A Study on Machine Learning-Based Approaches for PM$_{2.5}$ Prediction

V. Santhana Lakshmi and M. S. Vijaya

Abstract Clean air and water is the fundamental need of humans. But people are exposed to polluted air produced due to several reasons such as combustion of fossil fuels, industrial discharge, dust and smoke which generates aerosols. Aerosols are tiny droplets or solid particles such as dust and smoke that floats in the atmosphere. The size of the aerosol also called as particulate matter ranges from 0.001–10 μm which when inhaled by human affects the respiratory organs. Air pollution affects the health of 9% of the people every year. It is observed as the most important risk factor that affects human health. There is a need for an efficient mechanism to forecast the quality of air to save the life of the people. Statistical methods and numerical model methods are largely employed for the predicting the value of PM$_{2.5}$. Machine learning is an application of artificial intelligence that gives a system capability to learn automatically from the data, and hence, it can be applied for the successful prediction of air quality. In this paper, various machine learning methods available to predict the particulate matter 2.5 from time series data are discussed.

Keywords Aerosols · Particulate matter · Machine learning (ML) · Forecast · Predicting · Artificial intelligence (AI) · Time series data

1 Introduction

People when exposed to polluted air for a long time are affected with lung disease. This is due to the presence of harmful substances in the air. This section explains the role of polluted air in affecting human morbidity and mortality. Dust also called as Aeolian dust present in the dry region is injected into the atmosphere by high velocity wind. This forms aerosol in troposphere. It produces radiative forcing and

V. Santhana Lakshmi (✉) · M. S. Vijaya
Department of Computer Science (PG), PSGR Krishnammal College for Women, Coimbatore, India

M. S. Vijaya
e-mail: msvijaya@psgrkcw.ac.in

is responsible for the change in temperature, ocean cooling and alters the rainfall amount. Aerosols are small particles that float in the air. The open ocean is a significant source of natural aerosols. It produces 1015–1016 g of sea-salt aerosols annually. Sea-salt aerosols, along with wind-blown mineral dust, forms natural tropospheric aerosols. The radiative processes in the clean atmosphere, namely reflection, transmission and absorption, are affected by the physical and chemical properties of background aerosols. The primary environmental risk factor that poses the greatest threat to human life is air pollution. A person when inhales the polluted air for a long time is affected by the disease such as asthma, bronchitis, ventricular hypertrophy, Parkinson's diseases and Alzheimer [1]. Forecasting the air quality is essential so that people receive the information earlier about the increase in pollution level, thereby saving the life of the people. Time series data that are used for identifying $PM_{2.5}$ value includes meteorological data, geographical data, traffic data and satellite data in the form of images. A good choice of data and methodology plays a main role in the accurate prediction of air quality.

2 Air Pollution and Air Quality Prediction

Air gets contaminated when some solid particles and gases suspends in the air. This section introduces various factors that cause pollution and harmful substances in air. Natural sources of air pollution include dust from earth surface, sea salt, volcano eruptions and forest fire. Man-made sources of air pollution include industry emission, transportation emission and agriculture [2].

2.1 Natural Causes of Air Pollution

Volcano during eruption will emit sulphur dioxide. It produces respirable acid after reacting with water vapour and sunlight. Finally, it results in vog which is visible haze. The lava enters into the ocean to make sea water boil. As a result, it creates thick clouds of "laze". These clouds are filled with hydrochloric acid. The level of air pollution depends on the speed and direction of wind, vog and ash. This affects the respiratory organs of humans [3].

Forest fire will affect the meteorology of the surrounding area. It has severe impact on the quality of the air. Toxic gases are emitted while burning of biomass in the forest. These gases move several miles from their point of origin. The air quality in the surrounding areas also suffers as a result of this. During the fire period in Uttarakhand, the concentrations of contaminants such as nitrogen oxide, nitrogen dioxide, carbon monoxide, PM_{10} and $PM_{2.5}$ increased tremendously [4].

2.2 Man Made Pollution

The water bodies are polluted due to unauthorized release of industrial wastewater. The waste water includes toxic chemicals which is mainly responsible for health hazards in living beings. Untreated waste water contributes to the existence of toxins in water sources such as lakes, rivers and groundwater detrimental to the health of plants, livestock and human beings. Such toxins are mostly responsible for polluting sacred rivers such as the Ganga, Yamuna and others. The water was contaminated, and people could not use it for drinking, bathing, and other purposes [5].

