SO <sub>2</sub> | Before | 434 | 3.884   | 2.801   | 0.233    | 1163 | 0.816    |
|                 | During | 731 | 3.845   | 2.762   |          |      |          |
| NO <sub>2</sub> | Before | 749 | 243.473 | 347.533 | 11.022   | 1573 | 0.000    |
|                 | During | 826 | 79.384  | 237.693 |          |      |          |
| NO <sub>X</sub> | Before | 808 | 35.161  | 44.874  | 17.542   | 1645 | 0.000    |
|                 | During | 839 | 6.051   | 16.903  |          |      |          |
| O <sub>3</sub>  | Before | 808 | 104.121 | 154.679 | 16.277   | 1638 | 0.000    |
|                 | During | 832 | 13.621  | 41.747  |          |      |          |

of air pollution, including PM<sub>2.5</sub>, PM<sub>10</sub>, NO<sub>x</sub>, NO<sub>2</sub>, SO<sub>2</sub> and O<sub>3</sub> were decreased during the COVID-19 outbreaks. The findings are consistent because similar declines have been seen in different parts of the world. For instance, the level of PM (particulate matter) NO<sub>2</sub> decreased by 31% and 60%, respectively, in 34 countries during the lockdowns (Venter et al., 2020). According to the same study, the ground level of PM<sub>2.5</sub> and NO<sub>2</sub> level declined considerably while the concentration of O<sub>3</sub> increased in 34 countries. At the peak of the COVID-19 pandemic outbreak in China, the level of NO<sub>2</sub> fractional reduction was 93% in Wuhan (Le et al., 2020). There was a huge decrease in air pollution in terms of CO<sub>2</sub> emissions, which declined by a quarter following the lockdown in China (Dutheil et al., 2020).

The key findings of the study demonstrated that the obtained *t*-test results successfully measured the pollution before and during the COVID-19 lockdown. For the Marmara region, we found that the PM<sub>10</sub> pollution mean value (M=52.64, SD=32.33) before the pandemic and (M=47.07, SD=28.97) during the pandemic, which was significantly lower. Similar values were obtained for Istanbul city. For Istanbul, we found that the mean level of PM<sub>10</sub> pollution before the lockdown was significantly higher than during the pandemic (M=50.49, SD=28.66) and (M=45.98, SD=28.68), respectively. Based on the findings, we can conclude that air pollution has declined sharply in Istanbul as well as the Marmara region because of COVID-19 lockdowns in Türkiye (Efe, 2022).

The experimental results of the study demonstrated that the various error rates measured, including RMSE, MAE, MAPE, and  $R^2$  (the difference between observed and predicted values), were in an acceptable range. We found the MAPE error value to be between 20–50% for all stations. This finding was supported by Lewis (1982) because he emphasized that MAPE values between 20–50% are acceptable for a given model. Based on the different stations, the estimation of PM-10 with ANN, RBFreg, and SMOreg models measured between 0.50 and 0.958 in terms of the  $R^2$  which showed that most of the observed variables could explain the models' inputs successfully.

This methodology provides a straightforward but reliable framework for exploratory analysis and intervention detection in air quality studies. In particular, it aids in the interpretation of the air quality

improvement linked to the COVID-19 pandemic's impact on mobility and the pressing need to implement urban mobility plans like electrification of the vehicle fleet to reduce NO<sub>x</sub> concentrations. In addition, the resulting trained models can provide helpful insights for governmental decision-making in health.

## Limitations

Although the presented methodology could effectively measure air pollution before and during the COVID-19 pandemic lockdown, there were a few limitations to our study that need to be addressed. The first limitation was missing data in the training and the test data. The missing data were removed from the datasets before performing the machine learning algorithms. This can affect the distribution of the gas concentration at given time intervals. The second limitation was that our training and test data were limited to the Marmara region in Türkiye. Only performing experiments for one region in Türkiye was not enough to conclude the same results for the other six regions in Türkiye. Thus, we are planning to extend our experiments to test other regions in Türkiye as well. If it is possible, we will perform a test case for various regions in different parts of the world as well. In an ANN classifier, the obtained error values on the sample mean training were completed, but sometimes the obtained results were not optimal. The SMOreg classifier was susceptible to over-fitting depending on the kernel parameter used, and the error value could not control the noise all the time. For the next study, it is aiming to increase the performance of the ANN and SMOreg by changing the parameters, and using more machine learning algorithms to fit the data more efficiently. For training and testing, the k-fold=10 cross-validation method has been used. For the next study, we would like to use a holdout method (percentage split) as well.

## Conclusion

In this study, we investigated the amount of air pollution before and during the COVID-19 lockdown. As far as we know, this study is one of the few research studies that measure air pollution before and during

the COVID-19 outbreak. The tests were performed on real data that was provided by the General Directorate of Meteorology in Türkiye. Since the city of Istanbul and the Marmara region are the most crowded and industrialized regions, the study was focused on Istanbul city and the Marmara region.

