ISBN: 978-628-95207-8-1. ISSN: 2414-6390. Digital Object Identifier: 10.18687/LACCEI2024.1.1.1938

Air quality measurement and prediction system using artificial neural networks

Jacob Astocondor-Villar[1], Doctor, Carlos Canales-Escalante[1], Master, Raul Vilcahuaman-Sanabria[1], Doctor, Roberto Solis-Farfan[1], Maestro, Miguel Benites-Gutierrez[2], Doctor, Daniel Ipince-Antunez[1], Master, Nestor Gomero-Ostos[1], Doctor
[1]Universidad Nacional del Callao, Perú, jastocondorv@unac.edu.pe, cacanalese@unac.edu.pe, rcvilcahuamans@unac.edu.pe, resolisf@unac.edu.pe, daipincea@unac.edu.pe, ngomeroo@unac.edu.pe
[2] Universidad Nacional de Trujillo, Perú, mbenites@unitru.edu.pe

Abstract*– The purpose of this study is to measure the level of air pollution in the district of Ventanilla and Mi Peru, in Peru. The concentrations of suspended particles range from 2.5 g to 10 g, also known as PM10, and the concentrations of suspended particles less than 2.5 g, also known as PM2.5. The task is to measure CO2, PM2.5 and PM10 pollution to protect the health of people living in the region under study. A system was implemented to measure the concentrations of PM10 and PM2.5 pollutants in CO2 polluted air. The measurement system includes a dust and CO2 sensor, as well as an ambient temperature and humidity sensor, a DHT11 sensor for these measurements, and an ESP8266 module for wireless recording and cloud recording. An Arduino Uno R3 and ESP8266 board use wifi to process the sensor values. a Google Sheets spreadsheet and a PaaS cloud computing service provided by Google. An ANN was chosen because it has proven to be effective in air quality predictions. Compared to other similar works, only one network was created, but several prototypes were developed and evaluated to avoid arbitrariness in design decisions. Data normalization, architecture selection, and activation function selection were three specific components of the NR design that were examined. Finally, artificial neural networks are used to predict PM10 and PM2.5 particulate matter concentrations.

Keywords—Neural networks, pollution measure, air quality, MATLAB.

I. Introduction

Air pollution is defined as the presence in the atmosphere of substances produced by human activity [1] or natural processes [2] in adequate concentrations during a given period and under conditions that may have an impact on the environment. Pollution is a major environmental health problem that affects all countries in the world, both developed and developing [3]. In this case, we have presented the district of Ventanilla and Mi Perú because in the study area we can appreciate large quantities of gases and particles are released that can be harmful to the environment and human health like Fig. 1. show us [4], [5].

Digital Object Identifier: (only for full papers, inserted by LEIRD).
ISSN, ISBN: (to be inserted by LEIRD).
DO NOT REMOVE

Fig. 1. Photographic evidence of latent contamination in Mi Peru, Ventanilla, Callao [6]

Lead and cadmium are the two main metals that are breathed every day [7] by the population of Mi Peru and Ventanilla in various sectors where even educational centers such as the Arturo Padilla school, located in the Industrial Park of Ventanilla, are located. For this reason, it would have generated some consequences among children, including lack of concentration, respiratory and gastrointestinal problems [8], [9], all of them frequent according to the World Health Organization (WHO).

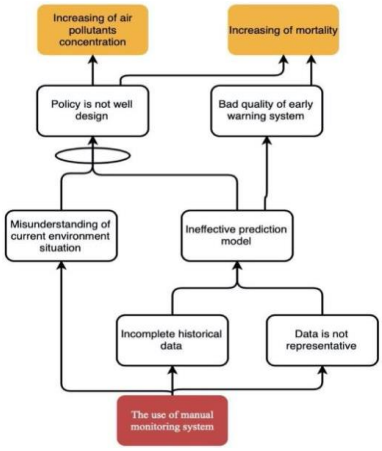

Fig. 2. Photographic evidence of latent contamination in Mi Peru, Ventanilla, Callao [10]

The project consists of the following parts. The first part presents the problem of air pollution in the areas under study: Ventanilla and Mi Peru districts. In the second part, the design, simulation and testing of the whole system including the programming algorithms are carried out. Finally, the third

chapter documents all the tests performed and analyzes the results obtained.

