IOT BASED AIR QUALITY PREDICTION USING ADABOOST ALGORITHM

R.Kanmani[1], D.Madhumitha[2], G.Thamaraiselvi[3], Mr.G.Anand[4]
[1, 2, 3] Dept of Electronics and Communication Engineering
[4]Assistant Professor, Dept of Electronics and Communication Engineering
[1, 2, 3, 4] N.S.N college of Engineering and Technology, Karur, India

Abstract- *Air pollution in smart cities in the world has been drastically increasing lately and the increase in the concentration of particulate matter in the air is a threat for the country and citizens as it can out-turn unbearable consequences such as cardiovascular disease and worsen asthma.PM2.5 is a deadly air pollutant that is a mixture of solid and liquid coarse particles and has a diameter of 2.5 micrometers. This project aims to implement machine learning algorithms to find the accuracy of the prediction of particulate matter, in air pollution in smart cities of INDIA. To test the implementation of machine learning in this prediction, Ada boost algorithm is chosen and using the smart city air Pollution dataset. The outcome of this research is that Ada boost gave the best accuracy in prediction of Particulate Matter, Air Pollution Index by using Raspberry pi PICO hardware in smart cities of INDIA. This project implements in real-time hardware setup with Raspberry pi Pico and IoT supporting node.*

Keywords- Air Pollution, Particulate Matter (PM2.5), Multi-layer Perceptron, Random Forest

I. INTRODUCTION

Pollution of the air is currently one of the most pressing issues facing humanity on a global scale, particularly in highly developed cities. A material from any atmospheric source that exists in any form, whether liquid, solid, or gas phase, and is capable of causing destruction or the capability to alter the typical properties of the atmosphere and increasing the health risk to living things or causing the environment and ecosystem to be out of balance is referred to as polluting the air. PM2.5 is the air pollutant that the United States Environmental Protection Agency (US-EPA) identifies as being the most dangerous to human health and the leading cause of death worldwide. The concentration of air pollutants, particularly PM2.5 during the haze season, is the primary factor that determines the API readings in Malaysia. PM2.5 is a very tiny particle that is able to diffuse into respiratory systems and has a negative impact on human lungs. The weight of PM2.5 is a fine and tiny particle in the air, and its diameter is 2.5 micrometers, as seen in PM2.5 is a very tiny particle that can be visualized in PM2.5. Pollution from vehicles and industrialization are the two primary sources of PM2.5 emissions. Cities that use smart city technology can be observed to be experiencing this stage of the PM2.5 problem. Problems with air pollution are plaguing a number of smart cities. According to Ameer, a "smart city" is an urban municipality that uses information technologies (ICT) to provide its residents with adequate health, transportation, and energy-related facilities, and that also assists the government in making efficient use of its available resources for the benefit of its people. In terms of information and communications technology (ICT) and the number of people living in urban areas, Kuala Lumpur and Johor Bahru are two of the developed smart cities in Malaysia. Both the process of industrialization and the movement of people from rural to urban areas have contributed to the rapid growth of urban populations in the modern world. The rise in the city's population has resulted in an increase in the number of people who use various modes of transportation and consume various forms of energy, both of which have contributed to the expansion of the city's industrial capacity and its vehicle population. As a result, the findings of a number of empirical studies have led researchers to the conclusion that the issue of air quality in smart cities has been one of the city's primary challenges, and that machine learning has provided a better and more strategic solution to the problem of air quality prediction. In contrast to the rest of the world, the application of machine learning in Malaysia to the forecasting of air pollutants and air pollutions has not been widely recognized. Since there has been significant development in the prediction of air pollution all over the world over the course of the last few decades, it is possible that the concentration of air pollutants in smart cities in Malaysia that are predicted using ML techniques will be accurate. According to the World Health Organization (WHO), air pollution is a contributing factor in approximately 1.3 million deaths each year around the world. The release of pollutants into the atmosphere has many negative effects, one of which is deterioration in the quality of the air. Other negative effects, such as acid rain, global warming, the production of aerosols, and photochemical smog, have also worsened over the course of the past few decades. Many researchers have been motivated to investigate the underlying pollution-related conditions that are contributing to COVID-19 pandemics in different countries as a result of the recent rapid spread of COVID-19. Air pollution has been linked to significantly higher COVID-19 death rates, and patterns in COVID-19 death rates mimic patterns in both areas with a high population density and areas with a high PM 2.5 exposure. This is evidenced by several pieces of circumstantial evidence. Because of everything that has been discussed up to this point, it is absolutely necessary to forecast and prepare for changes in pollution levels

