Smart Air Pollution Monitoring System with Smog Prediction Model using Machine Learning

Salman Ahmad Siddiqui*, Neda Fatima, Anwar Ahmad
Department of Electronics and Communication Engineering
Jamia Millia Islamia, New Delhi, India

Abstract—**Air Pollution is a harsh reality of today's times. With rapid industrialization and urbanization, the polluting gases emitted by the burning of fossil fuels in industries, factories and vehicles, cities around the world have become "gas chambers". Unfortunately, New Delhi too happens to be among the most polluted cities in the world. The present paper designs and demonstrates an IoT(Internet of Things) based smart air pollution monitoring system that could be installed at various junctions and high traffic zones in urban metropolis and megalopolis to monitor pollution locally. It is designed in a novelistic way that not just monitors air pollution by taking varied inputs from various sensors (temperature, humidity, smoke, Carbon monoxide, gas) and but also presents it on a smart mirror. Its unique feature is the demonstration of a smog prediction model by determining PM10 (Particulate Matter 10) concentration using the most efficient machine learning model after an extensive comparison by taking into account environmental conditions. This data generated can also be sent as a feedback to the traffic department to avoid incessant rush and to maintain uniform flow of traffic and also to environmental agencies to keep pollution levels under check.**

Keywords—*Air pollution; IoT; machine learning; smart mirror; temperature and humidity sensor*

I. INTRODUCTION

Air pollution is one of the most significant causes of deaths worldwide. It claims seven million lives a year, mostly premature, and is a major driver of fatal non-communicable diseases like lung cancer, bronchitis asthma, heart attacks etc. [1]. It is a potential cause for allergies and causes irritation to the eyes, throat and skin. It also adversely affects climate change and a major cause for global warming that is responsible for drastic climatic catastrophes in the world. With rapid industrialization, urbanization, emergence of huge metropolises and megalopolises, pollution is ever increasing in the cities with cities in India especially Delhi NCR consistently featuring in the top most polluted cities of the world. The problem is further exacerbated in the winter season when the temperature, humidity and prevailing wind conditions provide favorable conditions for development of smog like conditions in most parts of North, North West and West India.

Air pollution is measured as the concentration of various gases in parts per million or ppm. Pollutants are of two kinds: primary and secondary pollutants. Primary pollutants are those which are released directly into the atmosphere in the form smoke from the industries and factories or exhausts from vehicles. Primary pollutants include ammonia, sulfur dioxide, carbon monoxide and nitrogen dioxide. Secondary pollutants

are those which are comparatively harmless in the atmosphere, however, turn into toxic pollutants on reaction with atmospheric conditions. Some types of secondary pollutants are ground-level ozone, acid rain and nutrient enrichment compounds.

Multiple ways have been devised to measure air pollution. Currently, it is monitored with the help of Air Quality Index(AQI) that defines the air quality in terms of ranges of pollutants in air in parts per million. AQI can be considered as a yardstick that runs from 0 to 500. The higher is the AQI value, the greater is the level of air pollution and the greater is the health concern. The following Table I demonstrates the air quality suitability with the corresponding range of pollutant concentration.

In line with this system of measurement of Air pollution, the present paper implements an IoT based Smart Air pollution monitoring system that depicts the localized air pollution levels along with temperature and humidity levels on Smart Mirror. It also employs the most efficient Machine learning model for Smog Prediction and this predictive analysis can serve as a feedback to the traffic department to forecast possible traffic jams and environmental agencies to keep the pollution levels in check.

The paper has been logically arranged in different sections. Section II describes the previous literature in this theme with their relevant shortcomings that led to the improvement in this project. Section III describes the methodology of the system in detail with special emphasis on the different machine learning algorithms employed in the system. Section IV presents the results obtained and its extensive analysis describing the resolution of the most efficient algorithm and the consequent results obtained. Section V discusses the advantages of the developed system and discusses about its real time deployment. Finally, the paper sums up with a futuristic conclusion and applications in Section VI.

TABLE I.   AIR QUALITY INDEX

Range (PPM)	Status
0-50	Good
51-100	Moderate
101-150	Unhealthy for sensitive groups
151-200	Unhealthy
201-300	Very unhealthy
301 and above	Hazardous

*Corresponding Author

II. REVIEW OF LITERATURE

The section describes the different works in the broad thematic area of research. While many types of research have been carried out in this broad area, they suffer from some inadequacies and limitations. For instance, paper [2] proposes and develops an IoT based Air Quality Monitoring System for Smart Cities using Raspberry Pi. However, it does not calibrate the sensors which results in drastic difference in values as opposed to the values yielded in Air Quality Index (AQI). While the calibration of sensors is employed in [3], it only utilizes one sensor i.e. MQ135 gas sensor thus excluding an important primary pollutant Carbon Monoxide from its scope of study. The paper [4] also suffers from similar limitation though it has employed humidity and temperature sensor that was not utilized in [3].

