Rocznik Ochrona Środowiska

Volume 27 Year 2025 ISSN 2720-7501 pp. 152-163

https://doi.org/10.54740/ros.2025.013 open access

Received: September 2024 Accepted: January 2025 Published: March 2025

Study of AQI Monitoring System of Indoor Environment Using Machine Learning Model and IoT Device

Isha Talati[1], Kunjan Shah[2], Om Patel[2], Jaymin Tanna[2], Akshat Jain[2], Ankit D. Oza[3], Amruta Arun Yadav[4], Mohammed J. Alshayeb[5], Mohammad Amir Khan[6**], Saiful Islam[7]*

[1]*Department of Mathematics, School of Technology, Pandit Deendayal Energy University, Raisan, Gandhinagar, India*

[2]*Unitedworld Institute of Technology, Karnavati University, Gandhinagar, Gujarat, India*

[3]*University Centre of Research and Development, Chandigarh University, Mohali, Punjab, India*

[4]*Department of Civil Engineering, Yeshwantrao Chavan College of Engineering, Nagpur, India*

[5]*Department of Architecture, College of Architecture and Planning, King Khalid University, Abha, Saudi Arabia*

[6]*Department of Civil Engineering, Galgotias College of Engineering and Technology, Greater Noida, India*

https://orcid.org/0000-0003-1550-0393

[7]*Civil Engineering Department, College of Engineering, King Khalid University, Abha, Saudi Arabia*

[*]*corresponding author's e-mail: ishashah1111@gmail.com*

[**]*corresponding author's e-mail: amir.khan@galgotiacollege.edu*

Abstract: Indoor air quality has a direct impact on human health. Thus, it's essential to comprehend the various aspects of indoor air quality. It supports both the implementation of preventative measures and the monitoring of indoor air pollution. Monitoring and forecasting air pollution is extremely essential, especially in developing countries like India. This study proposes a system that employs ESP8266 (NodeMCU) data sent to the cloud to monitor the levels of air pollutants such as ozone, particle matter, carbon monoxide, carbon dioxide, temperature, and total volatile organic compounds. Our sensors include the ozone sensor MQ-131, the dust sensor GP2Y1010-AU0F, the TVOC sensor AGS02MA, the carbon monoxide sensor MQ-9, the carbon dioxide sensor MQ-135, and the humidity sensor DHT11. The IoT device continuously shows the indoor air quality level (IAQL). The next step was to accurately anticipate the Internal Air Quality Level (IAQL) and pollution levels from dangerous gases for the next seven days using the LSTM, Seasonal ARIMA, and Linear Regression models. The Authors could accurately predict the observations of the following seven days after using data from the previous ninety days to create our best model. This implies that our model can accurately predict the values for each parameter with an accuracy of at least 95%. Therefore, we believe such a solution would be advantageous if a large-scale installation were implemented. If consumers can remotely verify the air quality in their homes, the pollution in the interior atmosphere will decrease. This has the potential to make civilization healthier.

Keywords: Internet of Things, Indoor Air Quality, Linear Regression, LSTM, SARIMAX

1. Introduction

In 2020, household air pollution was linked to over 237000 fatalities worldwide, including over 237000 deaths of children under five, according to the World Health Organization. The Central Pollution Control Board reported that Ahmedabad had the fourth-most polluted air in the nation. Different air pollutants like $PM_{2.5}$, PM_{10}, temperature, humidity, CO_2, CO, O_3, VOCs etc. Here, the authors discuss the causes of many contaminants and their effects on health.

Particulate Matter ($PM_{2.5}$, PM_{10}) combines liquid droplets and solid particulate debris in the atmosphere. Ash, smoke, dust, and dirt are a few examples. Though lung damage, respiratory problems, etc., might be made worse by PM_2. Conversely, inhaling large airborne particles can result in heart attacks, bronchitis, high blood pressure, and asthma episodes. The typical summer temperature ranges from 23 to 25.5°C, whereas the typical winter temperature ranges from 20 to 23.5°C due to differences in humidity and climate. Stroke, heart failure, and respiratory infections can all be brought on by frequent temperature fluctuations. Mist, fog, and poorly functioning ventilation systems are a few instances that indicate the amount of humidity both inside and outside, weariness, lethargy, restlessness, and skin damage caused by humidity.

Carbon dioxide is an odorless, colorless, slightly acidic-tasting, non-flammable gas at ambient temperature. Cement manufacturing, deforestation, and other processes release CO_2. One of the main health problems resulting from breathing in too much carbon dioxide is headaches and breathing difficulties. Colorless, tasteless, and odorless, carbon monoxide is a combustible gas that is marginally less dense than air. Examples include the smoke from cigarettes, burning charcoal, and running automobiles. Overexposure to carbon monoxide can result in heart failure, neurological system failure, and brain damage. It is a pale blue gas with a strong odor

composed of three oxygen atoms. Ultraviolet radiation contains ozone. Ozone causes congestion, throat irritation, coughing, and chest pain. VOCs are often defined as organic compounds with a high vapor pressure at standard room temperature. Moth repellents, aerosol sprays, wood preservatives, and air fresheners are a few examples. VOCs can produce nausea, recurrent headaches, and irritation of the nose, eyes, and throat.

As seen in the following section, several researchers have employed various methods and strategies to address these health concerns and deaths.

2. Literature Review

Various air pollutants, such as CO_2, CO, $PM_{2.5}$, PM_{10}, and volatile organic compounds (VOCs), influence internal air quality, which can have detrimental health effects. Poor air quality has been associated with various health issues. It is a major factor in the spread of COVID-19 Agarwal et al. (2020). 5 million people die +yearly from illnesses caused by poor indoor air quality. Medical care costs due to poor indoor air quality exceed $150 billion in the USA. Poor IAQ is in the top five environmental threats to health and well-being worldwide Saini et al. (2020). Indoor air quality (IAQ) is a critical factor affecting health and well-being, given that humans spend 90% of their time indoors Liu et al. (2021). Indoor air pollution is overlooked compared to outdoor air pollution, even though indoor air pollution levels are twice as high as outdoor air pollution levels Javier et al. (2021).