This research project allows us to develop an alternative system to monitor (measure) and predict the air quality in the area of Ventanilla and Mi Peru in which low level sensors are used; the signal obtained previously by the sensor and trained artificial intelligence algorithm will generate the level of pollution in the polluted area. It is intended to report what has been done in the experimentation, use and training of artificial intelligence systems that is useful to determine the air quality in the area and generate a signal that is easy to read for the personnel.

II. Methods

For the initial conception of the project, the following list was taken into consideration, which details the selection process of the technical characteristics and prior analysis required for its execution. Selection of physical components and computer programs.

- Analysis and interpretation of physical and physiological aspects.
- Development of an expert algorithm
- Spectral analysis of air quality and noise reduction methods.
- Development of a method for predicting air quality.
- Simulation of system behavior
- Valuation of the system by experts

System design

The developed system is evaluated by testing pollution conditions in the urban areas of Ventanilla and Mi Perú districts. Studies by the World Health Organization and national entities have pointed out an increase in the concentrations of pollutants in the air in these areas [11]. The evaluation is carried out in areas of high influx of people, such as main avenues, educational institutions, markets and other important places. References [12] and [13] should be consulted for more information on the subject in question.

The measurement areas are determined considering the number of pollutants produced by nearby vehicles and adjacent factories, as indicated in reference 12.

When measuring the air pollution index, it is chosen considering the concentration level of suspended particles [2], [14]. These theories are based on the principle of electrochemical response. Research by [15] and [16] has shown that voltages are related to the concentration of pollutants and greenhouse gases that are sent to the microcontroller platform.

- Identification of necessary hardware and software prerequisites
- The development and execution of the electronic circuit
- Electrochemical sensor programming
- Analyze the data stored in the cloud
- Data analysis
- Data visualization
- System validation by experts

Design of the air pollution measurement system

The elements used to carry out the project are as follows:

- Arduino UNO Microcontroller Board
- SD Data Logger Shield
- MG811 Carbon Dioxide Sensor
- MQ135 Air Quality Sensor
- ESP8266 Wi-Fi Module
- 4x20 I2C LCD Display
- DHT-11 Temperature and Humidity Sensor
- DS3231 RTC Clock Module
- Audible Buzzer

In addition, it was decided to use electrochemical sensors to measure variables related to air quality [13]. The MQx electrochemical sensors show variations in their resistance when exposed to certain gases. These sensors have an internal heater that increases the temperature, allowing the sensor to react with the gases and modify the resistance value [18,19]
The heater may require a specific voltage, which varies depending on the model, within a range of 5 to 12 volts. The sensor operates as a resistor and requires a load resistor (RL) to complete the circuit. This allows a voltage divider to be established that facilitates the reading of the sensor by a microcontroller, as illustrated in Figure 3.

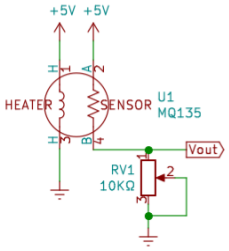

Fig. 3. MQ135 Sensor Diagram

The MQ135 gas sensor module simplifies connections and facilitates application by having a digital output that operates internally with a comparator. Using a potentiometer, it is possible to calibrate the threshold to interpret the presence or absence of gas. These sensors are more sensitive to certain gases than to others, always detecting more than one type of gas.

Sensor calibration

Calibration of the MQ135 sensor for the measurement of carbon dioxide (CO2) where the analysis of air quality in the area is performed.

The characteristic sensitivity curve of the MQ-135 sensor is shown in blue for the air concentration, where the maximum value is 200 ppm, the connection diagram and the characteristic curve shows in the Fig. 4.

