Int. J. Nonlinear Anal. Appl. 12 (2021) No. 2, 1497-1511
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/IJNAA.2021.5278

Improving air pollution detection accuracy and Status monitoring based on supervised learning systems and Internet of Things

D. Saravanan[a*], K. Santhosh Kumar[b]

[a] *Department of CSE, IFET College of Engineering, Villupuram, India.*

[b] *Department of IT, Annamalai University, Chidambaram, India.*

(Communicated by Madjid Eshaghi Gordji)

Abstract

In recent decades air pollution and its associated health risks are in growing numbers. Detecting air pollution in the environment and alarming the people may accomplish various advantages among health monitoring, telemedicine, and industrial sectors. A novel method of detecting air pollution using supervised learning models and an alert system using IoT is proposed. The main aim of the research is manifold: a) Air pollution data is preprocessed using the feature scaling method, b) The feature selection and feature extraction process done followed by performing a Recurrent Neural Network and c) The predicted data is stored in the cloud server, and it provides the end-users with an alert when the threshold pollution index exceeds. The proposed RNN reports enhanced performance when tested against traditional machine learning models such as Convolutional Neural Networks (CNN), Deep Neural Networks(DNN) and Artificial Neural Networks(ANN) for parameters such as accuracy, specificity, and sensitivity.

Keywords: Internet of Things, Convolutional Neural Networks (CNN), Deep Neural Networks (DNN), Artificial Neural Networks (ANN), Recurrent Neural Network

1. Introduction

Growth in science and technology has enabled development of cities and also sub- urban and rural areas. However, these developments go hand in hand with an increase in environmental issues like land, water, air and sound pollution. Air pollution has proven to have a direct impact on humans and other living beings. The particulate matters in the air are known to cause the pollution. Presence of these matters in the air that we breathe draws special attention for many researches

*Corresponding author: D. Saravanan
Email addresses: saranmds@gmail.com (D. Saravanan[a*]), santhosh09539@gmail.com (K. Santhosh Kumar[b])

Received: April 2021 *Accepted:* July 2021

works [14, 10, 24]. The most commonly known causes of air pollution are factories and industrial exhaust waste, household heating, natural disasters like forest fires while burning of fossil fuels has largely contributed to it. After the formation of Clean Air Act program, the United States (US) has taken initiative to assess the Air status. The study of air is conducted from the past thirty years. Implementation of this program has certainly improved air status in the past decades however complete and absolute results are not yet received. [5]. US has reported approximately 200,000 premature deaths in one year caused by combustion emissions that contain particulate matter 2.5 (PM2.5). US also reports 10,000 deaths in a year reason being, changes in the ozone surface. A study conducted by the American Lung Association stated that illnesses related to air pollution result in approximate expense of 37 billion dollars every year in United States alone and California ads up to $15 billion [13]. With this increase in the air pollution and its harm caused to the environment and living beings, several researches have been conducted among which air pollution forecast is of great significance. Now that increase in air pollution and its related issues are confirmed, it has given importance to accurate assessment of air pollution forecasting. This is an important step for managing the air status and prevention from pollution hexes. The atmosphere is polluted when harmful and increased substances like gases, particulates, and biological molecules are let into the air. Air is said to be polluted whenever toxic substances especially in a large quantity is introduced in the air. These include gases, particulates, and biological molecules. Such mixtures in the air have definite results, causing diseases and harm to living organisms and impairing the growth and yield of crops. Air pollutants consists of solid particles, liquid droplets, and gases, these are categorized as: Primary pollutants, these pollutants are released directly to the atmosphere from the origin. The origins can vary from natural disasters like sand storm, forest fires or human made activities like industrial emission, vehicles. Sulfur dioxide (SO2), particulate matter (PM), nitrogen dioxide (NOx), and carbon monoxide (CO) are primary contaminants highly responsible for pollution. However, other category of pollutants is created in the air, by chemical or physical reactions among the primary pollutants. These are called secondary pollutants. Some of these contaminants are photochemical oxidants and secondary particulate matter. Pollutants that highly causes pollution are called criteria pollutants, these are related to health hazards, like CO, SO2, lead, ground-level ozone (O3), NO2, and PM. Air status is estimated by US Environmental Protection Agency (EPA) Recent studies have shown that the co-relation between exposure of these pollutants and health issues. These include unable to respond to increased oxygen supply during exercises, respiratory tract infections in healthy individuals and exacerbating manifestations for individuals with asthma, respiration casualties specially among young and the old aged groups [33]. Several environmental agencies like EPA, EU have come up with air status guidelines and standards for permissible levels of pollutants. Air status index (AQI) is measured on a daily basis, and indicates how polluted the air is and it's associated health problems especially among children and elderly people. It mainly throws light on the symptoms that will be experienced shortly after exposure to polluted air. This is estimated by calculating the number of individually registered AQI for the criteria pollutants stated above. Generating prediction technique, according to individual pollutant levels can anticipate air status on an hour-to-hour basis. This feature also enables AQI to be flexible and also benefit the population's health. The system will be enabled to produce warning on the basis of air status. This system is beneficial because it will assist the health care systems by producing health vigilances in situations when pollution levels exceed the threshold values. It may also merge with the currently functional emission control programs, like letting the environmental regulators to opt for "on-demand" decreased releases, systematic plan and emergency measures [6]. The main sections of this study are summarized as:

2. Problem Statement

The capital city of our country has been facing air pollution since several years. The air status variables have been higher than the threshold values thus hampering the health. Even though the government has implemented rules against air pollution the favorable outcome has not been obtained yet. Active western disturbance causes presence of moisture in high index over the north Indian zone that includes Delhi. Geographical conditions like rains caused by western disturbance leads to poor range of National Air Status Index [32] however rains will alter the factors used as a means to forecast weather conditions by increasing the AQI. This study works towards developing Artificial intelligence system for AQI forecast. Figure 1 illustrates AQI data prediction design. At stage one, the data is imported. Recurrent Neural Network is forecasting the future information. The forecasted AQI value is transferred to the SVM in the second step where the precision of forecasted data and real-time data are assessed.