Growth of motor vehicles and rapid urbanization results in the deterioration of air quality. As per [6] in 2020, over 21.5 million vehicles were sold domestically. Exhaust from vehicles such as hydrocarbons, sulphur dioxide, lead/benzene, nitrogen dioxide, carbon monoxide and particulate matter are the major sources of outdoor air pollution worldwide. Public were affected by respiratory diseases and cardio-vascular diseases in India. As per the report of India Times, the vehicle tailpipe emission were linked to about 361,000 premature deaths from PM$_{2.5}$ and ozone in 2010 and about 38,500 in 2015.

2.3 Air Pollutants

Air pollution occurs when harmful substances are suspended in the atmosphere. Harmful substances include lead, carbon monoxide, ozone, nitrogen dioxide and sulphur dioxide.

Carbon monoxide is an odourless and colourless gas that is present in the air. It is toxic for humans to ingest more carbon monoxide. The normal carbon monoxide concentration that does not affect humans is about 0.2 parts per million (ppm).When the level of carbon monoxide consumed by a person increases, it lowers the amount of oxygen in red blood cells that is transported by haemoglobin throughout the body. As a result, essential organs such as the heart, brain and nervous tissues are deprived of sufficient oxygen to function. Bush fires and volcanoes are the natural sources of carbon monoxide.

Lead is a heavy metal present in the air. Humans are exposed to lead through inhalation process. Once enters within the body, it flows through the bloodstream until it reaches the bones. Lead has an impact on the nervous system, kidneys, digestive system and reproductive system, depending on the level of exposure. Lead poisoning reduces the blood's ability to contain oxygen. Motor vehicle waste and certain manufacturing processes are the primary sources of lead and carbon monoxide.

Nitrogen dioxide has an unpleasant odour. Plants contain a small amount of nitrogen dioxide. Sometimes, it is formed naturally in the atmosphere by lightning. Nitrogen dioxide has a larger role in the production of photo-chemical smog, which has serious health consequences. Breathing nitrogen dioxide will have significant

impact on people who are already affected with respiratory diseases. Older people and children with asthma and heart diseases are most at risk.

Ozone gas contains three atoms of oxygen (O_3). There are two types of ozone. One that is created in the upper atmosphere shields us from ultraviolet rays, and another type is formed in the ground level. Ground-level ozone is one of the harmful air pollutants. Ozone is created by chemical reactions between nitrogen oxides (NOx) and volatile organic compounds (VOC). The pollutants emitted by the cars, industrial boilers, power plants, refineries, chemical plants and other sources react in the presence of sunlight [7]. Breathing ozone can cause coughing, throat irritation, chest pain and airway inflammation among other things. Patients with bronchitis, emphysema and asthma will get severe infection when exposed to it for a long time.

Particulate matter is abbreviated as PM. Particle contamination is another name for it. It represents a combination of solid particles contained in the air, such as ashes, ash and soot, as well as liquid droplets. Some particles are visible to the naked eye and some can be seen only through electron microscope. Particle pollution includes **PM_{10}** and **$PM_{2.5}$**. Both are inhalable particles with diameters of 10 and 2.5 micrometres, respectively. Exposure to such particles affects the lungs, heart and leads to many respiratory problems.

Sulphur dioxide is a noxious and translucent substance with a pungent odour. Sulphuric acid, sulphurous acid and sulphate particles are formed when it interacts with other chemicals. Major sources of sulphuric acid are humans and industrial waste. Burning of fossil fuels also creates sulphuric acid. It creates irritation in nose, throat and causes coughing. It gives tight feeling around the chest. Asthma patients are at greater risk of being severely affected after being exposed to sulphuric acid.

When carbon monoxide, lead, nitrogen dioxide, ammonia, particulate matter and sulphur dioxide are mixed with air, it becomes polluted. Identifying the amount of these pollutants in air helps to solve many research problems. Correlation between amount of these pollutants in air and human health can be identified. Lot of researches are going on to identify the contribution of these factors in affecting the environment. It is possible to predict the quality of vegetation by identifying the amount of these pollutants in air. Many countries are trying to identify the correlation between vehicle emission and deterioration of air quality. Lot of computational methods are available for identifying these pollutants.

