Pollution 2022, 8(3): 904-933
DOI: 10.22059/POLL.2022.336044.1300

REVIEW PAPER

A Review of the Application of Machine Learning and Geospatial Analysis Methods in Air Pollution Prediction

Alireza Zhalehdoost[1,*] , Mohammad Taleai[1,2]

[1] GIS Department, Faculty of Geodesy & Geomatics Engineering, K. N. Toosi University of Technology, P.O.Box 16315-1355, Tehran, Iran
[2] School of Built Environment, Faculty of the Arts, Design & Architecture, University of New South Wales (UNSW).P.O.Box 259, Sydney, Australia

Received: 22.12.2021, Revised: 18.02.2022, Accepted: 01.03.2022

Abstract
During the past years, air quality has become an important global issue, due to its impact on people's lives and the environment, and has caused severe problems for humans. As prevention to effectively control air pollution, forecasting models have been developed as a base for decision-makers and urban managers during the past decades. In general, these methods can be divided into three classes: statistical methods, machine learning methods and hybrid methods. This study's primary intent is to supply an overview of air pollution prediction techniques in urban areas and their advantages and disadvantages. A comparison has also been made between the methods in terms of error assessment and the use of geospatial information systems (GIS). In addition, several approaches were applied to actual data, and the findings were compared to those acquired from previous published literatures. The results showed that forecasting using machine learning and hybrid methods has provided better results. It has also been demonstrated that GIS can improve the results of the forecasting methods.

Keywords: Air pollution forecast, Statistical methods, Neural Network, Hybrid methods, GIS

INTRODUCTION

Air pollution studies began with the deaths of large numbers of people. The incident in the Meuse Valley in 1930 and in December 1952 in London is evidence of this. In 1946, the city ordinance for air pollution was adopted in Pittsburgh. Still, at that level, the impact of pollution and its effects were not scientifically stated. Over time, and with discovering the more significant effects of air pollution on humans, controlling air pollution became a global problem. Therefore, countries are thinking of finding a solution to air pollution, which forecasting can help to tackle this problem. (Arthur, 1977).

Air pollution is one of the problems in the present century. It occurs when large amounts of harmful substances enter the Earth's atmosphere. Nowadays, air pollution is a grave danger to the world. In such a way, millions of tons of toxic substances enter the environment every year (Piraino et al., 2006). This phenomenon exists in many cities around the world.

Due to health and environmental problems caused by air pollution, it is necessary to be aware of the concentration of air pollutants in urban environments to assess the air quality condition. When accurate assessments of air conditions are available, we can create plans to control air quality. The need for forecasting requires that concentrations of pollutants can be modelled for areas at risk of contamination in sensitive situations. It is also helpful for the administration in this regard to prevent potential future losses. Many countries are attempting to reduce the concentration of pollutants. For this aim, they imposed several restrictions. For example, even

* Corresponding author Email: ar.jalehdoost@gmail.com

and odd number plate design and Tuesdays by bike (Núñez-Alonso et al., 2019).

In this study, the primary purpose is to investigate different introduced methods for predicting air pollution concentration. Since most of the review articles in the field of air pollution have not been specifically reviewed the research conducted in each method, in this study, in addition to reviewing the method, the advantages and disadvantages of each method were discussed. Besides, methods were compared in terms of error and the process of using GIS and its benefits for using data at different spatial resolutions to improve air pollution forecasting was also discussed.

The remainder of this paper is organized as follows. At first, the sources of air pollution and types of pollutants are mentioned. Then, information about the details of the reviewed papers is selected, and distribution of overviewed papers by year of publication and details about the papers work unit are given. In the next section, different prediction methods are examined. Then, methods are compared with each other in terms of error assessment. In the next Section, several modelling experiments were conducted on actual data to demonstrate the capability of various methodologies, and GIS for air pollution prediction. Then, discussion and conclusions are provided, and finally, recommendations for future work are given.

SOURCES OF AIR POLLUTION

Sources of air pollution can be classified into two major categories:

(i) Natural Sources

Natural air pollution sources include forest fires, volcanic eruptions, sandstorms, and gas leaks. These events are natural, and the pollution caused by them cannot be controlled. Humans do not play a significant role in the production of this type of air pollution.

(ii) Man-made Sources

Human-made plays a more significant role in air pollution caused by human activities. These sources of air pollution include (Bai et al., 2018; Nunez, 2019):

- Home resources: This type of pollution occurs due to indoor activities such as cooking, lighting a fireplace.
- Industrial resources: This type of pollution is caused by the activities of factories, power plants and petrochemicals factories.
- Dynamic resource: This type of pollution is caused by transportation and is affected by the development of cities.
- Agricultural resources: This type of pollution occurs when soil is plowed, and roots are exposed to air and oxidized to carbon dioxide. This is good for weed control but also releases carbon.

Various factors are affecting the spread of air pollution in cities described in Table 1.

Table l. List of affecting factors (Zickus et al., 2002; Cabaneros et al., 2019)

Types	Meaning of factors
Meteorological parameters	Meteorological parameters refer to variables that describe the chemistry of the atmosphere. These variables such as wind speed, wind direction and relative humidity have a wide impact on the dispersion and concentration of pollutants. Also, increasing the air temperature can cause the pollutants to move. Finally, just as air affects pollution, pollution also affects the climate.
City morphology	One of the most important factors in the spread of contamination in an urban environment is its morphology. Morphological parameters include: road network, population density, land-use type.
Types of pollutants	The type of pollutant can also be important, which is gases and particles.
Street traffic	The volume, speed and duration of traffic in each environment can cause different degrees of dispersion in pollutants.

Pollutants are suspended particles that darken the air in the city. The higher the number of particles in the air, the more pollution. The durability of these particles in the air is between a few instances to a few months. The types of air pollutants are as follows:

- CO (carbon monoxide): Carbon monoxide is toxic and has no color, smell or taste. The most common source of this gas entering the environment is the exhaust from cars. This gas blocks the stream of oxygen to the heart and brain by blocking the blood, which can eventually lead to death (Ochando et al., 2015; Ghadi et al., 2019).
- NO_x (Nitrogen oxide): Nitrogen oxide comes in two forms, nitrogen monoxide (NO) and nitrogen dioxide (NO2), which is gases, produced from natural sources, motor vehicles, and other combustion processes. Increased nitrogen dioxide levels can damage the human respiratory system and increase a person's vulnerability to respiratory infections and asthma. Large amounts of nitrogen dioxide are also harmful to plants, and they can damage the plant's leaves and reduce crop growth (Ochando et al., 2015).
- PM10: These suspended particles have an aerodynamic diameter of fewer than 10 micrometers. These particles reduce visibility in the city. They are produced mainly by industrial activities, which increase in relatively humid weather. These particles enter the body through the throat and nose, cause serious side effects on the lungs, and increase cancer risk (Alkasassbeh et al., 2013; Vitolo et al., 2018).
- PM2.5: This pollutant is one of the most dangerous particles. The size of these particles is so small that the body's natural filters are unable to purify them. If a person is exposed to this particle for a long time, the probability of death will be high (Ochando et al., 2015).
- O_3 (Ozone): The ozone in the air can be detrimental to health, particularly on warm shining days when ozone attains toxic levels. Even relatively low levels of ozone can have health consequences. Deep breathing can be difficult, depending on the amount of ozone exposure. Chest pain occurs during deep breathing. The frequency of asthma attacks increases and the lungs are more predisposed to infection. These consequences have been detected even in healthy people. However, it can be more severe in people with lung illnesses such as asthma (Jenkin & Clemitsha, 2000).

COLLECTION OF REVIEWED PAPERS

We selected articles in this work from reputable journals and identified the relevant literature from 1996. Table 2 categorizes the articles based on the authors, Types of forecasting methods, work unit, and forecasting time examined. The number of studied articles by year of publication is shown in Fig. 1. The percentage of articles based on the type of pollutant studied is shown in Fig. 2.

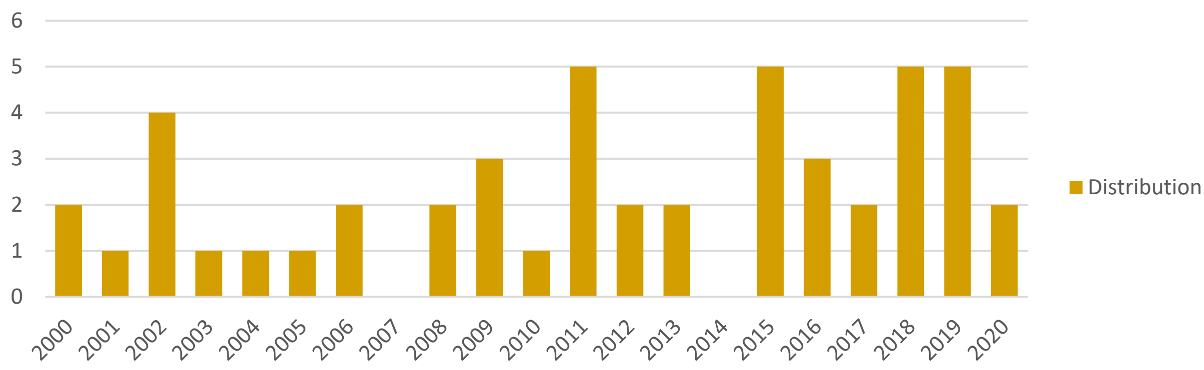

Fig. 1. Number of reviewed articles by year of publication

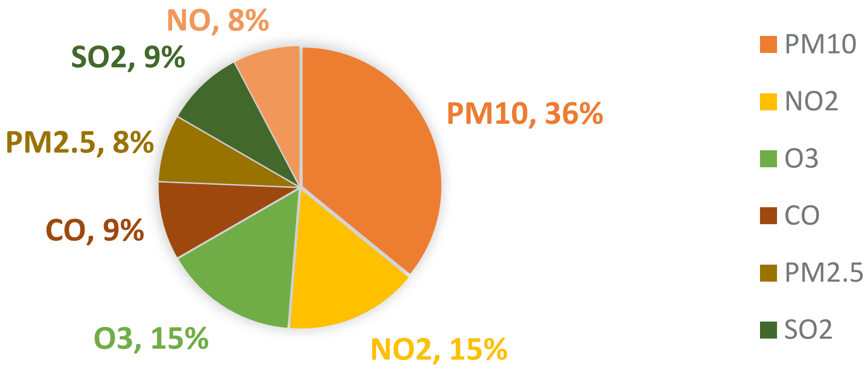

Fig. 2. Percentage of reviewed articles based on the type of pollutant

Table 2. Details of some papers based on work units

Authors (year)	Types of forecasting methods	Work unit	Forecasting time
Gardner & Dorling (2000)	Neural Network, MLR, Decision tree	Temporal	Hourly
Zickus et al. (2002)	Logistic Regression, Decision tree	Temporal	Daily average
Lu et al. (2002)	Neural Network, SVM	Temporal	-
Chelani et al. (2002)	Neural Network	Temporal	Real Time
McKendry (2002)	Neural Network, MLR	Temporal	Mean and maximum one day ahead
Chaloulakou et al. (2003)	Neural Network, MLR	Temporal	one-day
Lu et al. (2004)	Neural Network	Temporal	3 days ahead
Grivas & chalolaka (2006)	Neural Network+ Genetic	Temporal	1 day ahead
Papanastasiou et al. (2007)	Neural Network, MLR	Temporal	1 day ahead
Diaz-Robles et al. (2008)	Neural Network, MLR	Temporal	1 day ahead
Hrust et al. (2009)	Neural Network	Temporal	Hourly
Cai et al. (2009)	Neural Network+ MLR	Temporal	Hourly
Shad et al. (2009)	Fuzzy + Kriging + Genetic	Spatial	-
Kurt & Okay (2010)	Neural Network+ Geographical Factors	Temporal	3-days ahead
Paschalidou et al. (2011)	Neural Network, MLR	Temporal	1 hour ahead
Feng et al. (2011)	Neural Network+ Genetic + SVM	Temporal	Hourly
Sánchez et al. (2011)	Neural Network, SVM	Temporal	-
Kumar & Gloyal (2011)	MLR	Temporal	Daily
Nejadkoorki & Baroutian (2012)	Neural Network	Temporal	1 day ahead
Ahmad et al. (2012)	Neural Network+ GIS	Spatial	-
Alkasassbeh et al. (2013)	Neural Network	Temporal	-
Mishra & Goyal (2015)	Neural Network, MLR, PCA	Temporal	Hourly
Liu et al. (2015)	Regression	Temporal-spatial	Daily
Cortina–Januchs et al. (2015)	Neural Network	Temporal	1 day ahead
Abdullah et al. (2017)	MLR	Temporal	Daily
Aditya et al. (2018)	Logistic Regression, MLR	Temporal	Daily
Nunez-Alonso et al. (2019)	PCA + Kriging+ Cluster analysis	Spatial(city)	-
Leong et al. (2020)	SVM	Temporal	-

POLLUTION PREDICTION METHODS

Various methods and multiple factors have been used for monitoring and predicting air pollution concentration (Fig. 3). One of the critical points is the input variables that are different for each method. This issue limits the comparison between the methods. However, in this paper, we are tried to mention some positive and negative points of the methods and the usability of GIS to improve forecasting methods.