Researchers have studied IAQ prediction and forecasting methods to address this issue, using machine learning and sensor technologies to develop accurate models for real-time monitoring and warning systems. Wei et al. (2019) conducted a literature review about Machine Learning and statistical models for predicting IAQ. $PM_{2.5}$ and PM_{10} were the air pollutants that were studied the most. The most popular statistical models are ANN, Multiple Linear Regression, partial least squares, and decision trees. Krishan et al. (2019) proposed an LSTM algorithm to predict O_3, $PM_{2.5}$, NO_x, and CO concentrations in an area of NCT-Delhi. The LSTM model was applied and found to be more effective in handling complexities and accurately forecasting air quality. Many deep-learning models can be used to determine indoor air pollutants. Artificial Neural Networks (ANN) and Reinforcement Learning (RL) models are good at modeling the complex relationships between inputs and outputs in non-linear systems, even when those systems are not fully understood Agarwal et al. (2020) and Nan et al. (2021), proposed an ANN model to forecast concentrations of pollutants, namely O_3, NO_2, $PM_{2.5}$ and PM_{10} for the current and next 4 days. The model is also fitted with real-time correction to change forecasts dynamically based on data from the past few days.

Computational fluid dynamics (CFD) is a method to analyze and predict the behavior of fluids and gasses when they flow through an environment. It can be accomplished by utilizing energy consumption and regulating thermal comfort. To model the time-series data of $PM_{2.5}$, Dhakal et al. (2021) proposed a deep LSTM model with parameters including dew, ambient pressure, wind speed, humidity, maximum ambient temperature, and minimum ambient temperature. Kalaivani and Mayilvahanan (2021) review many papers using the same and predicting IAQ using ML algorithms. A particle swarm optimizer based on CFD combined with a Back-propagation neural network (BPNN) is used to predict concentrations of CO_2 and $PM_{2.5}$ Li et al. (2022). Ventilation can be used to create environments with IAQ. Tian et al. (2022) also use a BPNN to predict indoor environment indicators. They used parameters such as predicted mean vote, draft rate, air age, and air change efficiency to predict energy performance and IAQ. In urban areas, the AQI reaches very low levels due to several factors, such as vehicular emissions, high traffic, meteorological conditions, and other natural factors. $PM_{2.5}$ is a very harmful pollutant in Kathmandu Valley, Nepal. ML algorithms are used to predict indoor air quality when continuous monitoring is not possible through smart sensors. Kapoor et al. (2022) trained ten algorithms to predict IAQ using CO_2 concentrations. They concluded that the optimized Gaussian process regression (GPR) outperforms the other algorithms.

Internet of Things (IoT) based sensors are used nowadays to measure IAQ dynamically. Sensors can yield incorrect values when trying to predict IAQ. To tackle this, Zhao et al. (2019) proposed an IAQ detector that measures the IAQ data, which can be transmitted using wired communications, short range wireless communications and directly to the cloud. End users can monitor the IAQ of their homes and offices everywhere. Stefano et al. (2019) proposed an HVAC system that considers user habits and IAQ provided by IoT sensors. The data from the sensors is then used to improve the accuracy of estimation of the occupancy rate in the building to prevent discomfort. William et al. (2019) propose an ML algorithm based on a Deep Reinforcement Learning (DRL) Artificial Intelligence algorithm to control air conditioners and ventilation fans to maintain thermal comfort and air quality while minimizing energy consumption. Warning systems can be deployed to alert people if the air quality index (AQI) reaches a critical level. Balram et al. (2019) also estimated $PM_{2.5}$ values using a Bayesian regularized neural network. They used a Support Vector Machine (SVM) classifier on the estimated $PM_{2.5}$ concentrations and proposed an air quality warning system. Vagner et al. (2020) conducted sensor validation through three ML algorithms to classify predicted values as correct or incorrect.

The algorithms were Random Forest (RF), K-Nearest Neighbor (K-NN), and Multi-Layer Perceptron. This highlights the importance of air quality control systems and choosing important pollutants for IAQ prediction. Similarly, Ha et al. (2020) used the Kalman filter to combine the IAQ index (IAQI) and humidex data to form an enhanced indoor air quality index (EIAQI). IAQ readings from the sensors can be sent directly to the cloud. Kodali et al. (2020) monitor IAQ and send alerts to the end user over the Internet. Heating, ventilation, and air conditioning systems (HVAC) are also used to regulate thermal comfort and acoustics. However, they can be the most energy-consuming among the air quality control systems. Shanmugaraja et al. (2021) proposed a system that reads data from sensors and uploads it to the Thingspeak cloud using the Thingspeak API. The data is then monitored and analyzed on the Thingspeak platform. Liu et al. (2021) used a Zigbee wireless network to transmit the data to a database in the cloud through a collector gateway. The data is stored in Modbus RTU format. Tagliabue et al. (2021) propose a system consisting of an ANN trained on monitored data that triggers ventilation through IoT communication. Xie et al. (2021) proposed a Bayesian network (BN) model to forecast AQI and warn users about the risk of poor air quality. The authors concluded that exhaustively using the proposed BN model can achieve a monitoring and early warning accuracy rate of 90%. Rastogi and Lohani (2024) combined temperature, humidity, CO, PM_{10}, $PM_{2.5}$ and CO_2 for the same.

Furthermore, they used an extended Kalman Filter to clean up the data and make it more reliable by identifying and removing inconsistencies such as missing data points, errors, and outliers. Majdi et al. (2024) developed a neural network that takes inputs such as temperature, humidity and CO_2 from a control and a monitoring system and outputs the VOCs in the air. Many pollutants can be fused together to predict IAQ.

According to a literature review, no author has examined interior air quality while considering every contaminant. The majority of procedures focus on $PM_{2.5}$ and PM_{10}. However, research suggests that $PM_{2.5}$ and PM_{10} are not the only pollutants that affect health. This article discusses temperature, humidity, CO_2, CO, O_3, VOCs, $PM_{2.5}$, and PM_{10} air pollutants. The authors have created an Internet of Things device to measure the amounts of different pollutants. Furthermore, the authors have forecasted pollution levels for the next seven days using machine learning.