3   Computational Methods for Air Quality Prediction

Artificial intelligence, machine learning, deep learning, big data analytics, cloud and Internet of things (IoT) are the technologies causing revolutions in various industries. Applications developed by combining these technologies perform efficiently well and execute without the human intervention. These technologies help to analyse large datasets of complex data types with less computational difficulty. Apart from traditional statistical methods, machine learning and deep learning algorithms are

effective in predicting the airborne pollutants. This section describes various methods that are used for predicting air pollutant concentration.

3.1 Statistical Methods

The statistical methods use historical data to forecast the future value. For predicting air quality, statistical methods such as simple moving average, exponential smoothing and auto regressive moving average can be used. Simple moving average method is used to identify the long-term trend whether upward or downward in the time series data. It calculates the average of finite number of values from the historical data to forecast the new value. Air quality can be predicted using historical meteorological and geographical data.

Exponential smoothing predicts the future value by calculating weighed average of past values. Exponential smoothing assigns decreasing weights for older observations and more weights for recent observation. It smoothens the time series data. There are three types of exponential smoothing methods. It includes simple exponential smoothing, double exponential smoothing and triple exponential smoothing. Simple exponential smoothing was best suited for univariate data without trend. It uses smoothing parameter alpha. The speed at which the effect of previous time steps' observations decays exponentially is controlled by this parameter. The value of this parameter lies between 0 and 1. As the name implies, double exponential smoothing uses two smoothing parameters such as trend component and level component at each period. Triple exponential smoothing method also called as Holt Winters method uses third parameter when the data has seasonality in it.

Autoregressive moving average model also uses historical data to forecast the future value, but it expects the data to be stationary. If the time series data is not stationary, differencing is performed to make it stationary. In order to fit the ARIMA model three important values p, d, q has to be identified. p and q represent order of auto regression and order of moving average and d is the order of differencing. ARIMA and exponential smoothing are the most popular statistical methods used for forecasting.

3.2 Machine Learning

Machine learning is a form of artificial intelligence that builds an intelligent model based on historical data to predict the target. Machine learning algorithms are used to train models which can perform decision making without being explicitly programmed. Regression, auto regression and support vector machine are machine learning techniques that can be used to forecast pollution concentration.

Forecasting the quality of air is very much essential since it has more impact on the human health. Among all the pollutants, particulate matter plays main role in

increasing the mortality rate. Lot of methods are available to predict the $PM_{2.5}$ level in air. The simplest method is to build regression model.

Regression is a statistical technique used for identifying and analysing the relationship between dependent variables and independent variables. It is a predictive modelling technique used to predict the real value. Linear regression and logistic regression are the two types of regression algorithms. In simple linear regression, the relationship between single dependent variable and independent variable is analysed. Multiple linear regression is a variation of simple linear regression in which the relationship between single dependent variable and multiple independent variables are analysed. Regression algorithms are used to forecast the $PM_{2.5}$ value based on the multiple independent variables such as wind speed, temperature, precipitation, humidity and rainfall.

A time series model called an auto regression (AR) uses observations from previous time steps as input to the value at the next step. It is used to make predictions because there is a connection between the values in a time series and the values that come before and after them. The behaviour is modelled using historical evidence, thus the term autoregressive.

The support vector machine (SVM) is a new kind of classification algorithm that uses statistical learning theory to operate. Support vector regression (SVR) is used to investigate the concentration variation of air pollutants. The basic principle behind support vector regression is to use a non-linear mapping unknown function to map the original data x into a feature space F with high dimensionality, and then use linear regression in this space. SVR permits the user to specify the minimum amount of error that is accepted while prediction. This is one of the advantages of SVR, and the computational complexity does not depend upon the dimensionality of the dataset. So, when the time series data of previous years are used for training the model, the algorithm considers these factors as independent variables and tries to fit the line to predict the dependent variable which is air quality index.

3.3   Deep Learning

Deep learning is a subset of machine learning that simulates human problem-solving and decision-making. Deep learning algorithms do not need human supervision for extracting features. They automatically extract essential features from the dataset. Deep learning architectures such as multi-layer perceptron (MLP), convolutional neural network (CNN), long short-term memory (LSTM) a type of recurrent neural network (RNN) is used for time series forecasting.

Multi-layer perceptron is a feed-forward neural network with multiple perceptrons and many layers. Perceptron is a linear classifier. The algorithm specially deals with binary classification. It is widely used for speech recognition, image recognition etc. Multilayer perceptron neural network works efficiently in classifying whether the air quality index is dangerous to human health or safe.