Application of neural network has a restrictive scope, for instance, the study demonstrated in [5] is a predictive analysis through neural network to demonstrate the harmful effects of air pollution on the human body and larger scope of the environment is not covered. In paper [11] application of artificial neural network for predicting air pollution levels has been presented whereas paper [6] employs an algorithm uses RFID technology to track down vehicles that cause vehicular pollution with higher emissions and reports them to environmental agencies. Thus, a thorough comparative modeling has not been performed. Further, the system in [7] demonstrates a crude set up to display air pollution however its accuracy is compromised to balance the cost of the set up.

With quite a few papers in the area of monitoring, the paper [8] uses a remote server to store data related to air pollution levels. In [9], combinations of wireless sensor networks and electrochemical toxic gas sensors with RFID (Radio Frequency Identification) tagging have been utilized to understand the pollution levels in vehicular emissions. The study proposed in [10] uses seismic sensor to predict earthquakes and light and humidity to analyze the location weather conditions. Thus, these researches are limited to the 'monitoring' aspect of air pollution.

Predictive analysis through machine learning has been demonstrated in some of the researches that have been surveyed. In [12], the prediction of air pollution is restricted to the supervised learning algorithms. In [13], the authors have utilized deep learning but with a highly imbalanced dataset and have predicted generic air pollution that too with a lower accuracy without confirmation with a real time dataset. In [14] and [15], the research is limited to different variants of Long Short Term Memory (LSTM) models for Air quality prediction in Delhi and South Korea respectively than the system proposed in this paper of specific smog prediction. A hybrid model was tested for monitoring air pollution at Iran's combined cycle power plant in paper [16]. In [17], a single machine learning model SVM (Support Vector Machine) has been employed to detect air quality in California with an already available dataset from the internet. While an improved weighted LSTM model has been utilized in [18], significant resources have been utilized for real world application and the performance parameters such as speed and machine cycles have been compromised for accuracy. Finally in [19],

AdaBoost machine learning has been employed in Taiwan to monitor air quality.

The present paper hones upon the existing research and addresses the shortcomings of the previous research such as lack of sensor calibration, limited number of sensors, incomplete datasets, compromised accuracy of predictive modeling, lack of comparative analysis of predictive modeling, compromised speed and machine cycles and inefficiency of resources. The paper presents a comprehensive approach that monitors air pollution on a real time basis, by utilizing input data from a variety of sensors to cover all possible sources of air pollution. It translates the data through calibration of sensors and averaging the value of sensors for ambient air quality index. Finally, it displays the Air pollution data along with smog levels on a Smart Mirror that can be employed at various junctions and high traffic zones in the city. Taking this data as input as well as dataset from Internet, the most efficient trained machine learning model is employed for Smog prediction. Furthermore, based on the data presented, corrective measures can be taken by traffic department environmental agencies to improve air quality and reduce air pollution.

III. METHODOLOGY

The system has been designed as a multi input multi output system. The sensory data from variety of sensors employed are fed into the microcontroller which monitor the localized atmosphere and generate a dataset that is displayed on the dynamic Smart Screen i.e. Magic Mirror. The dataset acts as an input for the Smog Prediction Model that employs the most efficient Machine learning model for the predictive analysis.

As demonstrated in the block diagram in Fig. 1, the present paper utilizes Arduino Uno R3 employing ATMega328P as microcontroller for implementation of the system. Arduino is an interactive open source platform characterized by low cost and flexible hardware and software. Arduino Uno R3 is the reference model and widely used. It has ATmega328 microcontroller chip (8- bit) at 16 MHz, with 14 digital I/O pins and 6 analog input pins. It is usually powered through USB connection but can also be powered by DC power socket from batteries.

The inputs to the Arduino are fed from the various sensors to determine the local environmental conditions. The most important sensor is the MQ135 Gas sensor. It can sense Ammonia (NH3), Nitrous Oxides (NOx), alcohol, Benzene, smoke, CO2 and some other gases. It gives the output in the form of voltage levels.

Since carbon monoxide is a significant primary air pollutant and MQ135 has the limitation of not measuring CO data, MQ7 carbon monoxide sensor has also been employed. This further hones the air pollution data received by Arduino Uno and can establish better results. The MQ-7 sensor can measure CO concentrations ranging from 20 to 2000ppm. It possesses faster response time and a high sensitivity. The sensor's output is an analog resistance.
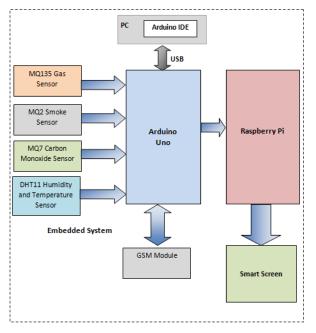
Fig. 1.   Block Illustration of the Technique.

