iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 Issue: V Month of publication: May 2022

DOI: https://doi.org/10.22214/ijraset.2022.42186

Adaptive Air Quality Sensing using Machine Learning

Dr. Vijayakumar K[1], Rahul M M[2], Dhamodara Prasath G[3], Aravinth V[4]
[1] Head of the department, Electronics and Instrumentation Engineering, Dr. Mahalingam College of Engineering and Technology, Pollachi
[2, 3, 4] BE. Electronics and Instrumentation Engineering, Dr. Mahalingam College of Engineering and Technology, Pollachi

Abstract: Air pollution is a worldwide problem having impacts on both local and global scales. According to the World Health Organization (WHO), air pollution causes 7 million deaths every year, with 4.2 million attributed to exposure to outdoor air pollution.

Compared to the reference methods defined in the Air Quality Directive, the use of low-cost air quality sensors for monitoring ambient air pollution would reduce air pollution monitoring costs and would also allow larger spatial coverage especially in remote areas where monitoring with traditional facilities is uneasy. Theses multi-sensors were either calibrated against standard gas mixtures or using artificial neural network under field conditions. The later method resulted in mixed results either satisfactory for short periods or generally weak for longer data series. This project mainly focuses on the adaptive calibration of the low-cost sensors using the trained machine learning model.

The performances of several field calibration methods for low-cost sensors, including linear/multi linear regression and supervised learning techniques will be compared. A cluster of ozone, nitrogen dioxide, nitrogen monoxide, carbon monoxide and carbon dioxide sensors will be operated. Subsequently, the accuracy of the predicted values will be evaluated for about a period. These predicted values are fed through the training model and the model will run accordingly and it will adapt all the error situations. This will be useful to the automobile industry in the current situation for smoke emission control and also for many refinery industries in which low-cost sensors can be used with the following modelling for the higher accuracy. This would reduce the cost and also yields more accuracy beyond the varying situation.

Keywords:

1) ***Low-Cost Sensors (LCS)***
2) ***Internet of Things (IoT)***

I. INTRODUCTION

Air pollution is a significant public health problem that has long been a source of anxiety for citizens. An air pollutant is described as any substance that can affect humans, animals, plants, or materials. In the case of humans, an air pollutant may cause or lead to an increase in mortality or serious illness, as well as pose a current or potential health risk. Measurements of air emissions are critical for epidemiology and air quality control, but the scope of ground-based air pollution observations has limitations.

Traditionally, concentrations of air emissions have been monitored by air monitoring stations equipped with standard equipment, allowing for highly reliable monitoring results.

However, the high costs of equipment and servicing make meeting the demands of high-resolution surveillance and assessing the extent of personal exposure impossible. Low-cost air quality sensors have been widely used in air monitoring in recent years due to the benefits of low cost, low power usage, quick operation, and rapid response.

Advances in sensor networks and Internet of Things (IoT) technologies have created new epoch in environmental monitoring facilitating the collection of high-resolution spatiotemporal dataset and filling the gaps that existed within current dataset. Although, IoT technology presents plausible tools to expand current capacity in environmental monitoring, the introduction of Low-Cost Sensors (LCS) is critical to have wider scope and adoption for this purpose. The application of LCS in environmental monitoring, however, has raised several concerns especially pertaining to their accuracy, reliability, in-field applicability and performance. LCS are less precise and less sensitive to compound or variables of interest as their response is largely influenced by cross-sensitivities in the case of gas sensors, particle properties as with particulate matter sensors or environmental factors in both cases, amounting to trade-offs when being used to replace or to supplement existing monitoring solutions

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

II. PROJECT OVERVIEW

As the low-cost sensors are not much accurate than the high-cost sensor, we are implementing trained machine learning model in the calibration program of the low-cost gas or air quality sensor with the help of Artificial Neural Network. First, we are focusing the creation of the training model which consists of the errors caused in the LCS during various environment conditions such as in the moving vehicles. When the LCS is placed in the moving vehicle there will be a vibration caused, due to that there will be some distortion in the electrical output beyond its actual value. So, the accuracy of LCS is changed which causes the error in the output.