II. LITERATURE SURVEY

A deep learning model for air quality prediction in smart cities:

In recent years, Internet of Things (IOT) concept has become a promising research topic in many areas including industry, commerce and education. Smart cities employ IOT based services and applications to create a sustainable urban life. By using information and communication technologies, IOT enables smart cities to make city stakeholders more aware, interactive and efficient. With the increase in number of IOT based smart city applications, the amount of data produced by these applications is increased tremendously. Governments and city stakeholders take early precautions to process these data and predict future effects to ensure sustainable development. In prediction context, deep learning techniques have been used for several forecasting problems in big data. This inspires us to use deep learning methods for prediction of IOT data. Hence, in this paper, a novel deep learning model is proposed for analyzing IOT smart city data. We propose a novel model based on Long ShortTerm Memory (LSTM) networks to predict future values of air quality in a smart city. The evaluation results of the proposed model are found to be promising and they show that the model can be used in other smart city prediction problems as well.

Comparative Analysis of Machine Learning Techniques for predicting r Quality in Smart Cities:

Dealing with air pollution presents a major environmental challenge in smart city environments. Real-time monitoring of pollution data enables local authorities to analyze the current traffic situation of the city and make decisions accordingly. Deployment of the Internet of Things-based sensors has considerably changed the dynamics of predicting air quality. Existing research has used different machine learning tools for pollution prediction; however, comparative analysis of these techniques is required to have a better understanding of their processing time for multiple datasets. In this paper, we have performed pollution prediction using four advanced regression techniques and present a comparative study to determine the best model for accurately predict quality with reference to data size and processing time. We have conducted experiments using Apache Spark and performed pollution estimation using multiple datasets. The Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) have been used as evaluation criteria for the comparison of these regression models. Furthermore, the processing time of each technique through standalone learning and through fitting the hyperparameter tuning on Apache Spark has also been calculated to find the best-fit model in terms of processing time and lowest error rate.

A Machine Learning Model for Air Quality Prediction for Smart Cities:

Air quality of a certain region can be used as one of the major factor determining pollution index also how well the city's industries and population is managed. Urban air quality monitoring has been a constant challenge with the advent of industrialization. Air pollution has remained a major challenge for the public and the government all over the world. Air pollution causes noticeable damage to the environment as well as to human health resulting into acid rain, global warming, heart diseases and skin cancer to the people. This paper addresses the challenge of predicting the Air Quality Index (AQI), with the aim to minimize the pollution before it gets adverse, using two Machine Learning Algorithms: Neural Networks and Support Vector Machines. The air pollution databases were extracted from the Central Pollution Control Board (CPCB), Ministry of Environment, Forest and Climate change, Government of India. The proposed Machine Learning (ML) model is promising in prediction context for the Delhi AQI. The results show improvement of the prediction accuracy and suggest that the model can be used in other smart cities as well.

"Air Quality Prediction in Smart Cities Using Machine Learning Technologies Based on Sensor Data:

The influence of machine learning technologies is rapidly increasing and penetrating almost in every field and air pollution prediction is not being excluded from those fields. This paper covers the revision of the studies related to air pollution prediction using machine learning algorithms based on sensor data in the context of smart cities. Using the most popular databases and executing the corresponding filtration, the most relevant papers were selected. After thorough reviewing those papers, the main features were extracted, which served as a base to link and compare them to each other. As a result, we can conclude that: (1) instead of using simple machine learning techniques, currently, the authors apply advanced and sophisticated techniques, (2) China was the leading country in terms of a case study, (3) Particulate matter with diameter equal to 2.5 micrometers was the main prediction target, (4) in 41% of the publications the authors carried out the prediction for the next day, (5) 66% of the studies used data had an hourly rate, (6) 49% of the papers used open data and since 2016 it had a tendency to increase, and (7) for efficient air quality prediction it is important to consider the external factors such as weather conditions, spatial characteristics, and temporal features.