Statistical methods

Statistical methods are a particular type of mathematical analysis that requires input variables. These methods analyze events without knowing the mechanism of their change. Statistical methods are used to analyze and summarize data. Also, statistical methods do not model uncertainty and are not dependent on biological and physical processes. In general, statistical methods are split into two classes: parametric and non-parametric. Parametric models include linear regression models and principal component analysis. Non-parametric models do not have a net figure and function (Bai et al., 2018; Taheri Shahraiyni & Sodoudi, 2016).

Kriging

It is an interpolation method that belongs to the geostatistical. It can be used to analyze and predict unknown values. Kridge developed this model in 1951. The function of this method

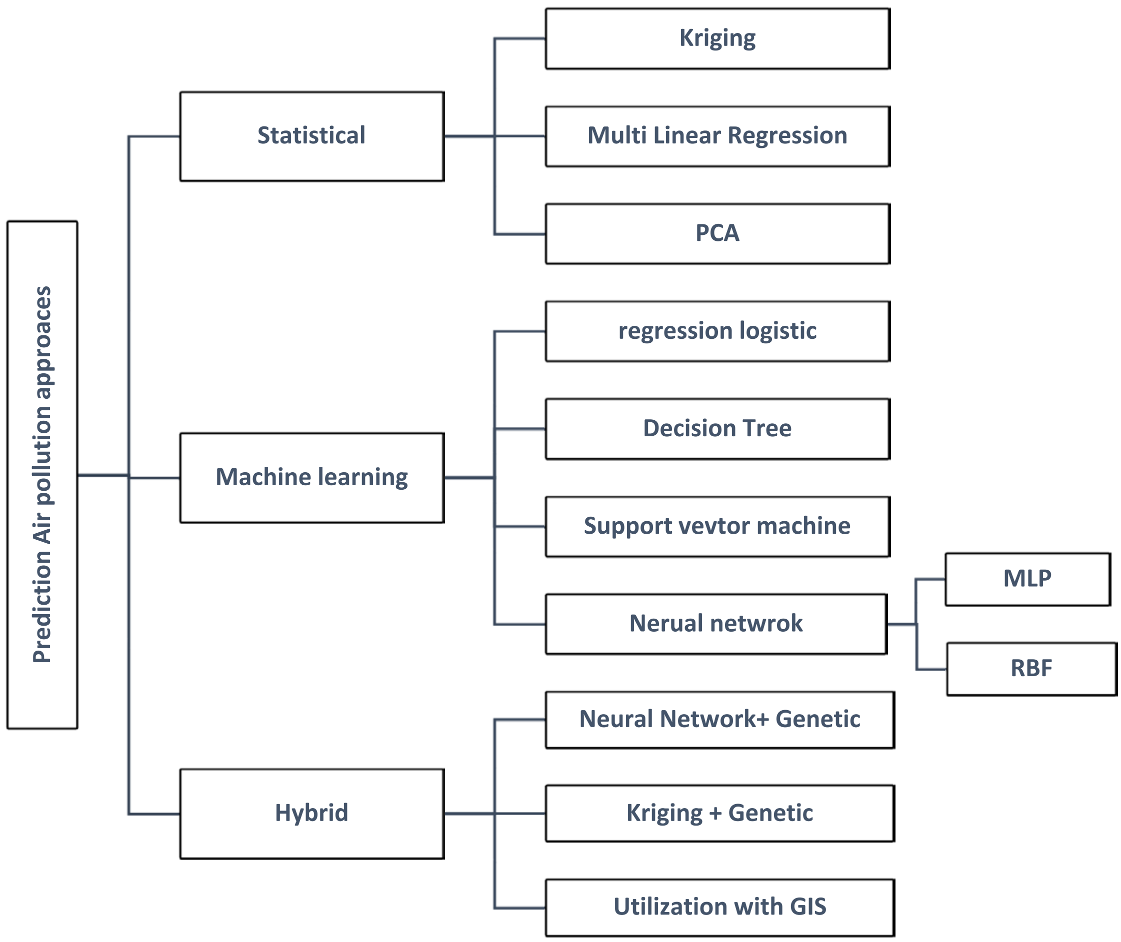

Fig. 3. Flowchart of air pollution prediction methods

is to minimize the variance by making an unbiased estimate of unknown situations. In other words, kriging is a statistical method for estimating random values at unknown points from observation points at known locations. Also, the primary tool of this method is Variogram, which relates the half average square difference between pair point data values to the distance between them (Shad et al., 2009; stein et al., 2001). Kriging is divided into three categories:

Ordinary Kriging: This method is used when the desired data level follows a constant μ. that means the average in the spatial domain is constant.

Universal Kriging: Universal kriging assumes that there is an overriding trend in the data that can be modeled by a deterministic function and a polynomial.

Indicator Kriging: Used when the data is binary. The basic Kriging formula is described by Equation (1):

$$Z(S_0) = \sum_{i=1}^{N} \lambda_i Z(S_i) \tag{1}$$

Where Z (S_i) is the measured value at the ith location, λ_i is an unknown weight for the measured value at the ith location and S_0 is the prediction location.

Shad et al. (2009) applied the Kriging method under three different scenarios to predict the concentration of PM10. First, they used ordinary Kriging, which directly uses spatial observations to model linear behaviors by linear predictors. However, they concluded that this method could not easily observe the change in concentration in the regions. Because this method has complex mathematical calculations, modelling uncertainty in this method is difficult, which is one disadvantages of this method. Second, they used the Indicator Kriging, so in order to be able to use this method, they converted the collected samples to binary (assumed a value of one for contaminated areas and zero for safe areas); Therefore, it is necessary to consider a threshold. The map obtained from this method shows significant clustering around the average, so it is easier to identify hazardous areas on this map. One of the disadvantages of this method is that several areas are generally considered pollution-free (due to threshold), which does not correspond to reality, so this method is inaccurate. In the third stage, they have used the fuzzy method to solve the index Kriging problem. In this way, three thresholds were considered. The results showed that this method has a better performance than the previous methods. However, this method of determining the thresholds depends entirely on the expert's opinion.

Remark1: In general, building and creating these types of models is simple and requires little time. It is better to use this method to show dangerous areas. Because it does not consider geographical factors such as wind, temperature and high dependence on the specialist, so this method is not too suitable.

Multi Linear Regression (MLR)

Regression is a statistical tool that examines the relationships between variables. There are two types of variables, namely dependent (y) and independent (x). Independent variables affect the occurrence of a phenomenon, but the dependent variable is the main parameter of the goal. MLR is widely used to predict air pollution (Bai et al., 2018; Kumar & Goyal, 2011- Yadav & Nath, 2019). Table 3 shows the various studies that have been performed using multiple regression to predict air pollution.

Remark 2: Since there is a significant correlation between the input parameters and the MLR cannot solve this problem, which uses principal component analysis to solve this problem. Besides, it is helpful to increase the independent variables in the regression model for improving the accuracy of predictions. However, increasing the independent variables will increase the computations, prolong the regression process, and complicate the forecasting. Consequently, the principal challenge for the regression equation is variable selection.

Principle Component Analysis (PCA)

PCA is a multivariate statistical analysis procedure that supported data compression and has an extraction. The extracted patterns contain information from the original data. This method reduces the number of predictor variables and makes them the main components. Using the normalized input correlation matrix can obtain the principal components (Bai et al., 2018;

Table 3. Various studies on the MLR method and their results

Types of pollutant	Country	Inputs	Result	Description	Ref.
PM10, NO_2, CO, O_3	China	The concentration of pollutants in 1st, 2nd and 3rd hour before, Traffic volume, Time of day, Day of the week, temperature, humidity, wind speed and direction, solar radiation and rainfall, Distance to road centerline, street direction, street aspect ratio	RMSE (CO) = 266.1 RMSE (NO_2) = 19.9 RMSE (PM10) = 17.9 RMSE (O_3) = 11.4	This Study distributed the factors affecting air pollution into four categories, which are the region's traffic, the concentration of existing pollutants, meteorological factors and geographical factors. Then they used the MLR but it didn't have good results, so they used other methods to predict.	Cai et al., (2009)
PM10	Chile	Maximum PM10, Minimum temperature, maximum temperature, relative humidity, wind speed, solar radiation, atmospheric pressure, and precipitation	RMSE = 28.39 MAE = 20.83 R^2 = 0.7786	In this study, MLR was used to predict air quality in urban areas. The results showed that, due to the inability to predict extreme events, they have limited accuracy and used a combined method to predict.	Mishra & Goyal, (2015)
PM10, PM2.5, O_3	Canada	PM10, PM2.5, NO, NO_2, CO, O_3, temperature, wind speed, direction and precipitation data	(R) $PM10_{mean}$ = 0.7 $PM10_{maximun}$ = 0.69 $PM2.5_{mean}$ = 0.69 $PM10_{maximum}$ = 0.63 O_{3mean} = 0.75 $O_{3maximum}$ = 0.68	In this study, data were obtained daily between 1995 and 1999 at the LVF station and the results showed that the regression method for predicting pollutants is not very satisfying. Therefore, the authors also used the neural network method in this study.	McKendry, (2002)
PM10	Mexico	PM10, wind speed, direction, temperature, relative humidity at the present time	(3 station) MSE (CR) = 0.0011 MSE (NA) = 0.0014 MSE (DIF) = 0.0016	In this study, data from three stations were used and a regression model was used to obtain a linear relationship between meteorological variables and pollutants. The results showed that because regression assumes that the behavior of the studied variables is linear, this model is not very accurate.	Cortina–Januchs et al., (2015)

Continued Table 3. Various studies on the MLR method and their results

Types of pollutant	Country	Inputs	Result	Description	Ref.
PM10	Greece	PM10, temperature, relative humidity, wind speed, direction, Barometric pressure, solar radiation and amount of rainfall	4 station RMSE (ARI) = 0.78 RMSE (GOU) = 1.06 RMSE (LYK) = 1.11 RMSE (UC) = 0.51	The present study considers four locations in Athens. Also measured the data was between 2001 and 2002. The results show that the model performance is better for the UC site because of the closeness of the meteorological station to this area.	Grivas & Chaloulakou, (2006)
PM10	Malaysia	PM10, ambient temperature, relative humidity, wind speed, Mean Sea Level Pressure and rainfall amount,	R^2 (NE) = 0.681 R^2 (Inter1) = 0.578 R^2 (SWM) = 0.570 R^2 (Inter2) = 0.628	Selected Observation of pm10 concentrations was from January 2005 to December 2011. In general, this study aims to create MLR models for different monsoon seasons with meteorological forecasting factors. The outcomes reveal that wind speed, relative humidity and rainfall are negatively correlated with PM10 concentrations, while atmospheric temperature and pressure are directly associated with PM10 concentrations. As a result, MLR models developed to predict PM10 concentrations locally for each monsoon are suitable.	Abdullah et al., (2017)

Mishra & Goyal, 2015). Also, Equation 3 can get the Eigenvalues of the C matrix.

$$|c - \lambda i| = 0 \tag{2}$$

Where C is a correlation or covariance matrix, λ is an Eigenvalues and i is the identity matrix (Yadav & Nath, 2019).