3. IoT-based internal Air Quality Monitoring System Prototype

3.1. Choice and application of sensors in this study

When computing the indoor air quality, the prototype considered the contaminants temperature, humidity, CO_2, TVOC, $PM_{2.5}$, PM_{10}, CO, and O_3. The sensor used to determine the pollution parameters for the indoor air quality is shown in Table 1. Figures 1 and 2, respectively, display the prototype and the simulation for the prototype.

Table 1. Sensors and their technical details

Pollutant	Sensor Used	Technical Details
O_3	MQ-131	Sensitivity: Rs (in 300 ppm O_3) / Rs (in air) ≥ 2 Range of measurement:10～1000 ppm Ozone
$PM_{2.5}$	$PM_{2.5}$ GP2Y1010-AU0F	Sensitivity: 0.1 mg/m^3 at 0.5 V Range of measurement: 500 µg/m^3. Particle Minimum Detection Level: 0.8 M
PM_{10}	$PM_{2.5}$ GP2Y1010-AU0F	Sensitivity: 0.1 mg/m^3 at 0.5 V Range of measurement: 500 µg/m^3. Particle Minimum Detection Level: 0.8 M
CO	MQ-9	Sensitivity: Rs (in 300 ppm O_3) / Rs (in air) ≥ 2 Range of measurement: 10～1000 ppm Ozone
CO_2	MQ-135	Sensitivity: 20-2000 ppm Range of measurement: 10 to 1,000 ppm
Humidity	DHT11	Humidity Accuracy(%): Â±5.0 Range of measurement (%): 20 to 90
TVOC	AGS02MA	Typical accuracy (%): 25% reading Range of measurement: 0-99999 ppb
Wifi Module	ESP8266(NodeMCU)	Typical accuracy: 0.3 cm Range of measurement: 2 cm to 400 cm
Display	LCD Screen	–

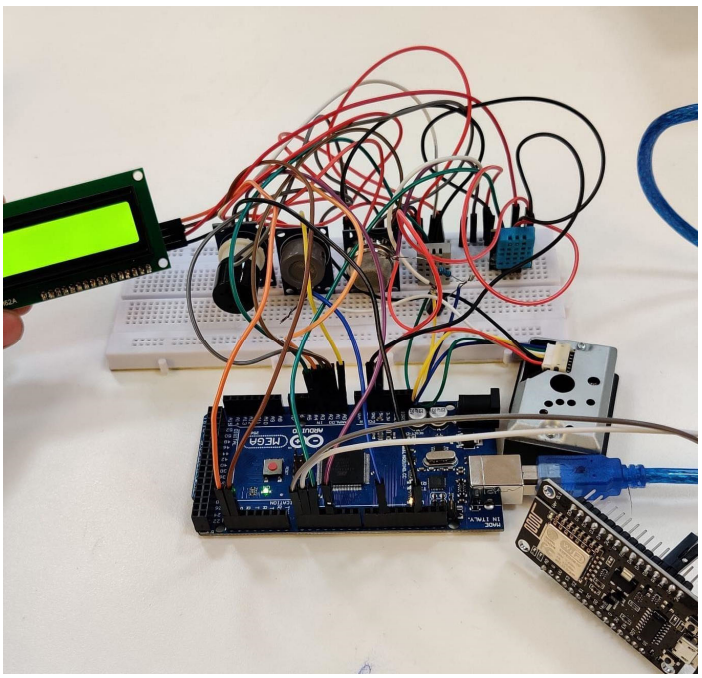

Fig. 1. Prototype to measure Internal Air Quality Index

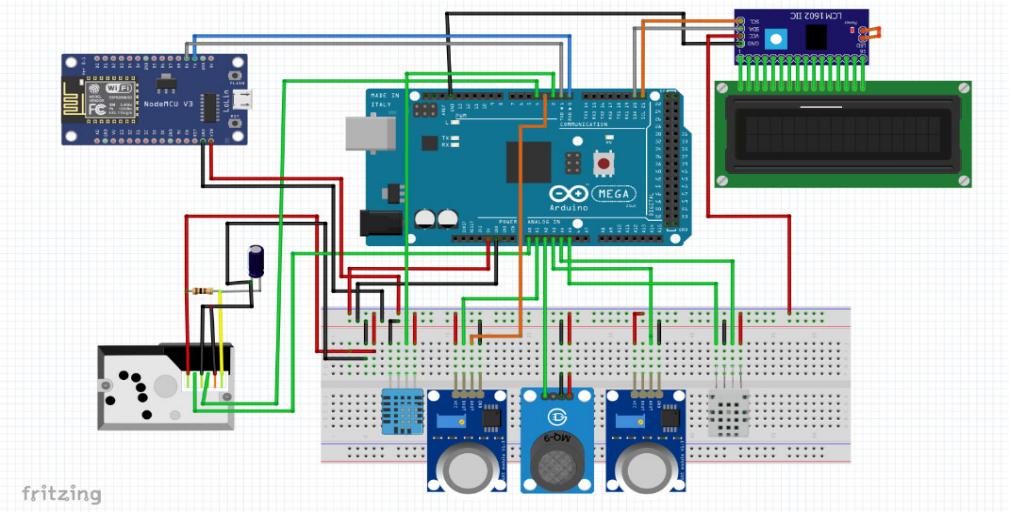

Fig. 2. Module Simulation

3.2. Internal air quality level calculation

The authors have evaluated the indoor air quality level (IAQL) in this study using the humidity level and the air quality index for pollutants inside the home.

The following linear interpolation technique can determine the indoor air pollution index for air quality (I_P).

$$I_P = ((Ii - Io / BPi - BPo)\ (C_P - BPo)\ ) + Io \quad (1)$$

where:
I_P – the Pollutant index P,
C_P – the abbreviated pollutant level P,
BP_i – The level threshold at which C_P is exceeded or remains constant,
BP_o – The level threshold at which Cp is equal to or less than,
Ii – the AQI value that is associated with BPi,
Io – the AQI value that is associated with BPo.

