Monitoring and Predicting African Rural Household Air Pollution Using Internet of Things and Artificial Intelligence

Samson O. Ooko[1*] and Enatha Rweyemamu[2]

[1]Adventist University of Africa, Kenya
[2]Dares Salaam Institute of Technology, Tanzania

https://doi.org/10.56893/ajhes2024v03i01.06

Abstract

According to a 2020 report from the World Health Organization (WHO), household air pollution has led to over 3 million deaths globally, with recent statistics showing a worsening situation in Africa. Integrating Internet of Things (IoT) and Artificial Intelligence (AI) technologies can help address this global challenge. IoT enables real-time data collection for monitoring pollution levels, while AI algorithms predict pollution before it reaches hazardous levels. However, existing solutions are not tailored to the African context, where wood fuel is a primary pollutant, and they predominantly focus on monitoring rather than prediction. This study presents the design and implementation of an IoT-based solution for monitoring and predicting indoor air pollution in rural African households. The system collects data in real time and transmits it to the cloud for storage, processing, and analysis, with alerts to users when pollution is detected. An AI model was successfully trained and tested to predict indoor air pollution based on the collected data. The results indicate that this approach significantly improves the accuracy and timeliness of pollution alerts, potentially reducing health risks associated with indoor air pollution. The successful implementation and testing of the system demonstrate its potential for broader applications in various indoor environments.

Keywords: Artificial intelligence (AI), household air pollution, Internet of Things, air pollution prediction

Introduction

People tend to spend more time indoors, which exposes them to indoor air pollution and increases their exposure to hazards connected with it. Premature deaths from indoor air pollution are estimated at 4 million annually (Saini et al., 2020a). Individual productivity is affected by indoor air pollution, which is estimated to cost between \$20 and \$200 billion each year (Saini et al., 2020a). Closed rooms may cause quicker accumulation of pollutants, resulting in greater indoor air pollution than outdoor pollution. The use of solid fuels such as wood fuel, coal, and biomass, which are common sources of fuel in African households when not entirely combusted in poorly ventilated dwellings, can result in elevated levels of hazardous gases such as carbon monoxide, nitrogen oxides, benzene, and particulate matter (Dey & Chattopadhyay, 2016). When people are constantly exposed to them, they might cause chronic health problems. Other sources of indoor pollution in

*Corresponding Author: ookos@aua.ac.ke

metropolitan areas include building materials, air conditioning systems, ventilation systems, heating systems, use of items containing chemicals, and other human-related activities.

The growing Internet of Things technologies (IoT) have capabilities that have been used to monitor pollution in various environments. First, a study by Gola et al. (2019) aimed to detect chemical air pollution in patients' rooms. Second, Ana et al. (2019) focused on the pollution conditions in a hairdressing saloon where many chemicals are used. Additionally, Saini et al. (2020b) suggested a mobile web service that allows people to share images of air pollution. Kodalli et al. (2020) presented a Smart Indoor Air Pollution Monitoring Station. Such solutions do not apply to the unique situations of rural African households, which is the focus of this study.

Some applicable solutions for indoor air pollution monitoring and alerting systems have been presented in recent studies (Al Ahasan et al., 2018; Jo et al., 2020; Janarthanan et al., 2022; Sá et al., 2022; and Rakib et al., 2022). Sensors are used in IoT to monitor pollution levels in real time, with the acquired data being communicated to the cloud or analyzed locally. Various notification mechanisms are used, with the majority focused on mobile user notifications. Obtaining real-time alerts may not be sufficient to solve this problem. Pollution prediction is required before an individual is exposed to high quantities of pollutants. Compared to monitoring the concentration of indoor pollutants and providing threshold-based alerts, prediction using AI-based technologies can be faster, cheaper, and noninvasive (Wei et al., 2019).

The proposed study, therefore, was aimed at providing first, an alert to users in the event poor air quality is detected in an African household and second, training and testing a prediction of indoor air pollution by integrating cloud-based technologies that provide powerful visualization and analytics to improve the efficiency and effectiveness of poor air quality by showing carbon monoxide (CO) level, particulate matter (PM) level, temperature and humidity of the household in real-time. The system has a dashboard to provide an Air Quality Index (AQI) in color codes. The following objectives guided the study; to investigate the available open source technologies that can be used in monitoring indoor air pollution; to prototype a real-time household pollution monitoring system using state-of-the-art IoT technologies; and to propose and develop an AI-based household pollution prediction algorithm and early warning system for rural African household. This was an improvement to current systems, which are not customized for the African household setting, where wood fuel is the main source of energy. Additionally, the use of alerts based on thresholds is not as effective and accurate as Artificial Intelligence (AI). The use of AI also enhances the possibility of predicting future occurrences.

The paper is structured as follows: The literature review is presented in the next section. The methodology is detailed in section three, and the results are presented and discussed before the conclusion is provided.

Literature Review

In this section, related studies are reviewed. First, Internet of Things-based air quality monitoring solutions are presented, followed by AI technologies for indoor air pollution prediction.