In high traffic zones such as highways, merging freeways and cross junctions, pollution probability is higher. This increases the probability of presence of smoke at the location. To measure the concentration of smoke in the location, MQ2 Smoke sensor has been utilized. The MQ2 gas sensor is an electronic sensor which is used to detect ambient gas concentrations such as smoke, LPG, propane, methane, hydrogen, alcohol and carbon monoxide in the range of 200-10000ppm. The gas sensor MQ2 is often referred to as a chemical resistor. It includes a sensing material whose resistance changes when the gas is in contact.

Finally, DHT 11 Humidity and Temperature sensor have been utilized to analyze the localized humidity and temperature conditions respectively. A requirement of digital temperature and humidity sensor that is simple and ultra-low-cost is fulfilled by the DHT11. A thermistor and a capacitive humidity sensor is used to test the air quality of surroundings and the result provided by the data pin in form of digital signal. It requires a careful timing in order to collect information but it's fairly easy to use.

The data from MQ7, MQ2 Smoke Sensor and DHT11 Humidity and Temperature Sensor is used to determine the smog levels in the location. Smog usually appears as haze in the air due to the mixture of smoke, gases, and particles. Smog formation is the result of the combination of stable atmospheric conditions, due to inversion, nitrogen oxides, and organic compounds reaction resulting in ozone and related compound and different kind of air pollution and the emissions from increasing number of cars. The most important step in this system design is the calibration of MQ135, MQ2 and MQ7 Gas sensor with respect to fresh air and then development of an equation that transforms output sensor voltage into corresponding PPM levels. For this, the average analog readings of the resistance from the sensor and converting it to the voltage are taken.

The data from Arduino R3 is fed into Raspberry Pi 3. The Raspberry Pi 3 is a Pi series development board that can be viewed as a single computer board. It works with the LINUX Operating System. It has a fast processing speed, and uses wireless Local Area Network (LAN) and Bluetooth and can set up a WIFI hotspot to connect to the internet. It has a dedicated Liquid Crystal Display (LCD) display port. All the above data are displayed on the smart screen/magic mirror specially designed for the same. Smart screen though previously developed multiple times provides basic facilities like displaying clock, news, weather using APIs. The proposed screen is especially designed to provide serial communication between Arduino and Raspberry Pi Screen.

Fig. 2 represents the schematic diagram of the air pollution monitoring system developed. It shows the pin connections and the schematic of the functioning of the system.
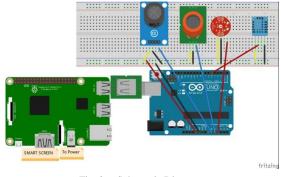

Fig. 2.   Schematic Diagram.

A. Equations

The sensor gives raw value and it needs to be converted into PPM. The relation for the same is obtained from the (Rs/Ro) versus PPM graph as shown in Fig. 3. The resistance of these sensors changes in response to the concentration of gases, viz. value of resistance decreases in response to increase in gas concentration; taking MQ135 sensor graph as reference for calculation.

Ro is the value of resistance in fresh air and the value of Rs is the value of resistance at various Gas concentrations. After preheating the sensor, the important step is to calibrate the sensor in fresh air to find Ro; from circuit diagram in Fig. 4 of MQ135, applying Ohm's law.
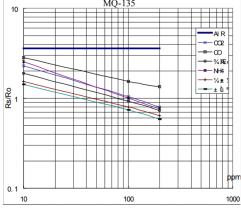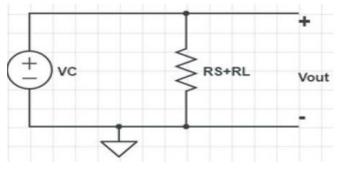

Fig. 3.   Graph showing Rs/Ro Ratio for MQ-135 Sensor.

Fig. 4.   Internal Circuit of MQ135.

$$V_{R_l} = \left[\frac{V_c}{R_s + R_l}\right] * R_L$$

$$V_{R_l} = \left[\frac{V_c * R_L}{R_s + R_l}\right] \tag{1}$$

$$V_{R_l} * (R_s + R_s) = V_c * R_L$$

$$(V_{R_l} * R_s) + (V_{R_l} * R_l) = V_c * R_L$$

$$(V_{R_l} * R_s) = (V_c * R_L) - (V_{R_l} * R_l)$$

$$R_s = \frac{(V_c * R_L) - (V_{R_l} * R_l)}{V_{R_l}}$$

$$R_s = \frac{(V_c * R_L)}{V_{R_l}} - R_l \tag{2}$$

Thus, the final equation in (2) is used to determine Rs.