These errors are made into the training set for training the artificial neural network. The ANN are very sophisticated modelling techniques able to model extremely complex functions well suited for the calibration of a cluster of sensors. In this project, two types of artificial neural network architectures were considered: radial based functions and Multi-Layer Perceptron (MLP). It consists of artificial units that receive several inputs (either from original data or from the output of other units in the neural network) and typically one hidden layer with hidden units. The activation signal is passed through an activation function to produce the output of the unit. With a defined number of layers and number of units in each layer, the network's weights and thresholds must be set in order to minimize the prediction error made by the network.

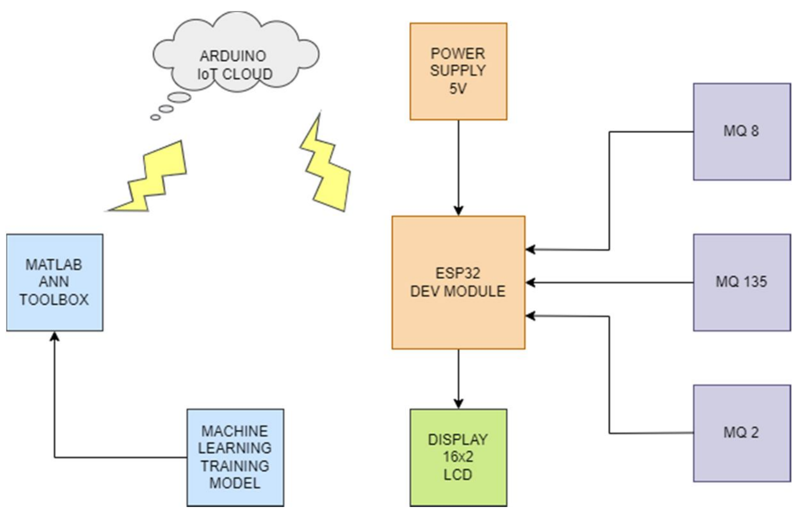

Fig-1 Block diagram

This is the role of the training algorithms which uses iterative techniques called backpropagation. The error of a particular configuration of the network can be determined by running all the training cases through the network, comparing the actual output generated with the desired or target outputs. The differences are combined by an error function which gives the network error as a sum squared error, where the individual errors of output units on each case are squared and summed together. As training progresses, the training error naturally dropped. However, when the selection error stopped dropping, or indeed started to rise indicating that the network was starting to over fit the data, the train was stopped. In this case, the number of hidden units was decreased.

The evaluation of sensor performances took into account hourly values. It was carried out using only values predicted by each calibration method. For each one, regression and difference-based analysis were conducted to evaluate their performance. These included the calculation of the coefficient of determination, comparing the slope and intercept of the regression line with objective values of 1 and 0 respectively. The Mean Bias error (MBE) and the Root Mean Squared Error (RMSE) standardized with the standard deviation of the reference measurements were used to train the model.

III. EXISTING SYSTEM

Wireless sensor networks connect a large number of fixed sensors in multiple places into a single network, enabling long-term, high-resolution surveillance of air contaminants. This uses the low-cost air quality sensors to implement the network of sensors which would require the highest safety measures for protecting its accuracy. Its applications are seen often in health-related studies and tracking individual exposure. By installing sensors on carriers such as vehicles, motorcycles, and drones, the mobile sensor network can provide more compact spatial data than fixed sensors and can achieve stereoscopic tracking of air quality and emission sources. But when it is fixed or mounted on the moving vehicles it may produce the errors due to vibration.

Despite the low-cost sensor's widespread usage, the data's precision has been challenged. Some countries have conducted sensor assessment and calibration tests, and recommendations for the use and evaluation of sensors have been written. However, just a few researchers have tested and calibrated low-cost sensors so far, and their performance under a variety of environmental conditions and time scales is still unknown.