Air-pollution prediction in smart city, deep learning:

Over the past few decades, due to human activities, industrialization, and urbanization, air pollution has become a life-threatening factor in many countries around the world. *Among air pollutants, Particulate Matter with a diameter* of less than 2.5µm2.5(PM2.52.5) is a serious health problem. It causes various illnesses such as respiratory tract and cardiovascular diseases. Hence, it is necessary to accurately predict the PM2.52.5 concentrations in order to prevent the citizens from the dangerous impact of air pollution beforehand. The variation of PM2.52.5 depends on a variety of factors, such as meteorology and the concentration of other pollutants in urban areas. In this paper, we implemented a deep learning solution to predict the hourly forecast of PM2.52.5 concentration in Beijing, China, based on CNN-LSTM, with a spatial-temporal feature by combining historical data of pollutants, meteorological data, and PM2.52.5 concentration in the adjacent stations. We examined the difference in performances among Deep learning algorithms such as LSTM, Bi-LSTM, GRU, Bi-GRU, CNN, and a hybrid CNN-LSTM model. Experimental results indicate that our method “hybrid CNN-LSTM multivariate” enables more accurate predictions than all the listed traditional models and performs better in predictive performance.

“**Graph-Deep-Learning-Based Inference of Fine- Grained Air Quality From Mobile IOT Sensors:**

Internet- of-Things (IOT) technologies incorporate a large number of different sensing devices and communication technologies to collect a large amount of data for various applications. Smart cities employ IOT infrastructures to build services useful for the administration of the city and the citizens. In this article, we present an IOT pipeline for acquisition, processing, and visualization of air pollution data over the city of Antwerp, Belgium. Our system employs IOT devices mounted on vehicles as well as static reference stations to measure a variety of city parameters, such as humidity, temperature, and air pollution. Mobile measurements cover a larger area compared to static stations; however, there is a tradeoff between temporal and spatial resolution. We address this problem as a matrix completion on graphs problem and rely on variational graph autoencoders to propose a deep learning solution for the estimation of the unknown air pollution values. Our model is extended to capture the correlation among different air pollutants, leading to improved estimation. We conduct experiments at different spatial and temporal resolution and compare with state-of-the-art methods to show the efficiency of our approach. The observed and estimated air pollution values can be accessed by interested users through a Web visualization tool designed to provide an air pollution map of the city of Antwerp**.**

III. EXISTING METHOD

Air quality should be monitored to keep it under control and improve the quality of life for everyone. Sensors placed in the air continuously report the levels of potentially dangerous gases and compounds to a central monitoring system. With this system, authorities will be able to keep tabs on noise pollution in sensitive areas like schools, hospitals, and no- honking zones. In order to swiftly and confidently implement decisions pertaining to air quality, it is crucial to have access to precise data from dependable devices in the appropriate location. Multiple wireless and wired sensors, a sensor node, gateway, and ESP 8266 Wi-Fi modules are utilised in the creation of an air quality monitoring system

DRAWBACKS

- Possible to track anytime difficult.
- Sensitivity of system is 92%
- More complex circuit

IV. PROPOSED METHOD

The purpose of this project is to use machine learning algorithms to determine how well particulate matter can be predicted in air pollution in INDIA's smart cities. The Adaboost algorithm is selected, and the Smart City Air Pollution dataset is used, to evaluate the use of machine learning in this prediction. This study found that Adaboost's prediction of PM2.5 and the Air Quality Index using Raspberry Pi PICO hardware in Indian smart cities was the most accurate. Raspberry Pi Pico and an Internet of Things (IOT) node are used in a real-world hardware implementation of this project.

ADVANTAGES

- Possible to track anytime from anywhere

- Real time capability
- Sensitivity o system is 90%
- Smaller in size

BLOCK DIAGRAM

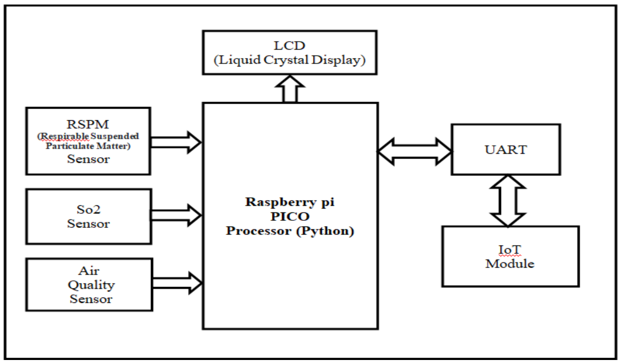


Fig 4.a Smart city air quality monitoring node

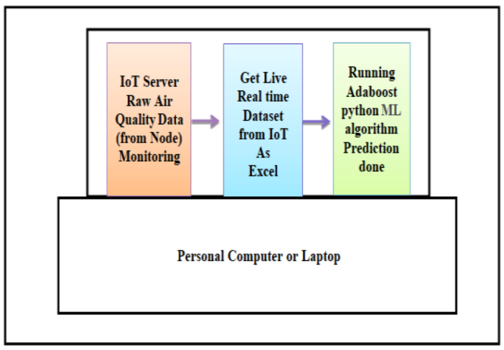


Fig 4.b Server with machine learning (prediction)