Remark3: The PCA method is not used independently to predict air pollution and is often used to solve the problem of correlation between input data in regression models.

Machine Learning Methods

Machine learning is a field of study derived from artificial intelligence. Machine learning was introduced in 1980 and is applied to develop complex models and algorithms such as prediction, classification, and clustering. Since 2010, the concept of deep learning in machine learning has occurred; which is a sub-branch of machine learning. Deep learning relates to algorithms motivated by the edifice and performance of the brain called artificial neural networks. The principle contrasts between machine learning, and deep learning is in the amount of data and problem-solving methods. Fig. 4. Shows the development stages of three methods of artificial

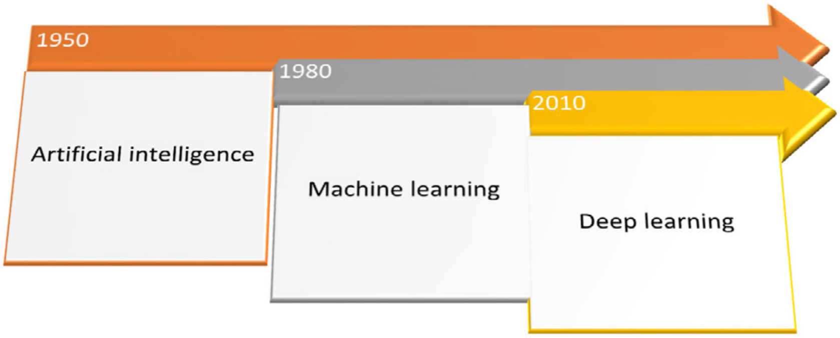

Fig. 4. Evolution steps of AI, ML and DL according to the years

intelligence, machine learning and deep learning according to its year (Copeland, 2016; Ayturan et al., 2018).

Logistic Regression

Logistic regression (LR) is a distribution algorithm that is used to estimate discrete values based on a given set of independent variables. This algorithm determines the probability of prediction, so its costs are between 0 and 1. Backward and forward stepwise variable selection algorithms are used to build the model (Aditya et al., 2018; Zickus et al., 2002). The logistic regression equation is as follow:

$$logit(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_{1,i} + \ldots + \beta_k x_{k,i} \tag{3}$$

Where x_1 to x_k are predictors and β is the model coefficients. Table 4 Table 3. Various shows the various studies that have been performed using Logistic regression to predict air pollution.

Remark 4: In LR-based methods, if the relationships between variables can be well described by the interaction between the two variables or the signal-to-noise ratio is very low, it is possible to identify the complex structure. It also describes data from a simple LR model better than a complex model. The outcomes also reveal that the execution of the model in the second data set is weak. Also, these models interpreted more easily and selected fewer input variables.

Support Vector Regression (SVR)

Among the various algorithms in machine learning, the SVR is one of the most well-known algorithms, which use for classification and regression. In (1999 Vapnik, V. N.), Vapnik et al. used SVR to solve nonlinear regression models, which is a more advanced version of backup vector machines for regression problems. This method has an advantage in high dimensions because in this case, it does not depend on the dimensions of the inputs. SVR has a function that transfers the input data to a higher-dimensional space, which is shown in Equation 4 (Bai et al., 2018).

$$f(x)=\left(\omega\times\varphi(x)\right)+b \tag{4}$$

Where f(x) is the prediction value, *w* is the weight vector, and *b* is the threshold.

Chen et al. Used SVR in 2010 to predict SO2 concentrations. They developed a model for studying air pollution to ensure people's health. They were able to pinpoint the mechanism of

Table 4. Various studies on the LR method and their results

Types of pollutant	Country	Inputs	Result	Description	Ref.
PM2.5	India	Temperature, wind speed, Dew point, Pressure and pm2.5 concentration	Mean accuracy = 0.998859 Standard deviation accuracy = 0.000612	In this study, logistic regression was applied to determine whether a region is polluted or not. They compared this method with other machine learning models used in the data set and concluded that Logistic Regression is the most appropriate model.	Aditya et al., (2018)
PM10	Finland	9 meteorological variables, 2 time variables and the mean PM10 concentration between 1-15 hours the day before.	Data set1 Overall accuracy =0.978 Data set2 Overall accuracy =0.975 Data set3 Overall accuracy =0.869	In this study, because many meteorological variables are firmly related, this algorithm cannot differentiate between statistical correlations and random correlations between inputs and outputs, so an initial screening was performed to avoid this problem. The results also determine that the execution of the model in the second data set is weak.	Zickus et al., (2002)

communication between pollutants. Also identified the factors that have the most significant impact on air quality (Qiao et al., 2010).

Remark 5: SVR is generally used to find a solution to nonlinear problems. This model also accurately shows the relationship between air pollutants. It also minimizes the prediction error and complexity of the problem at the same time. SVR has many advantages. First, there is a universal excuse for teaching it. Second, it does not get stuck in local minima like other methods. Third, it has a lighter and more automatic design. Finally, it controls the training error and complexity in the model.

Decision Tree

The decision tree is used to represent complex data structures and rules hierarchically and divides the data backward. This method displays the outputs of a decision branching. The decision tree receives the training data set from the input, and the variables are selected based on the training method and the purpose. First, all the data used to determine the rule, then the Gini index is applied using the most appropriate division criterion for each node, and by minimizing it, the measurement of the node impurity in the presence of the node is reduced to zero. The simplification process is then applied to the tree to describe the amount of error on the test set (Gardner & Dorling, 2000).

Gardner and Dorling used the decision tree in (2000) to predict ozone concentrations. They considered five sites, including the meteorological and monitoring site and their distance. After that regression tree was calibrated using an educational dataset and then evaluated with a test dataset.

Remark 6: The decision tree is easy to interpret, but its performance is less accurate than regression models. In contrast, the coefficients in the regression model need to interpret accurately, but in the decision tree method, there is no such problem.

Support Vector Machine (SVM)

In 1963, Vladimir Vapnik and Alexey Ya. Chervonenkis discovered the original support vector machine (SVM) algorithm. This method is one of the essential and developing topics in the discussion of machine learning that can solve regression problems and time series prediction (Wang et al., 2005). In general, this method minimizes the operational risk of classification or modeling instead of reducing errors. SVM is a supervised learning model that operates with learning algorithms that interpret data and recognize patterns. The SVM algorithm assigns samples based on training data and their batch type. The main idea of SVM is based on the kernel functions and its parameters, so it finds a hyperplane in the space and divides the samples into categories. The hyperplane completely separates the two categorize of data. Then maximize the distance between the nearest two points. This approach, in addition to leading to a unique response, also improves model performance. This hyperplane leads to solving a convex optimization problem, according to Equation 6, which is computationally cost-effective (Cortes & Vapnik, 1995; Djemai et al., 2016; Sánchez et al., 2011; Molina-Gómez et al., 2020).

$$min_{w,\zeta} \frac{1}{2}\|W\|^2 + C\sum_{i=1}^{n}\zeta_i \quad (5)$$

Where x is an independent variable that depends on the space R^p (p is the number of features of each sample), y is a dependent variable, w is a vector of coefficients, b is a constant value or bias, ζ is the kernel function and n is the number of training samples. Table 5 represents the various studies about SVM.

Table 5. Various studies on the SVM method and their results

Types of pollutant	Country	Inputs	Result	Description	Ref.
Respirable suspend particles	Hong Kong	RSP Concentration	MAE (for a day) = 0.0178 MAE (for a week) = 0.0204	In this study, they selected the months of June and December. The reason for choosing December is because RSP is usually high this month due to the cold weather. Also, the weather conditions are less effective this month so that they can ignore them. The results showed that SVM could be an alternative to time series prediction	Lu et al., (2002)
Air pollution index (PM, CO, O_3, SO_2, NO_2)	Malaysia	Pollutant Concentration	R^2 (before removal of outliers) = 0.9677 R^2 (after removal of outliers) = 0.9803	In this study, before developing the SVM model, a regression model was used to ensure the accuracy of the model because additional data can enhance the modeling error, and this can affect the model results. It has clearly shown that removing the outliers improves the execution of the model.	Leong et al., (2020)

Remark 7: The SVM algorithm minimizes structural risk so it can have good accuracy for forecasting. In this method, if the number of training data is large, the calculation time increases. It is also essential to choose the type of kernel function because it is highly influential in the final results.

Neural Network

In many air pollution studies that aim to predict, the neural network method is often used to model the behavior of air pollutants. They have a high ability to extract patterns from the data and are also very suitable for solving complex problems. Neural networks are computational networks that simulate biological neural network cells (human or animal). This simulation is a cell-to-cell (neuron-to-neuron) simulation (Graupe, 2007). Neural networks depending on the type of training, can be supervised and unsupervised.

Neural networks, like the neural system of the human brain, are trained with the help of examples and, by processing experimental data, transfer the law behind the data to the network structure. That is why these methods are denominated intelligent. The general process of neural network function is in the form of equations 6 and 7.

$$y_i^m = f\left(v_i^m\right) \tag{6}$$

$$v_i^m = \sum_{j=1}^{L} w_{ji}^m y_j^m + b_i^m \tag{7}$$

Where y_i^m is the model inputs, v_i^m is the output of the mth layer, f represents the transfer function, L is the number of interfaces with the previous layers, w_{ji}^m is equal to the weight of each interface, b_i^m represents the bias.

Multi-layer Perceptron (MLP)

MLP, as its name implies, is a multilayer network; each layer has its weights, inputs, biases and outputs. The multilayer neural network is the most widely used artificial neural network because if this neural network is trained correctly, it can solve many functions with an intermediate layer. However, using multiple layers increases network flexibility (Kröse et al., 1993). MLP also specifies the transfer functions used in each layer. These functions can be sigmoid, linear, radial, and so on. The function of the sigmoid is as follows.

$$f\left(v_i^m\right) = \frac{1}{1+\exp\left(-\theta v_i^m\right)} \tag{8}$$

Where θ is the slope parameter of the transfer function. The operation of the MLP network is that first the weights are randomly selected, and after applying the input values and adding random biases, they are compared with the output values, and the average squares of the error are calculated for them. Suppose the error value is less than the desired error specified for the goal. Training is stopped, otherwise weights and biases are corrected to reduce errors (Zangooei et al., 2016). Table 6 Table 6. Various Table 6. Various shows the diverse research in the field of MLP.