One of the deep learning classification algorithms is the convolutional neural network which is specially designed to analyse images. The image is fed into convolutional layer where the feature extraction is performed. The processes such as filtering, padding and pooling are performed, and the output is fed into fully connected layer after flattening it. Face recognition is the most popular application of CNN. CNN helps to predict PM$_{2.5}$ while using remote sensing images fetched from the satellite.

Long short-term memory (LSTM) is a form of artificial recurrent neural network that is commonly used to classify long-term data dependencies. This algorithm is well suited for handling time series data. LSTM architecture has more loops in it which makes the information to persist. It can automatically identify temporal dependencies and structures such as trend, seasonal and cyclic.

A cell, an input gate, an output gate and a forget gate make up a typical LSTM unit. The three gates control the flow of information into and out of the cell, and the cell remembers values across arbitrary time periods.

Because there might be delays of uncertain duration between key events, LSTM networks are well-suited to categorising, analysing and generating predictions based on time series data.

4 Literature Survey

Yves Rybarczyk and Rasa Zalakeviciute [8] created multiple regression model for predicting the concentration of PM$_{2.5}$. Along with meteorological data, traffic data was used to identify its role in polluting the air. They identified that the traffic was positively correlated with PM2.5 value. The higher the traffic, the higher the PM$_{2.5}$ value. Traffic data collected from the Google Map was used for analysis. Two images collected from same place were used, one with traffic and one without. The difference was calculated in terms of pixels for red, green and orange. Those values were given as input to predict the PM$_{2.5}$ concentration. Rather than creating one model to identify the PM$_{2.5}$ concentration, different models were created for a day because the parameters (traffic) were not same at all the time in a day. The accuracy of the model was evaluated using correlation coefficient and root mean square value.

Aditya C R, Chandana and et al. predicted particulate matter value using auto regression. The value was predicted using previous readings of PM2.5. The dataset had two attributes date and previous PM$_{2.5}$ concentration. Once the model was developed, it obtained the knowledge to predict PM$_{2.5}$ value for the given date. Mean squared error was used to evaluate the model [9].

Wei-Zhen Lu, A.Y.T Leung et al. used support vector machine and radial basis function for predicting the pollution concentration. Six pollutant values collected in hourly basis for 1 complete year were obtained from monitoring station. Radial basis function (RBF) method and SVM method was used to predict the concentration for a day, week and month. The performance of the model was evaluated using the statistical metric, mean absolute error (MAE) [10].

Hwee San Lim, Mohd Zubir, Mat Jafri et al. used remote sensing images and regression algorithm to identify the pollution concentration. Remote sensing is a method of gathering data about the earth using instruments without coming into close contact with it. Sensors measure the energy reflected from earth. Sensors are mounted on the satellites. Geometric correction was performed to avoid the distortion in raw digital satellite images. Optical depth and reflectance value were calculated. Reflectance value was the sum of the surface reflectance and atmospheric reflectance. AOD is a quantitative represents the amount of depletion that a beam of solar radiation undergoes as it travels through the atmosphere. Regression algorithm was used for determining the pollution concentration [11].

Mehdi Zamani Joharestani, Chunxiang Cao and et al. included more attributes to improve the accuracy of the model. In addition to meteorological data, satellite data, ground-measured $PM_{2.5}$, geographical data was used for the purpose of the modelling. Satellite remote sensing, in terms of aerosol optical depth (AOD) provides information with the required, spatial coverage for the areas under examination. The Moderate Resolution Imaging Spectro-radiometer (MODIS) is the AOD product which is installed in both Aqua and Terra satellites. In each and every city, air pollution monitoring sites record the concentration of PM_{10}, $PM_{2.5}$, CO, O_3, NO_2 and SO_2. Geographic data used includes altitude, latitude and longitude details of the sites. Meteorological time series data such as maximum and minimum air temperature (T-max, T-min), relative humidity (RH), daily precipitation, visibility, wind speed (Windsp), sustained wind speed (ST-windsp), air pressure and dew point were combined with satellite data. The time series data was applied on algorithms such as random forest, XGBoost, deep neural network LSTM and CNN to predict the $PM_{2.5}$ concentration [12].