From the graph above in Fig. 3, the resistance ratio Rs/Ro for fresh air is constant i.e. 3.6. To calculate Ro, Rs is to be found in fresh air, which is done by averaging the sensor raw reading and converting the same in volts and then using this volts value in equation (2) to obtain Rs, which is then used to find the Ro using Rs/Ro ratio of fresh air.

Now this Ro is used to find Ratio in presence of gas concentration. Using the Resistance ratio value in presence of gas, equivalent value of PPM for that particular gas can be determined from the graph.

To determine PPM from graph using resistance ratio, equation of line was applied on this log graph:

$$log_{10}y = m * log_{10}x + b \tag{3}$$

To find the slope, two points were selected from the graph. The points (200, 2.6) are and (10000, 0.75) from the LPG line. The formula becomes:

$$m = \frac{logy - \log(y_0)}{logx - \log(x_0)} \tag{4}$$

After solving (3) and (4), the values obtained were m= -0.318 and b= 1.13. Thus, the gas concentration using any ratio can be found using:

$$lo\,g(x) = \frac{lo\,g(y) - b}{m} \tag{5}$$

B. Smart Mirror

This paper presents the design and development of a smart mirror for displaying the air pollution based on Arduino sensor output. It provides a novel approach for establishing transmission between Arduino sensors and the Pi based Smart mirror which is based on node.js languages. The configuration file of the mirror was modified to incorporate the necessary parameters (port address, list of sensors and their output format) to make Arduino sensors value visible on magic mirror. The sensors value from Arduino is sent via serial communication in a standard format (understandable by mirror).

The mirror provides a natural means of interaction through which the commuters can control their movement based on the prevailing environmental condition in that area. The data displayed gives the real time conditions. The module for displaying the output is developed keeping in mind all the requirements. Fig. 5 shows the proposed Smart Mirror.
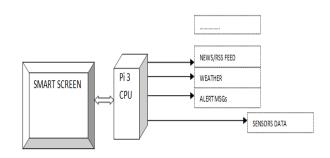

Fig. 5.   Schematic Diagram of Smart Mirror.

C. Smog Prediction Model

Smog is one the important cause of severe air pollution. It is formed as a result of combination of smoke, fog and water vapor. Sudden appearance of smog and/or fog on the highway more often than not causes serious and sometimes fatal accidents. It can also aggravate health problems including problems with breathing and sleeping, as well as it can adversely damage the surrounding flora and fauna.

Smog Prediction analysis has been performed by testing and comparing six different Machine leaning models. In the first model, a general learning framework based on an ensemble strategy and artificial neural networks (ANNs) has been employed. Thus, ANN has been trained to predict the PM 10 concentrations which is the main cause of the occurrence of smog phenomena. For the neural network presented, carbon monoxide, temperature, relative humidity, PM concentration of previous day and smoke were fed as an input and the target is the present day PM10 concentration.

Various types of neural network architectures with varied number of hidden nodes were formed and tested to get the best network for each measurement station. Among the different types of ANN, feed forward network is employed. It is a type of artificial neural network in which nodes' connections do not form a loop. As a result, it differs from its descendant, recurrent neural networks. The information in this network flows exclusively in one direction: forward, from the input nodes to the output nodes, passing through any hidden nodes (if any). Multi-layer perceptron (a class of feed forward ANN) as shown in Fig. 6 demonstrated the best result.
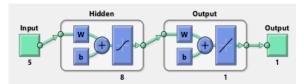
Fig. 6.   Artificial Neural Network Trained using MATLAB.

The next modeling was performed using Support Vector Regression. It is a supervised learning algorithm that is used to predict discrete values. Support Vector Regression uses the same principle as the support vector machine (SVMs). The basic idea behind SVR is to find the best fit line. SVR was tested with different kernel functions. The function of kernel is to take data as input and transform it into the required form.

Linear Regression Algorithm was also trained and tested for the same dataset to determine accuracy and other comparison parameter. It is a supervised machine learning algorithm that carries out a regression task. Based on independent variables, regression models a goal prediction value.

An AdaBoost regressor was performed on the given dataset. It is a meta-estimator that starts by fitting a regressor on the original dataset, and then fits new copies of the regressor on the same dataset, but with the weights of instances modified based on the current prediction's error.

