Environ Res Tec, Vol. 5, Issue. 1, pp. 56–71, March 2022

Environmental Research and Technology
https://ert.yildiz.edu.tr - https://dergipark.org.tr/tr/pub/ert
DOI: https://doi.org/10.35208/ert.1000739

Research Article

Developing a GMDH-type neural network model for spatial prediction of NO_x: A case study of Çerkezköy, Tekirdağ

Can Burak ÖZKAL*[1], Özkan ARSLAN[2]

[1]Department of Environmental Engineering, Tekirdağ Namık Kemal University, Tekirdağ, Türkiye
[2]Department of Electronics and Communication Engineering, Tekirdağ Namık Kemal University, Tekirdağ, Türkiye

ARTICLE INFO

Article history
Received: 25 September 2021
Revised: 06 January 2022
Accepted: 24 January 2022

Key words:
Air pollutant; Artificial intelligence; GMDH; NO_x; NO_x prediction; Self-organizing deep learning

ABSTRACT

Air pollution-induced issues involve public health, environmental, agricultural and socio-economic aspects. Therefore, decision-makers need low-cost, efficient tools with high spatiotemporal representation for monitoring air pollutants around urban areas and sensitive regions. Air pollution forecasting models with different time steps and forecast lengths are used as an alternative and support to traditional air quality monitoring stations (AQMS). In recent decades, given their eligibility to reconcile the relationship between parameters of complex systems, artificial neural networks have acquired the utmost importance in the field of air pollution forecasting. In this study, different machine learning regression methods are used to establish a mathematical relationship between air pollutants and meteorological factors from four AQMS (A-D) located between Çerkezköy and Süleymanpaşa, Tekirdağ. The model input variables included air pollutants and meteorological parameters. All developed models were used with the intent to provide instantaneous prediction of the air pollutant parameter NOx within the AQMS and across different stations. In the GMDH (group method of data handling)-type neural network method (namely the self-organizing deep learning approach), a five hidden layer structure consisting of a maximum of five neurons was preferred and, choice of layers and neurons were made in a way to minimize the error. In all models developed, the data were divided into a training (80%) and a testing set (20%). Based on R^2, RMSE, and MAE values of all developed models, GMDH provided superior results regarding the NOx prediction within AQMS (reaching 0.94, 10.95, and 6.65, respectively for station A) and between different AQMS. The GMDH model yielded NOx prediction of station B by using station A input variables (without using NOx data as model input) with R^2, RMSE and MAE values 0.80, 10.88, 7.31 respectively. The GMDH model is found suitable for being employed to fill in the gaps of air pollution records within and across-AQMS.

Cite this article as: Özkal CB, Arslan Ö. Developing a GMDH-type neural network model for spatial prediction of NOx: A case study of Çerkezköy, Tekirdağ. Environ Res Tec 2022;5:1:56–71.

***Corresponding author.**
*E-mail address: cbozkal@nku.edu.tr

Published by Yıldız Technical University Press, İstanbul, Turkey

INTRODUCTION

Air pollutants can be categorized as primary (SO_2, NO_x, CO, hydrocarbons, particle matter) and secondary (ozone) pollutants. NOx can be represented by the most commonly found NO_2 form generated as a result of reaction between combusted hydrocarbons due to industrial and traffic emissions and atmospheric oxygen. NO_x are responsible for the formation of tropospheric ozone (bad ozone). Nitrogen oxides likewise SO_2, ammonia and volatile organic carbons are responsible for the formation of $PM_{2.5}$. and have been subject a to numerous research on the basis of monitoring, prediction [1, 2] and mitigation methodologies development [3]. The spread and transport of air pollutants that are released into the atmosphere is influenced by weather and climatic factors [4, 5]. Interaction of meteorological factors, mainly wind speed and wind direction with the air pollutants bring along variations in spatial distribution, spatiotemporal variation of air pollutants. Numerous research have demonstrated regional transport of air pollutants interpretation with the main meteorological factors; Humidity, temperature, pressure, wind direction and wind speed, dew point [6].

Recently, a city air quality determination study has gained importance. Thus, the periodic and spatial dimensions of pollution necessitate accurate determination. In order to overcome observed deficiencies, higher maintenance and repair expenses in air quality monitoring stations (AQMS) and provide supplementary data, solutions including increasing prevalence of AQMS, use of mobile AQMS and low cost sensors (LCS) are widely applied [7, 8]. Up to date, research have emphasized the modeling approaches for temporal and spatial prediction [9] and forecasting of air pollution [10].

Monitoring of air quality and air pollution parameters is performed by static, mobile AQMS and establishment of low cost sensors (LCS). Regulatory pollutants (carbon monoxide, nitrogen oxides, ozone and particulate matter are measured by certified reference instruments at static AQMS. Those stations and sensors are large and expensive, also necessitate strict calibration and maintenance routines in order to provide high quality data and comparability between different region and stations [7, 8, 11]. Sensors and low cost sensors are able to monitor a range of air pollutants but mostly they are unable to meet the Air Quality Directive - Data quality objectives criteria and under effect of chemical interferences and environmental conditions [7, 12]. According to framework and legal requirements described in air quality directive 2008/50/EC for ambient air quality assessment and management, the reference measurement methods are applied in the stable AQMS in Europe. The data provided by LCS are usually less accurate than AQMS [11]. The Air Quality Directive also pave the way for alternative and supplementary techniques such as air quality models for air quality and air pollution management.

Recent research accordingly have hypothesized to estimate air pollutants concentrations through their association/ interpretation with meteorological parameters [13], landscape data, environmental information [14] and other measured air pollutants [15, 16]. Underlie the spatiotemporal correlations between air pollutants emission and diffusion mechanisms, mechanistic or a non-linear model must be considering those correlation and/or able to realize non-linear mechanism of air pollutants spread, diffusion, transport and interaction under environmental, meteorological and atmospheric conditions [16].

Air quality management includes monitoring and timely application of foreseen preventions related to extreme air quality scenarios. Therefore, short and mid-term forecasting of air pollution and air quality indexes became the focus of numerous research [2]. There are many gaps encountered in air pollution data series of stable AQMS, during their observation period [17]. The research method for predictive determination of air pollution parameters and air quality mainly based on statistical or deterministic approaches [16]. The emission and diffusion of pollutant is related both with interaction of pollutants and the meteorological factors. As a result it is prerequisite to use sufficient number of meteorological and air quality parameters in an air pollution prediction model. Similar approach was applied via Pearson correlation, support vector regression (SVR) with or without principal component analysis, for the purpose of decision making on keeping the most correlated pollutant parameters in the data set [9, 16]. According to recent literature findings, modelling approaches including artificial neural network (ANN) methods were found eligible to be employed to fill in the gaps of air pollution records by deep learning based prediction of $PM_{2.5}$, LSTM based estimation of air pollutant concentrations that cannot be directly measured by the air quality monitor and ANN based forecasting of the spatial–temporal profile of pollutant concentrations and air quality determinants in specific cases of power outages, negative and wrong records of pollutants [11, 13, 16].

Most generally, machine learning (ML) methods and mostly ANN application are proposed to forecast the spatial-temporal pollutant concentration profiles during extreme scenarios be it; power outage, maintenance, sensor repairmen and replacement, negative and faulty pollutant records. Also recent research have reported successful application of various modelling techniques developed with intent to eliminate high number of monitoring stations requirement and become an alternative for advanced representation of air pollution spatial variability; where Feed forward neural networks and Long short-term memory deep learning techniques were applied with intent to provide data at currently unmonitored locations [11, 18].

Traditional linear models or deterministic models descriptive of chemical dispersion and transportation remain limited as a result of the high degree of non-linearity between different air pollutants and weather conditions [2]. As it was reported in recent literature, the nonlinear mechanism of atmospheric phenomenon can be realized by ANN [17, 19] and excellent prediction performance can be achieved [16]. For the very reason air pollutant parameters prediction using ANN is found superior to multi-linear regression [17, 20].

Conventional recurrent neural networks (RNN) and long short-term memory (LSTM) is suitable option to be applied on time series thereby applied in various research topics from various disciplines [11]. These methods mostly conform with the cases where the values of concern are related to their previous situation like traffic flow prediction, air pollution prediction, solar irradiation scenarios [11]. For the case of interpolation and extrapolation based estimation, model inputs are selected from a group of monitoring stations. Model training is performed by historical data of limited number of station with LSTM method, therefore the model was proposed to be suitable in small cities where only a few monitoring stations are established [11]. Besides the reported high performance of LSTM application in research related to air pollution, concentration prediction is the common field of research area [16, 21].