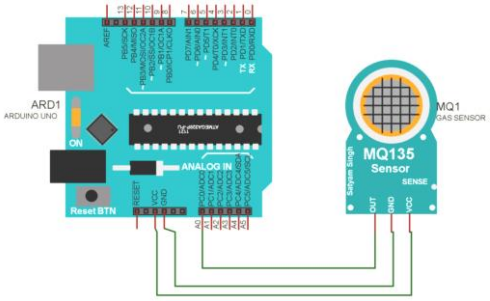

Fig. 4. Diagram of microcontroller and MQ135 sensor

Calibration of MQ135 Sensor

The points have been recorded and classified in their corresponding dimensions in order to determine and calculate the associated potential equation, as well as to perform the analysis to calculate the margin of error, which is presented in TABLE I. Based on this information, the corresponding graph is drawn up, locating the points in their respective coordinates (Figure 5) in order to analyze the concentrations in parts per million (ppm).

The equation that defines the concentration is expressed by the following formula:

$$\text{Concentration}_{CO_2} = 110{,}17\left(\frac{R_s}{R_o}\right)^{-2{,}64}$$

Where:

- R_s: Sensor resistance obtained in first programming
- R_o: Sensor resistance MQ135

TABLE I
MQ-135 SENSOR SENSITIVITY CURVE COORDINATES.

Rs/Ro	Concentration value (ppm)
2.4	10
1.9	20
1.59	30
1.49	40
1.39	50
1.25	60
1.2	70
1.14	80
1.09	90
1.04	100
0.78	200
0.78	200

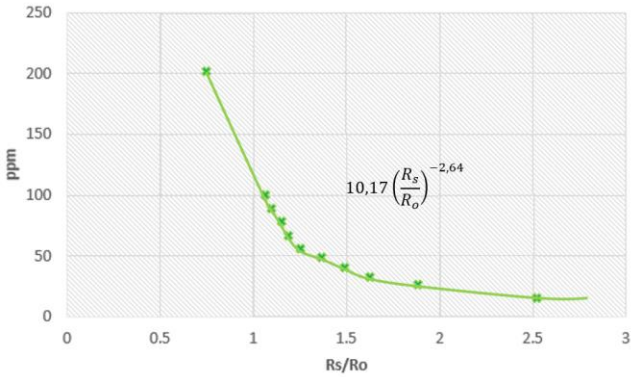

Fig. 5. MQ135 Calibration Curve

III. RESULTS

System diagram

In the measurement of polluted air, sensors (MQx) designed to detect air quality parameters were used. These sensors are connected to the Arduino UNO processor through the available analog channels.

Arduino then processes the information using an algorithm and transmits the data wirelessly through the Wifi module (ESP8266) that is compatible with the TCP/IP protocol. This module establishes the connection with the Google Sheets database. Wifi, whose abbreviation corresponds to Wireless Fidelity, is a technology that uses radio waves for the transmission and reception of data between devices, eliminating the need for wired connections. Wifi wireless technology is integrated into the vast majority of electronic devices, such as mobile phones and computers, allowing these devices to connect quickly to the Internet.

Google Sheets is a free online tool that allows the creation of spreadsheets. The cloud is a contemporary alternative that allows numerous users to access data easily. Its functions include the possibility of sharing documents to facilitate collaborative work between multiple participants. The final tests of the air measurement system were carried out taking into account the diagram presented in Figure 6, which includes various electronic components and chips.

The DHT11 module is responsible for measuring temperature and humidity, with a temperature measurement range of 0 to 50°C and a resolution of 1°C. As for humidity, its measurement range goes from 20% to 90% relative humidity, with a resolution of 1% relative humidity. In this case, to carry out the programming of the ESP8266, instead of using the default firmware of the module, it is programmed manually using the Arduino integrated development environment (IDE) and the Arduino programming language. In addition, some additional libraries are used to manage WIFI connectivity.

The ESP8266, with its hardware firmware, enables the connection to the Internet from the Arduino by linking both modules through the serial port. This allows the execution of AT commands on the ESP8266 and the reception of responses on the Arduino..