The prototype of the wearable device that uses a sensor system actively measures the ACL strain during the flexion made in knee while movement and then warns the user when they reach a threshold level. For observing every one of the contributing variables to TSF, there are sensors for estimating every one of the elements like knee flexion, shank angle shank acceleration, foot acceleration and the earth ground rection forces. The Android app can be used by the user and the persons related to the user, to check the actual TSF level of the athlete.

IV. PROPOSED SYSTEM

Compared to the existing methods defined in the Air Quality Directive, the use of low-cost air quality sensors for monitoring ambient air pollution would reduce air pollution monitoring costs and would also allow larger spatial coverage especially in remote areas where monitoring with traditional facilities is uneasy. Theses multi-sensors were either calibrated against standard gas mixtures or using artificial neural network under field conditions. The later method resulted in mixed results either satisfactory for short periods or generally weak for longer data series. This project mainly focuses on the adaptive calibration of the low-cost sensors using the trained machine learning model

The performances of several field calibration methods for low-cost sensors, including linear/multi linear regression and supervised learning techniques will be compared. A cluster of ozone, nitrogen dioxide, nitrogen monoxide, carbon monoxide and carbon dioxide sensors will be operated. Subsequently, the accuracy of the predicted values will be evaluated for about a period of time. These predicted values are fed into the training model in the ANN and the model will run accordingly with the help of the MCU and it will adapt all the kind of predefined error situations.

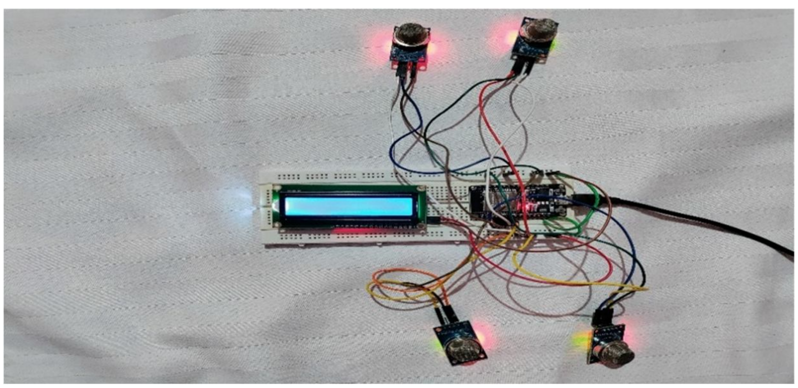

Fig - 2 Proposed System

This will be useful to the automobile industry in the current situation for smoke emission control and also for many refinery industries in which low-cost sensors can be used with the following modelling for the higher accuracy. This would reduce the cost and also yields more accuracy beyond the varying situation and reduces the cost of installation in any kind of industries or domestic purposes.

V. IMPLEMENTATION

A. Hardware Implementation

In hardware implementation, initially Node MCU- ESP8266 was installed in the Arduino IDE Software and tested using the Arduino IDE. Next MQ–135 gas sensor is connected with Node MCU ESP8266 and output is checked based on the atmospheric values. The program is dumped in controller to measure the change in gas levels from atmosphere by spraying some perfumes and igniting smoke near it.

1) Installation of Node MCU ESP8266 in Arduino IDE
2) Testing of Node MCU ESP8266
3) Configuration Node MCU ESP8266 with MQ – 135, MQ – 2, MQ – 8

B. Software Implementation

Testing and Monitoring the MQ – 135

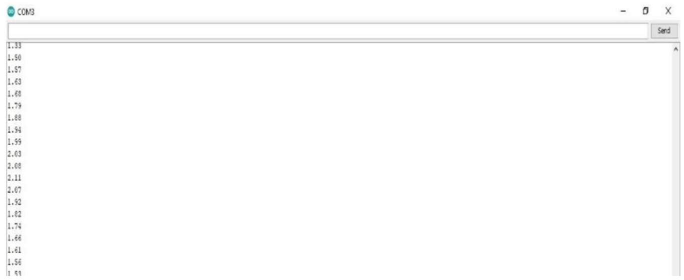

Fig - 3 Voltage Reading

Fig - 4 PPM Reading

VI. CREATING TRAINING SET

The training set was created by keeping the gas sensor in the various situations or conditions like vehicle vibration. The reading is taken in instance of 10 readings per second and average is taken from that, this is repeated for 15 readings and the readings are tabulated to create the sample training model. The readings are taken as PPM and corresponding voltage level. The sample training set for 15 readings are shown in the table.