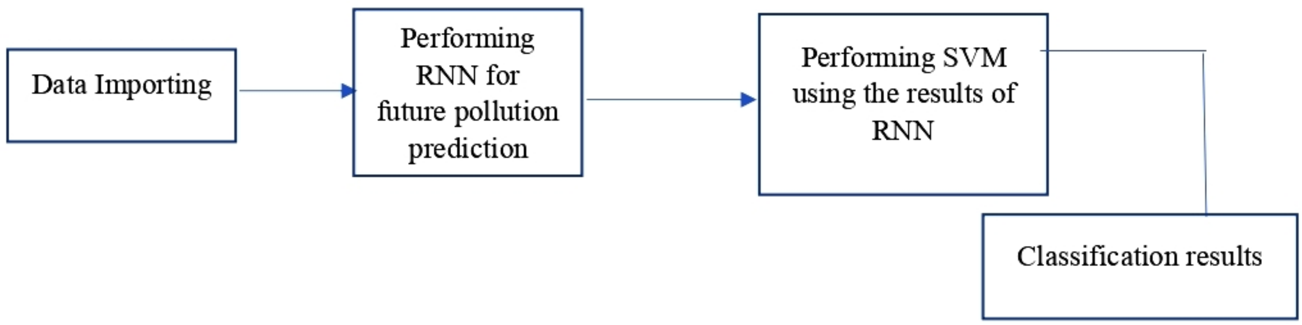

Figure 1: AQI Prediction

The two methods used for forecasting the air status are: deterministic and statistical approaches. Meteorological emissions and chemical systems are used with the help of techniques to simulate diffusion and expulsion pollutants process by the use of data received from monitoring areas, [2]. Saide et.al. used CMAQ and WRF-Chem for forecasting air status in urban areas, [28]. However, it gives lower prediction accuracy because of alterations in pollutant data, varying atmospheric situations and inadequate abstract data. Statistical methods forecast the air status with the help of statistical designing. This is a data-driven technique. Auto regression [3] and multiple linear regression (MLR) [20] techniques are used. The reason of obtaining limited results is their incompetence to design non-linear data, making air quality forecast inaccurate [11]. Other substitute is support vector regression design and artificial neural networks (ANN). Recent research by Prybutol et al.proved that non-linear designs such as ANN generate explicit results compared to linear systems due to the presence of clearer nonlinear pattern in air status data [26]. Several studies were performed by combining both the systems which resulted in better prediction results rather than using a single design, [8]. Shi et al. put forward RNN and LSTM for assessing rain in the coming two hours, which was standardized to aspatial-temporal issue, [31]. Air status estimation resembles climate gauge, but air status forecast is challenging due to the following reasons: 1) Time taken for air status conjecture is greater compared to climate estimate, 2) additional influential factors must be paid attention to, in air status gauge, like increase in pollutants and their association with climatic parameters. Ong et al. suggested using RNN to forseePM2.5 with natural sensor data, that improved outcome's precision [23]. Kurt and Oktay conducted a study forecasting air pollution with neural methods portrayed strategies importance and accomplishment [16]. Liang et al. forwarded a data that contained an estimation of PM2.5 which was assessed in Beijing [18]. Eventually, Liang et al. presented data to divide poison variable into five different urban places of China [19]. The data used here is stated in points, hence cannot be demonstrated in spatially express way.

3. Dataset

The dataset used here is AirNet 21. The data is obtained from China National Environmental Monitoring Center (CNEMC) having 1498 stations, consists of six variables. Meteorological data is obtained from Global Forecasting System (GFS). Every variable comprises of 1038240 values in a single point. GFS consists of a 3-D matrix that is formed by NOAA 6 hourly. The 3-D matrix is incorporated in the 2-D data to create a four-D matrix in order to produce AirNet data. The four-D database comprises data regarding latitude, longitude, timesteps, features. A single time frame consists of 6 GFS features and 7 air status indicators. The complete dimensions of data are (132, 228, 7072, 13).:

1. Temperature (K)
2. Relative Humidity (%)
3. U-wind component (m/s)
4. V-wind component (m/s)
5. Precipitation Rate ($kg/m^3/s$)
6. Total cloud cover (%)

The pollutants considered for prediction are:
PM10, PM2.5, NO2, CO, O3, SO2, AQI
The plot shown in figure 1 describes that the features required for forecasting are inter-related and hence these can be kept in consideration in order to conduct training of the design.