IoT in Indoor Air Quality Monitoring

To begin with, Smart Air is proposed by Jo et al., (2020). This solution monitors indoor air quality by using sensors to measure carbon dioxide concentrations, aerosols, Volatile Organic Compounds (VOC), humidity, and temperature. Data are transmitted to the cloud using long-term evolution (LTE) technologies, where the user can view the data from a web portal. This solution lacks an automatic alert method considered essential for air quality monitoring, similar to the e-nose (Taştan & Gökozan, 2019), a proposed real-time air quality monitoring system. The system can monitor selected air parameters such as nitrogen dioxide, carbon monoxide, particulate matter, temperature, and humidity. The ESP32 WiFi microcontroller is used to send data to the Blink cloud platform. The blink platform limits the ability to get alerts to only users within the same network as the system.

In a systematic literature review by Saini et al. (2020b), it was noted that there is a need for more studies that focus on real-time automatic alerts for indoor air quality conditions and the need to integrate automatic control of ventilation systems. The study recommends using the latest calibrated sensor technology in designing the real-time monitoring application to improve accuracy and efficiency. A study by Lapshina et al. (2019) also lacked an automatic real-time alert system. It is based on an Arduino microcontroller that monitors the level of carbon monoxide indoors using an infrared gas module and displays it on a screen. Al Ahasan et al. (2018) proposed a prototype that uses an Arduino microcontroller and an MQ135 sensor to measure a variety of toxic gases and display them on a screen. However, their study lacks automatic alert functionality and does not take into consideration other causes of indoor air pollution.

Abdullah et al. (2019) proposed a system for monitoring the total volatile organic compound (TVOC) concentration in an indoor setting. However, the system can notify users of TVOC levels in both outdoor and indoor environments. It is based on Arduino and a CCS811 sensor. This solution does not consider other major courses of pollution in an indoor environment, which are essential for determining air quality. This is the same for IDust (Marques et al., 2018), which is a system aimed at monitoring the particulate matter in the air. Another study focused on monitoring the concentration of toxic gases and not any other parameters (Maulana Azad et al., 2017). The solution uses Raspberry Pi with an array of MQ gas sensors. Data is sent to a web portal for visualization and analysis. This study is similar to a study by Husain et al. (2016), except that an Arduino is used instead of a Raspberry Pi. Other studies that did not monitor the causes of indoor pollution in African households include those by Yoon et al. (2024), Marques and Pitarma (2016), Rocha et al. (2016), as well as Dey and Chattopadhyay (2016).

A study by Gola et al. (2019) aimed to monitor chemical air pollution in patients' rooms. These unique conditions may not apply to other general indoor settings, rendering the proposed study inapplicable. This also applies to a study by Ana et al. (2019), which focuses on hairdressing saloon pollution conditions. Firdhous et al. (2017) proposed a prototype for tracking ozone concentrations around a photocopy machine to solve a unique indoor scenario.

Indoor Air Pollution Prediction

Wei et al. (2019) reviewed the techniques that have been used to predict indoor air pollution has been given. Artificial Neural Networks (ANN), regressions, decision trees, and PCR models have been used to effectively predict indoor air quality. A study by Mad Saad et al. (2017) classified sources of indoor air pollutants. Saad et al. (2017) used supervised learning to predict indoor pollution. The study shows the great potential of AI in predicting pollution. However, given the varying environment and the need to monitor pollution from PM 2.5, existing algorithms may not be applicable. A systematic literature review by Adil and Ahmed (2024) on the application of AI in predicting pollution showed a growing interest in the use of AI. From the reviewed studies, there is, therefore, a need to collect data from an African setting and develop a model an AI model that can be used to predict pollution in the same setting.

Methodology

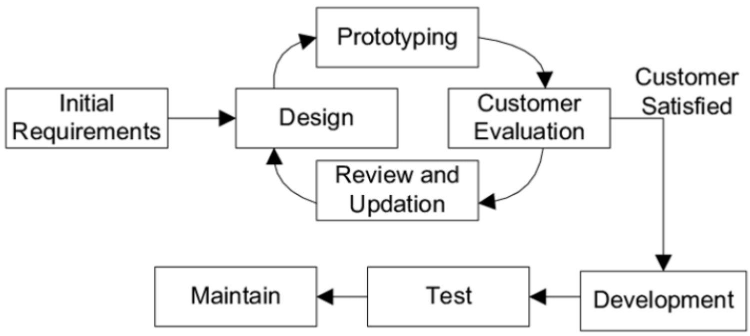
This section details the methodology used in the study and the design and implementation of the proposed system. The system design includes the sensor, processing, communication, power supply, monitoring, prediction components, and the hardware and software employed.

System Development Method

A prototype software development model was used to develop the system. This model was selected given the researchers' need to understand the problem to meet the system requirements at the early stage of development. Moreover, this methodology enabled the researchers to get users' feedback and better understand their needs to refine the prototype. The process was iterative to ensure that all user needs were met, as shown in Figure 1.