The level of pollution breakpoints are given in Table 2.

Table 2. Pollutant concentration breakpoints

Breakpoints						Level of Health	IAQI
Ozone (ppm) 8-hour	$PM_{2.5}$ ($\mu g/m^3$) 24-hour	PM_{10} ($\mu g/m^3$) 24-hour	CO_2 (ppm) 8-hour	CO (ppm) 8-hour	Total VOC (mg/m^3) 8-hour		
0-0.054	0-12	0-54	400-600	0-4.4	0-0.3	Excellent	0-50
0.055-0.070	12.1-35.4	55-154	700-1000	4.5-9.4	0.3-1	Good	51-100
0.071-0.085	35.5-55.4	155-254	1100-1500	9.5-12.4	1-3	Moderate	101-150
0.086-0.105	55.4-150.4	255-354	1600-2000	12.5-15.4	3-10	Poor	151-200
0.106-0.200	150.5-250.4	355-424	>2100	15.5-30.4	10-25	Unhealthy	201-300

3.2.1. Humidex

When relative humidity exceeds about 90% in hot conditions, sweat ceases evaporating to cool the body; hence, heat from interior sources can elevate body temperature and cause disease. Canadian meteorologists developed the humidex as a dimension-based attribute that utilized dew point theory that included the effects of humidity and heat with breakdowns provided by Agarwal (Agarwal et al. 2020) to represent how hot or chilly an average individual feels during different seasons. Here is the calculation for the humidex.

$$H = T + \left(\frac{5}{9}\right)\left(6.112 \times 107.5 \cdot \left(\frac{T}{237}.7 + T\right) \cdot \left(\frac{RH}{100}\right) - 10\right) \quad (2)$$

where:
T – temperature from Sensor,
RH – Relative Humidity from Sensor.

Table 3. Humidex Range

Range	Comfort level
16-29	Very good
30-39	Good
40-45	Poor
46-54	Very poor
55-60	Dangerous

3.2.2. IAQL calculation

Based on the calculation of IAQI and Humidex, the authors have given weights for both IAQI and Humidex from -2 to 2. After giving weight, the authors added both weights and decided on the indoor air quality level range. According to all calculations, the display continuously displays the IAQL and will change each after 8 hours. Figure 3 shows the complete computation of the Internal Air Quality Index based on weight, which depends on comfort and health level.

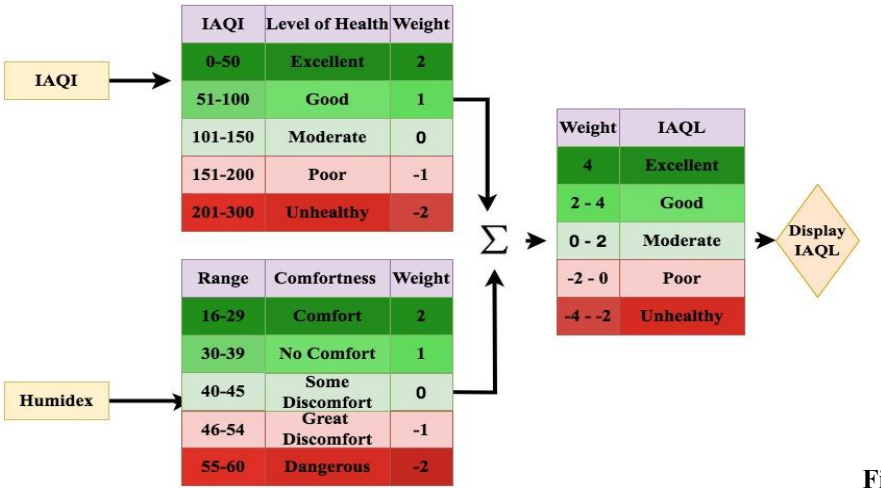

Fig. 3. IAQL Calculation

3.3. Data collection

In Ahmedabad, Gujarat, India, data on pollutants in the air readings were gathered between May 19, 2023, and July 19, 2023. Information collected locally using the designated sensors over 24 hours makes up the testing dataset for the models.

3.4. ML algorithms

3.4.1. Linear regression

Models based on linear regression are among the most widely used types of predictive analysis and one of the most basic types of statistical approaches. They employ an equation and a linear approach to show how two variables relate to one another. Several explanatory variables are used in multiple linear regression, or just "multiple regression", a statistical technique that predicts the values of a response variable. Multiple linear regression (MLR) aims to model the linear connection between response variables and explanatory factors. The multiple linear regression equation is as follows:

$$x_i = \beta_0 y_{i0} + \beta_1 y_{i1} + - - - - - - - + \beta_n y_{in} \tag{3}$$

where β_0, β_1, ---- β_n are coefficient, y_{i0}, y_{i1}, --- y_{in} are features.

3.4.2. SARIMAX

A set of observations that are consistently made over time is called a time series. Time series analysis requires understanding multiple aspects of the series' underlying structure to provide more accurate and insightful projections. Different time series models are accessible. The model the authors have employed is the SARIMAX (Exogenous variables plus Seasonal Autoregressive Integrated Moving Average). Three parameters make up a SARIMAX model: n, r, andρ. Where n represents the degree of stationary, r is the autocorrelation coefficient, ρ is the correlogram. The ARIMA model, in this instance, has moving-average order Q, integration order D, and autoregression order P. For each time step t, the SARIMAX(n, r, ρ) (P,D,Q,s) model is represented as for n exogenous variables.

$$y_t = (\sum_{t=1}^{p} \phi_n y_{1-n} + \varepsilon_t) \tag{4}$$

3.4.3. LSTM

An LSTM cell is a component that can be used to construct a bigger neural network. The LSTM module is far more sophisticated than popular construction blocks like fully-connected layers, which are just matrix multiplication of the input and the weight tensor to produce an output tensor.