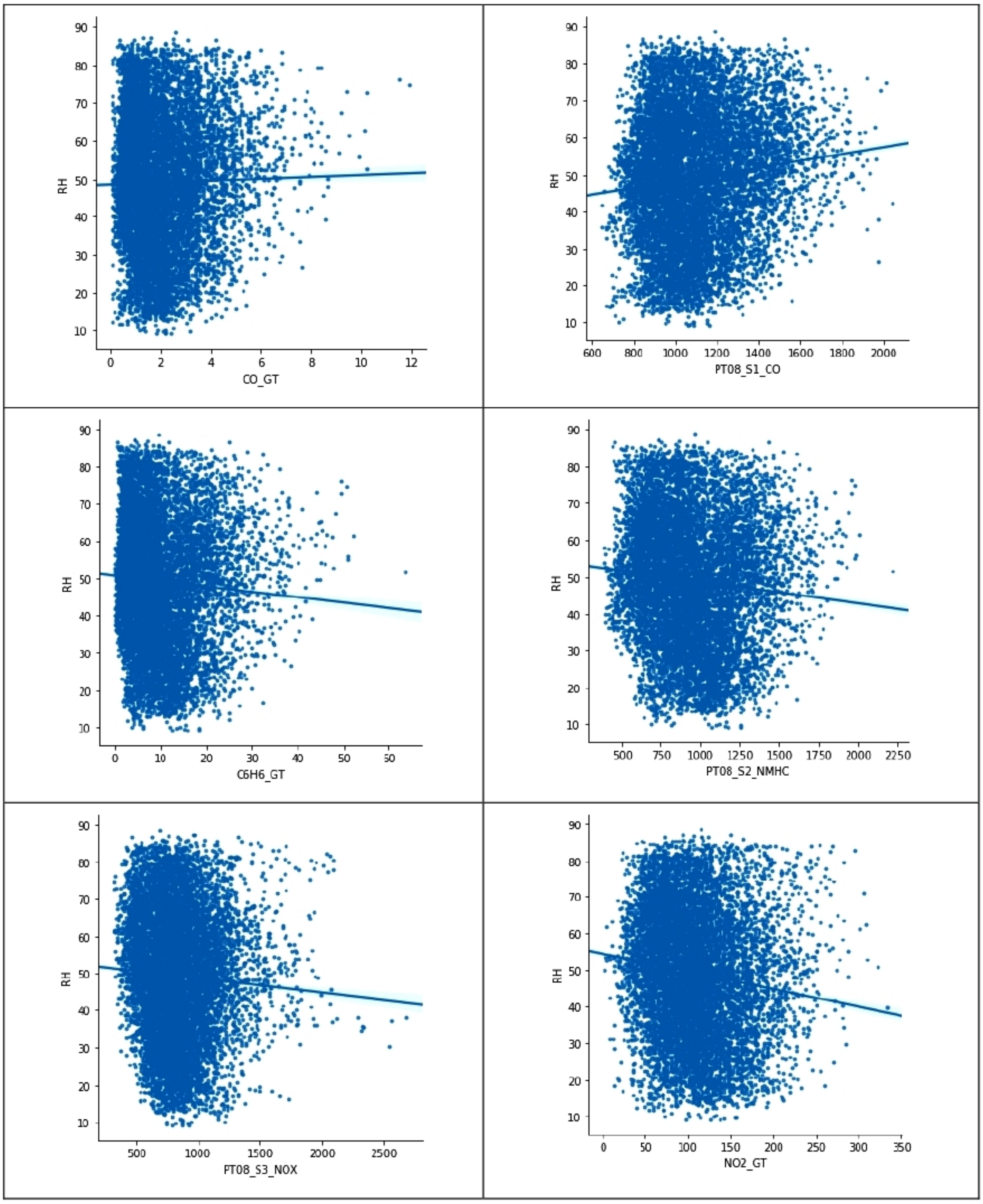

Figure 2: UCI data repository for air pollution

4. Pre- processing of the data: neighbor-based feature scaling technique

Scaling of features possesses an important influence in calculating pattern recognition and forecasting difficulties according to the prediction precision and low calculation expenses. Scaling systems are usually used to compress data at every interval. Due to compression process, calculation expense is minimized and class discrimination is escalated. In this research, the initial step of Euclidean distance of each variable is calculated. Next, the data values are acquired by dividing data to its total Euclidean distance. This technique is known as neighbor-based feature scaling [25].
This technique functions in the following way: "A" dataset comprises of n features and m data.

Y denotes the output variable. Feature set is $x_1(1),\ x_2(1), \ldots,\ x_n(1), \ldots,\ x_1(m), \ldots,\ x_n(m)$ and Y output variable is $Y_1(1),\ Y_2(2), \ldots,\ Y_n(m)$. After NBFS application, the new being features $X_1(1), \ldots,\ X_n(1), \ldots,\ X_1(m), \ldots,\ X_n(m)$ are calculated by:

$$X_{i,j} = \frac{x_{i,j}}{\text{The Total Euclidean Distancej (belong to each feature in dataset)}} \tag{1}$$

where, $X_{i,j}$ is the new feature value of dataset, $x_{i,j}$ is the old feature value of dataset for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$ (i: raw value of feature, j: number of features). The algorithm is illustrated below.

Set the values for input, output and distance value as
Train the dataset S with features as n and output as m
For every feature present in the dataset S
Perform the Euclidean distance for every feature
Assign the sum value as the feature value for every ratio
End for
The Table 1 shows the sample SO2 dataset Sample.

Table 1: Sample dataset containing ten raw SO2 data instances

Number of data	Feature 1	Feature 2	Feature 3	Feature 4	Output
1	5	72	1,019	2	107
2	6	75	1,019	13	160
3	4	86	1,020	14	71
4	4	83	1,018	26	71
5	11	77	1,020	10	108
6	2	81	1,017	3	110
7	3	80	1,013	5	88
8	5	94	1,016	8	36
9	4	95	1,016	10	53
10	4	86	1,024	9	57

5. Feature Extraction:

New features were obtained with the help of datetime components present in our dataset. These features were beneficial to identify series seasonality data. Properties that could be drawn out from a datetime group were assessed, further a few more variables were made: months [1-12], hours of a day [0-23], and weekends were attached as a Boolean feature. Hour belongs to cyclic feature and hence, two new variables were made with a trigonometric approach,

$$\text{hour\_sin} = \sin(2\pi \text{ hour}/24) \text{ and hour\_cos} = \cos(2\pi \text{ hour}/24), \text{ to map this behavior.}$$

Eventually, season variable was made, having four probable values Fall, Winter, Spring, and Summer. The overall features that were extracted enclose variables associated with pollutant and particulate measurements, meteorological conditions, lag features, rolling mean variables, season time, and time-related variables. Final compiled datasheet consists of 46 features Specifically, in this step, 10 lag parameters for pollutant, five rolling mean parameters, one parameter for the season, four trigonometric parameters (two for the month and two for the hour), one Boolean parameters for the weekend, and parameters for the date and time were established.

6. Feature Selection

After obtaining the 46 features the feature extraction step as mentioned earlier, in order to decrease data set dimensionality reducing co-linearity, selection of variables was done. As mentioned in [9], the concentration of the air pollutants and the ground-level ozone, PM2.5, and NO2, differs greatly based on environmental conditions and regional geography. Prevailing environmental conditions affect the concentrations, because of the complex inter connections among air pollutant sources like industrial gas release, motor vehicles, chemical transformations and waste disposals [7]. Therefore, all the suitable variables that are related to the environmental conditions are a part of the dataset. Where as in contrast, both filters and embedded techniques were implemented for sorting out all the additional variables. Filters are techniques that carry out feature selection irrespective to the forecasting design. Chosen learning method, in this study- SVR is used to carry out variable selection by embedded techniques. In accordance with the filters, as mentioned in [35] Pearson correlation-based feature selection was done. In order to confirm the presence of co linearity among the variables, Pearson correlation was used. It was found that few pollutants form a linear relationship between their observations. NO2, CO and CO, PM2.5 have such linear relations. Since the pollutants had a major co relation with their natural dependence, all pollutants were retained in the dataset. Pollutants have a strong correlation with their natural dependence, this was indicated by Cagliero et al. [4]. Pollutants' respective lag variables possesses high variance hence, concluded that only target pollutants' will be considered to minimize the probable co linearity among them. Moreover, even though SVR is considered strong with regards to co- linearity and multi co linearity [12], a few unessential variables were put aside to decrease the dataset dimensionality.