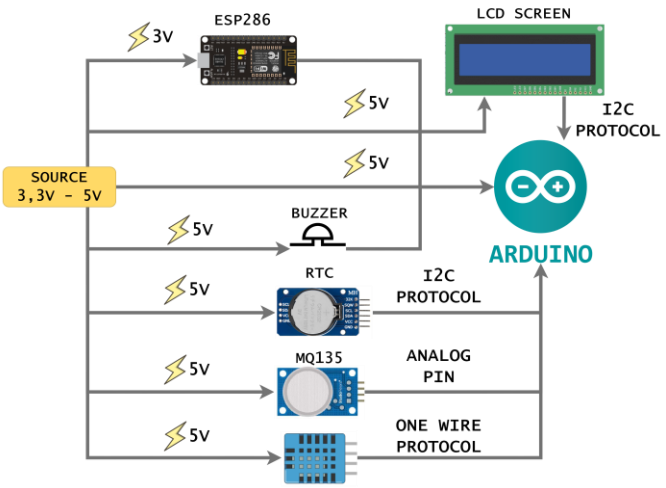

Fig. 6. Block diagram of the implemented system

B. Program flowchart

Figure 7 and 8 displays the flow chart of the executed program, illustrating the procedures conducted, which include monitoring the air quality (specifically CO2 levels) as well as assessing the temperature and humidity of the environment in the designated region of study. Google Sheets is a cloud-based tool that is accessible to anyone with a Google account. On our web platform, you can create and modify spreadsheets. In our work, we engage in the process of acquiring data in a spreadsheet format that is comparable to Microsoft Excel. Google Sheets is a web-based program that can be accessed using several web browsers such as Chrome, Firefox, Internet Explorer 11, Microsoft Edge, and Safari. Google Sheets is universally compatible with all desktop and laptop devices.

Google Sheets has several distinct advantages over alternative choices. One key asset is the ability to collaborate on the same document consistently, regardless of the number of devices, platforms, or locations involved. This is made possible by storing files in the cloud, namely in Google Drive. Modifications are automatically saved and the ability to edit without an internet connection is also provided (via the mobile application and Google Chrome web browser).

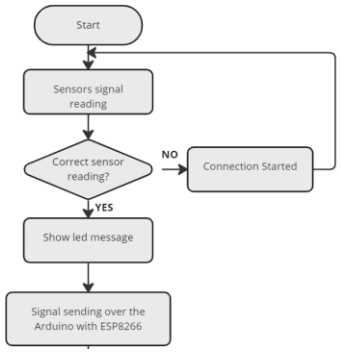

Fig. 7. Flowchart of the implemented program (1st part)

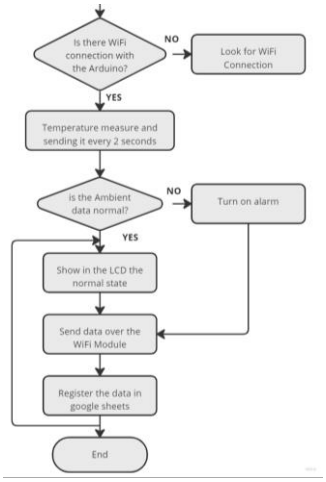

Fig. 8. Flowchart of the implemented program (2nd part).

C. Data measured in closed environment

The Figure 9. shows the minimum and maximum values of the sample taken at the Acho Bridge to determine the exposure ranges of each parameter to which people are exposed.

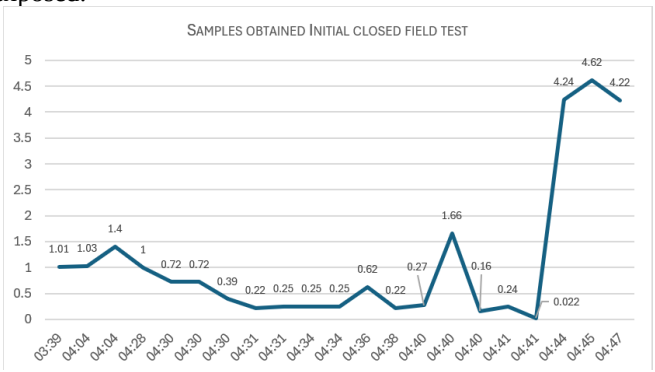

Fig. 9. Samples obtained initial closed field test

D. Data measured in open environment

The Figure 10 displays the minimum and maximum values of the sample collected in the Mi PERU area (main avenue), in order to determine the range of parameters that

individuals are exposed to when they travel through or reside in that location.