It involves a hidden state h, a cell memory c, and a one-time step of an input tensor x. Initially initializing the hidden state and cell memory to zero is possible. Then, inside the LSTM cell, x, c, and h will be multiplied by various weight tensors and will experience several activation functions. The final products are the concealed state and updated cell memory. These updated c and h will be used in the "next time step" of the input tensor. Until the final step is completed, the LSTM cell's output will be its hidden state and memory.

The forget gate (Fg) decides whether to keep or forget previous data depending on the network's dependencies. The input gate (Ig) chooses to update and store fresh data in its current state. The output is generated by the output gate (Og). Lastly, long-term past and future data is stored in the memory cell state Ct. The input and output information are indicated by the values xt and ht, respectively, in the LSTM unit. The dot product operation and the sigmoid activation function manage the LSTM unit's information transformation. The possible values of the sigmoid function are 0 and 1. All information is transmitted if the dot product of the sigmoid operation provides a value of 1, and no information is transmitted if it yields a value of 0.

One LSTM cell's equation is precisely as follows:

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \tag{5}$$
$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \tag{6}$$
$$O_t = \sigma_g(W_0 x_t + U_o h_{t-1} + b_0) \tag{7}$$
$$\sim c_t = \sigma_c(W_c x_t + U_c h_{t-1} + b_c) \tag{8}$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \sim c_t \tag{9}$$
$$h_t = o_t \odot \sigma_h(c_t) \tag{10}$$

where W is for weight, b is the offset term, and σ is the sigmoid activating function.

Here, equation (2) reduces the value of the forget gate memory cell by one at a time t - 1 to determine whether the information is discarded or kept. It computes the sigmoid range to accomplish this. The input gates in (3) and (4) scale the memory cell state value at a given time t, much like a forget gate does. Eq. (5) adds memories from the past and the present. Finally, the output of the cell state is given in (6) and the output gate is shown in (7).

4. Results and Discussion

The offered models are assessed, and the air pollutant concentrations are forecasted using the IoT-collected dataset over the following seven days. Root Mean Squared Error, or RMSE, is the statistic used in the present investigation to evaluate the efficiency of the different models. The authors have used the root mean square error to gauge the accuracy of data prediction. The RMSE calculates the variation between a vector of actual values and the vectors of expected values. The RMSE values of the models are shown in Table 4.

Table 4. RMSE values of Linear Regression, SARIMA, LSTM

Pollutants	Linear Regression	SARIMAX	LSTM
Humidity	41.82	5.4639	2.312
Temperature	54.494	11.017	4.023
$PM_{2.5}$	33.213	5.8748	5.0862
PM_{10}	11.263	6.4209	3.026
CO_2	87.11	8.8733	2.2136
CO	36.852	5.2997	2.6789
O_3	15.95	5.0862	3.2654
TVOC	14.86	5.099	3.5551

Regression coefficients can be easily understood by visualizing them as linear slopes. The information from the linear regression model is displayed in Figure 4, and the numerical output is in Table 4. A fitted line plot of the relationships between time and concentration of CO_2 (Figure 4(1)), time and concentration of CO (Figure 4(2)), time and concentration of humidity (Figure 4(3)), time and concentration of O_3 (Figure 4(4)), time and concentration of $PM_{2.5}$ (Figure 4(5)), time and concentration of PM_{10} (Figure 4(6)), time and temperature (Figure 4(7)), and time and TVOC (Figure 4(8)) is used to depict this graphically.

Figure 5 displays the obtained results: with the SARIMAX model, the predicting line (orange and green in Figure 5) nearly lies on the given values (blue in Figure 5). The differencing method was not even necessary. Using this model, the authors forecasted the values for seven days in the future. Here, subfigures 1 to 8 represented prediction and actual concentration for CO_2, CO, humidity, O_3, $PM_{2.5}$, PM_{10}, temperature, and TVOC, respectively.

A line plot of the test dataset (black) against the predicted outcomes (blue) is made in Figure 6 to demonstrate the context-appropriate persistence of the LSTM model forecast. Here, subfigures 1 through 8 showed the expected and actual concentrations for the following variables: temperature, TVOC, humidity, O_3, $PM_{2.5}$, PM_{10}, and CO_2.

The linear regression prediction is displayed in Figure 4. The predictions made with SARIMAX and LSTM are displayed in Figure 5 and Figure 6, respectively. This makes it abundantly evident that LSTM fits better than linear regression and SARIMAX, as shown quantitatively in Table 4.