7. Recurrent Neural Networks

RNN was established for the purpose of time series data designing. As depicted in in Fig.3, a circular loop is introduced to feed forward network. The function of this loop is to carry data of various time steps and carry out temporal task [17]. Consequently, the network comprehends temporal pattern and calculates the values immediately on the basis of the preceding and present values.

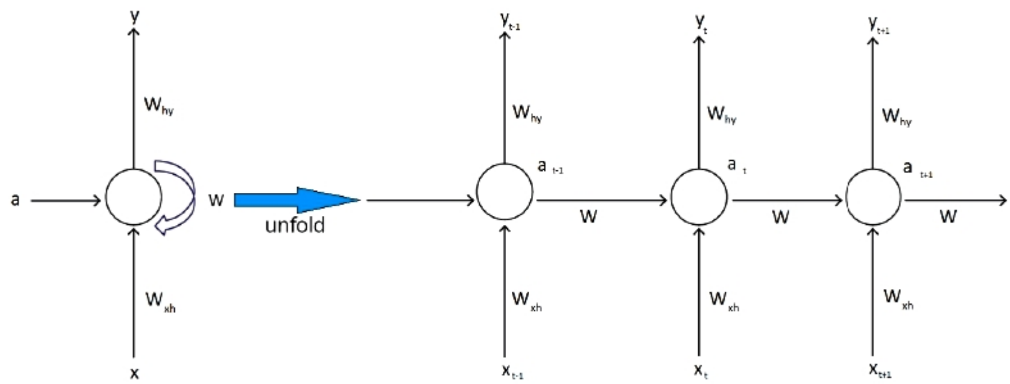

Figure 3: Recurrent Neural Networks (RNNs)

The image on the right-hand side is the un- rolled versions of the left sided network image, where

- **Wxh:** is weights for the connection of the input surface to the uncovered surface.
- **W:** is weights for the connection of the uncovered surface to the uncovered surface.
- **Why:** are the weights for the connection of the uncovered-surface-to-output-surface.
- **a:** is the activation of the surface.

The data is scanned from left to right by recurrent neural network. In this diagram, the variables used in each step are: variables **Wxh, Why** and **W** which are same for each time step. In the development of **RNN** with a prediction time **t**, the input at time **t** is **"xt"** and the data from the preceding input at time **t-1** is activated by parameter **"a"** and weights **"W"**. All these variables pass from previous uncovered surface to the present un-covered surface.

8. Air Status prediction using Recurrent Neural Networks

This paper presents RNN system for prediction of pollution levels by taking into consideration the temporal sequential data of every single pollutant. Hypothesis for this method: with the help of information such as pollutant concentration levels obtained from temporal sequence data and meteorological variables of a given area, system will obtain the dependencies present in the data and estimate and forecast the level of concentration for the coming next hour. The purpose is to provide a status representation of data by deriving conclusions from sequential features. This design is designed in a way that the hourly concentration values of pollutants are fed into the system in order to discover the uncovered trends in previous temporal values of the pollutants. Given, meteorological variables $v = \{v1,\ v2,\ v3,\ v4,\ v5,\ v6\}$ and pollutant concentrations xt , $t = 1, \ldots, T$ with paired input $S = \{(v, x_t)\}$, the aim of the design is to identify the trends and forecast x_{t+1}. RNN algorithm shown in Figure 3 is used for the design concentrations of air pollutants. Every time step of the design has processing input surface, recurrent surface and output surfaces.

Input Surface: This surface contains thick pollutant concentrations and meteorological variables by sequence.

$$Y_T^e = f(Y_T, v) \tag{2}$$

Where, for the given input of vectors Y_T^e function f produces embedding vector Y_T^e and n. Hence the resulting embedding vector is forwarded to the following recurrent surface.

Recurrent Surface: This surface produces factors representing uncovered sequential features from embedding vector Y_T^e. Uncovered states h_t are calculated with the help of input from the preceding time step h_{t-1} and present input Y_T^e . This value is updated with the below mentioned mathematical formula 3:

$$h_t = \Phi(Y_T^e, h_{t-1}) \tag{3}$$

Φ denotes the memory cell module of LSTM.

Output Surface: This last surface produces concentration values for the upcoming hour time with respect toyt. LSTM [6] is favorable for the algorithm since this RNN reduces the gradient vanishing issue and is capable of studying long term dependencies. This presented design, RNN-LSTM is important in order to improve forecasting precision of concentration values. Compared to the RNN surface cell module, RNN-LSTM possesses three gates to store long term dependencies. Important constituents of LSTM surface are

i. States: deliver data to un-covered and output surfaces.

ii. Gates: supervise the data flow of states.