1. So2 Sensor
2. Air Quality Sensor
3. IOT module
4. LCD display
5. Power Supply

HARDWARE DESCRIPTION

A. POWER SUPPLY

1. DEFINITION

A power supply (sometimes known as a power supply unit or PSU) is a device or system that supplies electrical or other types of energy to an output load or group of loads. The term is most commonly applied to electrical energy supplies, less often to mechanical ones, and rarely to others.

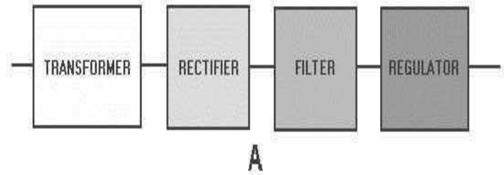


Fig4.1.A (I) Block diagram of a basic power supply

The transformer steps up or steps down the input line voltage and isolates the power supply from the power line. The RECTIFIER section converts the alternating current input signal to a pulsating direct current. However, as you proceed in this chapter you will learn that pulsating dc is not desirable. For this reason a FILTER section is used to convert pulsating dc to a purer, more desirable form of dc voltage.

Simple 5V power supply for digital circuits:

- Brief description of operation: Gives out well regulated +5V output, output current capability of 100 mA
- Circuit protection: Built-in overheating protection shuts down output when regulator IC gets too hot
- Circuit complexity: Very simple and easy to build
- Circuit performance: Very stable +5V output voltage, reliable operation
- Availability of components: Easy to get, uses only very common basic components

2. CIRCUIT DESCRIPTION

This circuit is a small +5V power supply, which is useful when experimenting with digital electronics. Small inexpensive wall transformers with variable output voltage are available from any electronics shop and supermarket. Those transformers are easily available, but usually their voltage regulation is very poor, which makes then not very usable for digital circuit experimenter unless a better regulation can be achieved in some way. The following circuit is the answer to the problem.

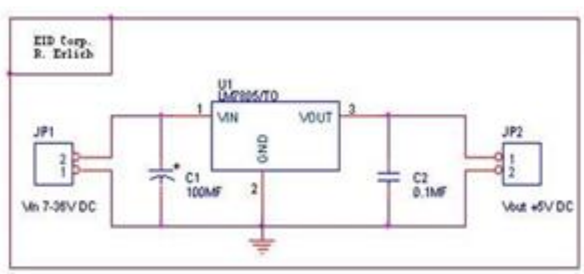
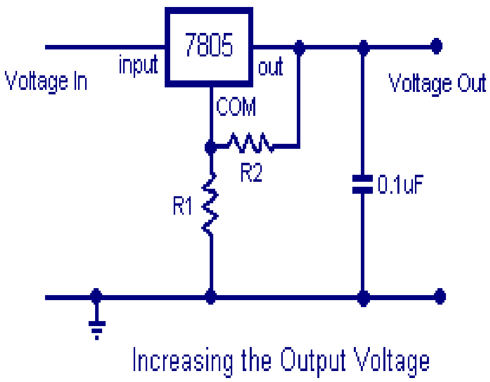


Fig 4.1.A(II) Circuit diagram of the power supply

More output current

If you need more than 150 mA of output current, you can update the output current up to 1A doing the following modifications.



A. 16×2 LCD

B. PRODUCT DESCRIPTION

This is an LCD Display designed for E-blocks. It is a 16 character, 2-line alphanumeric LCD display connected to a single 9-way D-type connector. This allows the device to be connected to most E-Block I/O ports. The LCD display requires data in a serial format, which is detailed in the user guide below. The display also requires a 5V power supply. Please take care not to exceed 5V, as this will cause damage to the device. The 5V is best generated from the E-blocks Multi programmer or a 5V fixed regulated power supply.

Fig 4.1.b 16×2 LCD

FEATURES

- Input voltage: 5v
- E-blocks compatible
- Low cost
- Compatible with most I/O ports in the E-Block range
- Ease to develop programming code using Flow code

C. MQ135 AIR QUALITY SENSOR

Fig 4.1.c MQ135 air quality sensor

A device that is used to detect or measure or monitor the gases like ammonia, benzene, sulfur, carbon dioxide, smoke, and other harmful gases are called as an air quality gas sensor. The MQ135 air quality sensor, which belongs to the series of MQ gas sensors, is widely used to detect harmful gases, and smoke in the fresh air. This article gives a brief description of how to measure and detect gases by using an MQ135 air quality sensor.