Recently, soft computation artificial intelligence (AI) techniques have been used successfully in the prediction of air pollution parameters such as NOx. Because of the many parameters that affect NOx, the results predicted from empirical models do not match well with the measured results. Therefore, it is necessary to develop models that provide more accurate prediction of NOx under various air pollution and meteorological conditions. Group Method of Data Handling (GMDH), an AI-based method, is a self-organizing technique that can be used to solve complex problems in nonlinear system with large degrees of complexity. The GMDH technique, which is a multi-layered structure, uses only neurons that can provide the most effective and accurate results unlike traditional machine learning methods. Thus, it is ensured that the most efficient input variables are used instead of using all input variables for the predictive model output. In this regard, the GMDH approach requires less data training compared to classical ANN methods and facilitates the interpretation of model input and output parameters. In air pollution prediction applications, it is very important to determine the pollutant (emission) sources and to determine the prevailing wind direction [19]. Demonstrated through preliminary research on AQMS of Tekirdağ and literature findings, deficiencies are observed in the AQMS data, with measurements either not being performed or not being shared on numerous occasions during the course of the year [13, 18, 22]. In order to overcome these shortcomings, GMDH-type neural networks using a non-linear structure can be preferred for the estimation of the NOx parameter.

In this study, we hypothesize the employment of GMDH-type neural network as an alternative technique in prediction (forecast length equal to 0) of NO_x air pollutant concentrations within the AQMS and spatial prediction of air pollutant concentration between different stations. The developed model provided prediction of NOx within a certain station by using data from a total number of four different weather stations and taking one station as reference. It has been observed that the GMDH-type neural network model minimizes the error rates under certain NOx prediction states within a station and between different stations.

MATERIALS AND METHODS

Location of Interest

Tekirdağ is located at European part of Turkey, in the Thrace region surrounded by Marmara Sea, Greece and Bulgaria. The city is located within the Ergene basin and is the center of population growth (predominated by the industrial development in the eastern part of the city, Çerkezköy and Çorlu districts) and air pollution as a result of heating, traffic and industrial activity. Tekirdağ constitutes the western border of İstanbul and the northern border of the Marmara Sea. Süleymanpaşa is the biggest district of Tekirdağ. There many organized industrial zones (OIZ) located at Çorlu, Çerkezköy, Kapaklı, Velimeşe, Ergene districts, as indicated in Figure 1. Tekirdağ host over 1100 factories (with the frequency of occurrence; textile, paper, packaging, chemical and metal industries respectively). In Tekirdağ, there a total of 14 organized industrial zones, while 5 of them established around Çorlu district and its immediate surroundings. 4 of the organized industrial zones (OIZ) are lined up in the west-east direction along Çorlu, the Velimeşe OIZ is located between Çerkezköy and Çorlu (North south direction) and more than 500 facilities operate. Çorlu is the area where the new settlement is located and the traffic is concentrated while Çerkezkoy host an OIZ under which more than 270 facilities operate and it one of Turkey's largest OIZs [23, 24]. The topographical properties of the location of interest (as indicated in Figure 1) can be described as land appearance in the form of wavy plains and is uneven, with low to mid slope values. Çorlu is under the influence of a transition type climate where Black Sea, Mediterranean and continental climate characteristics are encountered together. Cold air masses descending from the north and humid-warm air currents coming from the south, the Mediterranean and the Aegean affect the climate structure of the region. Typically, the wind blows at Tekirdağ dominantly from the directions of NNE-NE and rarely from the directions of SW-SSW [24].

Air Quality Monitoring Stations

In this study, urban and industrial AQMS located on the Çerkezköy-Çorlu-Tekirdağ line were selected as

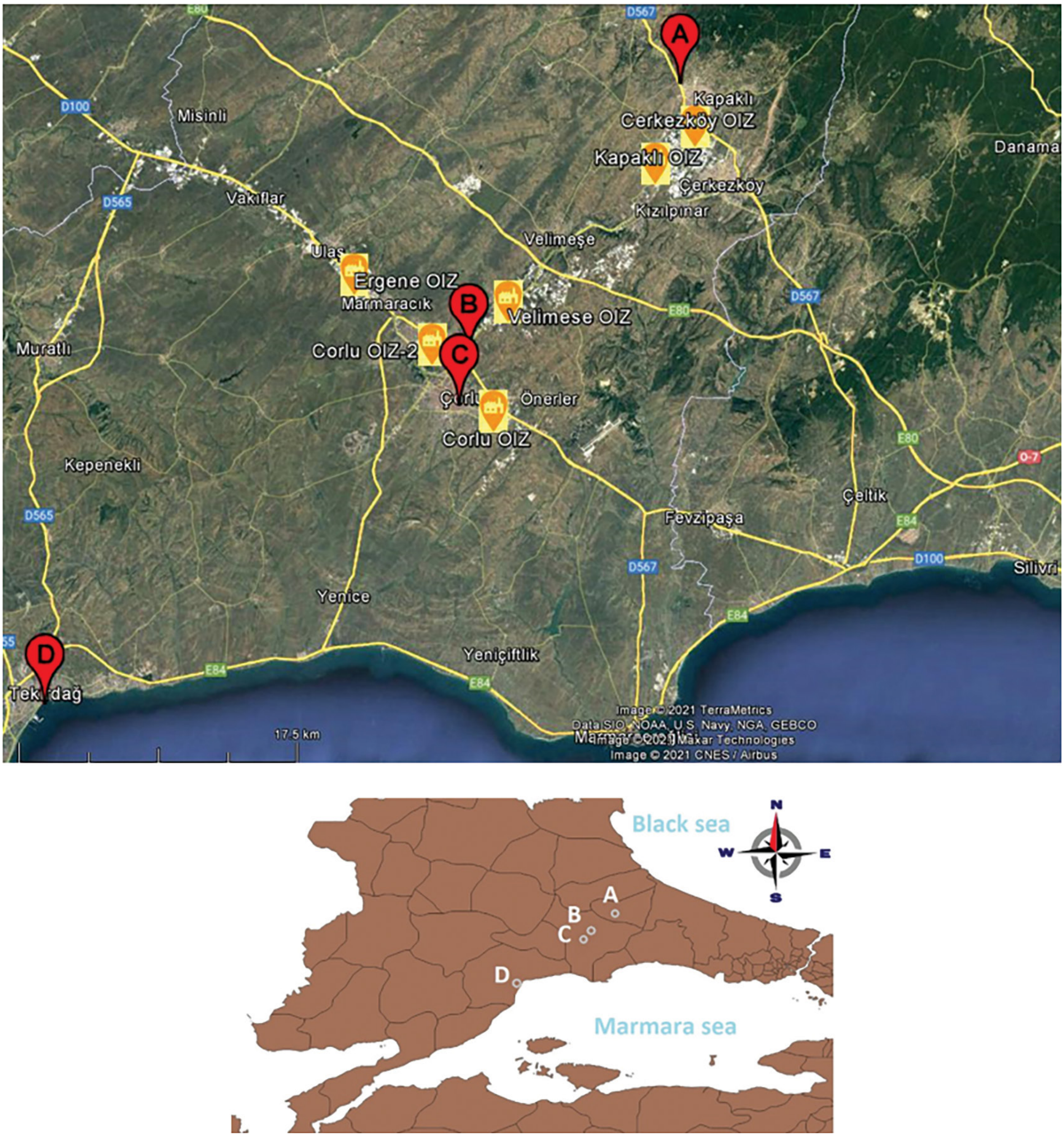

Figure 1. AQMS locations for stations; A: Çerkezköy, B: Çorlu, C: Çorlu Center, D: Tekirdağ center.

data source. Air quality of Çorlu is monitored through 2 AQMS; an urban AQMS located in town center (as a result of its distance from industrial activity) and an industrial AQMS located between town center and Çerkezköy (Fig. 1). For the case of Tekirdağ (Suleymanpasa center city) air quality is monitored at two different stations. The exact location, coordinates and bird flight (BF) distance between selected stations in Tekirdağ-Suleymanpasa, Çorlu and Çerkezköy are described in Figure 1 and AQMS distances between stations are given in Table 1.

Between studied dates, northern winds were prevalent around station A. For station B and C, the direction of wind could be described as spreaded over a wider range (as a result of E-SE winds prevalence), as it is demonstrated in Figure 2. Station B has distinctly strong winds where the average is above the meteorological upper limits reported for mild winds. Those strong winds are reported to trigger air pollutants transport and dispersal mechanisms [16]. Air pollutants at regions where mountain-valley and land-sea breezes cycles are dominant

Table 1. AQMS distances for stations

BF Distances between AQMS of interest	
AQMS # (stations separated by-)	**Distance in km**
Ed A-B	≈21
A-C	≈23
A-D	≈54
B-C	≈3
B-D	≈34
C-D	≈31

Letter Code/Name of AQMS. A: Çerkezköy; B: Çorlu; C: Çorlu Center; D: Tekirdağ center.

wind systems, are not easily transported from emission sources and accumulate. Also, WS over 5 m/s may create unstable weather conditions which increase PM and SO_2, NO_x distribution [4, 17].

Data Acquisition

The air pollutant and meteorological data source is the official Air Quality Monitoring Network website of the Ministry of Environment and Urbanization. Daily average values were used and taken as references all along the study. PM_{10}, SO_2, NO_x data as air pollutant parameters and air temperature (T), wind direction (WD), wind speed (WS), relative humidity (RH) and Air pressure (AP) were used as meteorological model variables for the period between December 2017 – December 2018.