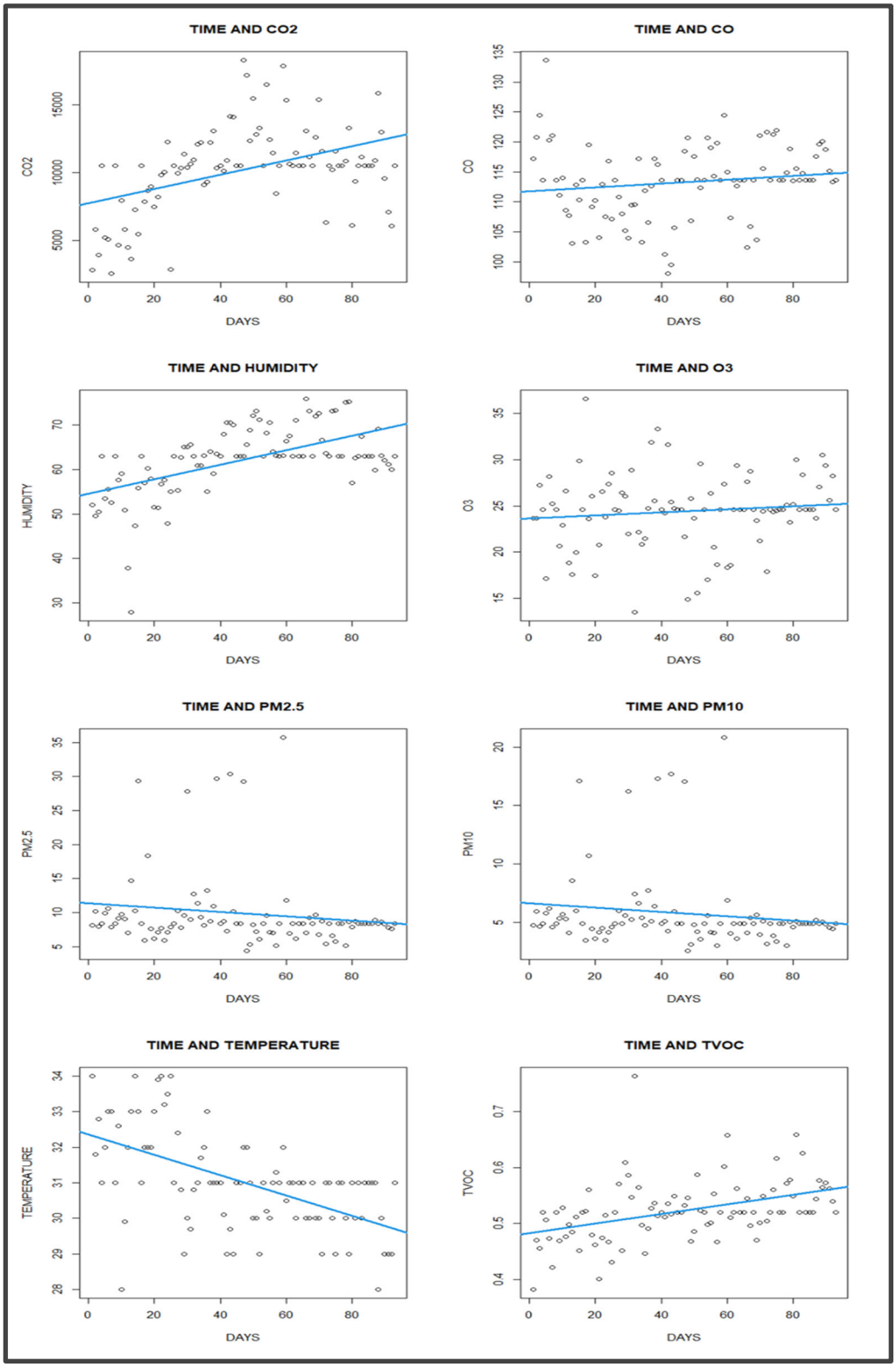

Fig. 4. Actual and prediction results from Linear Regression

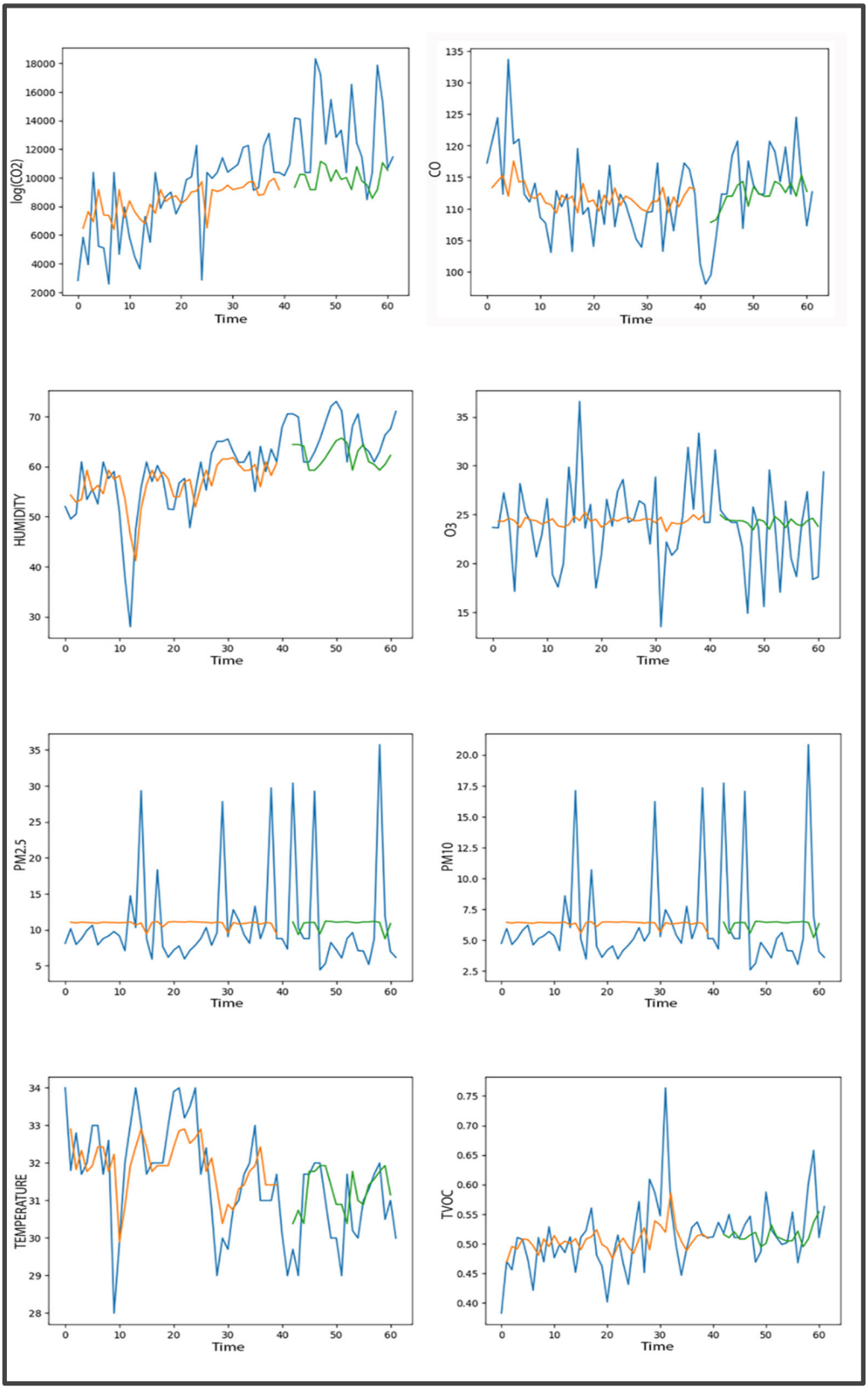

Fig. 5. Actual and prediction results from SARIMAX

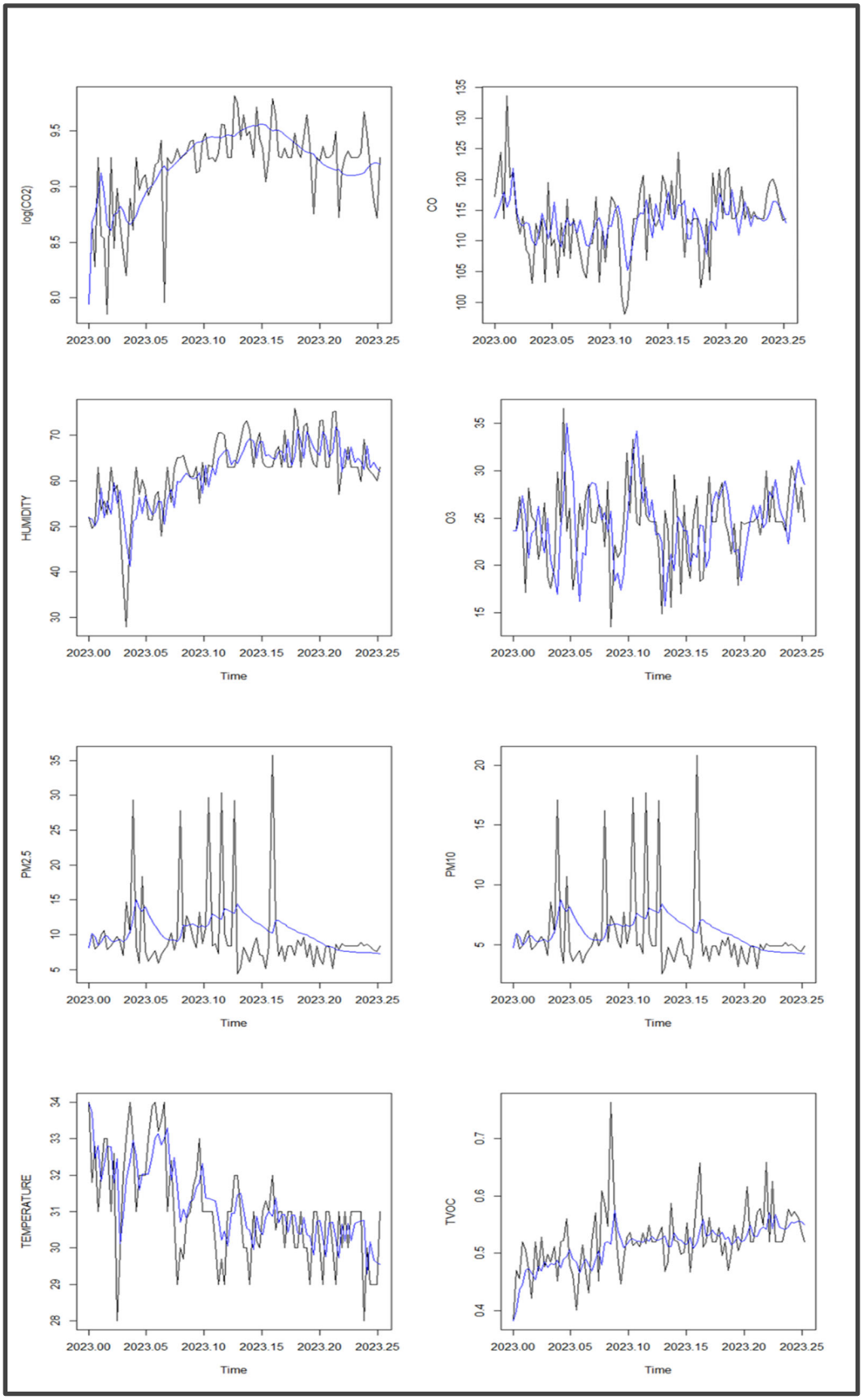

Fig. 6. Actual and Predicted Results using LSTM

5. Conclusions

It's important to stay aware of the potential problem to avoid circumstances when normal indoor concentrations could turn dangerous. Additionally, the idea of the Internet of Things (IoT) has enhanced environmental research by enabling the availability of low-cost sensors. All of the Internet of Things sensors for potentially hazardous pollution concentrations have been included in this study. The authors have employed SARIMAX, LSTM, and linear regression machine learning methods to predict the pollutant concentrations for the upcoming seven days. It is evident by comparing the RMSE values of each model that the LSTM model has the lowest RMSE values for temperature, humidity, $PM_{2.5}$, PM_{10}, CO_2, CO, O_3, and TVOC, respectively, at 2.312, 4.023, 5.0862, 3.026, 2.2136, 2.6789, 3.2654, and 3.5551. This model will be used to forecast the Internal Air Quality Index for the next seven days. A more powerful model that collects data hourly, extends the range of a wifi module or performs seasonal studies will be created in the future to obtain better findings. Additionally, new metaheuristic models will be used for this dataset.

The authors extend their appreciation to the Deanship of Research and Graduate studies at King Khalid University for funding this work through Large Groups Project under grant number RGP2/410/45.

References

Agarwal, N., Chandan S.M., Binju P.R., Saini, L., Kumar, A., Gopalakrishnan, N., Kumar, A., Balam, N.B., Alam, T., Kapoor, N., K, Aggarwal, V. (2021). Indoor air quality improvement in covid-19 pandemic: Review. *Sustainable Cities and Society*, *70*, 102942.