Stacking regression was also trained for this dataset. It is a technique for creating linear combinations of various predictors in order to enhance prediction accuracy. Under non-negativity restrictions, cross-validation data and least squares are used. It consists in stacking the output of individual estimator and use a regressor to compute the final prediction. Stacking allows using the strength of each individual estimator (base estimator) by using their output as input of a final estimator. RigdeCV and linear SVR was the base estimator for stacking regression and random forest was the final estimator.

Furthermore, Random Forest Regression, a supervised learning approach for regression was also trained and tested for predictive modeling with hyper parameter tuning. Hyper parameter tuning implies varying the hyper parameter (number of trees and number of features) selected with criterion parameter as "gini" and then "entropy". Random Forest Regression uses the ensemble learning method that combines predictions from several machine learning algorithms to get a more accurate prediction than a single model. The algorithm has been demonstrated in Fig. 7.

Certain parameters were compared to be able to predict the most efficient model. The comparison of the model was performed using accuracy, Root Mean Square Error (RMSE) and correlation coefficient as shown in Table IV.

Accuracy is the measurement used to determine which model is best at identifying relationships and patterns between variables in a dataset based on the input, or training, data. Mean Square Error (MSE) is one of the simplest metrics used in regression. It is defined as the sum of the squares of the difference of the actual value to the predicted value or it is average squared errors of the prediction made. RMSE is

defined as Square root of Mean square error. The square root minimizes the errors and RMSE is used instead of MSE as MSE does not give an accurate picture when data is noisy. Higher the value of RMSE lesser the accuracy of Supervised Learning methods, lower the value of RMSE higher the accuracy. For two variables, the Correlation Coefficient compares the distance of each data point from the variable mean and uses this to tell us how closely the relationship between the variables can be fit to an imaginary line drawn through the data.
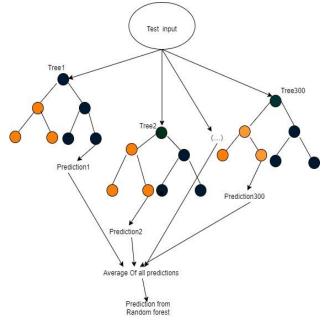

Fig. 7.   Random Forest Algorithm.

The dataset was split as 75% training subset, and 25 % testing subset. The dataset comprises of around 400 values collected in the interval of three months by taking measurements at different durations of the day for temperature, humidity, Carbon monoxide, smoke, PM10 (previous day) and PM (current day). The use of sensors for temperature, relative humidity, carbon monoxide concentration and smoke with PM10 concentration of previous day obtained from the official site using Python command ensures the real time prediction of smog and utility of the deployed system after network has been trained. This ensures the continuous working of the system.

Huge amount of data is being generated continuously on the social web in the form of air quality and weather forecast. Thus, to improve our machine learning models prediction further Big data analytics can be employed which will improve the prediction done by trained models using physical sensor by utilizing social web data like weather tweets, geographic information, meteorological records and air quality records.

IV. RESULTS

A. Sensors Measurement

1) After calibrations, Sensors were installed in an open terrace in the industrial area of Delhi NCR with reading taken at every 3 hours interval during the day as depicted in Fig. 8.

```
COM5

[status:setup:starting]
[status:setup:started]
DHTxx test!
Mq2 value=29.00
GAS SENSOR1 RAW VALUE = 342
Mq7 value in volts = 1.67
Rs of MQ7 value = 19.91
Resistance Ratio of MQ7 = 12.85
MQ7 desired PPM1 = 1.96
Mq7 Value in Percentage = 0.00
Mq135 value in volts = 1.08
Resistance Ratio of MQ135 = 5.42
MQ135 desired PPM1 = 0.89
  Humidity: 72.00%  Temperature: 18.80C 65.84F  Heat index: 18.62C 65.51F
[sensor:MQ135:0.89]
[sensor:MQ7:1.96]
[sensor:MQ2:29.00]
[sensor:DHT11:72.00]
[sensor:DHT11t:18.80]
Mq2 value=27.00
```

Fig. 8.   Sensor Measurement.

2) The readings from DHT and CO sensor were compared with values measured by standard real time AQI values and are tabulated in Table II.

3) The readings from MQ135 sensor were compared with Air quality monitor AM7000 and smoke sensor were compared with the standard values and tabulated in Table III.

4) The RMSE value for temperature is 0.644759 and for humidity is 0.953065. The RMSE value for CO is 0.5420.

5) The RMSE values for MQ135 sensor and Smoke is 0.8524 and 1.23051 respectively.

6) Lower values of RMSE obtained imply improved accuracy of Pollution monitoring than the previous research associated with monitoring.