It can be figured out from the correlation coefficients (depicted in Figure 4) of 0.79, 0.26, 0.43 and 0.91 for stations A, B, C, and D between targets output NOx and input variable PM, the PM parameter will be effective in NOx prediction for all stations. The PM_{10} and SO_2 are amongst the sole parameters continuously measured at each AQMS in common. Therefore, the choice of using PM10 and SO_2 as model inputs for NO_X prediction could be reasoned based on theory; the relation between air pollutant parameters as a result of their complex and interacting formation mechanisms. Nitrogen oxides likewise SO_2, ammonia and volatile organic carbons are responsible for the formation of particulate matter.

In the Nox prediction model, 174-day air pollution and meteorological data were used, which were recorded for 1 year. The statistical distributions of these data are shown in Figure 2 and can be summarized as; PM_{10}, SO_2 and NO_x data of station D shows a wider distribution compared to station A-

C. The distribution of annual temperature data did not differentiate between stations. Higher average wind-speed values were measured at Station C. A substantial difference of wind direction distribution was not observed at the stations. Average RH were reported to be the highest and lowest for station A and B respectively. Average AP values were distinctively higher at station D.

Regression Methods

Various machine learning regression methods are used to establish a mathematical relationship between the inputs which are air pollutants (PM_{10}, NO_X and SO_2) and meteorological parameters (T, WD, WS, RH and AP) and the target output. The predicted output is calculated by training the data in all models for different air quality measurement stations. In this study, various regression methods that provide a relationship between target NO_X and inputs are given below.

Linear Regression

Linear regression is one of the simplest methods that provide a mathematical relationship between the input parameters and the target output. It is often preferred because of the simple and convenient mathematical structure. In this regression method, the mathematical equation of the target based on the inputs is obtained with a slope and intercept value. The relationship between the target NO_X and the input variables is expressed by linear regression as:

$$Y = a_0 + a_1x_1 + a_2x_2 + \cdots + a_nx_n \tag{1}$$

where Y is output, $x_1, x_2, \ldots, x_n$ are input variables and $a_1, a_2, \ldots, a_n$ are coefficients obtained from the model.

Random Forest Regression

Random forest (RF) regression is one of the machine learning models that can be effective in predictive analysis under conditions where the output and input parameters are in a non-linear relationship. In this method, which reduces over fitting in model training, the predictions of all decision trees are combined to obtain more accurate and stable results. The forest tree diversity increases the robustness of the model obtained by regression [25].

Multilayer Perceptron Regression

Multilayer perceptron (MLP) is an artificial neural network method frequently used in regression. MLP method, which is also considered as the early stage of deep learning, consists of an input, multiple hidden and an output layer. In MLP neural networks, the first layer contains the input parameters and the output layer makes a prediction about the input. Hidden layers are used as a computational tool between the input and output layers [26].

MLP regression method is often used in supervised learning applications that involve the training phase for certain target and input parameters are used in all hidden layer neurons. In this method, which is based on the training of inputs and output, a model based on correlation between input and output is learned. In the training phase, the parameters and weight coefficients of the model that minimize the error are obtained.

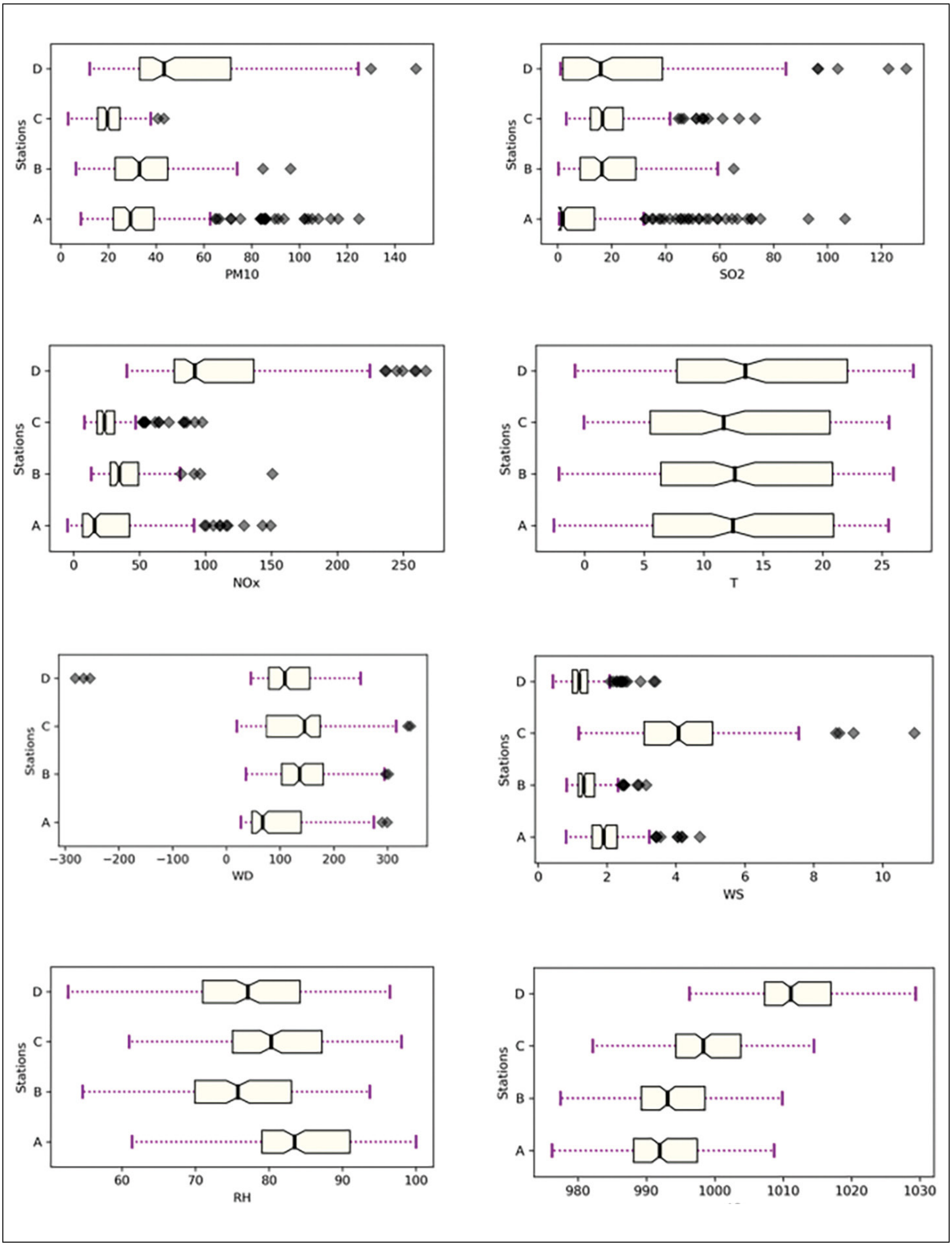

Figure 2. Boxplot data summary of PM_{10}, SO_2, NO_X, T, WD, WS, RH and AP parameters for stations A-WD: Wind direction (degrees), WS: Wind speed (m/s), RH: relative humidity (%), AP: Atmospheric pressure (hPa).