Figure 1

Prototyping Steps Used in the Development of the Solution

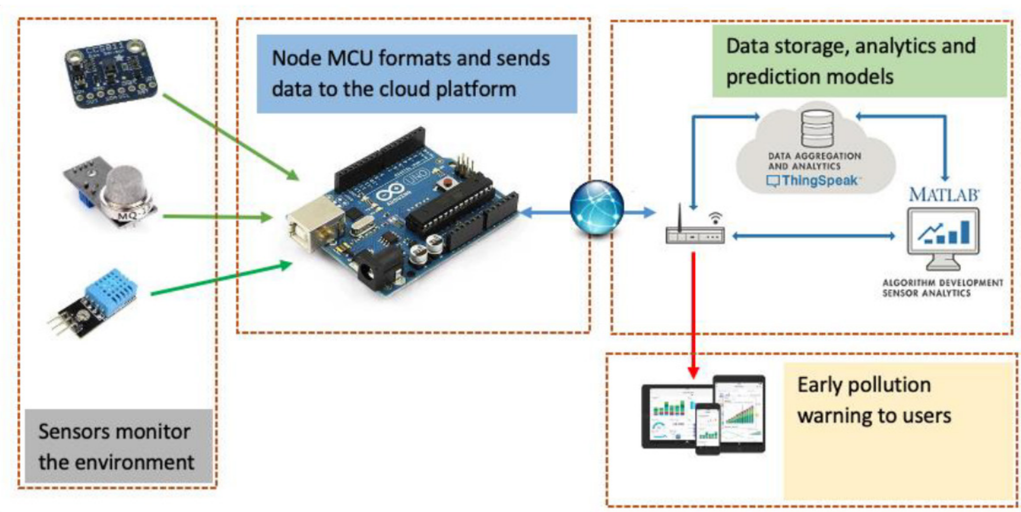


System Design

A PM2.5 particulate matter sensor, MQ7 carbon monoxide (CO) sensor, and DHT11 humidity and temperature sensor were connected to an Arduino UNO microprocessor to collect specific gas, matter, and environmental variables. Data was transferred through the internet to the cloud platform for IoT projects known as Thingspeak via the GSM module for storage and subsequent analysis. The pollution was predicted using an Artificial Intelligence (AI) approach. The system architecture is illustrated in Figure 2.

Figure 2

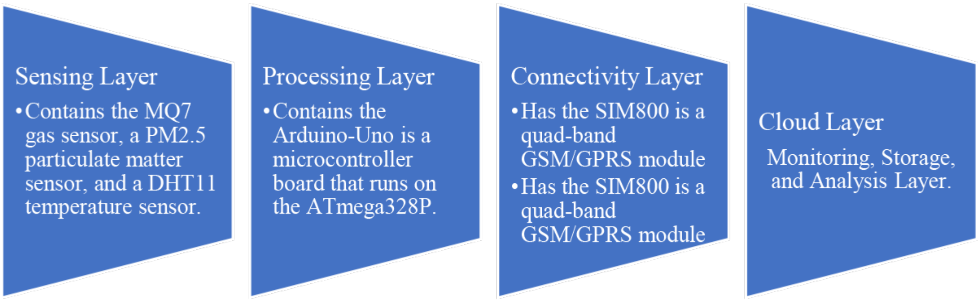
System Architecture Showing Different Components of the Systems



The system comprises four main layers, as shown in Figure 3. The specifications for the components are given in Table 1.

Figure 3

The Four Layers Making Up the System



The system components and their specifications are as follows:

The Temperature/Humidity Sensor (DHT11) has an accuracy of ±2°C and ±5%RH and operates in conditions ranging from 0 to 50°C and 20-80% humidity. It has an active current of 0.3mA and a sleep current of 60uA, with a power supply range of 3.5-5.5V. The sensor has a response time of 80 µs and dimensions of 12.6 mm × 5.83 mm by 16 mm.

The MQ7 Gas Sensor has a resolution of 20-2000 and operates at -20°C ±2°C with 65% ±5% relative humidity. It requires a power supply range of 2.5V-5.0V and has a response time of less than 1 second. The dimensions are 40.0 mm × 21.0 mm.

The PM 2.5 Sensor has an accuracy of ±10% and operates in 0-300µg/m³ conditions. It had an active current of less than 200uA, a power supply range of 4.95-5.05V, and a response time of less than 10 s. The dimensions are 46 mm × 30 mm × 17.6 mm.

The Arduino-Uno Embedded Processor with AtMEGA328p features 32 KB of flash memory and 2 KB of RAM. It has an active current of 50mA and a sleep current of 20mA, with a power supply range of 3.3-5V and an input reference clock of 16 MHz. The dimensions are 68.6 mm by 58.4 mm.

The LCD 24 × 4 Display Module has an accuracy of 95% and operates at 25°C with an active current of 1.5mA and a sleep current of 25mA. It has a power supply range of 5V; data hold time of 20n, and data output delay of 120n. The dimensions are 125 – 39 mm × 14 mm.