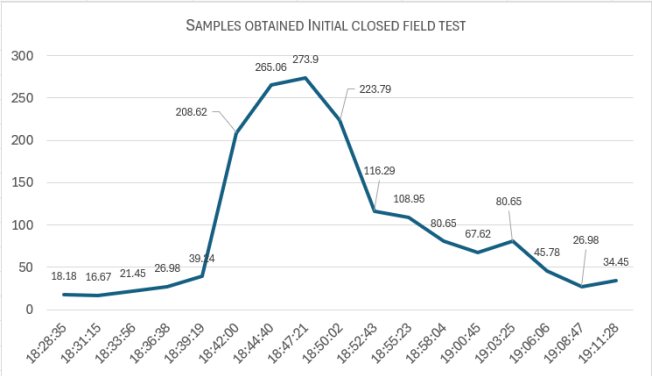

Fig. 10. Samples obtained initial closed field test

Prediction of ANN

Air pollution prediction is accomplished using artificial neural networks (ANNs) in the Matlab software. This software enables the manipulation of ANNs, and the parameters must be specified according to the desired type of neural network.

Total number of layers: The neural network utilized in this research comprises an input layer, an intermediate layer, and an output layer. Figure 8 displays the neural network that will be implemented.

- Neuron count of the input layer: Given that there are 8 daily measurements of particulate matter that describe the characteristics of a day, the input layer of the neural network is configured with 8 neurons to accommodate these data inputs.
- The number of neurons in the intermediary levels cannot be calculated using any existing approach. A frequently employed approach is trial and error, when neurons are systematically adjusted, either increased or decreased, until the optimal outcome is achieved.
- Number of neurons in the output layer: The output layer consists of a single unit, as it is necessary to determine the average contamination level for each day.

Network Performance Curve

Upon executing the program in Matlab, we generate a graph, depicted in Figure 11, which displays the training curves in blue, validation curves in green, and test curves in red. The training curve illustrates the relationship between the validation error and the network's training progress, measured in terms of interactions or epochs. The training of the network concludes either when the error reaches a minimum threshold or when the predetermined number of epochs, as specified in the neural network configuration step, is reached. In this case, it is noted that the error is decreased to 0.0001 in epoch 5, as determined during the configuration stage. If the test curve were to exhibit a substantial increase prior to the validation curve, it would indicate the presence of overfitting, which is not the situation in this case.

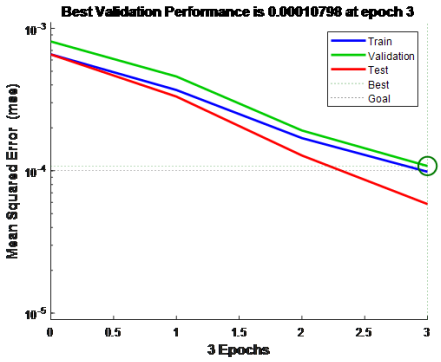

Fig. 11. Network validation and training

Results of the ANN´s training

Figure 12 depicts the anticipated result using a black dashed line. The solid line is the ideal linear regression that establishes a correlation between the outputs and the objectives.

The R value is a quantitative measure that assesses the correlation between the outputs and the objectives. When the value of R is 1, there is an exact correlation between outputs and objectives. A correlation coefficient (R) of 0 shows the absence of a linear relationship between the outputs and the targets.

The training data in this design does not exhibit a significant correlation as the solid line representing the linear regression does not align with the desired targets shown by the black segmented line.

The parameter R demonstrates substantial values throughout the training, validation, and testing stages. The scatter plot, represented by black circles for each occurrence, demonstrates the inadequate correlation among the data points.