Remark 8: One of the points to be considered in the design of this network is the selection of network input parameters, and on the other hand, the correct design of network parameters such as number of layers, number of hidden layer neurons, transmission functions, learning rate and coefficient, as well as data normalization, which increases Accuracy in the MLP based models.

Radial Basis Function (RBF)

Broomhead and Lowe introduced the RBF neural network in (1988). These networks offer a different approach to more popular neural networks such as MLP. RBF networks have been used to solve many various problems, especially classification, patterning, and time series analysis based on interpolation theories. These networks have a hidden layer and use radial basis functions as their activation function (Broomhead & Lowe, 1988; Haykin, 1998). The

Table 6. Various studies on the MLP method and their results

Types of pollutant	Country	Inputs	Result	Description	Ref.
Cr, Cd, Pb, Fe, Ni, Zn, PM10	India	Air temperature, wind speed, wind speed, relative humidity and temporal parameters	RMSE (Cr) = 3.7 RMSE (Cd) = 3.9 RMSE (Pb) = 5.8 RMSE (Fe) = 3.2 RMSE (Ni) = 3.8 RMSE (Zn) = 3.6 RMSE (PM10) = 7.9	In this study, 300 hidden neurons were considered, and the learning rate was between 0.1 and 1. The cross-validation method was used to estimate the error, and the results showed that the neural network could be an excellent way to predict pollutants.	Chelani et al., (2002)
NO_2, CO, O_3, PM10	Croatia	Air pressure, relative humidity, temperature, wind speed, wind direction and pollutants concentration	R (NO_2) = 7.56 R (CO) = 0.24 R (o_3) = 10.86 R (PM10) = 13.26	The main idea behind the model in this study was that the more neural networks could learn the relationship between the measured physical values and the concentration of future pollutants, the more accurate they become. Also, due to possible problems in measuring the concentration, the concentration of one pollutant was not used as an input parameter to predict another pollutant. In general, a good correlation was found between the concentrations of predicted and observed pollutants, which indicates the success of these prediction methods.	Hrust et al., (2009)
PM10	Greece	Model1 = relative humidity, temperature, wind speed, wind direction Model2 = parameters of the first model and the concentration of PM10	RMSE (MLP1) = 0.95 R^2 (MLP1) = 0.03 RMSE (MLP2) = 0.76 R^2 (MLP2) = 0.03	In this study, two neural network models have been used to predict pollution. They also suggest that for nonlinear and complex operations, it is better to use two hidden layers. The results show that artificial neural networks (ANNs), if properly trained, can be beneficial for predicting particle contamination.	Chaloulakou et al., (2003)
PM10	Iran	Mean temperature, maximum temperature, solar radiation, wind speed and direction, the concentration of PM10, CO, NO	R (Aghdasieh) = 0.72 R (Bazar) = 0.73 R (Fatemi) = 0.59 R (Geophysics) = 0.55 R (Golbarg) = 0.05 R (Masoudieh) = 0.64 R (Poonak) = 0.41 R (Shahrerey) = 0.51	In this study, an artificial neural network use to estimate the maximum PM10 concentration in 24 hours in Tehran. Meteorological and pollutant concentrations from different air quality monitoring stations and meteorological sites included in this model and the aim was to predict the maximum PM10 concentration for the next day using pollutants and meteorological parameters of the previous day. Also, the result of these NNs for each monitoring site in an air quality monitoring network can partially compensate for missing data at neighboring monitoring sites.	Nejadkoorki & Baroutian, (2012)

Continued Table 6. Various studies on the MLP method and their results

Types of pollutant	Country	Inputs	Result	Description	Ref.
PM10, PM2.5, O_3	Canada	The concentration of PM10, Pm2.5 and O_3, temperature, wind speed and direction	R (O_3 max) = 0.8 R (PM10 max) = 0.66 R (PM2.5 max) = 0.65	In this study, data were obtained daily between 1995 and 1999 at the LVF station. The results showed that MLP models could be useful for predicting PM but require particular approaches to model training.	McKendry, (2002)
PM10	Iran	Hourly concentration of PM10, Pm2.5, SO_2, CO and temperature	RMSE = 2.26	In this study, the best model, a two-layer network with sigmoid tangent and sigmoid log transmission use, and 28 and 27 neurons in the first and second layers, respectively, were selected. Data were normalized to improve network performance. This network succeeded in predicting validation data with a correlation coefficient of 0.88. The results showed that according to the proof of the correct operation of artificial neural networks in this study, this method could use to manage air pollution and provide appropriate solutions to reduce pollution.	Asadolah-Fardi & Zangoi, (2017)
NO_2, PM10	China	Temperature, relative humidity, air pressure, wind speed, cloud cover, percentage of haze, percentage of mist, percentage of rain, percentage of sun	R^2 = 0.89	In this study, seventy-four places for NO2 and thirty-six places for PM10 were used to construct the model. The results showed that essential spatial explanatory variables include main roads, residential lands and public lands, which shows that the spatial patterns of NO2 and PM10 are strictly related to traffic situations and human actions. The ability and achievement of the model were also proven.	Liu et al., (2015)

RBF neural network usually has three main layers. The first layer contains the input nodes, the second layer performs the non-linear transmission work and the third layer, which is the linear layer and announces the network response to the inputs (Haykin, 1998). The structure of this network is that the input layer can connect to any node of the hidden layer. Later the interval between the input vector and the weight of the vector is calculated. The calculated interval is then converted nonlinearly to the RBF function to obtain the output (Park & Sandberg, 1991 Park, J., &). Table 7 Table 8. Various Table 7. Various shows an example of RBF-related research.

Remark 9: In general, the benefits of neural networks are the ability to self-learn, adapt, and use nonlinear data. Neural networks also have disadvantages, such as low convergence speed, difficulty in local minima, and data dependence. However, it is more recommended to combine these methods with other methods to achieve better results.

Hybrid Methods

These methods are created by combining two or more specific methods. It is generally better to use this method to take advantage of each method and achieve accuracy. Some of these methods are concisely described below.

Neural Network + Genetic algorithm

The aggregate of the neural network and genetic algorithm is used in two cases. The first case is when the genetic algorithm is used to select the input variables for the neural network; in this case, the number of variables is reduced. The second case is when the genetic algorithm is used to optimize weights and in the neural network. This solves the problem at the local minimum. It also increases accuracy (Grivas & Chaloulakou, 2006; Feng et al., 2011). Table 8 Table 8. Various Table 7. Various shows an example of the integration of the NN with GA research.

Remark 10: The hybrid method, uses the advantages of both methods. The combination of NN and GA improves the prediction results and increases the accuracy of the model.

Kriging + Genetic

The genetic algorithm is utilized to select the appropriate threshold in the Indicator Kriging. In (2009), Shad et al. planned to use index Kriging to predict PM10. However, instead of using an expert for setting the data to binary with threshold, they used an optimal method, which was genetic. In this way, a population with a probability between 0.6 and 1 was considered. Finally, a mutation operator randomly altered each gene with a probability of less than 0.1, then new offspring were produced. This model improves forecasting efficiency and makes it easier to select and generate an optimal membership function to find areas with high PM10 levels.

Remark 11: In this method, the role of the expert is reduced to define the membership function, and the user interface is free from limiting the study area. Kriging-GA hybrid method also facilitates the implementation and execution of the algorithm in a GIS environment.

Table 7. Various studies on the RBF method and their results

Types of pollutant	Country	Inputs	Result	Description	Ref.
NO_2, NO_x, RSP	Hong Kong	The concentration of SO_2, NO_2, NO, CO, NO_x, RSP, Wind direction, wind speed, outdoor temperature, indoor temperature, solar radiation	RMSE (RSP) = 9.95 RMSE (NO_x) = 90.3 RMSE (NO_2) = 20.24 MAE (RSP) = 7.62 MAE (NO_x) = 68.98 MAE (NO_2) = 15.51	The RBF network is used to predict pollutant concentrations from an hourly time series based on an air pollutant database monitored in the downtown area of Mong Kook. This article also recommended that in order to achieve a more accurate result and not to lose specific information and more training speed, it is better to use a combination of this method with principal component analysis.	Lu et al., (2004)
PM10	Cyprus	The hourly concentration of PM10, Wind direction, wind speed, total solar radiation, barometric pressure, relative humidity, and amount of rainfall	RMSE (Larnaca) = 32.16 RMSE (Limassol) = 21.92 RMSE (Nicosia) = 34.78 RMSE (Paphos) = 19.51	In this study, four essential stations in a city in Cyprus consider predicting PM10 concentrations. The MLP method also used in this study, and the researchers found that the RBF method was slightly weaker here than the MLP method.	Paschalidou et al., (2011)

Table 8. Various studies on the integration of Neural Network with Genetic algorithm and their results

Types of pollutants	Type of integration	Description	Ref
PM10	select the input variables for the neural network with genetic algorithms	This study is based on measured data for two years from 2001 to 2002. They utilized a genetic algorithm to select the input variable. In general, this algorithm is created from a random population of possible solutions, then according to the fitness function, the answer is obtained, and a new generation is created using the cross over and mutation. This study showed that reducing the number of variables did not reduce the model's ability to predict and estimate. Because the researchers used other methods to measure the function of genetic composition to reduce variables, so they came to the same conclusion	Grivas & Chaloulakou, (2006)
PM10	optimize weights in the neural network with genetic algorithms	They used the genetic algorithm to optimize the fundamental behavior of ANN in both train and test, and calculated the initial weight genetically. They compared the two states of the neural network and genetics + neural network and concluded that accuracy is better in the combined state	Asghari & Nematzadeh, (2016)

Utilizing GIS

In various research in air pollution modelling and forecasting, Geographic Information System (GIS) has been utilized to reliable and quick access to the needed spatial and thematic data in the form of maps (Ahmad et al., 2012; GISLongue, 2019).

GIS has been used to display the results of methods and also to calculate some needed values (such as distance) of other methods. Equation 9 shows the formula for the IDW as a deterministic interpolation method.

$$\hat{Z} = \frac{\sum_{i=1}^{n}\left(\frac{Z_i}{d_i}\right)}{\sum_{i=1}^{n}\left(\frac{1}{d_i}\right)} \tag{9}$$

Where $\hat{Z}$ is the value to be estimated, Z_i are is known value and d_i is the distance from the n data points to the point estimated n (Atabi et al., 2013).

Also, GIS can be used to manage necessary data for forecasting and simulation methods. One of these tasks is to separate the data into different spatial resolutions and then used them to predict air pollution. Since data analysis from the perspective of different spatial resolutions can have many effects on the final accuracy of the results, differences in spatial resolutions provide different insights. Of course, this effect on accuracy does not always lead to better model performance. This can vary depending on the type of resolution (fine and coarse), the type of contaminant, and the case study area.

GIS can also be used to analyze data. These include analyzing patterns, which can affect pollution data. Because we know that spatial data are auto correlated, and if we conclude that there is a spatial relationship in the data, the motivation behind this can be achieved with the help of analyzing patterns. This can be effective in predicting air pollution. Because if we can identify the processes that cause something to happen in a particular place, it can help make decisions. For example, it helps us do better data collection because it identifies the event that caused air pollution, reducing the time and volume of work. The following is an example of this analysis (Ahmad et al., 2012; ESRI).

- **Average Nearest Neighbor**: This analysis works based on distance and examines whether the placement of the data follows a specific pattern or not. So it determines the data is clustered or dispersed.
- **Hotspot Analysis:** This analysis is used if we want to see the cluster data. Hotspot shows the patterns.
- **Global Moran's I:** This analysis examines data dependency and specifies whether the occurrence of data depends on location or not.