Agarwal, S., Sharma, S., Suresh R., Rahman, Md H., Vranckx, S., Maiheu, B., Blyth, L., Janssen, S., Gargava, P., Shukla, V.K., Batra, S. (2020). Air quality forecasting using artificial neural networks with real time dynamic error correction in highly polluted regions. *Science of The Total Environment*, *735*, 139454. https://doi.org/10.1016/j.scitotenv.2020.139454

Balram, D., Lian, K., Sebastian, N. (2019). Air quality warning system based on a localized $PM_{2.5}$ soft sensor using a novel approach of bayesian regularized neural network via forward feature selection. *Ecotoxicology and Environmental Safety*, *182*, 109386, https://doi.org/https://doi.org/10.1016/j.ecoenv.2019.109386

Dhakal, S., Gautam, Y., Bhattarai, A. (2021). Exploring a deep LSTM neural network to forecast daily $PM_{2.5}$ concentration using meteorological parameters in Kathmandu Valley, Nepal. *Air Qual Atmos Health*, *14*, 83-96. https://doi.org/10.1007/s11869-020-00915-6

Ha, Q.P., Metia S., Phung M.D. (2020). Sensing data fusion for enhanced indoor air quality monitoring. *IEEE Sensors Journal*, *20*(8), 4430-4441. https://doi.org/10.1109/jsen.2020.2964396

Javier, G., Norbertus, J., Richardus K., Cristina, P., Raquel, L., Raul, M. (2021). A state–of–the-art review on indoor air pollution and strategies for indoor air pollution control. *Chemosphere*, *262*, 128376.

Kalaivani, G., Mayilvahanan, P. (2021). Air quality prediction and monitoring using machine learning algorithm based IoT sensor – a researcher's perspective. In 2021 *6th International Conference on Communication and Electronics Systems (ICCES)*, pages 1-9. https://doi.org/10.1109/ICCES51350.2021.94891

Kapoor, N.K., Kumar, A., Kumar, A., Kumar, A., Mohammed, M.A., Kumar, K., Kadry, S., Lim, S. (2022). Machine Learning-Based CO_2 Prediction for Office Room: A Pilot Study. *Wireless Communications and Mobile Computing*, 9404807, 16. https://doi.org/10.1155/2022/9404807

Kodali, R.K., Pathuri, S., Rajnarayanan, S.C. (2020). Smart indoor air pollution monitoring station. *In 2020 International Conference on Computer Communication and Informatics (ICCCI),* pages 1-5. https://doi.org/10.1109/ICCCI48352.2020.910408

Krishan, M., Jha, S., Das, J., Singh A., Goyal M.K., Sekar, C. (2019). Air quality modelling using long short-term memory (LSTM) over NCT-Delhi, India. *Air Qual Atmos Health*, *12*, 899-908. https://doi.org/10.1007/s11869-019-00696-7

Li, L., Fu, Y., Jimmy, C., Fung, H., Kam, T.T., Alexis, K., Lau, H. (2022). Development of a back-propagation neural network combined with an adaptive multi-objective particle swarm optimizer algorithm for predicting and optimizing indoor CO_2 and $PM_{2.5}$ concentrations. *Journal of Building Engineering*, *54*, 104600.

Liu, Z., Wang, G., Zhao, L., Yang, G. (2021). Multi-points indoor air quality monitoring based on internet of things. *IEEE Access*, *9*, 70479-70492. https://doi.org/10.1109/ACCESS.2021.30736

Majdi, A., Alrubaie, A., Al-Wardy, A.H., Baili, J., Panchal, H. (2024). A novel method for indoor air quality control of smart homes using a machine learning model. *Advances in Engineering Software*, 173, 103253. https://doi.org/10.1016/j.advengsoft.2024.103253

Nan, M., Dorit, A., Hongshan, G., William, W.B. (2021). Measuring the right factors: A review of variables and models for thermal comfort and indoor air quality. *Renewable and Sustainable Energy Reviews*, *135*, 110436.

Rastogi, K., Lohani, D. (2024). Context-aware IoT-enabled framework to analyse and predict indoor air quality. *Intelligent Systems with Applications*. *16*, 200132. https://doi.org/10.1016/j.iswa.2024.200132

Saini, J., Dutta, M., Marques, G. (2020). Indoor Air Quality Prediction Systems for Smart Environments: A Systematic Review, *Journal of Ambient Intelligence and Smart Environments*, *12*(5), 433-453.

Stefano, R., Tagliabue, L., Ciribini, A. (2019). An IoT framework for the assessment of indoor conditions and estimation of occupancy rates: results from a real case study. *ACTA IMEKO*, *8*(70), 06. https://doi.org/10.21014/actaimeko.v8i2.647

Shanmugaraja, T., Sakthivel, M., Shaila, Shree, R., Sooraj, P., Vishnu, K. (2021). Analysis of Air Quality using IoT with machine learning prediction. *J. Phys.: Conf. Ser.* 1916 012188.

Tian, X., Cheng, Y., Zhang, L. (2022). Modelling indoor environment indicators using artificial neural network in the stratified environments. *Building and Environment*, *208*, 108581. https://doi.org/https://doi.org/10.1016/j.buildenv.2021.108581

Tagliabue, L.C., Cecconi, F.R., Rinaldi, S., Ciribini, A.L.C. (2021). Data driven indoor air quality prediction in educational facilities based on IoT network. *Energy and Buildings*, *236*, 110782. https://doi.org/10.1016/j.enbuild.2021.110782

Vagner, S., Ricardo, J.A., Richard, M. (2020). Sensor validation for indoor air quality using machine learning. https://doi.org/10.5753/eniac.2020.12174

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

William, V., Marco, G., Jorge, G., Wu-Chieh, W., Kuo, K.L., Jen-Chung, L., Kuang-Chin, L., Wang, C. (2019). Energy optimization associated with thermal comfort and indoor air control via a deep reinforcement learning algorithm. *Building and Environment*, *155*, 105-117. https://doi.org/10.1016/j.buildenv.2019.03.038

Xie, X., Zuo, J., Xie, B., Thomas, A.D., Selvarajah, M. (2021). Bayesian network reasoning and machine learning with multiple data features: air pollution risk monitoring and early warning. *Nat Hazards*, *107*, 2555-2572. https://doi.org/10.1007/s11069-021-04504-3

Zhao, L., Wu, W., Li, S. (2019). Design and implementation of an IoT-based indoor air quality detector with multiple communication interfaces. *IEEE Internet of Things Journal*, *6*(6), 9621-9632. https://doi.org/10.1109/JIOT.2019.2930191