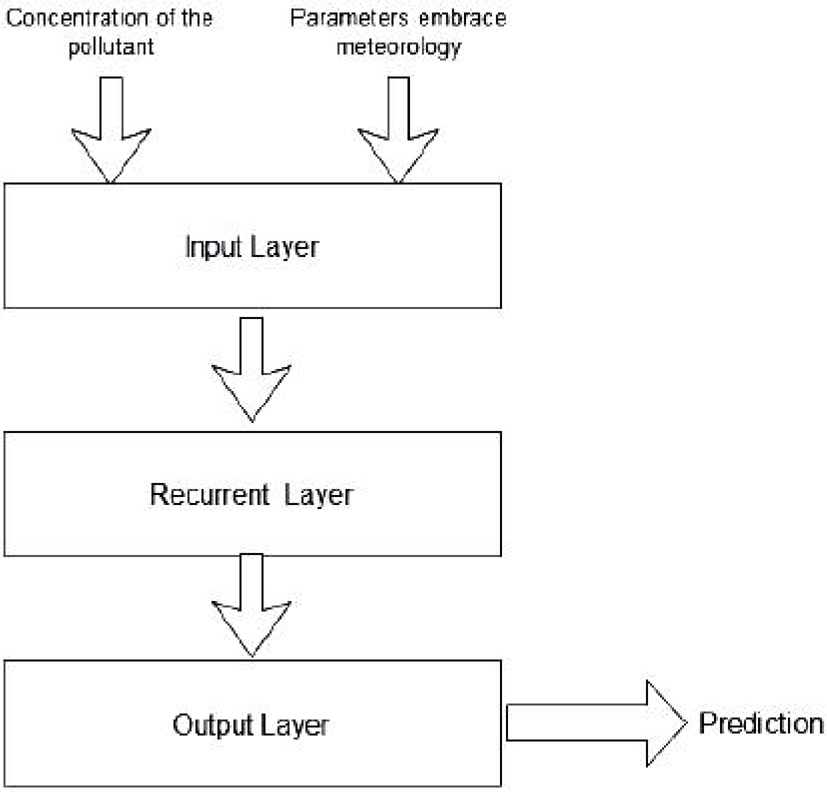

Figure 4: Prediction using RNN

States and gates calculation for time interval t is stated below: σ denotes the logistic sigmoid function or ReLU. i, f, o and c are input gate, forget gate, output gate, and cell activation vectors respectively. Vectors and the uncovered vector h are of uniform size.

$$i_t = \sigma(Y_{xi}\ x_i + Y_{hi}\ h_{i-1} + b_i) \tag{4}$$
$$f_i = \sigma(Y_{xf}\ x_t + Y_{hf}\ h_{t-1} + b_f) \tag{5}$$
$$\sigma_t = \sigma(Y_{xi}\ x_i + Y_{ho}\ h_{t-1} + b_o) \tag{6}$$
$$C_t = f_1 * C_{t-1} + i_t * tanh(Y_{xg} x_t + b_g) \tag{7}$$
$$h_t = \sigma_t * tanh(c_t) \tag{8}$$

$Y_{xi},\ Y_{xf},\ Y_{xo},\ Y_{xg}$ are projection matrices.
Whi ,Whf , Who are recurrent weight matrices.
Adam has developed this system to training and learning variables, a gradient descent optimization technique. Root mean square error (RMSE) represents the loss function. Ultimately, training process is done for using back propagation through time (BPTT).
In order to improve the efficiency, systems underwent training for 1000 epochs by shifting the surface

size. Weight matrices are stored in cases when the Mean Square Error (MSE) of the previous epoch is larger compared to present epoch. Once the systems are trained, every data point is tested and root Mean Square Error (RMSE) is estimated. Systems having smaller RMSE are considered for the final prediction design. MSE is stated by

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i}) \tag{9}$$

Air pollutants and environmental variables accounted for this research are displayed in Tables 2 and 3 respectively.

Table 2: List of pollutants

Sno	**Variables**	**Unit**
1	PM 2.5 (Particulate matter 2.5)	ug/m3
2	PM 10 (Particulate matter 2.5 to 10)	ug/m3
3	NO (Nitrogen oxide)	ug/m3
4	NO2 (Nitrogen dioxide)	ug/m3
5	NOX (Nitrogen oxides)	ug/m3
6	NH3 (Ammonia)	ug/m3
7	SO2 (Sulfur dioxide)	ug/m3
8	CO (Carbon monoxide)	ug/m3
9	Ozone ug/m3 10 Benzene	ug/m3

Table 3: Environmental variables

Sno	**Variables**	**Unit**
1	Ambient Temperature 0C	
2	Relative Humidity	%
3	Wind Direction	degree
4	Solar Radiation	W/mt2
5	Pressure	mmHg
6	Rain Fall	mm

The aim of this proposed research is to design an algorithm to forecast the level of air pollution for the upcoming hour based on air status of previous hour data. Performance evaluation has to be received by RNN-LSTM empirically [27]. The variables used for assessment are Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Coefficient of Determination (R2) [21]. 70% of the data was accounted for training purpose and remaining 30% is used to test the system. In the training phase cross validation is done in order to reduce the bias. Results received on evaluation of RNN-LSTM design is compared with the variables of base line regression method. Support Vector Regressor (SVR) [34]. Some recent studies [30] have focused on detecting and monitoring the air quality and pollution using modified deep neural network where the authors focused on classifying the attacked and clear data.

9. IoT based air status monitoring

For the purpose of assessing indoor air status, the air status monitoring based on IoT is used according to indoor Air Status Control Act. This was set into motion in the year 2007 by the Ministry of Environment, Korea in order to safe guard and control air standards leading to healthy body and healthy environment. According to this act, air is classified as good, moderate, or poor. Maximum levels were set. The thresholds levels can be altered from one area to another manually through web server on the basis of area inclination. Indoor air quality is highly affected by pollutant levels. Moderate air quality is defined in case temperature is neither good or bad. Never the less, pollutant thresholds are recommended parameters and these can be altered based on the user inclinations and the indoor situations. Figure 5 shows air quality monitoring on the basis of IoT.

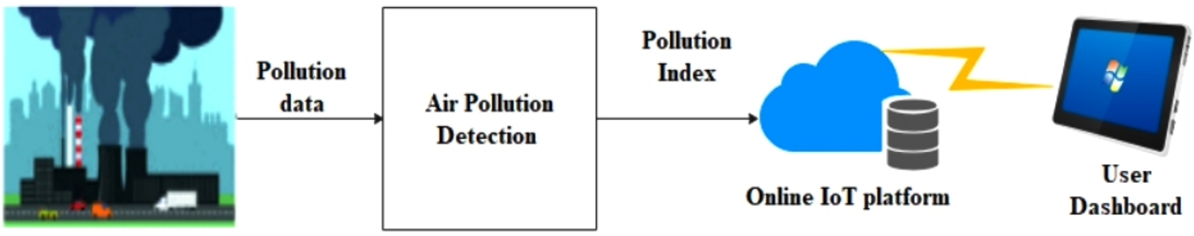

Figure 5: IoT based air quality monitoring

Real time assessment of air status is essential. Hence, a vigilance system is beneficial to vigilance changes and preserves the environment. With the vigilance system, the user platform may implement necessary required actions to normalize air status. Hence, AWS is enabled with an application known as Amazon Simple Notification Service for the vigilance system like a library for air status platform. The web server is developed in such a way so as to design a pop-up message in the program to signal the manager or the user in cases when condition is moderate or poor. an automatic vigilance message will be delivered once the system completes one year of successful usage. An auto prompt monitoring is generated with a pop-up message.