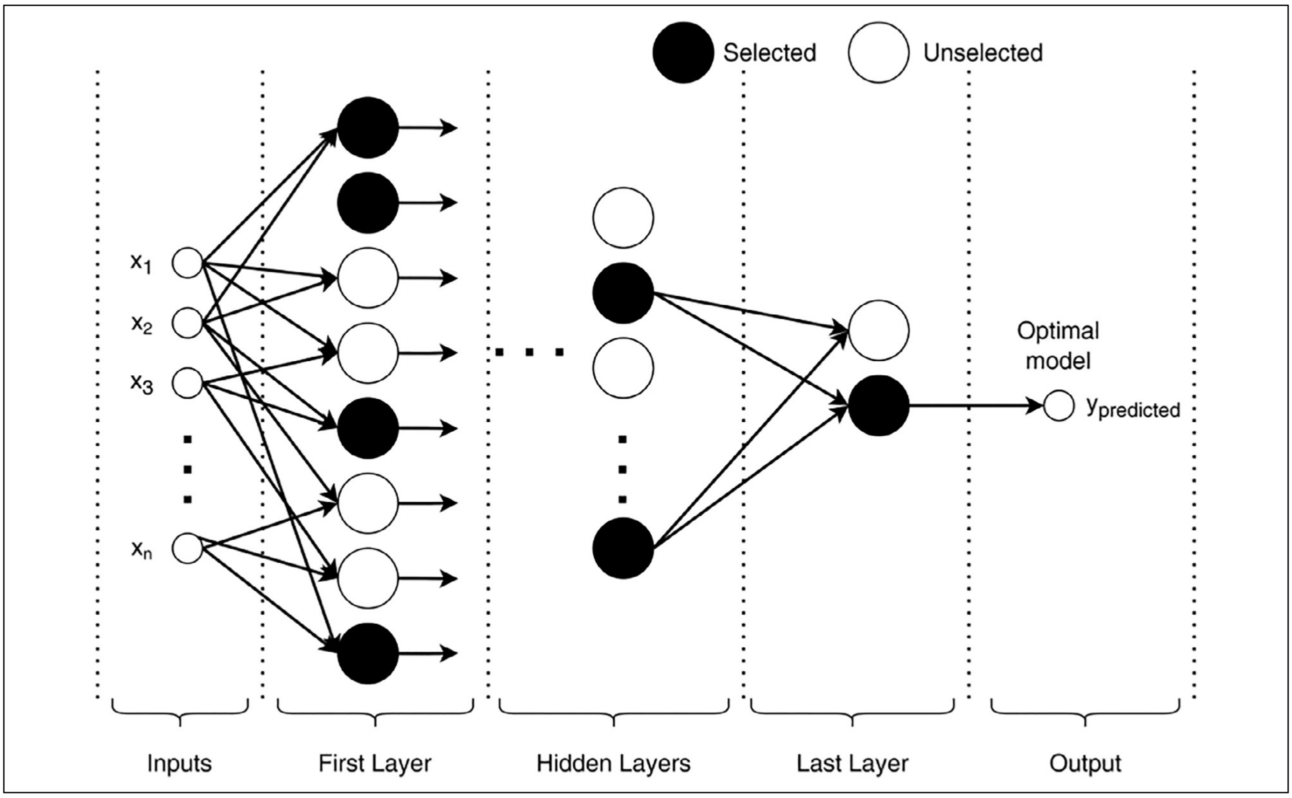

Figure 3. General structure of GMDH-type neural network.

GMDH-type Neural Network Regression

Group Method of Data Handling (GMDH), which is a self-organizing network model that behaves according to the input data, is preferred in regression analysis applications, unlike MLP artificial neural network models where the inputs are used on all neurons in the hidden layer [27]. The use of input parameters in all neurons may cause over fitting and performance degradation in regression models. In addition, there are difficulties and shortcomings in adjusting the bias and weight coefficients for a small number of datasets.

GMDH-type neural networks are one of the best methods for model estimation in complex structured problems. This neural network model is a multi-layered structure and uses only neurons that can provide the most effective and accurate results. Each layer consists of independent neurons and these neurons are used in pairs. In this network model, a quadratic polynomial function is used as the activation function. The neurons in the hidden layers work independently and neuron outputs which minimize the error rate are used. Thus, a multilayer neural network model consisting of optimal layers and neurons is designed instead of using all neurons in the layers [28]. Figure 3 shows the general structure of GMDH-type neural network.

GMDH-type neural networks are defined as a relationship between input and output parameters expressed in the form of a stepwise complex Kolmogorov-Gabor polynomial function. This relationship is expressed as a nonlinear form of the Kolmogorov-Gabor function [29].

$$\bar{y} = \alpha_0 + \sum_{i=1}^{n} \alpha_i x_i + \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_{ij} x_i x_j + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} \alpha_{ijk} x_i x_j x_k + \cdots \quad (2)$$

where $\bar{y}$ and α are predicted output and the coefficients of the quadratic polynomial, respectively. The number n is the degree of polynomial function and $i, j, k \in (1, 2, \ldots, n)$. In this study, the number n was chosen as 2. The polynomial operation is performed in three steps for i={0, 1 and 2}.

The Kolmogorov-Gabor polynomial, which has a nonlinear structure, is expressed in the form of a quadratic polynomial consisting of two variables as follows:

$$\bar{y} = G(x_i, x_j) = \alpha_0 + \alpha_1 x_i + \alpha_2 x_j + \alpha_3 x_i x_j + \alpha_4 x_i^2 + \alpha_5 x_j^2 \quad (3)$$

The GMDH-type neural network estimates the output for each set of input parameters (x_i and x_j) and is used to estimate the α_i coefficients that minimize the mean squared error between the predicted and the actual output. This process is called self-organization of models, and neurons with minimum error calculated by the least squares method are selected.

In the GMDG-type neural network model, the coefficient vector of the quadratic polynomial is calculated and the neurons that increase the error are eliminated. The objective function (OF), which is a selection criterion, is used for elimination process and OF is expressed as:

$$OF = \frac{1}{n}\sum_{i=1}^{n}\left(y_{pre} - y_{mea}\right)^2 \quad (4)$$

where y_{pre}, y_{mea} and n are the predicted, measured values and total number of dataset, respectively.

Evaluation of Models

The performance of a regression model is evaluated by calculating the error rate of the predicted output obtained by the model. In addition, the fit of the regression line to the data set is also used as a criterion in model evaluation. The correlation coefficient (R^2), root-mean square error (RMSE) and mean absolute error (MAE) are used to calculate between the predicted and actual values. The R^2 value is between 0 and 1, and a larger value indicates a better fit between the predicted and the actual values. The R^2 is a good measure to determine how well the model fits the dependent variables and is expressed as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_{mea} - y_{pre}\right)^2}{\sum_{i=1}^{n}\left(y_{mea} - y_{m}\right)^2} \quad (5)$$

The RMSE is calculated as the sum of the square of the error by subtracting the predicted from the actual value, then divided by the total number of data and the square root is taken. The RMSE, which is widely used in the evaluation of models, is expressed as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(y_{mea} - y_{pre}\right)^2}{n}} \quad (6)$$

The mean absolute error (MAE), a measure similar to the mean squared error (MSE), is defined as the sum of the absolute value of the error and is expressed mathematically as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_{mea} - y_{pre}\right| \quad (7)$$

where y_{mea}, y_{pre}, y_m and n represent the measured, predicted, the average of measured values and the total number of dataset, respectively.

Model Development Setup

Many empirical models have been developed to estimate the air pollution parameter NO_X values using air pollution and meteorological parameters. Recent research have focused on instant prediction of target pollutant parameter value [13]. Machine learning regression methods are extensively used to derive empirical equations. However, the application of ANN models in larger spatial dimensions would bring along a significant decline in the model's performance for places far away from the station (data of which was used for model training), as was hypothesized in recent studies [9, 11, 30]. This study aims to develop empirical models based on not only pollution conditions, but also depend on meteorological conditions for NOx prediction. In this study, different regression methods such as linear, RF, MLP, and GMDH-type NN were analyzed for the NOx prediction model. The influences of different parameters, including air pollutants (PM_{10}, SO_2) and meteorological parameters (T, WD, WS, RH, and AP), on the prediction of NOx within station and across different stations were investigated. In the experimental setup, the parameters of the regression methods were set as follows: the number of forest trees was taken as 500 for RF, five hidden layer structures consisting of 20 neurons was established for MLP and 0.01 learning rate, rectified linear unit activation function, Adam optimization were chosen. In the GMDH-type neural network, a five hidden layer structure consisting of a maximum of five neurons was preferred and, layers and neurons that minimized the error were used. In all developed models, the data were split into a training (80%) and a testing set (20%). R^2, RMSE, and MAE values were calculated to obtain the most effective and accurate empirical model that can be used in the prediction of NO_X.

RESULTS AND DISCUSSIONS

In this study, pollution and meteorological parameters were used as input variables in different regression methods for the NO_x prediction model and it was aimed to obtain a model that provides the best prediction. Figure 3 shows the correlation between the parameters used as input variables and the output NO_x, and the correlation values between these parameters. As can be seen from Figure 4, it has been observed that target output NO_x has a high correlation with PM and SO_2 parameters at all stations. It can be said that the obtained correlation results are compatible with the literature. A similar hypothesis was proposed in recent research in the Marmara region; were correlations between SO_2 and PM_{10} values reported for residential areas (with solid fuel use) were higher than it is determined for industrial areas [31].

There are correlation coefficients of 0.79, 0.26, 0.43 and 0.91 for stations A, B, C, and D between targets output NO_x and input variable PM, respectively. These results show that the PM parameter will be effective in prediction the output NO_x for stations A and D. Similarly, the correlation between output NOx and SO_2 is 0.83, 0.73, 0.11 and 0.89 for stations A, B, C, and D, respectively. It is concluded that the SO_2 parameter will have a negative effect on the prediction of the NO_X parameter for station C. In addition, it can be said that there is no high correlation between meteorological parameters and NO_x for all stations. However, the use of meteorological parameters in the prediction of target output NO_X is important to obtain a more robust model.