S.No	Volt	PPM
1	0.99	173
2	0.96	169
3	0.87	162
4	0.76	150
5	0.71	144
6	1.34	207
7	1.88	367
8	1.94	724
9	2.11	783
10	1.74	665
11	1.56	482
12	1.31	356
13	1.25	182
14	1.05	179
15	0.97	164

Table - 1 Training Set

A. ANN Training

The sample data set which is created in the above table is trained in the Artificial Neural Network toolbox in the MATLAB which uses the back propagation neural scheme in this project. There are generally four steps in the training process:

1) Assemble the training data
2) Create the network object
3) Train the network
4) Simulate the network response to new inputs

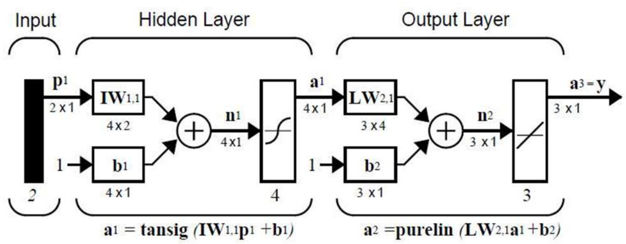

Fig - 5 ANN Training

The sample data set is assembled and fed into the backpropagation neural scheme program in the ANN toolbox by the following procedure and code.

The following code creates a training set of inputs p and targets t. For batch training, all of the input vectors are placed in one matrix.

p = [-1 -1 2 2;0 5 0 5];

t = [-1 -1 1 1];

Next, creation of the feedforward network. Here in this project, the function minmax to determine the range of the inputs to be used in creating the network

net=newff(minmax(p),[3,1],{'tansig','purelin'},'traingd');

At this point, we might want to modify some of the default training parameters.

- net.trainParam.show = 50;
- net.trainParam.lr = 0.05;
- net.trainParam.epochs = 300;
- net.trainParam.goal = 1e-5;

If you want to use the default training parameters, the above commands are not necessary.

Now it is ready to train the network.

- [net,tr]=train(net,p,t);
- TRAINGD, Epoch 0/300, MSE 1.59423/1e-05, Gradient 2.76799/1e-10
- TRAINGD, Epoch 50/300, MSE 0.00236382/1e-05, Gradient 0.0495292/1e-10
- TRAINGD, Epoch 100/300, MSE 0.000435947/1e-05, Gradient 0.0161202/1e-10
- TRAINGD, Epoch 150/300, MSE 8.68462e-05/1e-05, Gradient 0.00769588/1e-10
- TRAINGD, Epoch 200/300, MSE 1.45042e-05/1e-05, Gradient 0.00325667/1e-10
- TRAINGD, Epoch 211/300, MSE 9.64816e-06/1e-05, Gradient 0.00266775/1e-10
- TRAINGD, Performance goal met

The training record tr contains information about the progress of training. Now the trained network can be simulated to obtain its response to the inputs in the training set.