TABLE II.   COMPARISON OF MEASURED AND ACTUAL PARAMETERS

Temp$_M$ (°C)	Temp$_A$ (°C)	Humidity$_M$ (%)	Humidity$_A$ (%)	CO $_M$ (ppm)	CO $_A$ (ppm)
22.0	21.3	34.1	35.4	2.6	2.1
23.3	22.8	32.2	32.9	2	1.72
23.9	24.5	30.0	31.3	2.2	2.09
24.4	23.6	28.9	27.9	2.2	2.6
24.4	24.6	29.4	29.1	1.6	2.1
23.8	24.5	31.3	30.6	1.2	1.51
..	..	..	..	..	..
17.8	18.6	42.9	42.2	1.2	2.3

TABLE III.   COMPARISON OF MEASURED AND ACTUAL PARAMETERS CO2 AND SMOKE SENSORS

CO2$_M$ (ppm)	CO2$_A$ (ppm)	Smoke Sensor$_M$ (ppm)	Smoke Sensor$_A$ (ppm)
2.1	2.96	54.05	52.8
2.82	3.42	65.7	64.1
6.9	8.5	58.45	56.3
15.4	15.1	57.2	56.9
12.5	11.9	53.8	53.4
..	..	..	..
4.9	4.4	55	55.3

B. Model Training and Comparison

1) Six different machine learning models were trained and tested viz. MATLAB ANN, Support vector regression, linear regression classifier, random forest regression, Adaboost regression and Stacking regression for predicting the PM10 concentration based on fed input.

2) The following graph in Fig. 10 shows the test value (PM10) result and actual values (PM10) for ANN and the result is considerably accurate.

3) ANN is a statistical model and the choice of ANN architecture, including the number and type of neurons, and the selection of a learning algorithm is very important to improve the accuracy.

4) The ANN model was trained multiple times with different number of hidden layers to determine the best suited number of hidden layers as shown in Fig. 7. The best suited was hidden layer with 8 nodes.

5) The correlation between actual PM10 values and test values came out to be 0.76 which is normal.

6) While linear regression is simple to implement but the values obtained were worst amongst all the models tested though better than ANN. It was fitted to reach the best accuracy of 76.42%, the least among all.

7) The linear kernel is shown to be the best input transformation technique for SVR.

8) Stacking regressor performed with mediocre accuracy among the others.

9) For random forest, the best results were obtained for 300 number of trees (accuracy 78.97 and RMSE 24.65), however, in comparison of all parameter values, Adaboost regression proved most efficient.

10)Accuracy for Adaboost reaches 98.24% which is highest among all the models as presented in Table IV.

11)Thus, Adaboost Regression was finally selected for Smog Prediction Model.
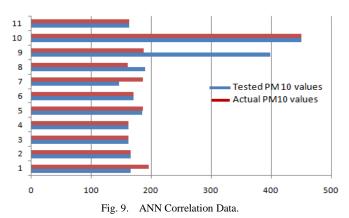
12)Parameters such as correlation and RMSE as shown in Fig. 9, 11, 12, 13, 14, 15 between measured and actual data for all models.

13)As demonstrated, multiple algorithms have been tested before determining the best and most efficient algorithm for Smog Prediction.

TABLE IV.   COMPARISON OF MODELS

Model	Accuracy	RMSE	Correlation
Linear Regression	76.42	50.69	0.906964
ANN (MATLAB)	58.156	66.41	0.765523
SVR	80.9	33.28	0.914284
Random Forest Regressor (300 Trees)	78.97	24.65	0.917268
Adaboost Regression	98.24518	12.37	0.991843
Stacking Regression	88.18798	31.3603	0.942327

14)A differential factor not found in the previous research was that the models have been developed in a robust manner for which each model was re-built 25 times using different random subsets of training and testing samples keeping the splitting proportion constant.
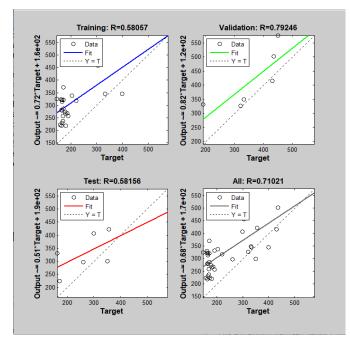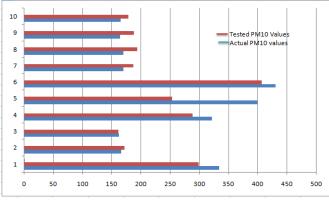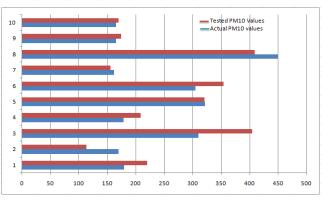

Fig. 9.   ANN Correlation Data.
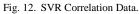

Fig. 10. ANN Accuracy.

Fig. 11. Linear Regression Correlation.