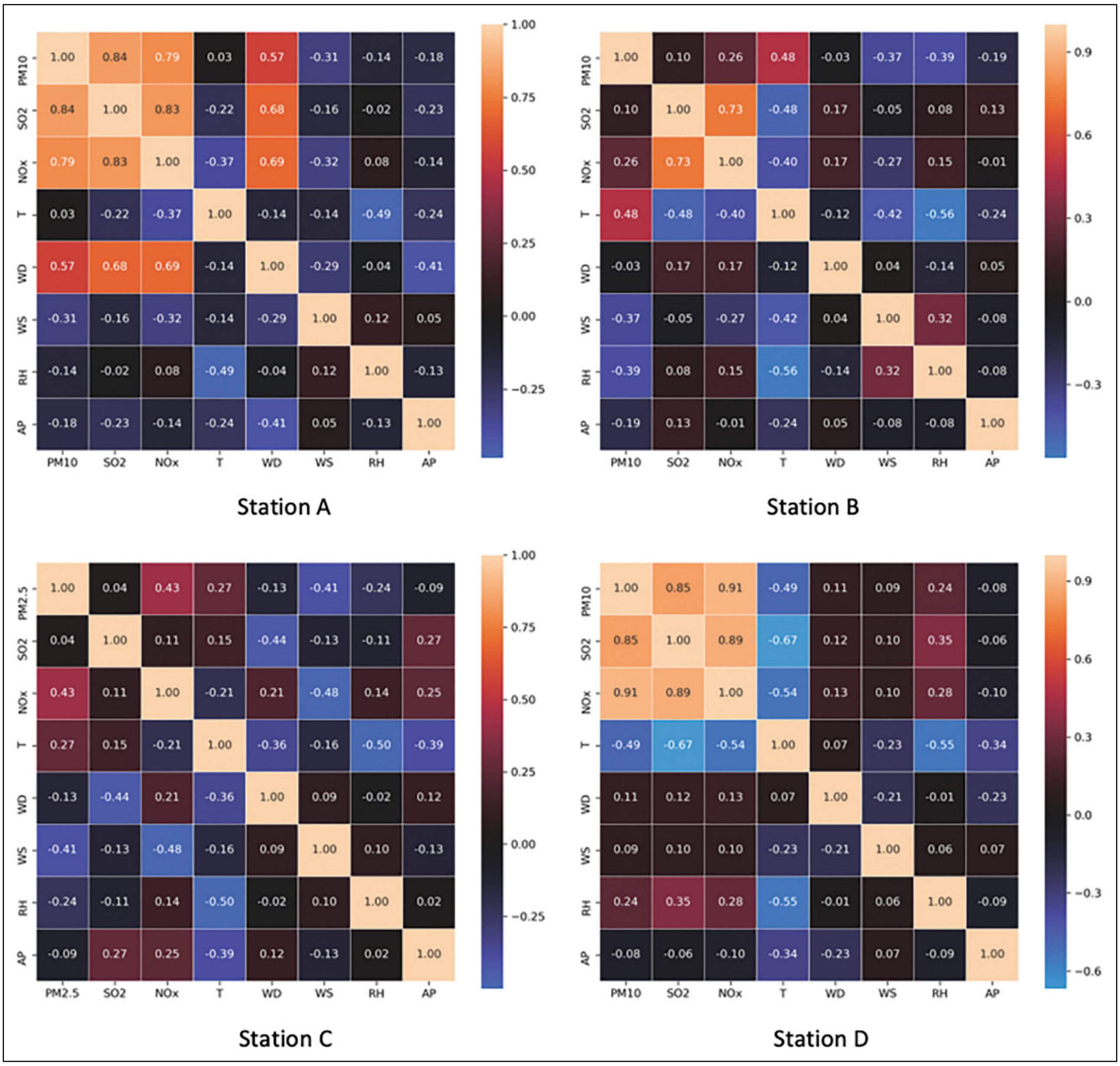

Figure 4. Pearson's correlation coefficients of air pollutant and meteorological parameters for stations A, B, C, D.

The Prediction Approach

Due to the non-linear relationship between NO_X and input parameters, linear regression methods were found ineligible for modelling. All the cells in the layers are used in MLP neural network-based modelling, resulting in an excessive training problem and a recession in model performance. RF modelling is one of the machine learning models that can be efficient in predictive analysis even in non-linear relations, have lower regression and higher error rates compared to the GMDH method. GMDH neural network prefer the most appropriate cells and pathways that minimize the error rate in the estimation.

As a result the GMDH method performed better for estimating NO_X as compared to other methods. In this method, the number of hidden layers and the number of cells in those layers are optimally obtained based on the input parameters. The NO_X prediction within a certain station by the GMDH-type neural network model resulted with regression coefficients (Table 2) ranging from 0.87 to 0.94. For the case of using meteorological and air pollution data of station A as model inputs; the regression coefficient of NO_x prediction by the GMDH-type models were found to be 0.85, 0.54, and 0.65 for station B, C, D (as shown in Table 4), respectively.

In another research, it was emphasized to use ANN algorithms for predicting hourly concentrations of O_3, NO_2, PM_{10}, $PM_{2.5}$, SO_2, CO with correlation coefficient (R^2) between measured and predicted values and root-mean-

Table 2. Prediction of each station's NOX for input parameters (PM_{10}, SO_2, T, WD, WS, RH and AP)

Station	Machine learning techniques											
	Linear regression			Random forest regression			MLP regression			GMDH regression		
	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
A	0.85	15.16	10.72	0.91	11.61	8.03	0.82	16.32	10.76	0.94	10.95	6.65
B	0.59	16.52	8.18	0.52	17.92	9.21	0.57	32.48	20.69	0.87	8.99	6.10
C	0.46	12.96	10.00	0.77	8.40	5.68	0.64	22.57	16.75	0.88	7.19	5.44
D	0.89	17.42	13.26	0.84	21.45	15.23	0.83	22.08	17.03	0.93	18.68	13.60

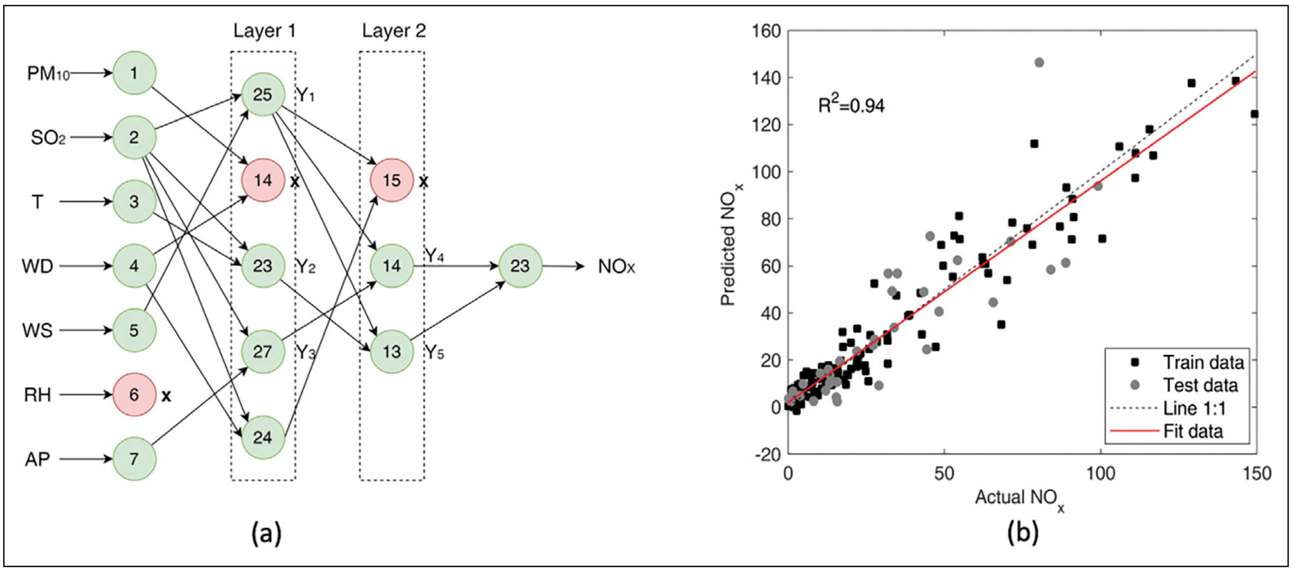

Figure 5. The prediction of NO_X in station A. **(a)** structure of double hidden GMDH layers; **(b)** actual and predicted NOX by GMDH model.

square (RMSE) values of 0.87 and 59.5 respectively [13]. It was reported by another research group that, using the WS, WD, and temperature as input variables, and ANN, ANFIS models have provided SO_2 prediction with R^2 values between 0.20 and 0.50 [32], and in another study with R^2 >0.70 [33]. Recent research have proposed using both the meteorological factors and air pollutant parameters as input variables and reported that the ANN model produced a $PM_{2.5}$ prediction with R^2>0.92 [34]. Another study revealed that the use of NO_X and meteorological parameters as input variables and the ANFIS model provided O_3 predictions with R^2>0.94 [35]. Based on findings of a recent study, using weather factors and air visibility as input variables is feasible for the ANFIS model and CO-NO_2, PM_{10}, SO_2 – O_3 were predicted with R^2 between 0.65–0.89 [10].

Prediction of NO_x Parameter Within a Station

In this study, the NO_X prediction was obtained by various regression methods using the PM_{10}, SO_2, T, WD, WS, RH, and AP of each station as input factors. Table 2 shows the calculated R^2, RMSE, and MAE values of linear, RF, MLP and GMDH-type regression algorithms for all stations. Regarding R^2, RMSE and MAE values, the prediction of NO_X within a certain station by the GMDH-type neural network have provided better results.