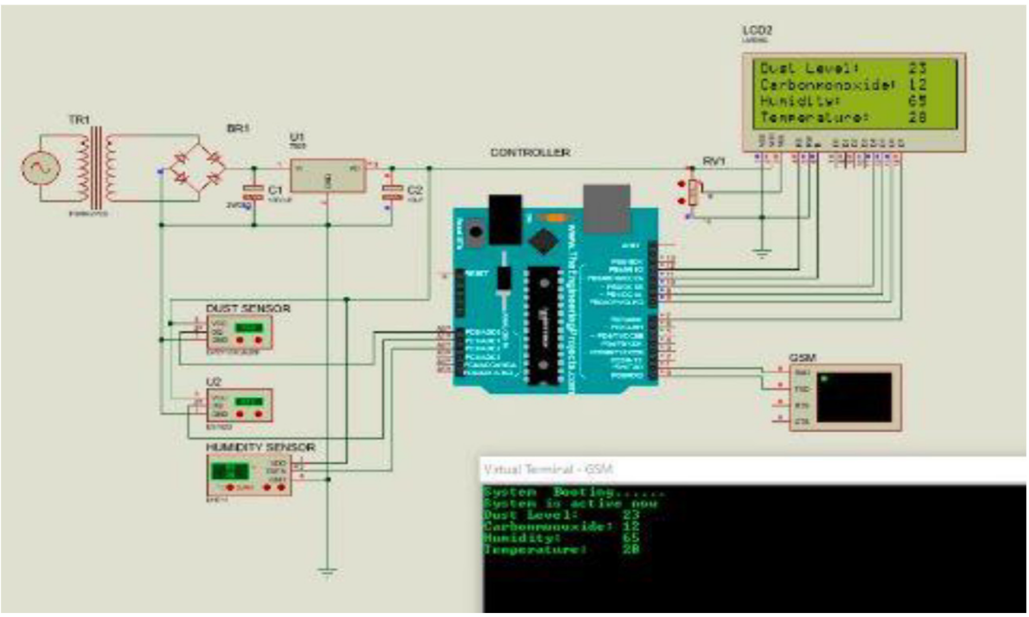
The SIM800C Communication Module features 24Mbit of Flash memory and 32Mbit of RAM. Its active current is 0.88mA, its power supply range is 3.4V-4.4V, and its performance speed is 85.6kbps for both uplink and downlink. The dimensions are 17.6 mm by 15.7 mm and 2.3 mm. The other components used in the proposed system were jump wires, a buzzer, and a case for enclosing the entire system.

System Implementation

Proteus software was used to design and simulate the system. The prototype was further extended by adding sensors to the microcontroller and sending data from the sensors to the ThingsSpeak cloud platform for storage and analysis using a GSM module. The MQ7 gas sensor was connected to the analog pin of the UNO via pin A2, the PM2.5 sensor was connected to the analog pin via pin A3. The LCD module was connected to the analog pins A4 and A5 of a microcontroller, while the DHT11 sensor was connected to the digital pin of the microcontroller via pin 2. The GSM module was connected to the digital pin of the UNO via pins 3 and 4, and LEDs were connected to the digital pin of the Arduino UNO. Figure 4 shows the implementation of the Proteus software.

Figure 4

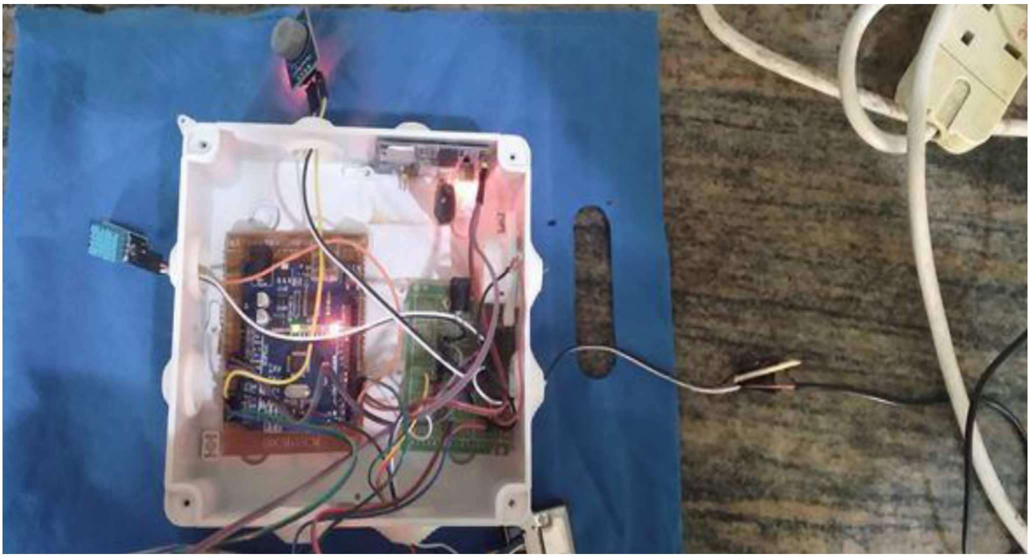
System Proteus Simulation Layout



The system was installed in a household with a single window and charcoal fire source. Within two months, data was continuously captured and transferred to the cloud using the GSM module. Since the system was situated inside the house, it was powered by the main grid energy. However, future implementations will involve designing and implementing a solar powering system. Figure 5 shows the system components in a casing.