Table 9 shows several reviewed studies that utilized the GIS tools.

Remark 12: With the help of GIS, the output of different pollution methods can be readily displayed in the form of a map and it is possible to quickly identify the areas involved in air pollution and divide/geo-visualize the areas into different classes of pollution. Data can also be indifferent spatial and temporal resolutions. Forecasting in various temporal resolutions such as daily, weekly, etc. and spatial resolutions can affect the accuracy of the results. Furthermore, utilizing spatial analysis such as pattern recognition of pollutants can improve the results of air pollution prediction.

RESULT AND SUMMARY

In this section, the results of utilizing the mentioned methods in the previous sections are compared in terms of error assessment. As a result of utilizing different parameters and conditions in the reviewed articles, different indexes have been utilized to evaluate the error and it is hard to make a truthful comparison between the methods used in various articles. Therefore, in this section, just articles that comprised two or more methods in the same conditions, including meteorological parameters, study area, etc., were considered.

To evaluate and compare the results of various methods, some evaluation metrics were used as explained below (Hiregoudar, 2020).

- **Root Mean Squared Error (RMSE):** In statistical modelling, a standard method for measuring the model's quality is the RMSE. If the predicted values are very close to the correct values, the RMSE will be small. If the predicted and correct values are very different, the RMSE will be enormous.
- **Mean Absolute Error (MAE)**: In statistics, the mean absolute error (MAE) is the measurement of the error between pairs of observations that reveals the same phenomenon. Its unit is the original data and is only comparable in models whose errors are measured in similar units. Its value is usually identical to RMSE but slightly smaller.
- **Correlation coefficient (R):** The correlation coefficient R measures the intensity and trend of a linear relationship between two variables in a disseminate. The value of R is always between +1 and -1. The closer this value is to +1, it means that there is a complete upward (positive) linear relationship, but the closer it is to -1, there is an entirely downward (negative) linear relationship, and if it is close to zero, it means that there is no linear relationship.
- **Coefficient of Determination (R^2):** R-squared (R^2) is an analytical measure that determines the proportion of variance for a dependent variable. While the R describes the relationship between an independent and dependent variable, the R^2 illustrates how much one variable's variance explains the second variable's variance. Consequently, if R2 is 0.50 for a model, nearly half of the observed inequality can be defined by the model's inputs.
- **The index of success (IS):** The index of success shows how properly the pollutant exceedances were predicted. Many correctly forecasted non-exceedances do not influence it, and also, it helps evaluate rare events.

- **Overall accuracy (A):** Overall accuracy indicates the portion of predictions that accurately predicted an event or a non-event.

MLP vs. MLR

As mentioned, RMSE and MAE are a way to identify a model's performance while low RMSE and MAE are preferred. Besides, the RMSE closer to zero and R2 to 1, indicate high accuracy

Table 9. Various studies on the utilization of GIS

Types of pollutants	Country	Kind of Utilizing GIS Tools	Description	Ref.
O_3	Pakistan	To calculate needed data (distance) for other methods	In this study, a combination of neural networks and GIS was used to predict ozone concentration. They used the IDW method for surface interpolation, which showed graphical variation in ozone concentration at different locations.	Ahmad et al., (2012)
PM10	Iran	To calculate needed data (distance) for other methods	In this study, they used kriging in different scenarios to predict PM10 concentrations, which is a flexible method for spatial forecasting.	Shad et al., (2009)
SO_2, NH_3, NO_x	Japan	Manage data	They compared emission maps of Osaka Prefecture at resolutions of 1*1 km and 3*3 km. They concluded that NH3 pollutants perform better at lower resolutions, but NOx has the opposite effect that means NOx performs better at higher resolutions, whereas in SO2, there is no difference between the two resolutions.	Nishikawa & Kannari, (2011)
O_3	France	Manage data	The purpose of this study is to investigate the effect of model results due to changes in spatial resolution on various sources of uncertainty. Modeling and simulation of ozone concentrations from 6 to 48 km indicate that increasing the resolution improves the model to some extent, but this improvement occurs to a certain amount.	Valari & Menut, (2008)
PM10, CO, NO, NO_2, SO_2	China	Manage data	Tan et al. used the WRF/CMAQ system in 2015 to investigate the effect of spatial resolution on air modeling. In this work, the effect of spatial resolution on air simulation in the industrial area was investigated. They considered the resolution to be 1 km and 3 km. They also used GIS to distribute emission sources. In this way, they determined the geographical location of industrial resources and power plants with the help of GPS and allocated the distribution to grid cells with the help of GIS. Their results showed that the emission inventory allocated at a finer resolution provided more accurate information about the area. They also made a comparison of the effect of spatial and temporal resolution in three industrial, urban and rural areas based on emission log data, which showed that the urban area showed better results in predicting in lower resolutions. On the other hand, in rural areas, due to the used emission inventory, the effect of resolution was not significant. Also, industrial areas were able to predict pollution at low resolutions but failed for the temporal variation.	Tan et al., (2015)

Continued Table 9. Various studies on the utilization of GIS

Types of pollutants	Country	Kind of Utilizing GIS Tools	Description	Ref.
O_3	Pakistan	Analyze data	They used Hot Spot analysis after prediction with a neural network. Because they believe that highlights the areas of high concentration is help to control it before it increases above alarming levels.	Ahmad et al., (2012)
PM10, SO_2	Turkey	Analyze data	This study aimed to determine the spatial variations of PM10 and SO2 pollutants in Turkey's Marmara region. The GIS method was used to monitor and model air pollution and spatial data analysis was used to measure pollutants' concentration. Spatial dispersion maps of those pollutants were formed to specify the emission template for the study areas using geo-statistical methods. Besides, two regression models were performed on the determined pollution and utilized to detect conceivable air quality facture in the area using standard ordinary least squares (OLS) and spatially autoregressive (SAR) models. Two regression types showed that all four critical meteorological variables (temperature, wind speed, humidity and atmospheric pressure) are utilized to describe the level of pollution about air quality. After defining the model parameters, Moran's I analysis was employed to explore the spatial correlation. When dependent variables have spatial correlation, the OLS method is more appropriate and valid than SAR.	Arslan & Akyürek, (2018)

between observed and predicted values. Table 10 presents the comparison of the results of MLP and MLR methods in terms of error evaluation in different articles. As shown in Table 10, the MLP model's accuracy is generally better than the MLR.

MLP vs. SVM

Table 11 presents the comparison of the results of MLP and SVM in terms of error evaluation in different articles. According to Table 11, it is clear that the SVM method has a better performance for predicting SO_2 and PM10, while the MLP method predicted CO better.

MLP vs. RBF

Table 12 presents the comparison of the results of MLP and SVM methods in terms of error evaluation in different articles. As shown in Table 12, the MLP method has less RMSE and more R2 than the RBF method and it can be concluded that the MLP method has more reliable performance.

RBF vs. SVM

Table 13 presents the comparison of the results of the MLP and SVM methods in terms of error evaluation in different articles. According to the MAE in the reviewed papers, SVM performs better in predicting air pollution than RBF.

Table 10. Comparison of MLP and MLR methods

	Types of pollutants	MLR	MLP	Ref
1	**PM10**	RMSE = 23.4 MAE = 17.98	RMSE = 21.19 MAE = 16	Chaloulakou et al., (2003)
2	**PM10**	RMSE = 11.25 R = 0.74	RMSE = 11.37 R = 0.78	Papanastasiou et al., (2007)
3	**O_3**	RMSE = 11.4 MAE = 8.1 R = 0.936	RMSE = 10.3 MAE = 7.5 R = 0.941	Cai et al., (2009)
4	**PM10**	RMSE = 28.39 MAE = 20.83 R^2 = 0.7786	RMSE = 28.57 MAE = 20.65 R^2 = 0.7770	Díaz-Robles et al., (2008)
5	**NO_2**	R^2 = 0.47	R^2 = 0.89	Liu et al., (2015)

Table 11. Comparison of MLP and SVM methods

Types of pollutants	MLP	SVM	Ref
SO_2	R = 0.5150	R = 0.5841	Sánchez et al., (2011)
CO	R = 0.6867	R = 0.6552	Sánchez et al., (2011)
PM10	R = 0.4185	R = 0.5359	Sánchez et al., (2011)

Table 12. Comparison of MLP and RBF methods

Types of pollutants	Station	MLP	RBF	Ref
PM10	Larnaca	RMSE = 27.56 R^2 = 0.79	RMSE = 32.16 R^2 = 0.54	Paschalidou et al., (2011)
PM10	Limassol	RMSE = 15.96 R^2 = 0.74	RMSE = 21.92 R^2 = 0.57	Paschalidou et al., (2011)
PM10	Nicosia	RMSE = 31.98 R^2 = 0.7	RMSE = 34.78 R^2 = 0.51	Paschalidou et al., (2011)

Table 13. Comparison of RBF and SVM methods

Types of pollutants	Type of prediction	RBF	SVM	Ref
Respirable suspend particles	One-day prediction	MAE = 0.0236	MAE = 0.0178	Lu et al., (2002)
Respirable suspend particles	One-week prediction	MAE = 0.0266	MAE = 0.0204	Lu et al., (2002)

Table 14. Comparison of Logistic regression and Decision tree methods

Types of pollutants	Type of data	LR	DT	Ref
PM10	Data set 1	A = 0.978 IS = 0.43	A = 0.962 IS = 0.26	Zickus et al.,(2002)
PM10	Data set 2	A = 0.975 IS = 0.25	A = 0.953 IS = 0.11	Zickus et al.,(2002)

Logistic Regression vs. Decision Tree

Table 14 presents a comparison of the results of Logistic regression and Decision Tree methods. Considering the index of success as a measure of performance, the Logistic regression method has better accuracy and performance than the decision tree method.

Hybrid

Table 15 presents the performance of the hybrid method. As shown in Table 15, hybrid methods have been presented better results. As a general result, it was found that machine learning methods have better accuracy and performance than statistical methods. It can also point out that hybrid methods significantly improve the model's accuracy, speed, and performance.

