The Bioscan
AN INTERNATIONAL QUARTERLY JOURNAL OF LIFE SCIENCES
20(2): S2: 184-188, 2025 www.thebioscan.com

An IoT-Enabled Real-Time System for Detection and Analysis of Toluene using Machine Learning

***Shishir* A. Bagal[1], *Nitin* K. Choudhari[2], *A.* R. Chaudhari[3], *Sushmita* A. Meshram[4], And *Amar* T. Banmare[5]**

1 Assistant Professor, Department of Electronics Engineering, KDK College of Engineering & Research Scholar at Dept. of Electronics & Computer Science Rashtrasant Tukadoji
Maharaj Nagpur University, Nagpur (Maharashtra), India
2 Principal, Priyadarshini Bhagwati College of Engineering, , Nagpur (Maharashtra), India
3 Associate Professor and Head of Chemistry Department Priyadarshini Bhagwati College of Engineering, Nagpur (Maharashtra), India
4 Assistant Professor, EXTC dept., PRMCEAM, Amravati and Research Scholar at dept. of PGTD, Computer science & Electronics, RTMNU, Nagpur (MH),
India
5 Department of Electronics & Computer Science Rashtrasant Tukadoji Maharaj Nagpur University Nagpur, India

DOI: 10.63001/tbs.2025.v20.i02.S2.pp184-188

KEYWORDS
Volatile Organic Compounds (VOCs),
World Health Organization (WHO),
Internet of Things (IoT),
Amazon Web Services (AWS),
Machine Learning (ML).

Received on:

12-02-2025

Accepted on:

15-03-2025

Published on:

26-04-2025

ABSTRACT

Background / Objective: According to World Health Organization (WHO) reports, Toluene is one of the volatile organic compounds (VOCs) that is released by industries and vehicles causing serious health concerns, including brain damage and respiratory problems. To reduce health risks, maintain regulatory compliance and help the regulatory agencies to frame the future policies, toluene concentrations must be accurately and consistently monitored. However, traditional detection methods, including gas chromatography-mass spectrometry (GC-MS), lack real-time monitoring capabilities and are very expensive from installation and maintenance point of view. **Methodology:** The presented work suggests a portable, low-cost, real-time toluene detection system that integrates inexpensive sensors with IoT architecture for real-time data Collection and monitoring on the Amazon Web Service (AWS) cloud platform. The system deals with Machine learning (ML) based analysis for trend identification and interpretation. **Findings:** The strong correlation between VOC grade and Toluene concentration (0.98) shows that the developed device is highly capable of detecting the toluene concentration. The dataset can be useful to the external agencies for the pollution audit in industrial settings. **Novelty / Improvement:** The device is developed using the low cost and portable sensors, collects the real time data using the AWS cloud. Therefore, this device can be easily installed in industrial settings that can ensure increased workplace safety and regulatory compliance by providing an affordable and scalable environmental monitoring solution.

INTRODUCTION

Deteriorating Air quality is one of the greatest threats to humanity; according to WHO research, air pollution endangers the health of 7 million people globally [1]. According to a World Health Organization (WHO) assessment, about 92% of the world's population lives in locations where ambient air quality is greater than recommended. According to the research, worsening air quality kills one in every nine people each year, making it the most serious environmental health concern. Furthermore, new studies reveal that outdoor air pollution is the most dominant component, directly leading to 3 million fatalities per year [2].

Toluene, a volatile organic compound (VOC) found in vehicle and industrial exhaust, causes serious health concerns depending on exposure concentration and duration. Although toluene is not recognized as a carcinogen, long-term exposure to toluene can cause cardiovascular impacts like arrhythmias, heart failure, and myocardial infarction. Also prolonged exposure to toluene results in hematologic disorders, immune dysfunction, hepatic and genetic toxicity and reproductive effects. It also affects the neurological alterations in brain structure, neurotransmitter levels, cognitive function, anxiety, impulsivity, and depression. Furthermore, toluene metabolism in the liver and kidneys might result in organ damage over time [3]. To minimize the danger of exposure, regulatory authorities such as OSHA and WHO have established safety limits. Continuous monitoring and emission control are critical for limiting the health risks associated with toluene pollution [4].