Fig 4.1.c MQ135 air quality sensor

A-pins: Here A-pins and B-pins can be interchanged. These are connected to the voltage supply.

B-pins: Here A-pins and B-pins can be interchanged. One pin is used to generate output while the other pin is connected to the ground.

H-pins: There are 2 H-pins, where one is connected to the voltage supply and the other is connected to the ground supply. The circuit has a variable resistor. The resistance across the pin depends on the smoke in air in the sensor. The resistance will be lowered if the content is more. And voltage is increased between the sensor and load resistor.

D. MQ2 SENSOR

MQ2 gas sensor can be used to detect the presence of LPG, Propane and Hydrogen, also could be used to detect Methane and other combustible steam, it is low cost and suitable for different application. The circuit has a heater. Power is given to heater by VCC and GND from power

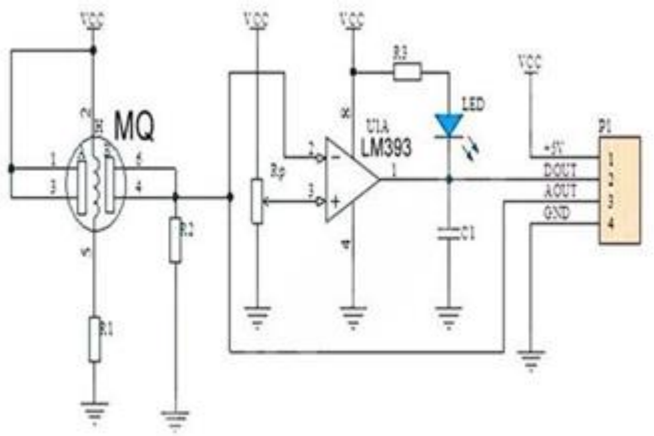


Fig 4.1.D Board Schematic

SPECIFICATIONS:

- Power Supply: 4.5V to 5V DC
- High sensitivity to Propane, Smoke, LPG and Butane
- Wide range high sensitivity to Combustible gases
- Long life and low cost
- Analog and Digital output available
- Onboard visual indicator (LED) for indicating alarm

E. RASPBERRY PI PICO

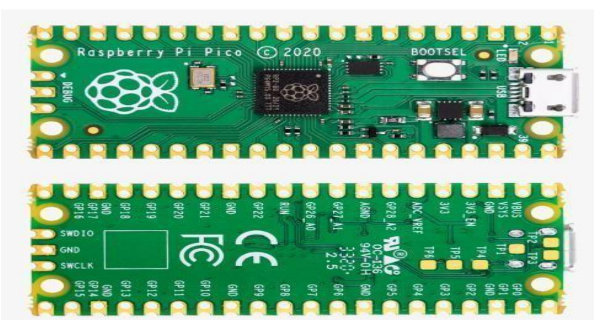


Fig 4.1.E Raspberry pi Pico

A Raspberry Pi Pico is miniature version of the Raspberry Pi that features the RP2040 microcontroller. Raspberry Pi Pico has been developed to be a low-cost and flexible develop me platform for RP2040

SOFTWARE RQUIREDMENT:

- Python IDE
- Python Language
- Webpage Server Php
- Adaboost Algorithm
- Php

SOFTWARE DESCRIPTION A.PYTHON

Python is an interpreted, high-level, general- purpose programming language. Python is dynamically typed and garbage-collected.

Python Features

- Easy-to-learn
- Easy-to-read
- Easy-to-maintain
- A broad standard library
- Interactive Mode
- Portable
- Extendable

Development

Python's development is conducted largely through the Python Enhancement Proposal (PEP) process, the primary mechanism for proposing major new features, collecting community input on issues and documenting Python design decisions. Python coding style is covered in PEP Outstanding PEPs are reviewed and commented on by the Python community and the steering council.