EXPERIMENT

Some of the approaches mentioned in the preceding parts are assessed in this section. They are tested using actual data to verify the results reported at previous researches. This section is

Table 15. Comparison with the hybrid method

Types of pollutants	Type of integration	Error assessment		Ref
PM10	MLP + Genetic Algorithm	MLP	RMSE = 42.1764 R^2 = 0.53932	Asghari & Nematzadeh, (2016)
		MLP + GA	RMSE = 41.9134 R^2 = 0.54889	
NO_2	RBF + PCA	RBF	RMSE = 20.24 MAE = 15.51	Lu et al., (2004)
		RBF + PCA	RMSE = 15.87 MAE = 12.16	
PM10	Kriging + Genetic	Kriging	MSE* = 12.29545	Shad et al., (2009)
		Kriging + GA	MSE = 3.74528	

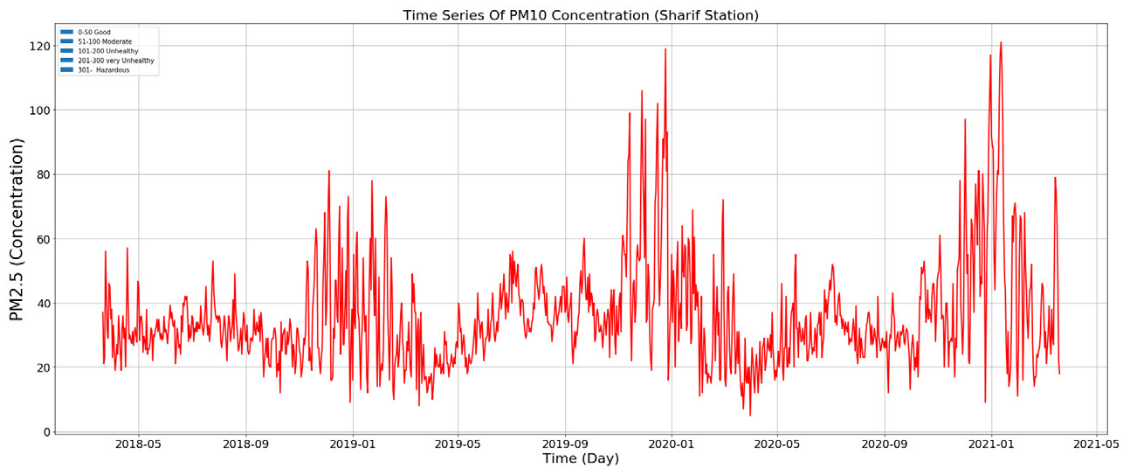

Fig. 5. time series plot of the PM2.5 pollutant

Table 16. Modeling outcomes of various methods

	INDICATORS			
METHODS	**MSE**	**RMSE**	**MAE**	**R^2**
MLP	37.655	6.136	4.352	0.85
RBF	59.138	7.690	5.510	0.73
Decision Tree	73.303	8.562	6.367	0.73
SVR	78.767	8.875	6.518	0.49
MLR	77.344	8.795	6.626	0.56

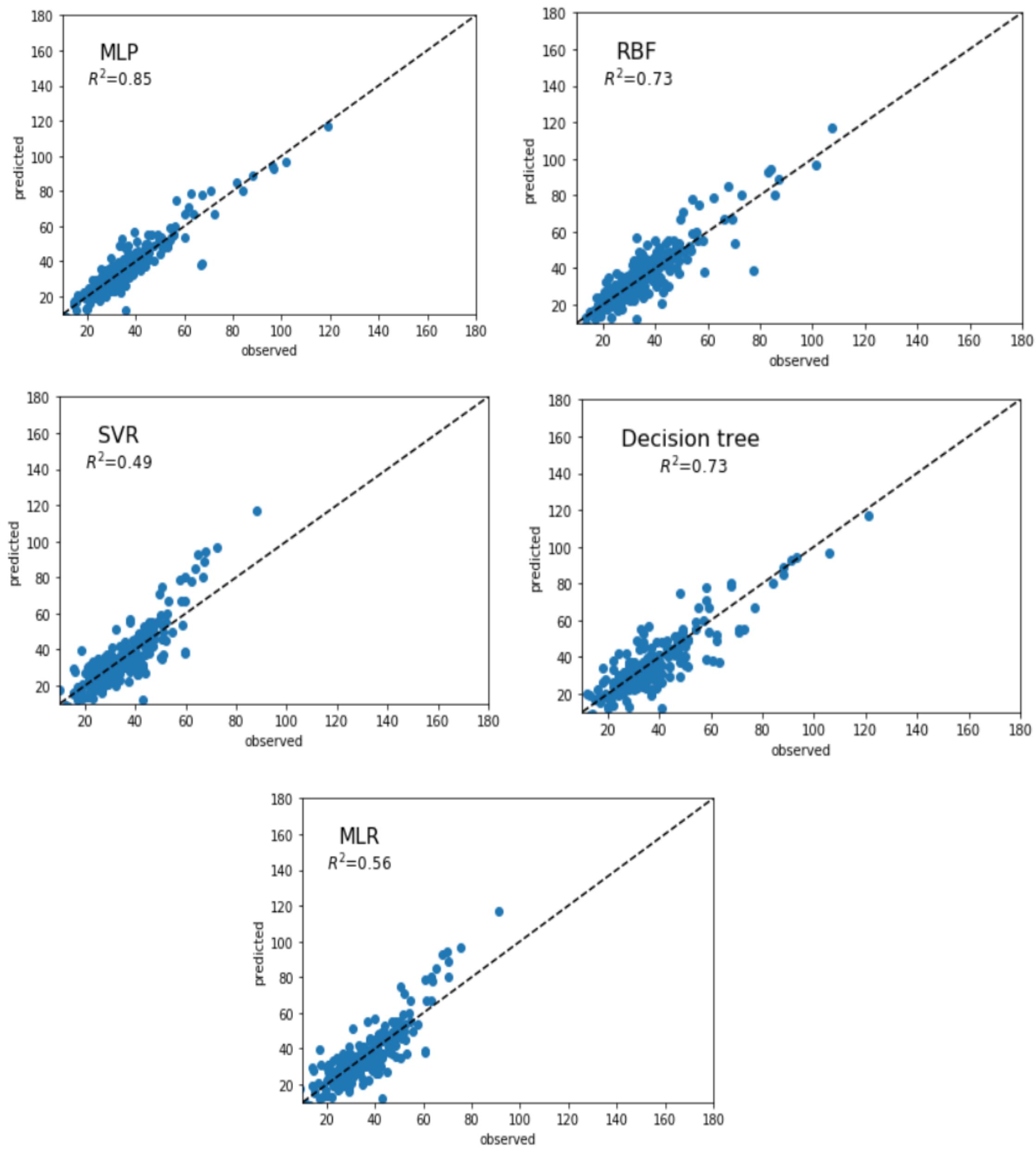

Fig. 6. Relationship between Real values and Predicted values in several methods

divided into two parts. The first, compares the five approaches: MLP, RBF, SVR, Decision Tree, and MLR, and the second part shows how the application of GIS could improve the outcomes of the models.

Comparison of five methods: MLP, RBF, SVR, Decision Tree, and MLR

The PM10 and PM2.5 pollutants used for modelling were obtained from Sharif Pollution Station in Region 2 of Tehran from 2018/03/21 to 2021/03/20. Also, Meteorological information, including temperature, humidity, wind speed, wind direction, pressure and rainfall, was obtained from the Geophysic Meteorological Station from 2018/03/21 to 2021/03/20. Furthermore, the number of days was evaluated as a criterion. As a result, meteorological data, days, and PM10 pollutants are input of the model, while PM2.5 pollutants are output. Figure 5 depicts a time series plot of the PM2.5 pollutant.

Table 16 shows the results of implementation of various methods on the case study data, based on several indicators.

Figure 6 shows the regression relationship between the observed and predicted values within the models.

According to Table 16 and Figure 6, machine learning approaches (MLP) frequently produce acceptable results for predicting pollutants. Among these approaches, the MLP method is more accurate since it can predict unknown values with extremely good R^2 and RMSE. These findings are entirely compatible with the findings of the reviewed papers, which reported in section 5 of the article.

Evaluating the capability of GIS to improve the outcomes of the models

GIS may be used to organize, display, and analyze data, as stated in previous sections. This part of experiment, will demonstrate how to utilize GIS to improve the results of air pollution prediction. The purpose of this step is to spatially forecast the concentration of PM2.5 pollutants using Decision Tree techniques. Meteorological data, such as temperature, humidity, wind speed, wind direction, pressure, and precipitation, are acquired on a monthly average from meteorological stations from 2020/03/20 to 2020/04/19. Other parameters include PM10 and PM2.5 pollutants gathered on a monthly average from air pollution monitoring stations in Tehran for the relevant period. Other statistics, such as location, altitude, and population density, are also considered.

To begin, for spatial modelling and forecasting, we utilize data from 14 air quality monitoring stations for training and test, followed by six stations for evaluating the results of the model. These models employ PM10, NOx, Ozone, meteorological data, position, elevation, and population density at each point as inputs, and the output is the PM2.5 concentration at each point. The data must also be computed in regular grids of points with resolutions of 500m * 500m, 1500m * 1500m and 3000m * 3000m. In order to do this, IDW is used in GIS software. Figure 7 shows these grids.

As illustrated in Figure 7, GIS allows the data to be readily shown and computed in multiple resolutions to be utilized in the model. Table 17 displays the outcomes of geographic modelling of PM2.5 concentrations using the Decision Tree approach within three grids 500 m * 500 m, 1500 m * 1500 m, and 3000 m * 3000 m. Table 17 shows that modelling accuracy improves with lower resolutions; hence, PM2.5 works better at lower resolutions.

Finally, using GIS, a spatial concentration map of PM2.5 pollutants as shown in Figure 7 within a grids of 500m * 500m.

DISCUSSION AND CONCLUSION

Based on the reviewed papers, different methods have been used to predict air pollution, however, there is different kind of air pollutants which has its characteristics and it is impossible

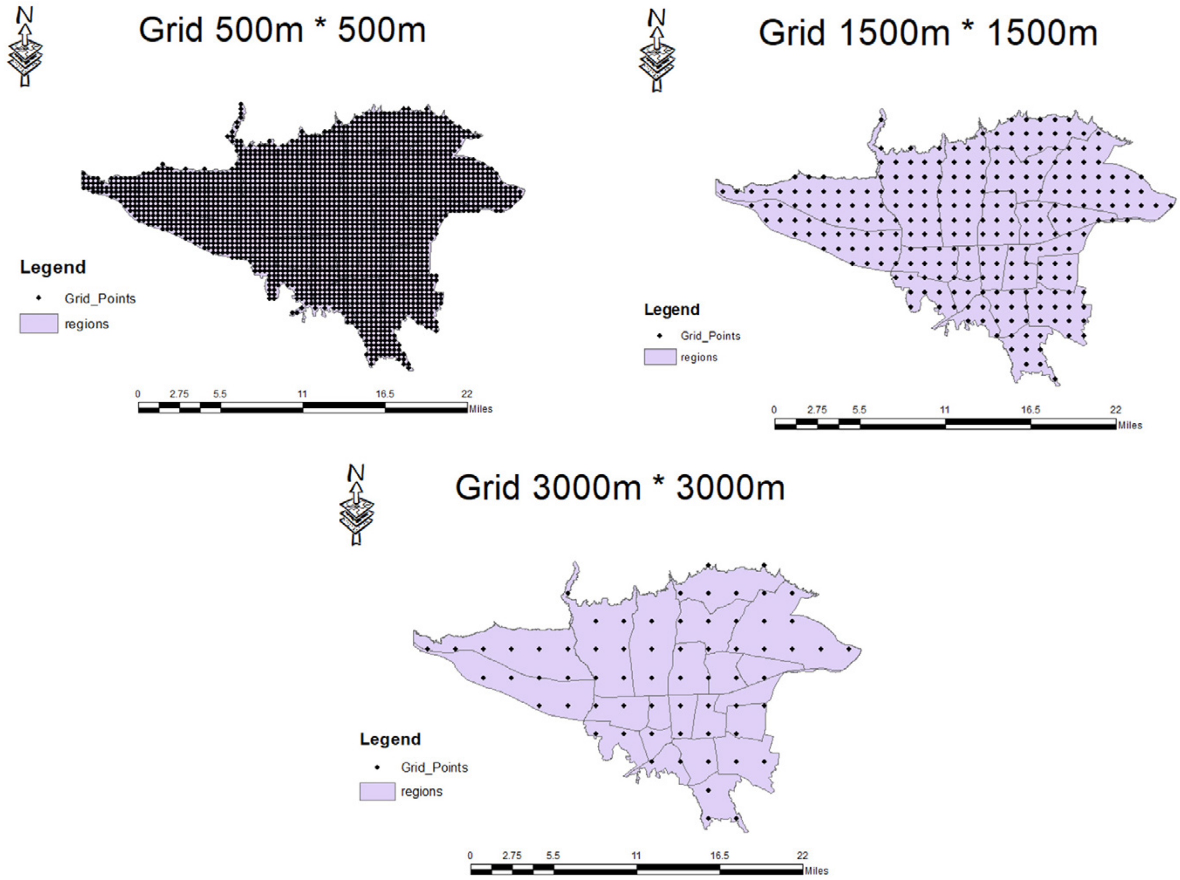

Fig. 7. Created grid points in Tehran

Table 17. the results of spatial modelling of PM2.5 concentration

Grid	Decision Tree					
	Test			Evaluation		
	MSE	RMSE	MAE	MSE	RMSE	MAE
500m *500m	0.052	0.227	0.099	5.565	2.359	2.012
1500m * 1500m	0.356	0.597	0.305	3.191	1.786	1.373
3000m * 3000m	2.075	1.440	1.101	11.949	3.475	2.589

to say precisely which method works best. Consequently, in this paper, the motivation was on considering the advantages and disadvantages of the methods.