- a = sim(net,p)
- a = -1.0010 -0.9989 1.00
- 18 0.9985

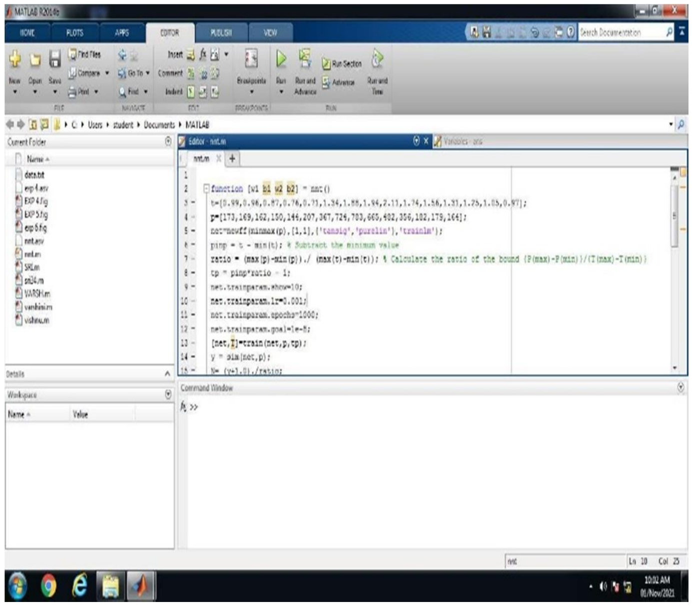

Fig - 6 Implementing Training Set into ANN Toolbox

Now while running the program which we implemented above in MATLAB it will go into new neural network training window. In which it will show up with details mentioned in the Fig 6. A neural network breaks down the input into layers of abstraction. It can be trained using many examples to recognize patterns in speech or images, for example, just as the human brain does. Its behaviour is defined by the way its individual elements are connected and by the strength, or weights, of those connections. These weights are automatically adjusted during training according to a specified learning rule until the artificial neural network performs the desired task correctly.

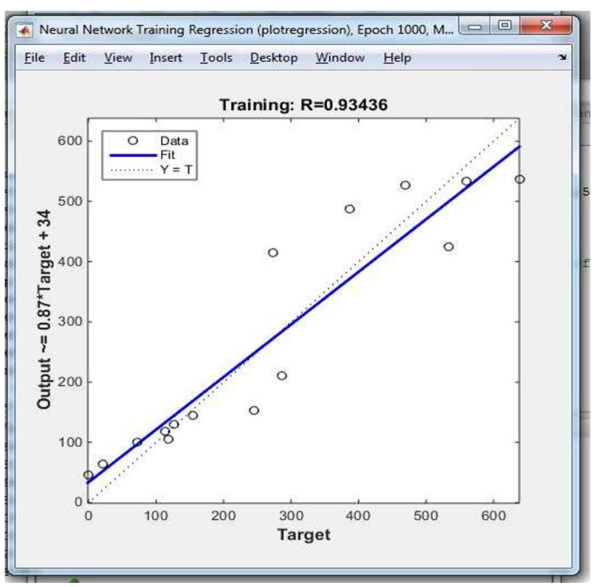

Fig - 7 Plotting the Regression Line

Regression analysis is a technique for modelling and analysing several variables when there is focus on one dependent variable with relationship to one or more independent variables. It provides an understanding on how the typical value of the dependent variable changes when one of the independent variables is changed while other independent variables remain constant.

Now creation of dataset and training is a completed. In this, a single-layer network, with tan-sigmoid transfer function in the hidden layer is used. This is a useful structure for function approximation situations. As an initial guess, five neurons are used in the hidden layer. The network should have single output neurons since there are single target. The Levenberg-Marquardt algorithm is used for training. The training stopped after 200 iterations because the validation error increased. It is a useful diagnostic tool to plot the training, validation and test errors to check the progress of training. The values are trained and extracted in the matrix sheet for single layer