10. Performance Evaluation

For evaluating the performance various machine learning models such as DNN [29], CNN [1] and ANN [22] are considered for evaluation. As shown in below graph, when compared to the other machine learning models, Recurrent neural networks reports enhance performance in terms of the following parameters:

a) Accuracy

b) Specificity

c) Sensitivity

Accuracy is the ratio of summation of true positive and true negative to the summation of true positive, true negative, false positive and false negative values that are acquired from various instances. Figure 6 shows the comparative graph for accuracy plotted for proposed method against CNN, ANN and DNN respectively. Likewise, specificity refers to ratio of true negative values to the summation of true negative and false positive. Figure 7 highlights the comparative analysis of proposed RNN for detecting the pollution content in air against various machine learning models. Sensitivity defined as the ratio of number of true positive values to the summation of true positive and false negative for various air pollution prediction values. Figure 8 illustrates the performance graph plotted for sensitivity of various machine learning models. Comparison graphs have shown that the proposed system reports enhanced performance.

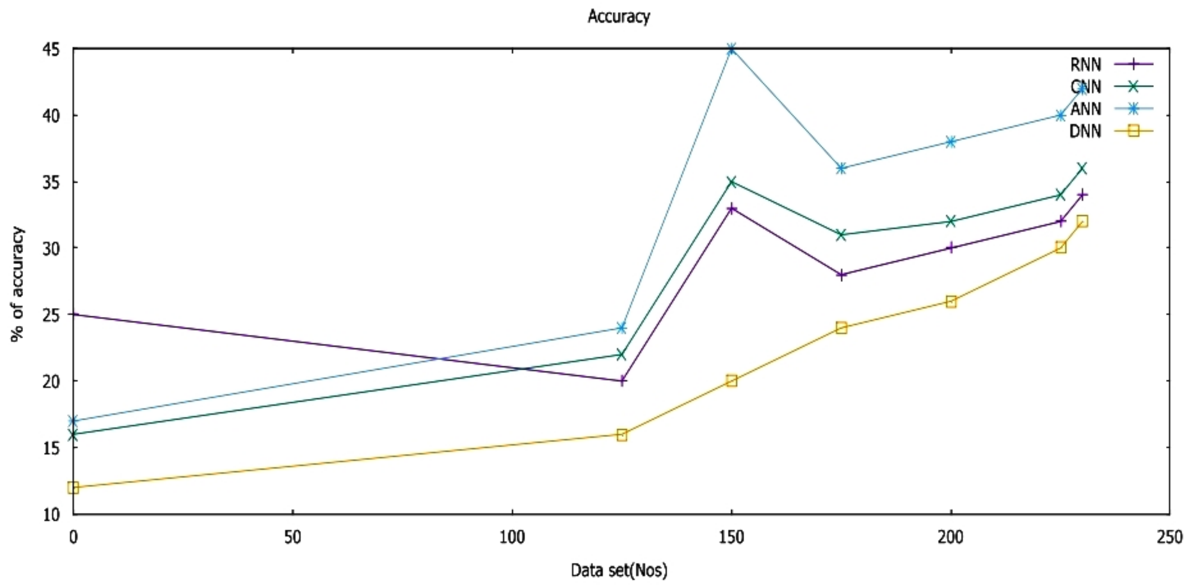

Figure 6: Accuracy Graph

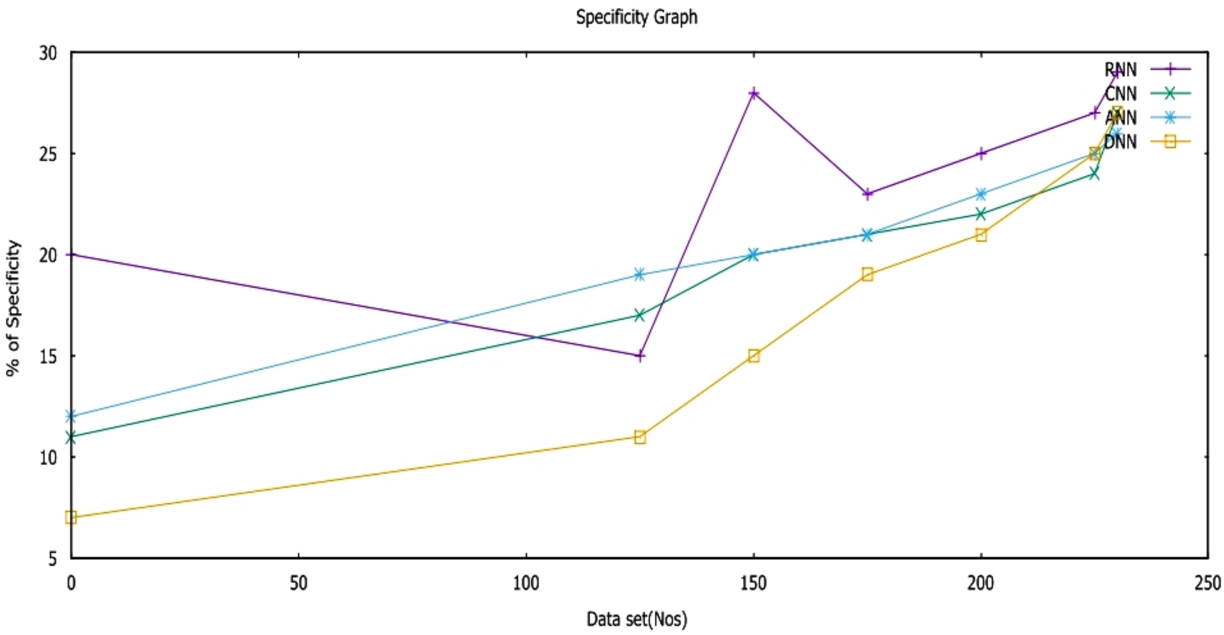

Figure 7: Specificity Graph

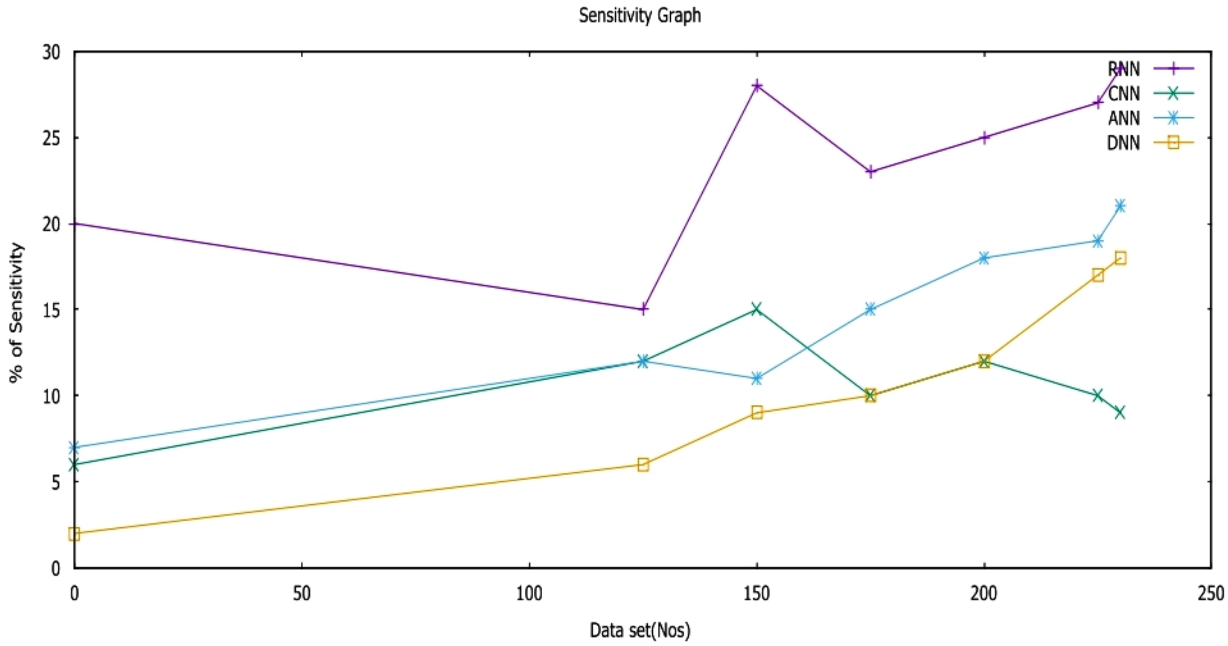

Figure 8: Sensitivity Graph

11. Conclusion:

Air pollution monitoring facilitates well-being status of people and it enhances human welfare in several ways. A Recurrent Neural Network based air pollution prediction proposed in this paper composed of data preprocessing using feature scaling method, Feature reduction and feature extraction process. More significantly, the predicted data is stored in cloud server and the status of air pollution index is updated. Using IoT technologies, the predicted data is sent to the users dash board and an alert message popped out when the threshold limit exceeds. Performance graphs have shown that the proposed RNN based air quality monitoring reports enhanced performance in terms of accuracy, sensitivity and specificity.