Result

Adaptive Boosting is a good ensemble technique and can be used for both Classification and Regression problems. In most cases, it is used for classification problems. It is better than any other model as it improves model accuracy which can be checked by going in sequence. One can first try decision trees and then go for the random forest to finally apply the boost and implement AdaBoost. Accuracy keeps increasing as we follow the above sequence. The weight-assigning technique after every iteration makes the AdaBoost algorithm different from all other boosting algorithms and that is the best thing about it

V. CONCLUSION

Since air pollution is such a serious problem in urban areas, this paper explores the performance of machine learning models for air pollution prediction in Smart Cities in Malaysia. Adaboost's Smart air quality forecasts were thus the most precise. Thus, further research can be done to develop a more effective algorithm for addressing this issue. Other air pollutants' predictions can be tested, too. In smart cities, where forecasts are made using machine learning, factors like temperature can be used to evaluate the accuracy of air-pollution forecasts.

VI. FUTURE SCOPE

In the future, we aim to investigate the performances of techniques. High-quality data is needed in examples of AdaBoost algorithm. It is also very sensitive to outliers and noise in data requiring the elimination of these factors before using the data.

REFERENCES

[1] J. Sentian, F. Herman, C. Y. Yin and J. C. H. Wui, "Long-term air pollution trend analysis in Malaysia," International Journal of Environmental Impacts 2(4):309-324, vol. 2, 2019.

[2] D. o. Environment, "Air Pollutant Index," [Online]. Available: http://apims.doe.gov.my/public_v2/aboutapi.html.

[3] United States Environmental Protection Agency, "Particulate Matter (PM) Pollution," 1 October 2020. [Online]. Available: https://www.epa.gov/pmpollution/particulate- matter-pm-basics.

[4] S. Ameer, M. Ali Shah, A. Khan, H. Song, C. Maple, S. U. Islam and M. N. Asghar, "Comparative Analysis of Machine Learning Techniques For Predicting Air Quality in Smart Cities," Urban Computing and Intelligence, vol. 7, p. 128325, 2017.

[5] U. Mahalingam, K. Elangovan, H. Dobhal, C.

[6] Valiappa, S. Shresta and G. Kedam, "A Machine Learning Model to Air Quality Prediction for Smart Cities," vol. 19, p. 452, 2019.

[7] R. M. Espana, A. B. Crespo, I. Timon, J. Soto, A. Munoz and J. M. Cecilia, "Air Pollution in Smart Cities through Machine Learning Methods," Universal Computer Science, vol. 24, 2017.

[8] Yi, Wei, Kin Lo, Terrence Mak, Kwong Leung, Yee Leung, and Mei Meng. "A survey of wireless sensor network based air pollution monitoring systems." Sensors 15, no. 12 (2015): 31392-31427.

[9] Y. Xing, Y. Xu, M. Shi, and Y. Lian, \The impact of PM2. 5 on the human respiratory system," vol. 8, no. I, pp. 69{74, 2016.

[10] M. M. Rathore, A. Paul, A. Ahmad, and S. Rho, \US CR," Comput. Networks, no. 2016, 2015.

[11] Asgari, Marjan, Mahdi Farnaghi, and ZeinabGhaemi. "Predictive mapping of urban air pollution using Apache Spark on a Hadoop cluster." In Proceedings of the 2017 International Conference on Cloud and Big Data Computing, pp. 89-93. ACM, 2017.

[12] D. Zhu, C. Cai, T. Yang, and X. Zhou, \A Machine Learning Approach for Air Quality Prediction: Model Regularization and Optimization," no. December,pp. 1{14, 2017.

[13] R. W. Gore, \An Approach for Classic_cation of Health Risks Based on Air Quality Levels," pp. 58{61, 2017.

[14] K. G. Ri, R. Manimegalai, G. D. M. Si, R. Si, U. Ki, and R. B. Ni, \Air Pollution Analysis Using Enhanced K-Means Clustering Algorithm for Real Time Sensor Data," no. August 2006, pp. 1945{1949, 2016.

[15] N. Zimmerman et al., \Closing the gap on lower cost air quality monitoring:machine learning calibration models to improve low-cost sensor performance," no. 2, pp. 1~36, 2017.

[16] I. Bougoudis, K. Demertzis, and L. Iliadis, \EANN HISYCOL a hybrid computational intelligence system for combined machine learning: the case of air pollution modeling in Athens," Neural Comput. Appl., vol. 27, no. 5, pp. 1191{1206, 2016.

[17] C. Yan, S. Xu, Y. Huang, Y. Huang, and Z. Zhang, \Two-Phase Neural Network Model for Pollution Concentrations Forecasting," Proc. - 5th Int. Conf. Adv. Cloud Big Data, CBD 2017, pp. 385{390, 2017.