NO_x prediction at station A has been provided by the GMDH-type neural network with the design demonstrated in Figure 5a. This designed network structure consists of an input layer with 7 neurons, two hidden layers with 5 and 3 neurons, respectively, and an output layer with a single neuron. In this network structure, neuron outputs that minimize the error rate between predicted and actual output were selected. As a result, at station A the NO_X was predicted with R^2=0.94, RMSE=10.95, and MAE=6.65, (Fig. 5b). Similarly, prediction of NO_x in stations B, C, and D by the GMDH-type neural network model have ended up with regression coefficients of 0.87, 0.88, and 0.93, respectively.

The prediction of NO_X is formulated using the optimal neuron outputs of the network structure shown in Figure 5. Each polynomial equation obtained with active neuron outputs and finally the NO_X prediction equation is given in Table 3. Relative humidity (RH), one of the input layer pa-

Table 3. Parameters and coefficients used in neuron equations for the prediction of NO_X in station A

Equation	No.
$Y_1 = 54.76 + 3.06\ SO_2 - 37.49\ WS - 0.02\ {SO_2}^2 + 6.47\ WS^2 - 0.17\ SO_2 \cdot WS$	
$Y_2 = 15.54 + 2.58\ SO_2 - 0.56\ T - 0.02\ {SO_2}^2 - 0.003\ T^2 + 0.04\ SO_2 \cdot T$	
$Y_3 = [3.99 - 0.0008\ SO_2 - 0.0081\ AP] \cdot 10^4$	
$Y_4 = -6.09 + 0.77\ Y_1 + 0.69\ Y_3 + 0.014\ {Y_1}^2 + 0.004\ {Y_3}^2 - 0.02\ Y_1 \cdot Y_3$	
$Y_5 = -6.96 + 0.75\ Y_1 + 0.8\ Y_2 + 0.02\ {Y_1}^2 + 0.009\ {Y_2}^2 - 0.04\ Y_1 \cdot Y_2$	
$NO_x = 0.26 - 0.43\ Y_4 + 1.39\ Y_5 + 0.04\ {Y_4}^2 + 0.03\ {Y_5}^2 - 0.08\ Y_4 \cdot Y_5$	(8)

Table 4. Prediction of B, C and D station's NOX for input parameters (PM_{10}, SO_2, NO_X, T, WD, WS, RH and AP) of station A

Station	Machine learning techniques											
	Linear regression			Random forest regression			MLP regression			GMDH regression		
	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
B	0.56	17.25	9.31	0.63	15.83	8.44	0.55	17.36	9.69	0.85	9.45	6.16
C	0.11	16.66	9.78	0.22	15.54	9.54	0.08	16.92	10.30	0.54	13.24	8.19
D	0.29	44.85	33.44	0.38	41.94	28.62	0.25	46.25	32.66	0.65	38.63	29.15

rameters, is deactivated because it increases the error rate in the NO_X prediction model. The output of PM_{10} and WD polynomial pair in the first hidden layer, and the output of SO_2 and WD polynomial pair and Y1 output in the second hidden layer are not included in the model because they increase the error rate in the NO_X prediction. Thus, the NO_X prediction is modelled as a polynomial function using the selected optimal neuron outputs.

Prediction of NOX Across Stations

The NOX prediction of stations located at different distances and directions were performed using station A data as input parameters via various machine learning regression methods. Table 4 shows the NO_X prediction results of stations B, C, and D. The NO_X values are also used as input data of station A which is accepted as a source station. The results show that the GMDH-type neural network provides higher R^2, lower RMSE, and MAE values than other methods for all stations. It has been observed that the prediction performance of the proposed model is more successful at station B than at other stations. This result is due to the fact that reference station A, whose data is used as input, is closer to station B than to other stations.

The GMDH-type neural network model designed for the prediction of NO_X values at station B using data from station A and the regression fit line of this model are shown in Figure 6. In the double hidden layer model designed in Figure 6a, the input parameters WD and RH were deactivated because they increased the error between the predicted and actual output. The PM_{10} and NO_X polynomial pair output from the data of station A in the first hidden layer and output polynomial pair Y_1 and Y_2 in the second hidden layer were not used because they adversely affected the NO_X prediction of station B.

The NO_X prediction at station B is modelled as a polynomial function via the remaining optimal neuron outputs and the fit line of model is shown in Figure 5b. The model equations were obtained by using the optimal neuron outputs of the GMDH-type neural network. The design is demonstrated in Figure 5a, and the equations are given in Table 5. As a result, Eq. (9) is obtained for the case that neuron outputs are used as polynomial pairs in the NO_X prediction model of station B.

Prediction of NOX Across Stations without Using NO_x as Input Variable

The NO_X prediction across stations that are located at different distances and directions was performed using meteorological and air pollution data of station A data (the NO_X is excluded) as input parameters via linear, RF, MLP and GMDH-type neural network regression methods. The R^2, RMSE and MAE results of these regression methods including all stations were listed in Table 6. According to the analysis of NO_X prediction results of stations B, C and D (without using the NO_X data of station A, that was taken as the reference station). It is apparent that GMDH-type

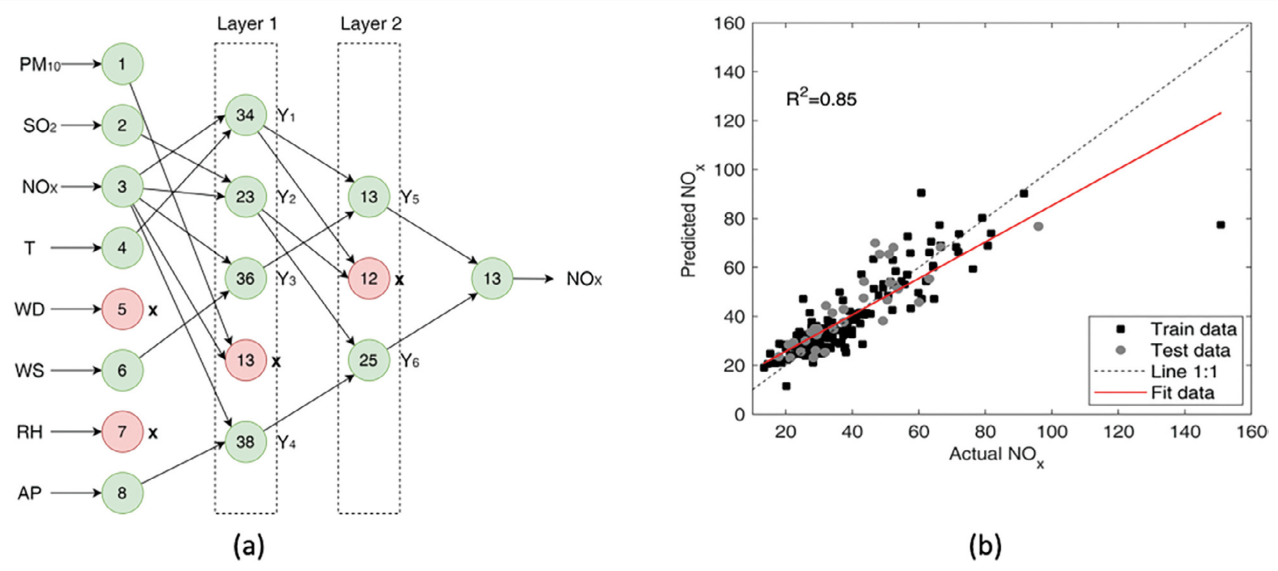

Figure 6. The prediction of NO_x in station B using NO_x of station A. **(a)** structure of double hidden GMDH layers; **(b)** actual and predicted NO_x by GMDH model.

Table 5. Parameters and coefficients used in neuron equations for the prediction of NO_x in station B using NOX of station A

Equation	No.
$Y_1 = 15.43 + 1.08\ NO_x + 0.76\ T - 0.03\ NO_x{}^2 - 0.01\ T^2 - 0.03\ NO_x \cdot T$	
$Y_2 = 21.5 - 0.52\ SO_2 + 1.003\ NO_x + 0.0002\ SO_2{}^2 - 0.005\ NO_x{}^2 + 0.04\ SO_2 \cdot NO_x$	
$Y_3 = 4.51 + 1.11\ NO_x + 16.59\ WS - 0.004\ NO_x{}^2 - 3.64\ WS^2 - 0.11\ NO_x \cdot WS$	
$Y_4 = [-1.05 + 0.0004\ NO_x + 0.0021\ AP] \cdot 10^4$	
$Y_5 = 0.27 + 0.94\ Y_1 + 0.05\ Y_3 - 0.06\ Y_1{}^2 - 0.05\ Y_3{}^2 + 0.11\ Y_1 \cdot Y_3$	
$Y_6 = 7.03 - 0.52\ Y_2 + 1.17\ Y_4 + 0.006\ Y_2{}^2 - 0.02\ Y_4{}^2 + 0.01\ Y_2 \cdot Y_4$	
$NO_x = -1.08 + 0.45\ Y_5 + 0.59\ Y_6 - 0.01\ Y_5{}^2 - 0.02\ Y_6{}^2 + 0.03\ Y_5 \cdot Y_6$	(9)

Table 6. Prediction of NO_x for station B, C and D station's by using (PM_{10}, SO_2, T, WD, WS, RH and AP) input parameters of station A

Station	Machine learning techniques											
	Linear regression			Random forest regression			MLP regression			GMDH regression		
	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
B	0.49	18.58	9.01	0.54	17.64	9.68	0.50	18.27	10.19	0.80	10.88	7.31
C	0.17	16.04	9.71	0.23	15.50	9.63	0.16	16.14	10.16	0.48	13.69	8.51
D	0.34	43.38	32.45	0.38	41.91	29.35	0.20	47.62	35.10	0.71	35.90	26.13

neural network has higher R^2 values; 0.80, 0.48 and 0.71 for stations B, C, and D compared to other methods, respectively. Therefore, empirical models for prediction of NO_x were obtained via GMDH-type neural networks.