1) **Statistical methods:** Statistical models have many advantages. In general, building and creating these types of models is simple and requires little time. These methods require more long-term data.

- **Kriging:** According to the results, it is complicated to model uncertainty in ordinary Kriging; because of the complex mathematical computation. In the indicator Kriging method, determining a proper threshold based on the experts/specialists' opinion is a challenge.
- **Multi Linear Regression:** In regression methods, the selection of variables is too difficult when there is a high correlation between input parameters. This method is easier to implement and interpret. When there is a linear relationship between independent and dependent variables,

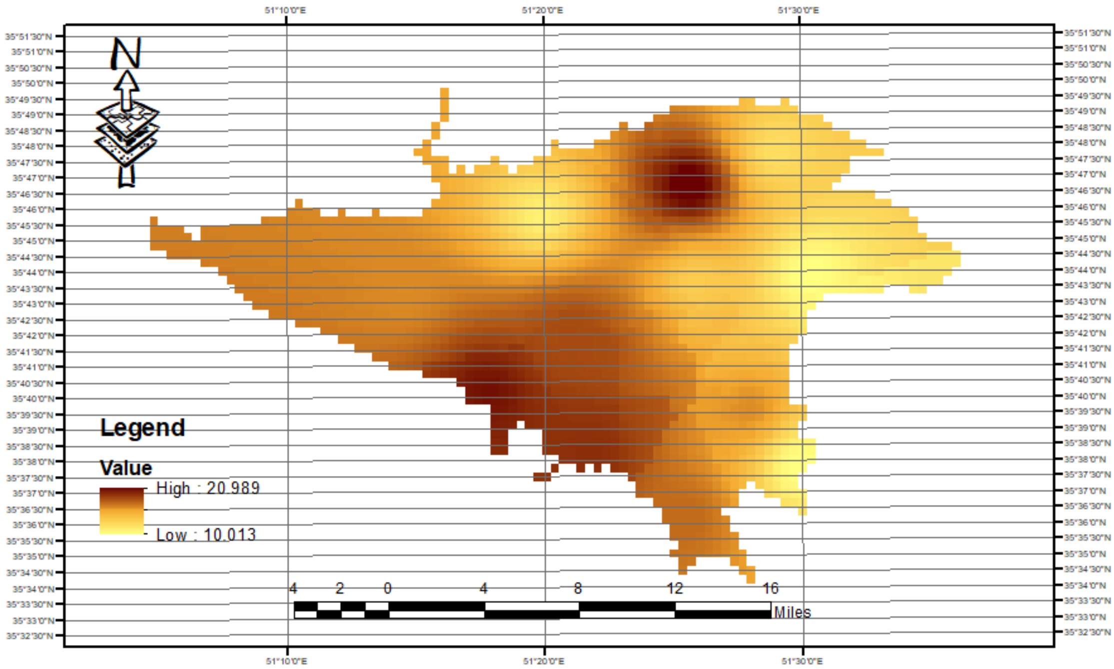

Fig. 8. PM2.5 concentration map

this algorithm is the best option because it is less complicated than other algorithms. In this method, over-fitting may also occur. However, this method is not recommended for most applications because it considers the relationship between variables to be linear, which does not correspond to reality

- **Principle component analysis:** This method is generally used to reduce the number of variables and solve the regression correlation problem and is also rarely used independently to predict air pollution.

2) **Machine learning methods:** These methods generally work well and their advantages are effortless to combine with other methods and the ability to identify the pattern of data. Machine learning analysis improves the performance of air quality predictions and helps to understand what factors affect pollutant concentrations. The disadvantage of machine learning methods is that the data may contain incorrect information and as a result an erroneous result. Collecting needed data to run the method is also costs a lot. Furthermore, different algorithms have to be implemented to choose the best one, which is too tedious and time-consuming activity. Among machine learning methods, the performance of the decision tree method is significantly lower than the others. Logistic regression models are more comfortable to interpret than others while has fewer input variables.

3) **Hybrid methods:** These methods take advantage of different techniques, so in most cases, has been recommended to use this method for prediction. The result of combining the Fuzzy and Kriging, Indicator Kriging and genetic algorithm proved that the prediction accuracy has improved; this makes it easier to select a suitable membership function to find areas with high levels of pollutants. This approaches also reduces the role of expert opinions and makes the forecasting process more automated. Combining the neural network with the genetic algorithm has proved improvement of the prediction accuracy, reducing the number of variables and increasing the speed of the model implementation.

Finally, utilizing machine learning methods has become very common in contemporary years and most researchers have resorted using of hybrid methods.

The utilization of GIS revealed that it could be used to analyze data in different spatial and temporal resolutions. The relationship between the location of the phenomenon and its other characteristics can be obtained using GIS analysis, which affect the prediction results.

A critical point in the reviewed articles was that some of them have used emission inventory instead of pollutant concentrations. The emission inventory contains a set of data that express the emission of different pollutants from different sources, which can significantly affect the results.

RECOMMENDATIONS

The main aim of this paper is to help researchers choose the appropriate method for predicting air pollution. Therefore, according to the contents and results are taken from the reveiwed articles, here are some tips:

- It is more helpful to use new hybrid models because of their excellent results.
- It is better to use the emission inventory to select the data because it's containing a structured collection of the emissions from a variety of air pollution sources.
- Evalaution of the effect of various spatial and temporal resolutions on the accuracy of the methods could be considered at future researches.

ACKNOWLEDGMENTS

The authors would like to acknowledge the I.R. of Iran Meteorological Organization, for their support and contribution to this study.

GRANT SUPPORT DETAILS

The present research did not receive any financial support.

CONFLICT OF INTEREST

The authors declare that there is not any conflict of interest regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/ or falsification, double publication and/or submission, and redundancy has been completely observed by the authors.

LIFE SCIENCE REPORTING

No life science threat was practiced in this research.

REFERENCES

Abdullah, S., Ismail, M. and Fong, S. (2017). Multiple linear regression (MLR) models for long term PM10 concentration forecasting during different monsoon seasons. *Journal of Sustainability Science and Management, 12*(1), 60-69. http://jbsd.umt.edu.my/wp-content/uploads/sites/51/2017/06/7-web.pdf

Aditya, C., Deshmukh, C. R., Nayana, D. and Vidyavastu, P. G. (2018). Detection and prediction of air pollution using machine learning models. *International Journal of Engineering Trends and Technology (IJETT), 59*(4). https://doi.org/10.14445/22315381/IJETT-V59P238

Ahmad, S. S., Aziz, N., Ejaz, M. and Ali, M. T. (2012). Integration of GIS and Artificial Neural Network for prediction of Ozone Concentration in Semi-rural areas of Rawalpindi and

Islamabad. *International Journal Of Computational Engineering Research, 2*. https://www.semanticscholar.org/paper/Integration-of-GIS-and-Artificial-Neural-Network-of-Ahmad-Aziz/f0c7357e7b7f0fa2f5a99474b390285877d3be3c

Alkasassbeh, M., Sheta, A. F., Faris, H. and Turabieh, H. (2013). Prediction of PM10 and TSP air pollution parameters using artificial neural network autoregressive, external input models: a case study in salt, Jordan. *Middle-East Journal of Scientific Research, 14*(7), 999-1009. http://www.idosi.org/.../20.pdf

An overview of the analyzing patterns toolset. From ESRI. Website at:https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/an-overview-of-the-analyzing-patterns-toolset.htm

Arslan, O. and Akyürek, Ö. (2018). Spatial modelling of air pollution from pm10 and so2 concentrations during winter season in marmara region (2013-2014). *International Journal of Environment and Geoinformatics, 5*(1), 1-16. https://doi.org/10.30897/ijegeo.412391

Arthur C. Stern, editors. The Effects of Air Pollution; Third edition. Academic Press: New York, NY, USA.1977.

Asadolah-Fardi, G. and Zangoi, H. (2017). PM10 AIR POLLUTION IN MASHAD CITY USING ARTIFICIAL NEURAL NETWORK AND MAKOV CHAIN MODEL. Journal of Applied researches in Geographical Sciences.; 17(47):39-59. https://www.sid.ir/en/journal/ViewPaper.aspx?ID=608455

Asghari, M. and Nematzadeh, H. (2016). Predicting air pollution in Tehran: Genetic algorithm and back propagation neural network. *Journal of AI and Data Mining, 4*(1), 49-54. https://doi.org/10.5829/IDOSI.JAIDM.2016.04.01.06

Atabi, F., Moattar, F., Mansouri, N., Alesheikh, A. and Mirzahosseini, S. (2013). Assessment of variations in benzene concentration produced from vehicles and gas stations in Tehran using GIS. *International Journal of Environmental Science and Technology, 10*(2), 283-294. https://doi.org/10.1007/s13762-012-0151-6

Ayturan, Y. A., Ayturan, Z. C. and Altun, H. O. (2018). Air pollution modelling with deep learning: a review. *International Journal of Environmental Pollution and Environmental Modelling, 1*(3), 58-62. https://api.semanticscholar.org/CorpusID:201904082

Bai, L., Wang, J., Ma, X. and Lu, H. (2018). Air pollution forecasts: An overview. *International journal of environmental research and public health, 15*(4), 780. https://doi.org/10.3390/ijerph15040780

Broomhead, D. and Lowe, D. (1988). Multivariable functional interpolation and adaptive networks, complex systems, vol. 2. https://sci2s.ugr.es/keel/pdf/algorithm/articulo/1988-Broomhead-CS.pdf

Cabaneros, S. M., Calautit, J. K. and Hughes, B. R. (2019). A review of artificial neural network models for ambient air pollution prediction. *Environmental Modelling & Software, 119*, 285-304. https://doi.org/10.1016/j.envsoft.2019.06.014

Cai, M., Yin, Y. and Xie, M. (2009). Prediction of hourly air pollutant concentrations near urban arterials using artificial neural network approach. *Transportation Research Part D: Transport and Environment, 14*(1), 32-41. https://doi.org/10.1016/j.trd.2008.10.004

Chaloulakou, A., Grivas, G. and Spyrellis, N. (2003). Neural network and multiple regression models for PM10 prediction in Athens: a comparative assessment. *Journal of the Air & Waste Management Association, 53*(10), 1183-1190. https://doi.org/10.1080/10473289.2003.10466276

Chelani, A. B., Gajghate, D. and Hasan, M. (2002). Prediction of ambient PM10 and toxic metals using artificial neural networks. *Journal of the Air & Waste Management Association, 52*(7), 805-810. https://doi.org/10.1080/10473289.2002.10470827

Copeland, M. (2016). What's the Difference between Artificial Intelligence, Machine Learning, and Deep Learning? https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligencemachine-learning-deep-learning-ai/, retrieval date: 24.04.2018.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning, 20*(3), 273-297. https://doi.org/10.1007/BF00994018

Cortina–Januchs, M. G., Quintanilla–Dominguez, J., Vega–Corona, A. and Andina, D. (2015). Development of a model for forecasting of PM10 concentrations in Salamanca, Mexico. *Atmospheric Pollution Research, 6*(4), 626-634. https://doi.org/10.5094/APR.2015.071

Díaz-Robles, L. A., Ortega, J. C., Fu, J. S., Reed, G. D., Chow, J. C., Watson, J. G. and Moncada-Herrera, J. A. (2008). A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile. *Atmospheric Environment, 42*(35), 8331-8340. https://doi.org/10.1016/j.atmosenv.2008.07.020

Djemai, S., Brahmi, B. and Bibi, M. O. (2016). A primal–dual method for SVM training. *Neurocomputing, 211*, 34-40.