Chitrini Mozumder and K. Venkata Reddy and Deva Pratap used vegetation index to identify air pollutants. It was found that vegetation was negatively correlated with the air pollution index. Air Pollution Index (API) can be identified using vegetation indices as well as some other image extracted parameters. Remotely sensed data was available as image in IRS P6 (Resourcesat 1) Linear Imaging Self-scanning Sensor (LISS) IV. The images were captured at a resolution of 5.8m using three spectral bands in the visible and NIR (near-infrared) regions: green, red and blue. The air pollution parameter considered was API or AQI (Air Quality Index) The image parameters considered were Normalized Difference Vegetation Index (NDVI), Vegetation Index (VI), Transformed Vegetation Index (TVI) and Urbanization Index (UI). Using all the parameters, multivariate regression model was developed with API as dependent variable and the features extracted from IRS and Landsat images as independent variables. Root mean squared error (RMSE) was used to assess the performance of the model [13]. The research works carried out to predict the air pollutant are summarized in the Table 1 and the results are summarized correspondingly in Table 2.

Table 1 Summary of techniques and algorithms used for prediction

S No	Author name	Title of the paper	Dataset	Objective	Technique	Algorithm used for model construction
1	Yves Rybarczyk and Rasa Zalakeviciute	Regression models to predict air pollution from affordable data collections	Traffic data meteorological data	To predict the PM2.5 concentration	Regression analysis	Linear regression model and multiple regression model
2	Aditya C R, Chandana R Deshmukh, Nayana D K, Praveen Gandhi Vidyavastu,	Detection and prediction of air pollution using machine learning models	Date and PM 2.5 level	To predict the PM2.5 concentration	Time series analysis	Auto regressive model
3	Wei-Zhen Lu, A.Y.T Leung	Air pollutant parameter forecasting using support vector machines	Date and six pollutants such as sulphur dioxide nitrogen oxides nitric oxide, nitrogen dioxide carbon monoxide and respirable suspend particles (RSP)	To predict the RSP concentration	Regression analysis	Support vector regression
4	Hwee San Lim, Mohd Zubir Mat Jafri Khiruddin Abdullah Wong Chow Jeng,	Air pollution determination using remote sensing technique	Digital satellite images	To find aerosol optical depth (AOD) and PM$_{10}$ concentration	Regression analysis	New algorithm proposed based on regression
5	Mehdi Zamani Joharestani, Chunxiang Cao, Xiliang Ni, Barjeece Bashir and Somayeh Talebiesfandarani, atmosphere, July	PM2.5 prediction based on random forest, XGBoost, and deep learning using multisource remote sensing data	Meteorological data, geographical data, satellite images, ground measured values	To predict PM$_{2.5}$ concentration	Ensemble learning and deep learning	Random forest, XG Boost, LSTM, CNN

Table 1 (continued)

S No	Author name	Title of the paper	Dataset	Objective	Technique	Algorithm used for model construction
6	Chitrini Mozumder, K. Venkata Reddy, Deva Pratap	Air pollution modelling from remotely sensed data using regression techniques	Normalized difference vegetation index (NDVI), vegetation index (VI), transformed vegetation index (TVI) and urbanization index (UI) from satellite images	To predict air quality index of all pollutants and design as image	Regression	Multivariate regression model

Table 2 Summary of results and findings

S No	Algorithm used for model construction	Evaluation measure	Result	Findings
1	Linear regression model and multiple regression model	Correlation coefficient Root mean square value	R = 0.8, RMSE = 5.3	Traffic data plays a main role in predicting the PM$_{2.5}$ concentration
2	Logistic regression auto regressive model	Mean accuracy mean squared error	Mean accuracy = 99% SE = 27	Auto regression suits best for predicting PM$_{2.5}$ value
3	Support vector regression	Mean absolute error	MAE = 0.0178(day) MAE = 0.0204(week)	Compared to feed forward neural network SVR suits better for prediction since the parameters to be set by the user is only 2
4	New algorithm proposed based on regression	correlation coefficient Root mean square value	R = 0.8 RMSE = 16 µg/m^3	Landsat TM satellite data provides very useful information for estimating and mapping air pollution
5	Random forest, XG Boost, LSTM, CNN	Mean absolute error Root mean square value	MAE = 10.20 µg/m^3, RMSE = 14 µg/m^3	Satellite-derived AODs have no significant effect on forecasts. Images with high spatial resolution have to be used
6	Multivariate regression model	Correlation coefficient Root mean squared error	R2 = 0.79 RMSE = 7.77	Landsat is better than IRS in providing air pollution information of the study area

5 Observations and Discussion

Forecasting the air quality is considered as important because of the serious impact the airborne pollutants creates on human health. It is essential to build a predictive model to take some precautionary measures. Air pollutants are identified using techniques such as machine learning and deep learning. Machine learning algorithms such as regression, auto regression and SVM are broadly used. The observations made from the study are summarized below.