The GMDH-type neural network model designed for the prediction of NO_x values at station B using station A data (not including NO_x) and the regression fit line of this model are shown in Figure 7. In the double hidden

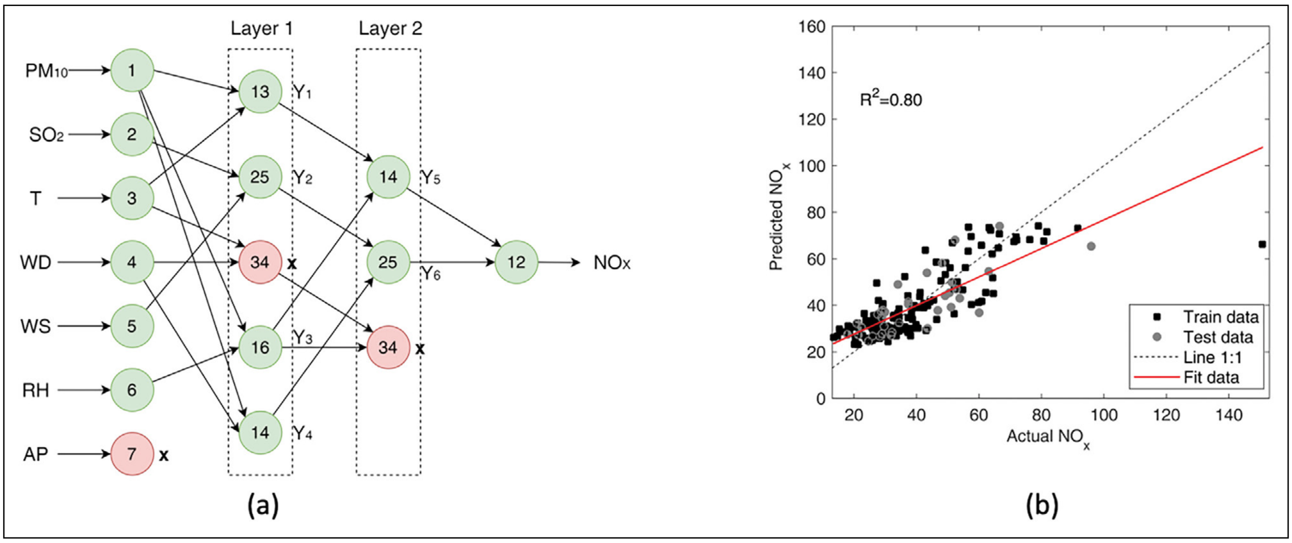

Figure 7. The prediction of NO_x in station B without NO_x of station A. **(a)** structure of double hidden GMDH layers; **(b)** actual and predicted NO_x by GMDH model.

Table 7. Parameters and coefficients used in neuron equations for the prediction of NO_x in station B without NO_x of station A

Equation	No.
$Y_1 = 13.51 + 1.47\ \mathrm{PM}_{10} + 0.28\ \mathrm{T} - 0.003\ \mathrm{PM}_{10}{}^2 + 0.01\ \mathrm{T}^2 - 0.06\ \mathrm{PM}_{10} \cdot \mathrm{T}$	
$Y_2 = 49.94 + 1.49\ \mathrm{SO}_2 - 11.5\ \mathrm{WS} - 0.01\ \mathrm{SO}_2{}^2 + 0.79\ \mathrm{WS}^2 + 0.009\ \mathrm{SO}_2 \cdot \mathrm{WS}$	
$Y_3 = 307.35 - 1.73\ \mathrm{PM}_{10} - 6.48\ \mathrm{RH} - 0.0002\ \mathrm{PM}_{10}{}^2 + 0.04\mathrm{RH}^2 + 0.03\ \mathrm{PM}_{10} \cdot \mathrm{RH}$	
$Y_4 = 27.84 - 0.16\ \mathrm{PM}_{10} + 0.1\ \mathrm{WD} - 0.005\ \mathrm{PM}_{10}{}^2 - 0.0008\ \mathrm{WD}^2 + 0.007\ \mathrm{PM}_{10} \cdot \mathrm{WD}$	
$Y_5 = -1.58 + 1.12\ Y_1 - 0.1\ Y_3 - 0.01\ Y_1{}^2 - 0.007\ Y_3{}^2 + 0.02\ Y_1 \cdot Y_3$	
$Y_6 = 1.61 + 1.42\ Y_2 - 0.5\ Y_4 + 0.003\ Y_2{}^2 + 0.01\ Y_4{}^2 - 0.02\ Y_2 \cdot Y_4$	
$\mathrm{NO}_x = -13.96 + 0.82\ Y_5 + 0.8\ Y_6 - 0.02\ Y_5{}^2 - 0.02\ Y_6{}^2 + 0.04\ Y_5 \cdot Y_6$	(10)

layer model shown in Figure 7a, the AP input parameter was not used because it increases the error between the predicted and actual NO_x values. The T and WD polynomial pair output was not used in the first hidden layer. In the second hidden layer, as they may pose an adverse effect on NO_x prediction, the polynomial pair output and the polynomial output formed by the Y_3 output were not used.

The NO_x prediction of station B is modelled in terms of polynomial function by using the optimal neuron outputs and the data fit line is shown in Figure 7b. The equations of the model obtained by using the optimal neuron outputs of the GMDH-type neural network, are given in Table 7. And the network structure was demonstrated in Figure 7a. As a result, Eq. (10) is obtained for the case that the neuron outputs are used as polynomial pairs in the NO_x prediction model of station B.

The results show that the predicted NO_x output at stations B, C and D varies depending on the distance and direction from the source station A. The stations B, C and D are located 21, 23 and 54 km from source station A, respectively. The correlation coefficient R^2 of the GMDH model proposed for stations B, C and D is 0.85, 0.54 and 0.65, respectively when source station A NO_x parameter is used. The station B, which is closest to the source station, has the highest correlation coefficient. Similarly, when the source station A station NO_x parameter is not used, the correlation coefficient R^2 of the GMDH model proposed for the stations B, C and D is 0.80, 0.48 and 0.71, respectively (prediction accuracy negatively affected for station D). Although the station D is at the farthest distance from source station A, the correlation coefficient results of the prediction model are higher than station C. That can be interpreted to the influence of local air pollution sources (point and/or linear sources like industry and traffic) on measured NOX values rather than the level of effect ascribed to near surrounding environment. In order to carry out an ascendant evaluation and sort out such entangled issues, air pollution trajectory and dispersion model and their outputs can be used as helpful tools and useful source of information for specific periods of time [4, 36].

The developed model provided prediction of NO_x within a certain station by using data from a total number of four different weather stations and taking one station as reference. For the specific cases of NO_x prediction within a station and across different stations, the GMDH like neural network model results have been increased as a means of enhancing accuracy and minimizing error rates (lower error rates obtained).

CONCLUSIONS

This study presents a case study of application of machine learning algorithms to predict NO_X concentrations using both air pollution and meteorological parameters. To accurately predict the NO_X parameter, data of a certain station and across different stations were compiled and new models were derived. The GMDH model produced a precise prediction of NO_X within stations and across stations (station to station) with/without using NO_X as the model input variable. The key findings can be emphasized below:

(1) The results show that meteorological parameters significantly affect the NO_X air pollution parameter and that the effects of meteorological parameters change with distance between stations.

(2) The proposed empirical models provide a rapid assessment of air quality and the prediction of NO_X with an acceptable range of accuracy (R^2={0.94, 0.85, 0.80} within station A, B via source station data and B via source station not including NO_X).

(3) Results obtained through GMDH models exhibit a high degree of accuracy for NO_X prediction values and significantly outperform conventional methods. The proposed model provides the opportunity to evaluate the effect of each input parameter on the model output. It has been observed that relative humidity (RH) increases the error rate and is disabled in the derivation of empirical models.

(4) The proposed GMDH-type neural network model uses air pollution parameters (PM_{10} and SO_2) and meteorological parameters (T, WD, WS, RH and AP) as inputs to estimate NO_X air pollution values.

(5) The obtained model is region specific, but for a wide range of spatial representations and validity, data can be gathered from multiple stations at certain distance.