Figure 5

System Components in a Casing



System Results and Discussion

Data Transmission, Storage, and Monitoring

The system was installed in the kitchen of a house, and data were collected every 5 minutes for 24 hours. This project used a SIM800 GSM module to transmit data from the system to the cloud. The data acquired had noise to the point of negative values, mostly from the MQ7 gas sensor. The ThingsSpeak platform was utilized, and several graphs were created from the acquired data. From the collected data, the room temperature is normal when there are no activities in the kitchen; however, when the stove is turned on and the kitchen is active, the room temperature rises. It was noted that humidity increases when a kitchen is busy too and then goes back to normal.

Figure 6

A Plot of CO Over Time from the Collected Data

Figure 6 depicts the detection of CO from the kitchen; CO was found to be normal until the time of the day when cooking was on when the levels went beyond acceptable levels. This observation was similar in the particulate matter plots, and we can conclude that the particulate matter or the room air quality changes whenever there is a rise in particles in the air, that is, whenever the stove is on or when the number of people in the room increases, and the readings were higher in the morning, noon, and evening, with the highest value being 63, which is moderate and dangerous to people and leads to pollution-related health problems.

AQI Prediction Using AI

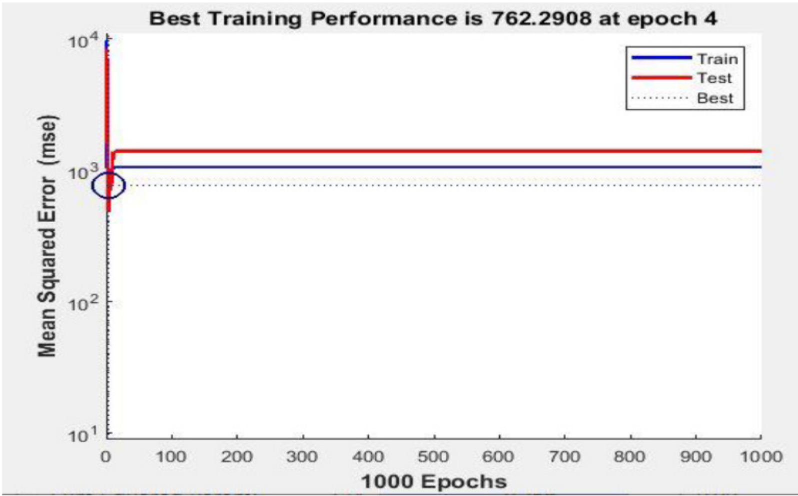
In a related study (Sahoo et al., 2021), data were collected for one week for 8 hours daily. A systematic review of indoor air quality prediction (Saini et al., 2020b) showed that such predictions involve measuring humidity, temperature, and air velocity, and determining how they affect carbon monoxide and particulate matter. Monitoring particulate matter, which is a major contributor to indoor air pollution in African households, is also important.

Therefore, the dataset used for training was collected from a system installed in a household within two months. Data was captured continually and transferred to the cloud through the GSM module, and data was sent to the cloud every five minutes daily for two months. The collected data included humidity, temperature, carbon monoxide, and particulate matter in the targeted household.

The Air Quality Index (AQI) was calculated using the data obtained from the system. AQI prediction was carried out using MATLAB software, which is used to address complex problems, such as prediction applications. It uses dynamic time series software of a nonlinear autoregressive with external (exogenous) input (NARX) to forecast series y(t) given d previous values of y(t) and other series of x (t). The network design that created a NARX neural network was tested numerous times with different numbers of hidden neurons, and an output of 9.2 performance was obtained with 18 hidden neurons and 2 delays. The algorithm was developed using Bayesian Regularization by dividing the dataset into three groups: training, validation, and testing. The algorithm randomly divided the dataset into 19 target time steps, where the training group had 13 target time steps with 70% of the dataset adjusted according to its error. The validation group had three target time steps with 15% of the dataset, which helped to measure network generalization and to halt training when generalization stopped improving. The testing network had three target time steps, with 15% of the dataset having no effect on training, thus providing an independent measure of network performance during and after training. Following the training, several plots were created and examined, with the mean squared error (MSE), which is the number of mistakes in statistical models identified at 18 hidden layers, with two indicating that the error acquired was quite modest. The smaller the mean square error, the better the prediction accuracy. When a model had no errors, the MSE was zero, indicating that the match between the actual and anticipated datasets was significant, as illustrated in Figure 7.