Even though the traditional methods such as simple moving average, ARIMA and exponential smoothing were available to identify air quality, machine learning methods and deep learning algorithm do the prediction with ease and accuracy. When considering air quality prediction, the algorithm used and the features selected decide

the effectiveness of the model. Among all the pollutants, $PM_{2.5}$ plays an important role in affecting the health of the human.

The concentration of $PM_{2.5}$ can be identified using machine learning algorithms such as regression, support vector machine, neural network, time series algorithm and deep learning. It is also found that developing one model for a day will not be accurate since the concentration of pollutants varies at different times in a day. Different models should be created to forecast the concentration in the peak time and non-peak time. To improve the accuracy of prediction, traffic information and vegetation index can also be included.

The features such as meteorological information and geographical information are used to identify the $PM_{2.5}$ value. In addition to these parameters, remote sensing images can be included in which aerosol optical depth helps to identify the pollution level. It is also identified that the concentration of pollutants such as sulphur dioxide, carbon monoxide, ozone and nitrogen dioxide contributes more for the prediction. From the observations, it is concluded that using only meteorological factors and geographical information is not sufficient to identify the concentration of $PM_{2.5}$. It is necessary to include the features such as traffic, vegetation index and remote sensing images.

6 Conclusion

In this paper, need for air quality prediction and the research directions in building efficient prediction models are reviewed. The different airborne pollutants and sources of their emission is presented, and their importance is reviewed. The significance of machine learning methods in air quality prediction is studied. Various computational methods involved in predicting the quality of the air are discussed and their results are analysed. Research on air quality prediction carried out by implementing machine learning and deep learning algorithms on time series meteorological data, geographical data, remote sensing images, traffic data and vegetation data is deliberated. Various observations made from the analysis are highlighted. It is necessary to design and develop an efficient air quality prediction model by including more features that contribute in identifying the air contamination. Further study can be carried out by identifying the impact of air pollution in vegetation and health.

References

1. A. Ghorani-Azam, B. Riahi-Zanjani, M. Balali-Mood, Effects of air pollution on human health and practical measures for prevention in Iran. J. Res. Med. Sci. (2016)
2. C. Pénard-Morand, I. Annesi-Maesano, Air pollution: from sources of emissions to health effects. Breathe **1**(2), 108–119 (2004)
3. http://magde.info/main/texts/PhD/Chap2.PDF
4. https://www.thoracic.org/patients/patient-resources/resources/volcanic-eruptions.pdf

5. S. Singh, Implications of forest fires on air quality—a perspective. Bull. Environ. Sci. Res. **5**(3–4), 1–4, (2016)
6. R.K. Lohchab, J. Kumar Saini, Industrial Pollution Management (IAHRW Publications, 2018)
7. A.A. Desai, A review on assessment of air pollution due to vehicular emission in traffic area. Int. J. Curr. Eng. Technol. **8**(2) (March/April 2018)
8. https://www.epa.gov/ground-level-ozone-pollution/ground-level-ozone-basics
9. Y. Rybarczyk, R. Zalakeviciute, Regression Models to Predict Air Pollution, (IntechOpen, 2018)
10. C.R. Aditya, C.R. Deshmukh, D.K. Nayana, P.G. Vidyavastu, Detection and prediction of air pollution using machine learning models **59**(4) (2018)
11. W.-Z. Lu, A.Y.T Leung, et al, Air pollutant parameter forecasting using support vector machines. Int. Joint Conf. Neural Network (2002)
12. H.S. Lim, M. Zubir, M. Jafri, K. Abdullah, W.C. Jeng, Air pollution determination using remote sensing technique. Adv. Geosci. Remote Sens.
13. M.Z. Joharestani, C. Cao, X. Ni, B. Bashir, S. Talebiesfandarani, PM2.5 prediction based on random forest, XGBoost,and deep learning using multisource remote sensing data, Atmosphere (2019)