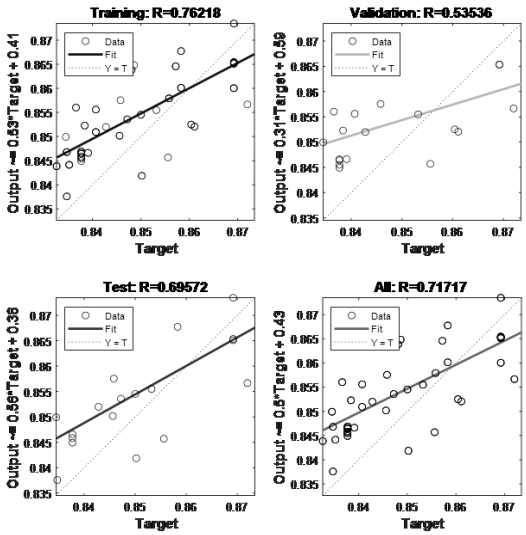

Fig. 12. Linear Regression for ANN Training

Linear Regression for ANN Training

Fig. 13 shows the target values represented by the blue line and the actual output values of the neural network represented by the orange line. From this graph it can be deduced that the neural network model composed of 8 input neurons, 16 neurons in the intermediate layer and 1 neuron in the output layer, fails to meet the expected results, because the two graphs are not similar.

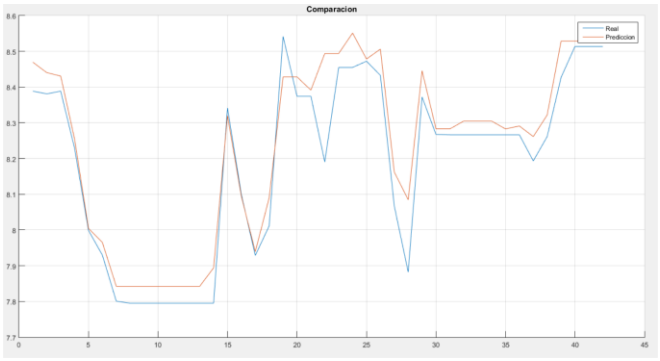

Fig. 13. Plot of actual vs. predicted values at the output of the 16-neuron model in the intermediate layer

III. CONCLUSIONS

It is observed that the increase of air pollution in this case of particulate matter PM10, the humidity factor decreases and temperature increases the pollution with PM10 registering the highest levels at midday, there is also a greater movement of people and vehicles in the area under study.

When implementing the ANN structure and analyzing the data obtained from the prediction with artificial neural networks (ANN), better results are obtained with the network model by varying or increasing the input neurons, the neurons in the intermediate layer and one neuron in the output layer.

In the measurements of air pollution with CO2 changes are observed when measurements are made in a closed environment and then in an open environment in the same area and it is seen that there is pollution.

An air measurement system has been developed with low-cost elements.

REFERENCES

[1] I. Eguiluz-Gracia et al., "The need for clean air: The way air pollution and climate change affect allergic rhinitis and asthma," Allergy Eur. J. Allergy Clin. Immunol., vol. 75, no. 9, pp. 2170–2184, 2020, doi: 10.1111/all.14177.

[2] I. Manisalidis, E. Stavropoulou, A. Stavropoulos, and E. Bezirtzoglou, "Environmental and Health Impacts of Air Pollution: A Review," Front. Public Health, vol. 8, 2020, doi: 10.3389/fpubh.2020.00014.

[3] R. Burnett et al., "Global estimates of mortality associated with longterm exposure to outdoor fine particulate matter," Proc. Natl. Acad. Sci. U. S. A., vol. 115, no. 38, pp. 9592–9597, 2018, doi: 10.1073/pnas.1803222115.

[4] M. Taştan and H. Gökozan, "Real-time monitoring of indoor air quality with internet of things-based e-nose," Appl. Sci. Switz., vol. 9, no. 16, 2019, doi: 10.3390/app9163435.