Three machine learning algorithms, including ANN, RBFreg, and SMOREg were performed on air pollution datasets created before and during the COVID-19 pandemic. The air pollution was measured based on six different gasses' concentration levels in the air, including  $PM_{10}$ ,  $PM_{2.5}$ ,  $SO_2$ ,  $NO_2$ ,  $NOX$  and  $O_3$ . The performance of the classifiers was compared based on the RMSE, MAE, MAPE, and  $R^2$  error criteria. The RBFRegressor algorithm outperformed the SMOREg and ANN algorithms for obtaining RMSE, MAE, MAPE, and  $R^2$  errors. Most of the observed variables could explain the models' inputs successfully. In this regard, on different stations, the estimation of PM-10 with ANN, RBFreg, and SMOREg models measured between 0.50 and 0.958 in terms of the  $R^2$ .

For the Marmara region, it was found that the  $PM_{10}$  pollution mean and standard deviation were measured lower during the pandemic when compared before the pandemic, with values of ( $M=47.07$ ,  $SD=28.97$ ) versus ( $M=52.64$ ,  $SD=32.33$ ). Similar performance results were obtained for Istanbul as well. In addition, a  $t$ -test was performed to compare the air pollution before and during the COVID-19 periods for given datasets. The results demonstrated that the PM value (in some way indicates the degree of air pollution) was reduced significantly in Istanbul city ( $t=3.188$ ,  $p<0.001$ ) as well as the Marmara region ( $t=11.45$ ,  $p<0.001$ ) during the pandemic lockdown. The concentration of other measured gases such as  $SO_2$ ,  $NO_2$ ,  $NOX$  and  $O_3$  decreased significantly as well.

Based on the experiment results, we concluded that the air pollution declined considerably in the Marmara region of Türkiye especially in Istanbul, during the COVID-19 pandemic lockdowns. We think that the main reasons for this decline were the reduced burning of fossil fuels, industrial emissions, vehicle transportation, and agricultural activities that were performed during the COVID-19 outbreak. For future study, we aim to extend our research to six more regions, including Ege, Akdeniz, Karadeniz, İç Anadolu, Doğu Anadolu, and Güneydoğu Anadolu in

Türkiye. It is also planned to test more machine learning algorithms on given datasets.

**Author contribution** The authors confirm their contribution to the paper as follows: study conception and design: Arpacı I., and Kılıçarslan S.; data collection: Arpacı I., and Kılıçarslan S.; analysis and interpretation of results: Kılıçarslan S., and Aslan O; draft manuscript preparation: Aslan O., and Ozturk I. All authors reviewed the results and approved the final version of the manuscript.

**Funding** The authors did not receive support from any organization for the submitted work.

**Data availability** The data were generated by the air quality stations belonging to the Ministry of Environment and Urbanization.

#### Declarations

**Conflict of interest** We, the authors, have no conflicts of interest and disclose any financial and personal relationships with other people or organizations that could inappropriately influence the presented work.

#### References

- Adam, M. G., Tran, P. T., & Balasubramanian, R. (2021). Air quality changes in cities during the covid-19 lockdown: a critical review. *Atmospheric Research*, 264, 105823. <https://doi.org/10.1016/j.atmosres.2021.105823>
- Adem, K., & Kılıçarslan, S. (2021). COVID-19 diagnosis prediction in emergency care patients using convolutional neural network. *Afyon Kocatepe Üniversitesi Fen Ve Mühendislik Bilimleri Dergisi*, 21(2), 300–309.
- Al-Sharafi, M. A., Al-Emran, M., Arpacı, I., Marques, G., Namoun, A., & Iahad, N. A. (2022). Examining the impact of psychological, social, and quality factors on the continuous intention to use virtual meeting platforms during and beyond COVID-19 pandemic: A hybrid SEM-ANN approach. *International Journal of Human-Computer Interaction*. <https://doi.org/10.1080/10447318.2022.2084036>
- Arpacı, I., Alshehabi, S., Mahariq, I., & Topcu, A. E. (2021). An evolutionary clustering analysis of social media content and global infection rates during the COVID-19 pandemic. *Journal of Information & Knowledge Management*, 20(3), 2150038. <https://doi.org/10.1142/S0219649221500386>
- Arpacı, I., Alshehabi, S., Al-Emran, M., Khasawneh, M., Mahariq, I., Abdeljawad T., & Hassanien, A. E. (2020). Analysis of Twitter data using evolutionary clustering during the COVID-19 pandemic. *Computers, Materials & Continua*, 65(1), 193–204. <https://doi.org/10.32604/cmc.2020.011489>
- BBC (2021). COVID: Türkiye enters first full lockdown. BBC News. 29 April 2021. Retrieved from <https://www.bbc.com/news/world-europe-56912668>. Accessed 14 Mar 2023.