References

[1] J. Amoh, K. Odame,*Deep Neural Networks for Identifying Cough Sounds, IEEE transactions on biomedical circuits and systems,* 2017.

[2] A. Baklanov, P. Mestayer, A. Clappier, S. Zilitinkevich, S. Joffre, A. Mahura and N. Nielsen, *Towards improving the simulation of environmental fields in urban areas through updated/advanced surface fluxes description,* Atmospheric Chemistry and Physics, 8 (2008)523-543.

[3] G. E. Box, G. M. Jenkins, G. C. Reinsel and G. M. Ljung, *Time series analysis, forecasting and control.* John Wiley and Sons, 2015.

[4] L. Cagliero, T. Cerquitelli, S. Chiusano, P. Garza, G. Ricupero and X. Xiao, *Designing correlations among air pollution-related data through generalized association rules,* in Proceedings of the 2016 IEEE International Conference on Smart Computing, SMARTCOMP 2016, St. Louis, MO, USA, May 2016.

[5] F. Caiazzo, A. Ashok, I. A. Waitz, S. H. L. Yim and S. R. H. Barrett, *Air pollution and early deaths in the United States. Part I: quantifying the impact of major sectors in 2005,* Atmospheric Environment, 79, 2013, pp. 198-208.

[6] CERN, *Air Status Forecasting,* CERN, Geneva, Switzerland, 2001.

[7] M. Demuzere, R. M. Trigo, J. Vila-Guerau de Arellano and N. P. M. Van Lipzig, *&e impact of weather and atmospheric circulation on O3 and PM10 levels at a rural mid-latitude site,* Atmospheric Chemistry and Physics, 9(8), 2009, pp. 2695-2714.

[8] L. Diaz-Robles, J. Ortega, J. Fu, G. Reed, J. Chow, J. Watson and J. Moncada-Herrera,*A hybrid ARIMA and artificial neural networks design to forecast particulate matter in urban areas, The case of Temuco, Chile*, Atmospheric Environment, 42 (2008) 8331-8340.

[9] D. Dominick, M. T. Latif, H. Juahir, A. Z. Aris and S. M. Zain, *An assessment of influence of environmental factors on PM10 and NO2 at selected stations in Malaysia,* Sustainable Environment Research, 22(5), 2012, pp. 305-315.

[10] Y. Gonzalez, C. Carranza and M. Iniguez et al., *Inhaled air pollution particulate matter in alveolar macrophages alters local pro-inflammatory cytokine and peripheral IFN production in response to mycobacterium tuberculosis,* American Journal of Respiratory and Critical Care Medicine, 195, 2017, p. S29.