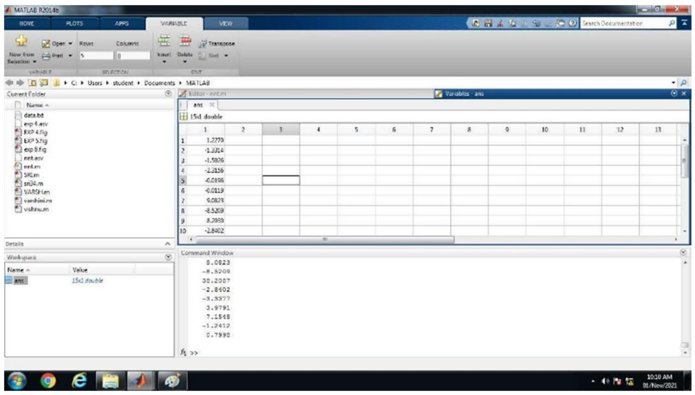

Fig - 8 ANN Single Layer Values

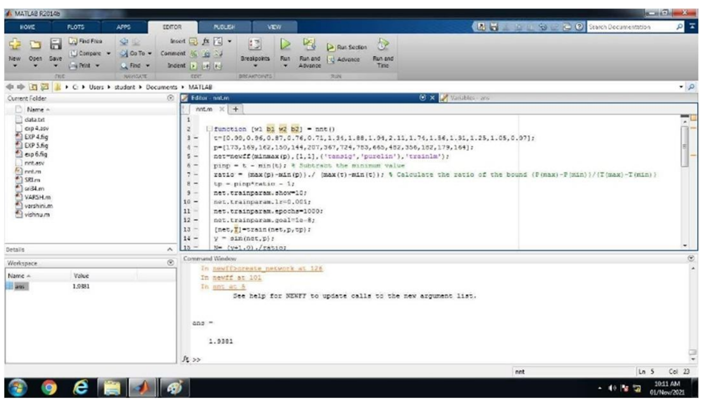

Fig - 9 Final Trained Data

B. *IOT Cloud Implementation*

The Arduino IoT Cloud is designed to help people create great things with ease. It is as simple as connecting a device, creating properties and a dashboard to monitor it. Whenever we create a thing, a sketch is automatically generated, and updates with all changes we make in the cloud.

The Arduino IoT Cloud allows you to register devices that you may control remotely from the dashboard. You'll be able to send data, change settings, as well as receive sensor data and here we are used MQ1345, MQ – 2, MQ – 8

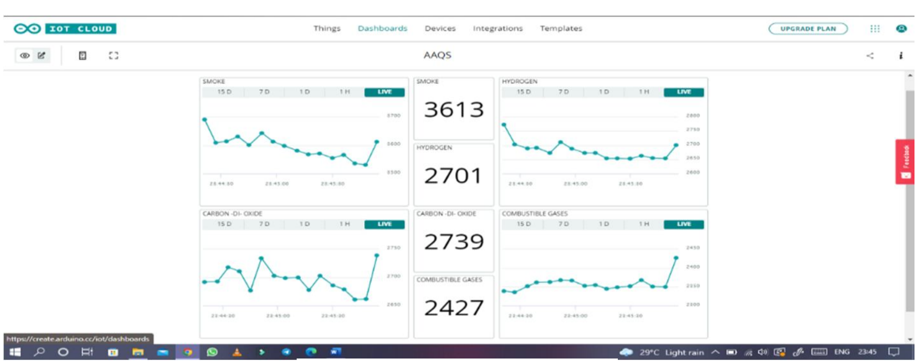

Fig-10 Real-Time Cloud Data

VII. CONCLUSION & FUTURE SCOPE

A. *Conclusion*

Low-cost gas sensors have the capacity to contribute to real time environmental monitoring, providing high resolution spatiotemporal dataset. But Low-cost sensors can be affected by environmental factors such as temperature, vehicle movement and humidity. It is therefore necessary to include these factors in the calibration model to account for their effects on the sensors response in order to ensure effective calibration. This project mainly focuses on making the performance of the low-cost gas sensors to the level of the high-cost sensor which is done through the machine learning training model in with the help of MATLAB. The sample dataset is trained using the backpropagation algorithm and it is implemented in the Node MCU which is connected interfaced with the low cost MQ – 135 gas sensors.

The result of this project showing that there is difference in the voltage levels of the MQ-135 sensor for the corresponding PPM levels it is identified from making it to the zero level. This training data is fed to ANN which produced the good trained values for the sample data set. When the values are averaged and given to the Node MCU it shows the corresponding results. This project also states that the low-cost sensor is capable of acting like high-cost sensor.