- Dutheil, F., Baker, J. S., & Navel, V. (2020). COVID-19 as a factor influencing air pollution? *Environmental Pollution*, 263, 114466.
- Efe, B. (2022). Air quality improvement and its relation to mobility during COVID-19 lockdown in Marmara Region. *Turkey. Environmental Monitoring and Assessment*, 194(4), 1–19.
- Fattorini, D., & Regoli, F. (2020). Role of the chronic air pollution levels in the COVID-19 outbreak risk in Italy. *Environmental pollution*, 264, 114732. <https://doi.org/10.1016/j.envpol.2020.114732>
- Fisher, R. A. (1922). The goodness of fit of regression formulae, and the distribution of regression coefficients. *Journal of the Royal Statistical Society*, 85(4), 597–612.
- Jephcote, C., Hansell, A. L., Adams, K., & Gulliver, J. (2021). Changes in air quality during COVID-19 ‘lockdown’ in the United Kingdom. *Environmental Pollution*, 272, 116011. <https://doi.org/10.1016/j.envpol.2020.116011>
- Kaastra, I., & Boyd, M. (1996). Designing a neural network for forecasting financial and economic time series. *Neurocomputing*, 10(3), 215–236. [https://doi.org/10.1016/0925-2312\(95\)00039-9](https://doi.org/10.1016/0925-2312(95)00039-9)
- Kiliçarslan, S., Közkurt, C., Baş, S., & Elen, A. (2023). Detection and classification of pneumonia using novel Superior Exponential (SupEx) activation function in convolutional neural networks. *Expert Systems with Applications*, 217, 119503.
- Le, T., Wang, Y., Liu, L., Yang, J., Yung, Y. L., Li, G., & Seinfeld, J. H. (2020). Unexpected air pollution with marked emission reductions during the COVID-19 outbreak in China. *Science*, 369(6504), 702–706. <https://doi.org/10.1126/science.abb7431>
- Lewis, C. D. (1982). Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting. Butterworth-Heinemann.
- Ministry of Environment and Urbanization. Accessed on 2022–04–10 from <https://sim.csb.gov.tr>.
- Moustris, K. P., Larissi, I. K., Nastos, P. T., Koukouletsos, K. V., & Paliatsos, A. G. (2013). Development and application of artificial neural network modeling in forecasting PM 10 levels in a Mediterranean city. *Water, Air, & Soil Pollution*, 224, 1–11.
- Pereira, F., & Aires-de-Sousa, J. (2018). Machine learning for the prediction of molecular dipole moments obtained by density functional theory. *Journal of Cheminformatics*, 10(43), 1–11. <https://doi.org/10.1186/s13321-018-0296-5>
- Sayegh, A. S., Munir, S., & Habeebullah, T. M. (2014). Comparing the performance of statistical models for predicting PM10 concentrations. *Aerosol and Air Quality Research*, 14(3), 653–665.
- Schalkoff, R. J. (1997). Artificial neural networks. McGraw-Hill Higher Education. ISBN:978-0-07-057118-1.
- Shevade, S. K., Keerthi, S. S., Bhattacharyya, C., & Murthy, K. K. (2000). Improvements to the SMO algorithm for SVM regression. *IEEE Transactions on Neural Networks*, 11(5), 1188–1193
- Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3), 199–222.
- TUIK. (2021). Address based population registration system results. Retrieved from <https://data.tuik.gov.tr/Bulten/Index?p=Adrese-Dayali-Nufus-Kayit-Sistemi-Sonuc-lari-2021-45500>. Accessed 14 Mar 2023.
- Venter, Z. S., Aunan, K., Chowdhury, S., & Lelieveld, J. (2020). COVID-19 lockdowns cause global air pollution declines. *Proceedings of the National Academy of Sciences*, 117(32), 18984–18990. <https://doi.org/10.1073/pnas.2006853117>
- Wang, P., Chen, K., Zhu, S., Wang, P., & Zhang, H. (2020). Severe air pollution events not avoided by reduced anthropogenic activities during COVID-19 outbreak. *Resources, Conservation and Recycling*, 158, 104814. <https://doi.org/10.1016/j.resconrec.2020.104814>
- Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1), 79–82.
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical Machine Learning Tools and Techniques* (2nd ed.). Morgan Kaufmann Publishers.
- Wu, X., Nethery, R. C., Sabath, M. B., Braun, D., & Dominici, F. (2020). Air pollution and COVID-19 mortality in the United States: Strengths and limitations of an ecological regression analysis. *Science Advances*, 6(45), eabd4049.
- Yadav, S., Praveen, O. D., & Satsangi, P. G. (2015). The effect of climate and meteorological changes on particulate matter in Pune, India. *Environmental Monitoring and Assessment*, 187, 1–14.
- Yöntem, M. K., Adem, K., İlhan, T., & Kılıçarslan, S. (2019). Divorce prediction using correlation based feature selection and artificial neural networks. *Neşehir Hacı Bektaş Veli University Journal of ISS*, 9(1), 259–273.
- Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with Artificial Neural Networks: The State of the Art”. *Inter. Journal of Forecasting*, 14, 35–62.
- Zhu, Y., Xie, J., Huang, F., & Cao, L. (2020). Association between short-term exposure to air pollution and COVID-19 infection: Evidence from China. *Science of the Total Environment*, 727, 138704. <https://doi.org/10.1016/j.scitotenv.2020.138704>

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.