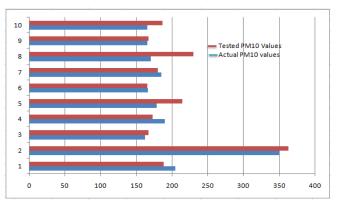Fig. 12. SVR Correlation Data.
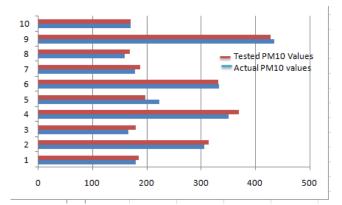

Fig. 13. Random Forest Correlation Data.

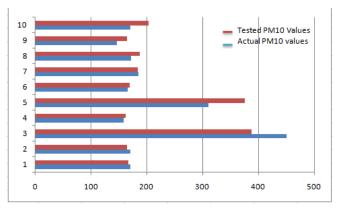Fig. 14. AdaBoost Correlation Data.

Fig. 15. Stacking Regression Correlation Data.

C. Magic Mirror

1) Fig. 17 shows the dynamic magic mirror screen output.

2) Fig. 16 shows the sensors output from Arduino and this output are taken as input to trained model for predicting smog.

3) Environmental conditions viz. sensors' output are displayed on the specially designed Magic Mirror/ Smart Mirror using Raspberry Pi and the same screen can be installed in the traffic vicinity or the traffic junctions in the metropolis/megalopolis for the common people to have 24 hours pollution and smog monitoring.

4) To make the demonstrated smart screen more attractive, additional features on the screen apart from air pollution data have also been incorporated like showing News, Calendar, and the Weather forecast of Delhi region as shown in Fig. 17.

5) Alert messages for maintaining social distancing and wearing mask for corona and air pollution has also been incorporated.

Fig. 16. Sensor Output on Magic Mirror Screen.

Fig. 17. Magic Mirror Display.

V. Discussion

The novelty of the system is the Smog Prediction model presented based on PM10 concentrations which is trained using Machine learning model. The system is based on constantly updated data and the finally displayed in a dynamic and interactive Smart Mirror. The machine learning model employed has been selected after extensive testing of six models with different parameters and the chosen model has the maximum efficiency and maximum stability. The dataset utilized is dynamic from both real time measurement as well as the internet.

The system presents an improved and calibrated approach towards data analysis by the sensor inputs and corrects PPM calculations thus making it accurate and ready to display in accordance with standard air quality index measurements. These values along with the obtained previous day PM10 values are fed as an input to the trained model to predict the formation of smog. Based on this analysis, SMS alerts can be sent by triggering communication with SIM800l using Python script when smog is detected to the designated traffic official to divert the traffic, especially at crowded junctions.

Thus, the system overcomes the limitations of the previous research by adopting sensor calibration, increasing number of sensors, varied datasets, improving accuracy of predictive modeling, presenting comprehensive comparative analysis of predictive modeling and implementing efficient use of resources.

The system is designed to be flexible and can be altered easily by adding new sensors. The system generates awareness among the masses and the government about air pollution and provides data to localized tackling of air pollution. A small compact kit can also be developed for indoor air pollution monitoring whereas for outdoors the entire kit would suffice for accurate results.

VI. Conclusion

The paper successfully implements a Smart IoT based Air pollution monitoring system that employs advanced machine learning models to implement a novel system with SMOG prediction modeling that helps in improving the health of the people and their environment by making them aware of their surroundings. The data is projected on a Dynamic smart screen/ Magic Mirror that can be displayed at various traffic junctions for general awareness. Also, the data and further graphical representations obtained can help the traffic department and environmental agencies to take corrective actions regarding the pollution levels as and when it is updated. This set up can further aid the local municipalities to take corrective steps and solve this rabid problem in a democratically decentralized manner.

An efficient, viable and affordable implementation of the system is presented that is functioning in real time conditions in New Delhi's environment. The System has been made flexible and to further hone the results of smog detection, ozone layer status sensor can also be utilized. PM 2.5 laser dust

sensor can also be utilized; however, it may take up the cost of the system. For the purpose of forecasting air pollution on a large scale, large-scale node location and data collection can be expected in the future.

REFERENCES

[1] Manisalidis, E. Stavropoulou, A. Stavropoulos, and E. Bezirtzoglou, "Environmental and Health Impacts of Air Pollution: A Review," Front. Public Heal., vol. 8, no. February, pp. 1–13, 2020, doi: 10.3389/fpubh.2020.00014.