[5] C.-J. Huang and P.-H. Kuo, "A deep cnn-lstm model for particulate matter (Pm2.5) forecasting in smart cities," Sens. Switz., vol. 18, no. 7, 2018, doi: 10.3390/s18072220.

[6] Y. Zhu and M. Costa, "Metals and molecular carcinogenesis," Carcinogenesis, vol. 41, no. 9, pp. 1161–1172, 2020, doi: 10.1093/carcin/bgaa076.

[7] F. Perera, "Pollution from fossil-fuel combustion is the leading environmental threat to global pediatric health and equity: Solutions exist," Int. J. Environ. Res. Public. Health, vol. 15, no. 1, 2018, doi: 10.3390/ijerph15010016.

[8] M. C. Turner et al., "Outdoor air pollution and cancer: An overview of the current evidence and public health recommendations," CA Cancer J. Clin., vol. 70, no. 6, pp. 460–479, 2020, doi: 10.3322/caac.21632.

[9] Purnomo, Muhammad & Anugerah, Adhe. (2020). Achieving Sustainable Environment through Prediction of Air Pollutants in Yogyakarta using Adaptive Neuro-Fuzzy Inference System. Journal of Engineering Science and Technology. 15. 2995 - 3012.

[10] J. Chen and G. Hoek, "Long-term exposure to PM and all-cause and cause-specific mortality: A systematic review and meta-analysis," Environ. Int., vol. 143, 2020, doi: 10.1016/j.envint.2020.105974.

[11] S. Chen, C. Guo, and X. Huang, "Air Pollution, Student Health, and School Absences: Evidence from China," J. Environ. Econ. Manag., vol. 92, pp. 465–497, 2018, doi: 10.1016/j.jeem.2018.10.002.

[12] L. Morawska et al., "Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: How far have they gone?," Environ. Int., vol. 116, pp. 286–299, 2018, doi: 10.1016/j.envint.2018.04.018.

[13] Y. Wang, Y. Yuan, Q. Wang, C. Liu, Q. Zhi, and J. Cao, "Changes in air quality related to the control of coronavirus in China: Implications for traffic and industrial emissions," Sci. Total Environ., vol. 731, 2020, doi: 10.1016/j.scitotenv.2020.139133.

[14] J. Lelieveld, K. Klingmüller, A. Pozzer, R. T. Burnett, A. Haines, and V. Ramanathan, "Effects of fossil fuel and total anthropogenic emission removal on public health and climate," Proc. Natl. Acad. Sci. U. S. A., vol. 116, no. 15, pp. 7192–7197, 2019, doi: 10.1073/pnas.1819989116.

[15] A. S. . R, G. Priya, S. Nishanth, P. Sai, and V. Kumar, "Smart IoT-Based Greenhouse Monitoring System," Internet Things, vol. Part F1851, pp. 261–271, 2024, doi: 10.1007/978-3-031-09955-7_15.

[16] A. J. Lakshmi, R. Dasari, M. Chilukuri, Y. Tirumani, and A. Pramodkumar, "IoT Based Smart Greenhouse Using Raspberry Pi," presented at the 2023 International Conference on Computer, Electronics and Electrical Engineering and their Applications, IC2E3 2023, 2023. doi: 10.1109/IC2E357697.2023.10262510.

[17] J. Rajagukguk and R. A. Pratiwi, "Emission Gas Detector (EGD) for Detecting Vehicle Exhaust Based on Combined Gas Sensors," presented at the Journal of Physics: Conference Series, 2018. doi: 10.1088/1742-6596/1120/1/012020.

[18] W. G. D. U. Wijerathne, M. L. M. P. Perera, R. H. C. Nuwandika, R. A. K. A. Ranasinghe, K. A. D. C. P. Kahandawaarachchi, and N. D. U. Gamage, "Proximity based intelligent air pollution alerts for garbage disposal sites," presented at the ICAC 2020 - 2nd International Conference on Advancements in Computing, Proceedings, 2020, pp. 500–505. doi: 10.1109/ICAC51239.2020.9357286.