B. *Future Scope*

In the future extension, the display is interfaced to the project for live monitoring in the various situation and environmental conditions. Thereby, considering preceding environmental conditions, observation frequencies and other environmental information such as hydro interference, smoky and dusty air to improve this project may happen. It also important to prepare datasets obtained from a range of environments for the practical applications in future work. The multi layers of training network will also added to this project for the improved accuracy. This is also will beneficial in incorporating gas sensors into the WSN network. In addition, it is useful to analyse automotive emissions with integrated mobile sensors in the vehicles and using the model's improvement from this project.

REFERENCES

[1] Ziyue Guan., Richard O. Sinnot., "Prediction of Air Pollution through Machine Learning Approaches on the Cloud", 2018. IEEE/ACM 5th International Conference on Big Data Computing Applications and Technologies (BDCAT)

[2] Anikender Kumar., Pramila Goyal., "Forecasting of Air Quality in Delhi using Principal Component Regression Technique" Atmospheric Pollution Research Volume 2, Issue 4, October 2011, Pages 436-444.

[3] Huixiang Liu., Qing Li, Dongbing Yu, and Yu Gu., "Air Quality Index and Air Pollutant Concentration Prediction Based on Machine Learning Algorithms," Applied Sciences, vol. 9, no. 19, p. 4069, Sep. 2019.

[4] Mohamed Shakir and N. Rakesh., "Investigation on Air Pollutant Data Sets using Data Mining Tool." 2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 480-485.

[5] Jambi Ratna Raja Kumar, Rahul K. Pandey, Biplab K. Sarkar., "Pollutant Gases Detection using the Machine Learning on Benchmark Research Datasets", Procedia Computer Science, Volume 152, 2019, Pages 360-366, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2019.05.005.\

[6] J. Mazumdar and R. G. Harley, "Recurrent Neural Networks Trained with Backpropagation Through Time Algorithm to Estimate Nonlinear Load Harmonic Currents," Industrial Electronics, IEEE Transactions on, vol. 55, pp. 3484-3491, 2008

[7] L. Sun, D. Westerdahl, Z. Ning, "Development and Evaluation of a Novel and Costeffective approach for Low-cost NO2 Sensor Drift Correction", Sensors 17 (8) (2017)

[8] Goh, C.; Kamarudin, L.; Shukri, S.; Abdullah, N.; Zakaria, A. "Monitoring of Carbon Dioxide (CO2) Accumulation in Vehicle Cabin", In Proceedings of the 2016 3rd International Conference on Electronic Design (ICED), Phuket, Thailand, 11–12 August 2016; pp. 27–432.

[9] Demuth, Howard B., and Mark H. Beale. "Neural Network Toolbox for Use with MATLAB", User's Guide. Natick, Mass: MathWorks, 2001.

[10] Tripathy J. R, Tripathy H. K, Nayak S. S., "Implementation of Artificial Neural Network Training Data in Micro-Controller Based Embedded System", Orient. J. Comp. Sci. and Technol;7

[11] Laurent Spinelle, Michel Gerboles, Maria Gabriella Villani, Manuel Aleixandre, and Fausto Bonavitacola, "Field Calibration of a Cluster of Low-cost Commercially available Sensors for Air Quality Monitoring". part b: No, co and co2. Sensors and Actuators B: Chemical, 238:706–715, 2017.

[12] Winsen, "Air Quality Gas Sensor", MQ - 135 datasheet, Oct. 2014 [Revised Sept. 2018].

[13] Espressif Systems, "Node MCU Development Board", ESP8266 datasheet, Jan. 2015 [Revised Nov. 2020].

[14] Arduino LLC, "Arduino Software Release Notes", Arduino Project. Retrieved September 25, 2019

[15] Natick, Massachusetts, "MATLAB and ANN Toolbox Release 2014b'', The MathWorks, Inc., United States.

[16] Skorpil, Vladislav & Stastny, Jiri. (2006), "Neural Networks and Back Propagation Algorithm", IEEE Transactions on, vol. 62, pp. 3501-3617, 2016.

[17] How2electronics, "IoT Based Air Quality Index Monitoring with ESP8266 & MQ135".