[11] P. Goyal, A. Chan and N. Jaiswal, *Statistical systems for the prediction of respirable suspended particulate matter in urban cities* , Atmospheric Environment, 40 (2006) 2068-2077.

[12] P. S. Gromski, E. Correa, A. A. Vaughan, D. C. Wedge, M. L. Turner and R. Goodacre, *A comparison of different chemometrics approaches for the robust classification of electronic nose data,* Analytical and Bioanalytical Chemistry, 406(29), 2014, pp. 7581-7590.

[13] B. Holmes-gen and W. Barrett,*Clean Air Future, Health and Climate Benefits of Zero Emission Vehicles,* American Lung Association, Chicago, IL, USA, 2016.

[14] U. A. Hvidtfeldt, M. Ketzel and M. Sørensen et al., *Evaluation of the Danish AirGIS air pollution designing system against measured concentrations of PM2.5, PM10, and black carbon,* Environmental Epidemiology, 2(2), 2018.

[15] P. Kakabutr, K. S. Chen, V. Wangvisavawit, P.Padungweang and O. Rojanapornpun, *Dog cough sound classification using artificial neural network and the selected relevant features from discrete wavelet transform*, 2017 9^{th}nternational Conference on Knowledge and Smart Technology(KST), Chonburi, 2017, pp. 121-125, doi,10.1109/KST.2017.7886118.

[16] A. Kurt and A. Oktay, *Forecasting air pollutant indicator levels with geographic systems 3days in advance using neural networks.* Expert Systems with Applications, 37 (2010)7986-7992 .

[17] Y. LeCun, Y. Bengio and G. Hinton, *Deep learning nature*, 521, 436, 2015.

[18] X. Liang, T. Zou, B. Guo, S. Li, H. Zhang, S. Zhang, H. Huang and S. Chen. *Assessing Beijing's PM2.5pollution, severity, weather impact, APEC and winter heating.* Proceedings of the Royal Society A, Mathematical, Physical and Engineering Science, 471 (2015) 20150257.

[19] X. Liang, S. Li, S. Zhang, H. Huang and S. Chen, *PM2.5data reliability, consistency, and air status assessment in five Chinese cities*, Journal of Geophysical Research, Atmospheres 121, 10, (2016) 220-10,236.

[20] C. Li, N. Hsu and S. Tsay, *A study on the potential applications of satellite data in air status monitoring and forecasting*, Atmospheric Environment, 45 (2011) 3663-3675.

[21] X. Li et al., *Long short-term memory neural network for air pollutant concentration predictions, Method development and evaluation [J],* Elsevier Ltd., Environmental Pollution, September, 231 (2017) 997-1004.

[22] S. A. Mahmoudi, P. D. Cunha Possa, T. Ravet, Th. Drugman, R. Chessini, T. Dutoit and C. Valderrama,*Sensor-based System for Automatic Cough Detection and Classification,* ICT Innovations 2015 Web Proceedings ISSN , (2015) 1857-7288.

[23] B. Ong, K. Sugiura and K. Zettsu, *Dynamically pre-trained deep recurrent neural networks using environmental monitoring data for forecasting PM2.5,* Neural Computing and Applications, 27(2015) 1553-1566.

[24] L. Pimpin, L. Retat and D. Fecht et al.,*Estimating the costs of air pollution to the National Health Service and social care, an assessment and forecast up to 2035,* PLoS Medicine, Article ID e1002602, 15(7), 2018, pp. 1-16.

[25] K. Polat,*A novel data preprocessing method to estimate the air pollution (SO2): neighbor-based feature scaling (NBFS),* Neural Comput and Applic, doi: 10.1007/s00521-011-0602-x.

[26] V. Prybutok, J. Yi and D. Mitchell, *Comparison of neural network systems with ARIMA and regression systems for prediction of Houston's daily maximum ozone concentrations*, European Journal of Operational Research 122, (2000) 31-40.

[27] Al-R. Rami et al. and Theano, *A Python framework for fast computation of mathematical expressions [R]*, arXiv preprint arXiv, 1605.02688, 2016.

[28] P. Saide, G. Carmichael, S. Spak, L. Gallardo, A. Osses, M. Mena-Carrasco and M. Pagowski, *Forecasting urban PM10 and PM2.5 pollution episodes in very stable nocturnal conditions and complex terrain using WRF-Chem CO tracer design,* Atmospheric Environment, 45 (2011) 2769- 2780.

[29] D. Saravanan, K. P. Hemalatha, B. Jinitha, Dr. K. Santhosh Kumar and P.Hemalatha, *Predict and Measure Air Quality Monitoring System Using Machine Learning,* 12(2) (2021) 2554-2561, https,//doi.org/10.17762/turcomat.v12i2.2217.

[30] D. Saravanan, K. Santhosh Kumar, R. Sathya and U.Palani, *An Iot Based Air Quality Monitoring And Air Pollutant Level Prediction System Using Machine Learning Approach - Dlmnn,* International Journal of Future

Generation Communication and Networking, 13(4), 2020, pp. 925-945.

[31] X. Shi, Z. Chen, H. Wang, D. Y. Yeung, W. K. Wong and W. C. Woo, *Convolutional LSTM network, A machine learning approach for precipitation nowcasting*, In Advances in neural information processing systems, (2015) 802-810.

[32] https://timesofindia.indiatimes.com/life-style/healthfitness/health-news/top-8-main-causes-for-air-pollution-indelhi/articleshow/61626744.cms.

[33] US Environmental Protection Agency (US EPA),*Criteria air pollutants,* America's Children and the Environment, US EPA, Washington, DC, USA, 2015.

[34] K. Veljanovska and A. Dimoski,*Air Status Index Prediction Using Simple Machine Learning Algorithms [J],* International Journal of Emerging Trends and Technology in Computer Science (IJETTCS), 7(1), January - February 2018.

[35] H. X. Zhao and F. Magoules, *Feature selection for support vector regression in the application of building energy prediction,* In Proceedings of the 9th IEEE International Symposium on Applied Machine Intelligence and Informatics,Smolenice, Slovakia, January 2011, pp. 219-223.

[36] S. Zhao, X. Yuan, D. Xiao, J. Zhang and Z. Li, *AirNet:a machine learning dataset for air status forecasting,* 2018.