Figure 7

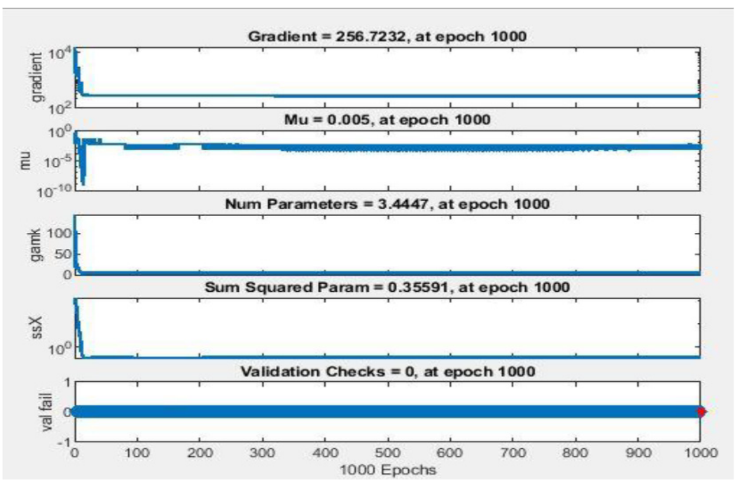
Graph of Mean Square



Because the gradient is a first-order iterative optimization algorithm for finding a local minimum of a differentiable function by taking repeated steps in the opposite direction of the gradient of the function at the current point, the gradient was 256.7232 at epoch 1000 with 3.447 parameters and 0.35591 as the sum squared parameter giving zero as the validation check, as shown in Figure 8.

Figure 8

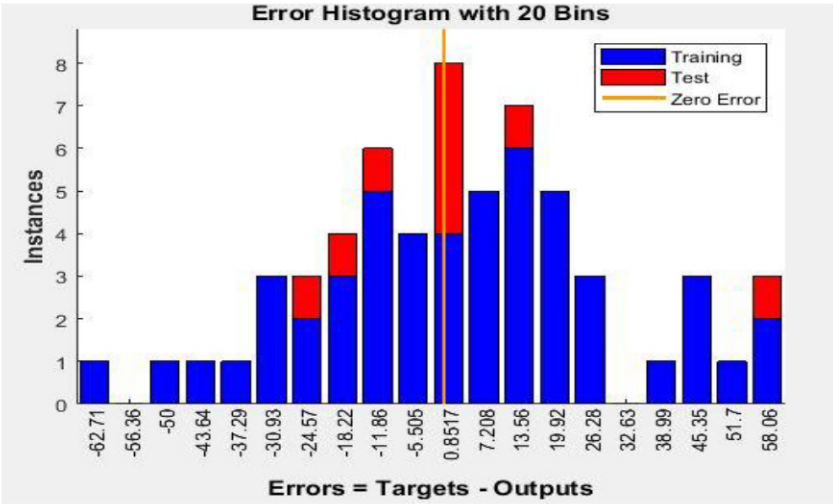
Graph of Gradient



As shown in Figure 9, the error histogram, which is a histogram of the errors between target values and predicted values after training a feedforward neural network, had zero error at 0.8517 and training between 0-4 with test lying between 4-8, indicating that many samples from the dataset used had an error range between 4 and 8.

Figure 9

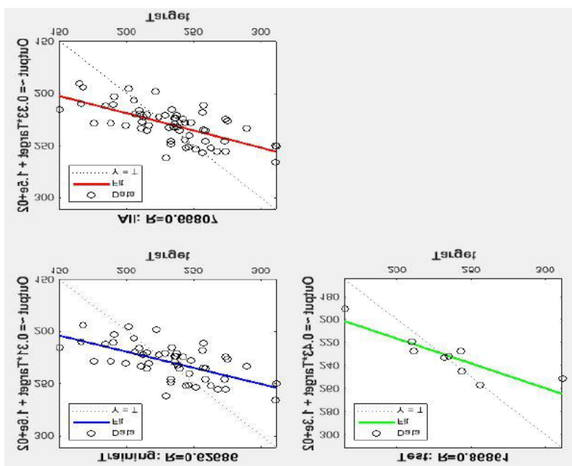
Graph of Error Histogram Plot



A linear regression value was calculated, which depicts the best-fit line or curve for the connection between the factors and anticipated data. The regression was 0.62586 during training and 0.868 during testing, yielding an overall regression value of 0.66807, indicating a close relationship, as shown in Figure 10. We observed that the testing regression improved, implying that there was no nose compared to the training procedure.