Feng, Y., Zhang, W., Sun, D. and Zhang, L. (2011). Ozone concentration forecast method based on genetic algorithm optimized back propagation neural networks and support vector machine data classification. *Atmospheric Environment, 45*(11), 1979-1985. https://doi.org/10.1016/j.atmosenv.2011.01.022

Gardner, M. and Dorling, S. (2000). Statistical surface ozone models: an improved methodology to account for non-linear behaviour. *Atmospheric Environment, 34*(1), 21-34. https://doi.org/10.1016/S1352-2310(99)00359-3

Ghadi, M. E., Qaderi, F. and Babanezhad, E. (2019). Prediction of mortality resulted from NO 2 concentration in Tehran by Air Q+ software and artificial neural network. *International Journal of Environmental Science and Technology, 16*(3), 1351-1368. https://doi.org/10.1007/s13762-018-1818-4

Graupe, D. (2007). Principles of Artificial Neural Networks: World Scientific. https://doi.org/10.1142/8868

Grivas, G. and Chaloulakou, A. (2006). Artificial neural network models for prediction of PM10 hourly concentrations, in the Greater Area of Athens, Greece. *Atmospheric Environment, 40*(7), 1216-1229. https://doi.org/10.1016/j.atmosenv.2005.10.036

Haykin, S. (1998). Neural Networks: A Comprehensive Foundation (2nd Edition): Prentice Hall. https://dl.acm.org/doi/book/10.5555/1213811

Hiregoudar S. (2020, August 5). Ways to Evaluate Regression Models. Website at: https://towardsdatascience.com/ways-to-evaluate-regression-models-77a3ff45ba70

Hrust, L., Klaić, Z. B., Križan, J., Antonić, O. and Hercog, P. (2009). Neural network forecasting of air pollutants hourly concentrations using optimised temporal averages of meteorological variables and pollutant concentrations. *Atmospheric Environment, 43*(35), 5588-5596. https://doi.org/10.1016/j.atmosenv.2009.07.048

Jenkin, M. E. and Clemitshaw, K. C. (2000). Ozone and other secondary photochemical pollutants: chemical processes governing their formation in the planetary boundary layer. *Atmospheric Environment, 34*(16), 2499-2527. https://doi.org/10.1016/S1352-2310 (99)00478-1

Kröse, B., Krose, B., van der Smagt, P. and Smagt, P. (1993). An introduction to neural networks. http://citeseerx.ist.psu.edu/viewdoc/similar?doi=10.1.1.18.493&type=cc

Kumar, A. and Goyal, P. (2011). Forecasting of air quality in Delhi using principal component regression technique. *Atmospheric Pollution Research, 2*(4), 436-444. https://doi.org/10.5094/APR.2011.050

Leong, W., Kelani, R. and Ahmad, Z. (2020). Prediction of air pollution index (API) using support vector machine (SVM). *Journal of Environmental Chemical Engineering, 8*(3), 103208. https://doi.org/10.1016/j.jece.2019.103208

Liu, W., Li, X., Chen, Z., Zeng, G., León, T., Liang, J., Huang, G., Gao, Z., Jiao, S., He, X. and Lai, M. (2015). Land use regression models coupled with meteorology to model spatial and temporal variability of NO2 and PM10 in Changsha, China. *Atmospheric Environment, 116*, 272-280. https://doi.org/10.1016/j.atmosenv.2015.06.056

Lu, W.-Z., Wang, W.-J., Wang, X.-K., Yan, S.-H. and Lam, J. C. (2004). Potential assessment of a neural network model with PCA/RBF approach for forecasting pollutant trends in Mong Kok urban air, Hong Kong. *Environmental Research, 96*(1), 79-87. https://doi.org/10.1016/j.envres.2003.11.003

Lu, W., Wang, W., Leung, A. Y., Lo, S.-M., Yuen, R. K., Xu, Z. and Fan, H. (2002). *Air pollutant parameter forecasting using support vector machines.* Paper presented at the Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290). https://doi.org/10.1109/IJCNN.2002.1005545

McKendry, I. G. (2002). Evaluation of artificial neural networks for fine particulate pollution (PM10 and PM2. 5) forecasting. *Journal of the Air & Waste Management Association, 52*(9), 1096-1101. https://doi.org/10.1080/10473289.2002.10470836

Mishra, D. and Goyal, P. (2015). Development of artificial intelligence based NO2 forecasting models at Taj Mahal, Agra. *Atmospheric Pollution Research, 6*(1), 99-106. https://doi.org/10.5094/APR.2015.012

Molina-Gómez, N., Díaz-Arévalo, J. and López-Jiménez, P. A. (2020). Air quality and urban sustainable development: the application of machine learning tools. *International Journal of Environmental Science and Technology*, 1-18. https://doi.org/10.1007/s13762-020-02896-6

Nejadkoorki, F. and Baroutian, S. (2012). Forecasting extreme PM10 concentrations using artificial neural networks. https://www.sid.ir/en/Journal/ViewPaper.aspx?ID=370188

Nishikawa, Y. and Kannari, A. (2011). Atmospheric concentration of ammonia, nitrogen dioxide, nitric acid, and sulfur dioxide by passive method within Osaka prefecture and their emission inventory.

Water, Air, & Soil Pollution, 215(1-4), 229-237. https://doi.org/10.1007/s11270-010-0472-3
Núñez-Alonso, D., Pérez-Arribas, L. V., Manzoor, S. and Cáceres, J. O. (2019). Statistical tools for air pollution assessment: multivariate and spatial analysis studies in the Madrid region. *Journal of analytical methods in chemistry, 2019*. https://doi.org/10.1155/2019/9753927
Nunez, C. (2019). Air pollution, explained. From the nationalgeographic. Website at: https://www.nationalgeographic.com/environment/global-warming/pollution/
Ochando, L. C., Julián, C. I. F., Ochando, F. C. and Ramirez, C. F. (2015). *Airvlc: An application for real-time forecasting urban air pollution.* Paper presented at the MUD@ ICML. https://dl.acm.org/doi/10.5555/3045776.3045786
Papanastasiou, D., Melas, D. and Kioutsioukis, I. (2007). Development and assessment of neural network and multiple regression models in order to predict PM10 levels in a medium-sized Mediterranean city. *Water, air, and soil pollution, 182*(1), 325-334. https://doi.org/10.1007/s11270-007-9341-0
Park, J. and Sandberg, I. W. (1991). Universal approximation using radial-basis-function networks. *Neural computation, 3*(2), 246-257. https://doi.org/10.1162/neco.1991.3.2.246
Paschalidou, A. K., Karakitsios, S., Kleanthous, S. and Kassomenos, P. A. (2011). Forecasting hourly PM 10 concentration in Cyprus through artificial neural networks and multiple regression models: implications to local environmental management. *Environmental Science and Pollution Research, 18*(2), 316-327. https://doi.org/10.1007/s11356-010-0375-2
Piraino, F., Aina, R., Palin, L., Prato, N., Sgorbati, S., Santagostino, A. and Citterio, S. (2006). Air quality biomonitoring: Assessment of air pollution genotoxicity in the Province of Novara (North Italy) by using Trifolium repens L. and molecular markers. *Science of the Total Environment, 372*(1), 350-359. https://doi.org/10.1016/j.scitotenv.2006.09.009
Qiao, C., Gen-niu, C. and Liu, C. (2010). Application of support vector machine to atmospheric pollution prediction. *Computer Technology and Development, 20*, 250-253. http://en.cnki.com.cn/Article_en/CJFDTotal-WJFZ201001064.htm
Sánchez, A. S., Nieto, P. G., Fernández, P. R., del Coz Díaz, J. and Iglesias-Rodríguez, F. J. (2011). Application of an SVM-based regression model to the air quality study at local scale in the Avilés urban area (Spain). *Mathematical and Computer Modelling, 54*(5-6), 1453-1466. https://doi.org/10.1016/j.mcm.2011.04.017
Shad, R., Mesgari, M. S. and Shad, A. (2009). Predicting air pollution using fuzzy genetic linear membership kriging in GIS. *Computers, environment and urban systems, 33*(6), 472-481. https://doi.org/10.1016/j.compenvurbsys.2009.10.004
Stein, A., Riley, J. and Halberg, N. (2001) Issues of scale for environmental indicators. Agriculture, Ecosystems & Environment. 87(2):215-32. https://doi.org/10.1016/S0167-8809(01)00280-8
Taheri Shahraiyni, H. and Sodoudi, S. (2016). Statistical modeling approaches for PM10 prediction in urban areas; A review of 21st-century studies. *Atmosphere, 7*(2), 15. https://doi.org/10.3390/atmos7020015
Tan, J., Zhang, Y., Ma, W., Yu, Q., Wang, J. and Chen, L. (2015). Impact of spatial resolution on air quality simulation: A case study in a highly industrialized area in Shanghai, China. *Atmospheric Pollution Research, 6*(2), 322-333. https://doi.org/10.5094/APR.2015.036
Valari, M. and Menut, L. (2008). Does an increase in air quality models' resolution bring surface ozone concentrations closer to reality? *Journal of Atmospheric and Oceanic Technology, 25*(11), 1955-1968. https://doi.org/10.1175/2008JTECHA1123.1
Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE transactions on neural networks, 10*(5), 988-999. https://doi.org/10.1109/72.788640
Vitolo, C., Scutari, M., Ghalaieny, M., Tucker, A. and Russell, A. (2018). Modeling air pollution, climate, and health data using Bayesian Networks: A case study of the English regions. *Earth and Space Science, 5*(4), 76-88. https://doi.org/10.1002/2017EA000326 Wang, L.-S., Xu, Y.-T. and Zhao, L.-S. (2005). A kind of hybrid classification algorithm based on *rough set and support vector machine.* Paper presented at the 2005 international conference on machine learning and cybernetics. https://doi.org/10.1109/ICMLC.2005.1527214
What is GIS?. (2019). From GISLongue. Website at: https://www.gislounge.com/what-is-gis/
Yadav, V. and Nath, S. (2019). Novel hybrid model for daily prediction of PM 10 using principal component analysis and artificial neural network. *International Journal of Environmental Science and Technology, 16*(6), 2839-2848. https://doi.org/10.1007/s13762-018-1999-x
Zangooei, H., Delnavaz, M. and Asadollahfardi, G. (2016). Prediction of coagulation and flocculation

processes using ANN models and fuzzy regression. Water Science and Technology, 74(6), 1296-1311. https://doi.org/10.2166/wst.2016.315

Zickus, M., Greig, A. and Niranjan, M. (2002). Comparison of four machine learning methods for predicting PM 10 concentrations in Helsinki, Finland. *Water, Air and Soil Pollution: Focus, 2*(5-6), 717-729. https://doi.org/10.1023/A:1021321820639