[2] B. Kalinga Institute of Industrial Technology, IEEE Singapore Section, Institute of Electrical and Electronics Engineers, IEEE Power Electronics Society, and IEEE Industry Applications Society, "1st IEEE International Conference on Sustainable Energy Technologies and Systems, IEEE-ICSETS 2019 : 26 February – 1 March 2019, Kalinga Institute of Industrial Technology, Bhubaneswar, Odisha, India," 2019 IEEE Int. Conf. Sustain. Energy Technol., pp. 173–177, 2019.

[3] K. B. K. Sai, S. Ramasubbareddy, and A. K. Luhach, "IOT based air quality monitoring system using MQ135 and MQ7 with machine learning analysis," Scalable Comput., vol. 20, no. 4, pp. 599–606, 2019, doi: 10.12694/scpe.v20i4.1561.

[4] D. Verma, S. Dhul, R. Saini, and R. B. Dubey, "IoT Based Air Pollution Monitoring System AND," Int. J. Sci. Eng. Res., vol. 8, no. 3, pp. 116–120, 2018.

[5] T. W. Ayele and R. Mehta, "Air pollution monitoring and prediction using IoT," Proc. Int. Conf. Inven. Commun. Comput. Technol. ICICCT 2018, no. Icicct, pp. 1741–1745, 2018, doi: 10.1109/ICICCT.2018.8473272.

[6] R. Rushikesh and C. M. R. Sivappagari, "Development of IoT based vehicular pollution monitoring system," Proc. 2015 Int. Conf. Green Comput. Internet Things, ICGCIoT 2015, pp. 779–783, 2016, doi: 10.1109/ICGCIoT.2015.7380568.

[7] S. Kumar and A. Jasuja, "Air quality monitoring system based on IoT using Raspberry Pi," Proceeding - IEEE Int. Conf. Comput. Commun. Autom. ICCCA 2017, vol. 2017-Janua, pp. 1341–1346, 2017, doi: 10.1109/CCAA.2017.8230005.

[8] K. Okokpujie, E. Noma-Osaghae, O. Modupe, S. John, and O. Oluwatosin, "A smart air pollution monitoring system," Int. J. Civ. Eng. Technol., vol. 9, no. 9, pp. 799–809, 2018.

[9] S. Manna, Souvik, sankar, "Vehicular Pollution Monitoring U sing loT," 2014.

[10] J. Magendie, J. Lange, And Dupin, "Sur un cas de rétrécissement péricolique pelvien (syndrome de Küss)," Bord. Chir., vol. 3, pp. 117–120, 1951.

[11] Małgorzata Pawul, Małgorzata Śliwka, "Application Of Artificial Neural Networks For Prediction Of Air Pollution Levels In Environmental Monitoring ",Int J. Eco Engg, DOI: 10.12911/22998993/64828.

[12] M. Madhuri, G. H. Samyama Gunjal, and S. Kamalapurkar, "Air pollution prediction using machine learning supervised learning approach," Int. J. Sci. Technol. Res., vol. 9, no. 4, pp. 118–123, 2020.

[13] J. Kalajdjieski et al., "Air pollution prediction with multi-modal data and deep neural networks," Remote Sens., vol. 12, no. 24, pp. 1–19, 2020, doi: 10.3390/rs12244142.

[14] A. Tiwari, R. Gupta, and R. Chandra, "Delhi air quality prediction using LSTM deep learning models with a focus on COVID-19 lockdown," 2021, [Online]. Available: http://arxiv.org/abs/2102.10551.

[15] T.-C. Bui, V.-D. Le, and S.-K. Cha, "A Deep Learning Approach for Forecasting Air Pollution in South Korea Using LSTM," 2018, [Online]. Available: http://arxiv.org/abs/1804.07891.

[16] A. Heydari, M. Majidi Nezhad, D. Astiaso Garcia, F. Keynia, and L. De Santoli, "Air pollution forecasting application based on deep learning model and optimization algorithm," Clean Technol. Environ. Policy, no. 0123456789, 2021, doi: 10.1007/s10098-021-02080-5.

[17] M. Castelli, F. M. Clemente, A. Popovič, S. Silva, and L. Vanneschi, "A Machine Learning Approach to Predict Air Quality in California," Complexity, vol. 2020, no. Ml, 2020, doi: 10.1155/2020/8049504.

[18] F. Xiao, M. Yang, H. Fan, G. Fan, and M. A. A. Al-qaness, "An improved deep learning model for predicting daily PM2.5 concentration," Sci. Rep., vol. 10, no. 1, pp. 1–11, 2020, doi: 10.1038/s41598-020-77757-w.

[19] Y. C. Liang, Y. Maimury, A. H. L. Chen, and J. R. C. Juarez, "Machine learning-based prediction of air quality," Appl. Sci., vol. 10, no. 24, pp. 1–17, 2020, doi: 10.3390/app10249151.