Figure 10

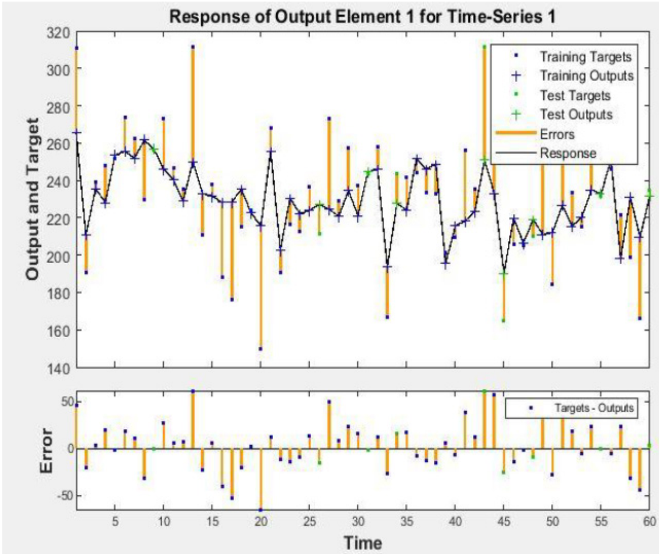
Graph of Regression Plot



The response of prediction was also obtained between 210 and 260 with few errors in between, as shown in Figure 11

Figure 11

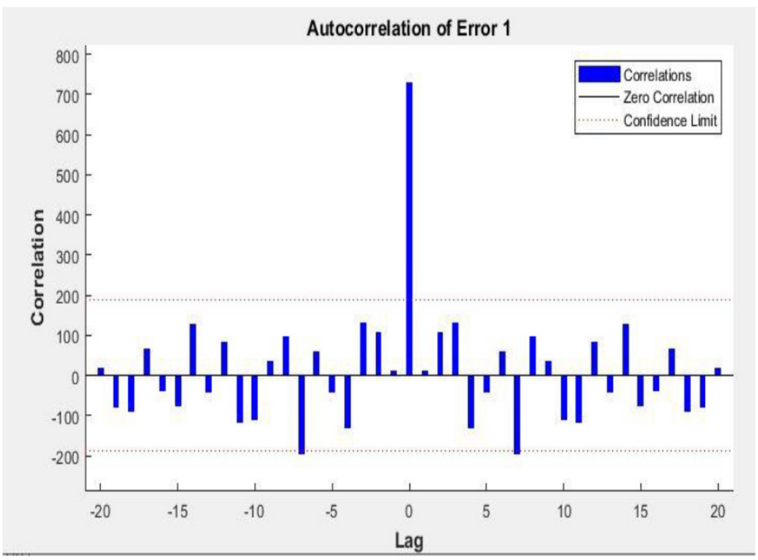
Response on Prediction



Finally, the autocorrelation was determined by utilizing the same time series twice: once in its original form and once after it had lagged. The computed autocorrelation of error was 1, indicating that the resemblance between a given time series and its lagged version across subsequent time intervals was perfect, as shown in Figure 12.

Figure 12

Autocorrelation



Carbon monoxide gas (CO), particulate matter, temperature, and humidity were collected from a kitchen that used charcoal as a fire source, and data were transmitted to the cloud platform using a GSM module and alerted through a beep sound from a buzzer. The results show that the air quality index (AQI) is high at night when windows and doors are closed, posing health risks in Africa. The success of the prediction model trained for this research highlights the potential for utilizing IoT and AI technologies in addressing indoor air quality issues. These results and data collected can be used as a starting point for further research and development of similar systems to improve indoor air quality and mitigate health risks. The successful implementation of this system shows the potential for using IoT and AI to monitor and predict air quality in other indoor spaces and environments, making the technology accessible and useful for a wide range of applications.

References

Al Ahasan, Md. A., Roy, S., Saim, A. H. M., Akter, R., & Hossain, Md. Z. (2018). Arduino-Based Real Time Air Quality and Pollution Monitoring System. *International Journal of Innovative Research in Computer Science & Technology*, *6*(4), 81–86. https://doi.org/10.21276/ijircst.2018.6.4.8

Alabdullah, A. J., Farhat, B. I., & Chtourou, S. (2019). Air Quality Arduino Based Monitoring System. *2nd International Conference on Computer Applications and Information Security, ICCAIS 2019*, 1–5. https://doi.org/10.1109/CAIS.2019.8769529

Ana, G. R., Alli, A. S., Uhiara, D. C., & Shendell, D. G. (2019). Indoor air quality and reported health symptoms among hair dressers in salons in Ibadan, Nigeria. *Journal of Chemical Health and Safety*, *26*(1), 23–30. https://doi.org/10.1016/j.jchas.2018.09.004

Dey, D., & Chattopadhyay, A. (2016). Solid Fuel Use in Kitchen and Child Health in India. *Artha Vijnana: Journal of The Gokhale Institute of Politics and Economics*, *58*(4), 365. https://doi.org/10.21648/arthavij/2016/v58/i4/153076

Firdhous, M. F. M., Sudantha, B. H., & Karunaratne, P. M. (2017). IoT enabled proactive indoor air quality monitoring system for sustainable health management. *Proceedings of the 2017 2nd International Conference on Computing and Communications Technologies, ICCCT 2017*, 216–221. https://doi.org/10.1109/ICCCT2.2017.7972281

Gola, M., Settimo, G., & Capolongo, S. (2019). Indoor air in healing environments: Monitoring chemical pollution in inpatient rooms. *Facilities*, *37*(9–10), 600–623. https://doi.org/10.1108/F-01-2018-0008

Husain, A. M., Rini, T. H., Haque, M. I., & Alam, M. R. (2016). Air Quality Monitoring: The Use of Arduino and Android. *Journal of Modern Science and Technology*, *4*(1), 86–96.