(6) As a future scope, GMDH-type neural network and proposed approach can be used to support decision makers and engineers in planning stages including but not limited to optimization of total number and spatial distribution of AQMS to be set-up in a specific region.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Y. Dokuz, A. Bozdağ, and B. Gökçek, “Use of machine learning methods for estimation and spatial distribution of air quality parameters,” Nigde Omer Halisdemir University Journal of Engineering Sciences, Vol. 9, pp. 37–47, 2020. (in Turkish)

[2] S.M. Cabaneros, J.K. Calautit, and B.R. Hughes, A review of artificial neural network models for ambient air pollution prediction, Environmental Modelling & Software, Vol. 119, pp. 285–304, 2019. [CrossRef]

[3] A. Yakın, and R. Behçet, “Effect of different types of fuels tested in a gasoline engine on engine performance and emissions,” International Journal of Hydrogen Energy, Vol. 46, pp. 33325–33338, 2021. [CrossRef]

[4] H. Zhang, Y. Wang, J. Hu, Q. Ying, and X.M. Hu, “Relationships between meteorological parameters and criteria air pollutants in three megacities in China,” Environment Reseach Vol. 140, pp. 242–254, 2015. [CrossRef]

[5] H.K. Elminir, “Dependence of urban air pollutants on meteorology,” Science of the Total Environment, Vol. 350, pp. 225–237, 2005. [CrossRef]

[6] E. Demirci, and B. Cuhadaroglu, “Statistical analysis of wind circulation and air pollution in urban Trabzon,” Energy Building, Vol. 31, pp. 49–53, 2000. [CrossRef]

[7] S. Munir, M. Mayfield, D. Coca, S.A. Jubb, and O. Osammor, “Analysing the performance of low-cost air quality sensors, their drivers, relative benefits and calibration in cities—a case study in Sheffield,” Environmental Monitoring and Assessment, Vol. 191, Article 94, 2019. [CrossRef]

[8] N. Castell, F.R. Dauge, P. Schneider, M. Vogt, U. Lerner, B. Fishbain, D. Broday, and A. Bartonova, "Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates?," Environment International, Vol. 99 pp. 293–302, 2017. [CrossRef]

[9] M. Castelli, F.M. Clemente, A. Popovič, S. Silva, and L. Vanneschi, "A machine learning approach to predict air quality in California," Complexity, Vol. 2020, Article 8049504, 2020. [CrossRef]

[10] K. Prasad, and A.K. Gorai, P. "Goyal, Development of ANFIS models for air quality forecasting and input optimization for reducing the computational cost and time," Atmospheric Environment, Vol. 128, pp. 246–262, 2016. [CrossRef]

[11] N. Liu, X. Liu, R. Jayaratne, and L. Morawska, "A study on extending the use of air quality monitor data via deep learning techniques," Journal of Cleaner Production, Vol. 274, Article 122956, 2020. [CrossRef]

[12] R. Dongol, "Evaluation of the usability of low-cost sensors for public air quality information," [Unpublished Master Thesis], Universitetet I Oslo Department of Informatics Programming and Networks, 2015.

[13] H. Maleki, A. Sorooshian, G. Goudarzi, Z. Baboli, Y. Tahmasebi Birgani, and M. Rahmati, "Air pollution prediction by using an artificial neural network model," Clean Technologies and Environmental Policy, Vol. 21, pp. 1341–1352, 2019. [CrossRef]

[14] S.R. Shams, A. Jahani, S. Kalantary, M. Moeinaddini, and N. Khorasani, "Artificial intelligence accuracy assessment in NO2 concentration forecasting of metropolises air," Scientific Reports, Vol. 11, pp. 1–9, 2021. [CrossRef]

[15] J. Xu, Y. Xu, H. Wang, C. Guo, H. Qiu, Y. He, Y. Zhang, X. Li, and W. Meng, "Occurrence of antibiotics and antibiotic resistance genes in a sewage treatment plant and its effluent-receiving river," Chemosphere, Vol. 119, pp. 1379–1385, 2015. [CrossRef]

[16] U. Pak, J. Ma, U. Ryu, K. Ryom, U. Juhyok, K. Pak, and C. Pak, "Deep learning-based PM2.5 prediction considering the spatiotemporal correlations: A case study of Beijing, China," Science of the Total Environment, Vol. 699, Article 133561, 2020. [CrossRef]

[17] H.K. Cigizoglu, K. Alp, and M. Kömürcü, "Estimation of air pollution parameters using artificial neural networks," Advances in Air Pollution Modeling for Environmental Security, pp. 63–75, 2005. [CrossRef]

[18] A. Alimissis, K. Philippopoulos, C.G. Tzanis, and D. Deligiorgi, "Spatial estimation of urban air pollution with the use of artificial neural network models", Atmospheric Environment, Vol. 191, pp. 205–213, 2018. [CrossRef]

[19] F. Kunt, Z.C. Ayturan, and S. Dursun, "Used some modelling applications in air pollution estimates," Journal of International Environment Applied Science Vol. 11, pp. 418–425, 2016.

[20] Y.A. Ayturan, Z.C. Ayturan, H.O. Altun, C. Kongoli, F.D. Tuncez, S. Dursun, and A. Ozturk, "Short-term prediction of pm2.5 pollution with deep learning methods," Global Nest Journal, Vol. 22, pp. 126–131, 2020.

[21] M. Krishan, S. Jha, J. Das, A. Singh, M.K. Goyal, and C. Sekar, "Air quality modelling using long short-term memory (LSTM) over NCT-Delhi, India," Air Quality, Atmosphere & Health Vol. 12, 899–908, 2019. [CrossRef]

[22] G. Varol, B. Tokuç, S. Ozkaya, and Ç. Çağlayan, "Air quality and preventable deaths in Tekirdağ, Turkey," Air Quality, Atmosphere & Health, Vol. 14, pp. 843–853, 2021. [CrossRef]

[23] Thrace Development Agency, Thrace Region Plan (Rep.), 2019. Available at: https://www.trakyaka.org.tr/upload/Node/33264/xfiles/trakya_bolge_%0Aplani_2014-2023.pdf. Accessed on Feb 2022, 06.

[24] A. Vardar, R. Okursoy, and Y. Tekin, "Local wind characteristics for east Thrace, Turkey," Energy Sources, Part B: Economics, Planning, and Policy, Vol. 7, pp. 1–9, 2012. [CrossRef]

[25] A. Liaw, and M. Wiener, "Classification and regres-sion by random forest," R News, Vol. 2, pp. 18–22, 2002.

[26] F. Murtagh, "Multilayer perceptrons for classification and regression," Neurocomputing, Vol. 2, 183–197, 1991. [CrossRef]

[27] T. Kondo, "GMDH neural network algorithm using the heuristic self-organization method and its application to the pattern identification problem," In Proceedings of the 37th SICE Annual Conference. International Session Papers, IEEE, pp. 1143–1148, 1998.

[28] S.-K. Oh, and W. Pedrycz, "The design of self-organizing polynomial neural networks," Information Sciences, Vol. 141, pp. 237–258, 2002. [CrossRef]

[29] S.J. Farlow, "Self-organizing methods in modeling: GMDH type algorithms," CRC Press, Florida, 2020. [CrossRef]

[30] J. Ma, Y. Ding, J.C.P. Cheng, F. Jiang, and Z. Wan, "A temporal-spatial interpolation and extrapolation method based on geographic Long Short-Term Memory neural network for PM2," Journal of Cleaner Production, Vol. 237, Article 117729, 2019. [CrossRef]

[31] Ö. Akyürek, O. Arslan, ve A. Karademir, "SO2eV PM10 hava kirliliği parametrelerinin CBS ile konumsal analizi: Kocaeli örneği,TMMOB Coğrafi Bilgi Sistemleri Kongresi, Cilt 12, 2013.

[32] M. Savic, I. Mihajlovic, and Z. Zivkovic, "An anfis–based air quality model for prediction of SO2 concentration in urban area," Serbian Journal of Management, Vol. 8, pp. 25–38, 2013. [CrossRef]

[33] A.B. Chelani, C.V.C. Rao, K.M. Phadke, and M.Z. Hasan, "Prediction of sulphur dioxide concentration using artificial neural networks," Environmental Modelling & Software, Vol. 17, pp. 159–166, 2002. [CrossRef]

[34] G. Asadollahfardi, H. Zangooei, and S.H. Aria, "Predicting PM 2.5 concentrations using artificial neural networks and markov chain, a case study Karaj City," Asian Journal of Atmospheric Environment, Vol. 10 pp. 67–79, 2016. [CrossRef]

[35] L. Rafati, M. Ehrampoush, A. Talebi, and M. Mokhtari, Z. Kheradpisheh, H. Dehghan, "Modelling the formation of Ozone in the air by using Adaptive Neuro-Fuzzy Inference System (ANFIS) (Case study: city of Yazd, Iran)," Desert, Vol. 19, pp. 131–135, 2014.

[36] S. Teixeira, P. Pereira, and F. Ferreira, "Co-creating data based on human nose perceptions to study odour nuisance from an oilseed industry," 16th International Conference on Environmental Science and Technology, 2019.