Janarthanan, A., Paramarthalingam, A., Arivunambi, A., & Vincent, P. M. D. R. (2022). Real-time indoor air quality monitoring using the Internet of Things. *Proceedings of the 2022 3rd International Conference on Intelligent Computing, Instrumentation and Control Technologies: Computational Intelligence for Smart Systems, ICICICT 2022*, 99–104. https://doi.org/10.1109/ICICICT54557.2022.9917990

Jo, J., Jo, B., Kim, J., Kim, S., & Han, W. (2020). Development of an IoT-Based indoor air quality monitoring platform. *Journal of Sensors*, *2020*, 13–15. https://doi.org/10.1155/2020/8749764

Lapshina, P. D., Kurilova, S. P., & Belitsky, A. A. (2019). Development of an Arduino-based CO2 Monitoring Device. *Proceedings of the 2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, ElConRus 2019*, 595–597. https://doi.org/10.1109/EIConRus.2019.8656915

Mad Saad, S., Andrew, A. M., Shakaff, A. Y. M., Mohd Saad, A. R., Kamarudin, A. M. Y., Zakaria, A., Saad, S. M., Andrew, A. M., Shakaff, A. Y. M., Dzahir, M. A. M., Hussein, M., Mohamad, M., & Ahmad, Z. A. (2017). Pollutant recognition based on supervised machine learning for Indoor Air Quality monitoring systems. *Applied Sciences (Switzerland)*, *7*(8), 11665–11684. https://doi.org/10.3390/s150511665

Marques, G., & Pitarma, R. (2016). An indoor monitoring system for ambient assisted living based on internet of things architecture. *International Journal of Environmental Research and Public Health*, *13*(11). https://doi.org/10.3390/ijerph13111152

Marques, G., Roque Ferreira, C., & Pitarma, R. (2018). A system based on the internet of things for real-time particle monitoring in buildings. *International Journal of Environmental Research and Public Health*, *15*(4). https://doi.org/10.3390/ijerph15040821

Maulana Azad (2017). *International Conference on Recent Innovations in Signal Processing and Embedded Systems (RISE-2017) : 27th-29th October 2017 : venue: Department of Electronics and Communication Engineering, MANIT, Bhopal, India - 462003*. 27–29.

Rakib, M., Haq, S., Hossain, M. I., & Rahman, T. (2022). IoT Based Air Pollution Monitoring & Prediction System. *2022 International Conference on Innovations in Science, Engineering and Technology, ICISET 2022*, 184–189. https://doi.org/10.1109/ICISET54810.2022.9775871

Rocha, Á., Correia, A. M., Adeli, H., Reis, L. P., & Teixeira, M. M. (2016). New advances in information systems and technologies. *Advances in Intelligent Systems and Computing*, *445*, 13–21. https://doi.org/10.1007/978-3-319-31307-8

Sá, J. P., Alvim-Ferraz, M. C. M., Martins, F. G., & Sousa, S. I. V. (2022). Application of the low-cost sensing technology for indoor air quality monitoring: A review. In *Environmental Technology and Innovation* (Vol. 28). Elsevier

B.V. https://doi.org/10.1016/j.eti.2022.102551

Saad, S. M., Andrew, A. M., Shakaff, A. Y. M., Dzahir, M. A. M., Hussein, M., Mohamad, M., & Ahmad, Z. A. (2017). Pollutant recognition based on supervised machine learning for Indoor Air Quality monitoring systems. *Applied Sciences (Switzerland)*, *7*(8). https://doi.org/10.3390/app7080823

Sahoo, L., Praharaj, B. B., & Sahoo, M. K. (2021). Air Quality Prediction Using Artificial Neural Network. *Advances in Intelligent Systems and Computing*, *1248*, 31–37. https://doi.org/10.1007/978-981-15-7394-1_3

Saini, J., Dutta, M., & Marques, G. (2020a). Indoor air quality monitoring systems based on internet of things: A systematic review. *International Journal of Environmental Research and Public Health*, *17*(14), 1–22. https://doi.org/10.3390/ijerph17144942

Saini, J., Dutta, M., & Marques, G. (2020b). Indoor air quality prediction systems for smart environments: A systematic review. *Journal of Ambient Intelligence and Smart Environments*, *12*(5), 433–453. https://doi.org/10.3233/AIS-200574

Taştan, M., & Gökozan, H. (2019). Real-time monitoring of indoor air quality with internet of things-based e-nose. *Applied Sciences (Switzerland)*, *9*(16). https://doi.org/10.3390/app9163435

Wei, W., Ramalho, O., Malingre, L., Sivanantham, S., Little, J. C., & Mandin, C. (2019). Machine learning and statistical models for predicting indoor air quality. *Indoor Air*, *29*(5), 704–726. https://doi.org/10.1111/ina.12580