The presented work deals with toluene detection in ambient air utilizing an ESP32 microcontroller, an AWS cloud service, and a Multi Sensor Array (MSA) Detection Module. The work highlights how effectively this integrated system monitors toluene concentrations, provides real-time monitoring capabilities, and enables effective data interpretation using machine learning algorithm. The main objective of this study is to overcome the disadvantages of traditional methodologies by developing a cheap, portable system capable of gathering and analysing data in real time.

Gap analysis

Literature review becomes evident to identify the research gaps, recent technological development and exploring the existing technology for assessing the requirements for the detection of toluene. Therefore, it is necessary to have a comprehensive understanding of existing technology that addresses recent developments in the detection of toluene.

An Internet of Things (IoT)-based infrastructure for monitoring indoor air quality is suggested by Mohamed Arabi et'al, which consists of a web server and an air quality-sensing gadget named "Smart-Air." This platform uses cloud computing and the Internet of Things to effectively monitor air quality and send data in real time over LTE to a web server [5] [6]. The system presented by Armaan Suhel Shaikh et'al detects VOCs in the air and sends automated alerts to users. It uses miniature MOS gas sensors in the 1-30 ppm range and a GSM module to send SMS alerts. The system consists of two identical sensor nodes that connect to the control side via ZigBee transceiver kits [7].

Air quality monitoring system have seen rapid technological developments from basic manual procedures to sophisticated real-time systems that use satellite imaging, cutting-edge Internet of Things (IoT) sensors, and efficient artificial intelligence (AI) algorithms. These systems send out continuous data streams, allowing for the identification of pollution sources, the assessment of regulatory compliance, and the development of effective mitigation strategies. The investigation dives into the current environment, obstacles, and potential solutions in the field of air pollution monitoring and analysis through the integration of sensors and technologies such as Zigbee and IoT [8][9][10]. Which use several methodologies for data collecting and analysis of common contaminants such as CO, CO_2, VOC, and PM.

Machine learning (ML), particularly deep learning (DL)-based models, are now widely employed for regression issues and have garnered a lot of attention for predicting air pollution [11], [12]. Amazon SageMaker's scalability allows for easy handling of large-scale deployment scenarios. SageMaker uses an expanded Machine Learning Operations (MLOps) framework to handle ML algorithms and Big Data on a corporate scale. SageMaker gives us the capacity to grow resources at a cheap cost. On the other hand, Lambda is best suited for low-cost event-driven operations. Combined with this, Elastic Container Services (ECS) guarantees that installed services grow smoothly in response to demand [13],[14]. In order to mitigate the consequences of toluene exposure on workshop workers, a system was built using an ESP32 microcontroller, an HCHO sensor, and a DHT11 sensor. Each node's average temperature and humidity readings are obtained using DHT11[15] [16]. Jungmo Ahn et'al presented deployable Internet of Things (IoT) based system for identifying and categorizing the many types of volatile organic compounds (VOCs) in the air which utilizes a paper-based fluorometric sensor in conjunction with integrated processing and camera modules makes up our suggested solution, VOCkit [17].

After exploring different literatures, it is observed that most of the existing work is associated with VOC as a single entity, analytical feature is also not supported in some existing systems. Though, some of these systems are IoT based systems but they are not capable of addressing any specific VOC. This limitation in the existing studies motivated for undertaking the presented work that deals with the development of an IoT based system for detection and analysis of Toluene in an ambient air.

Methodology

The presented work employs integration of Multi Sensor Array (MSA) Module ZPHS01B with ESP-32-an IoT Controller and AWS Cloud for the detection and analysis of toluene, as shown in the Figure-1. The MSA consists of multiple sensors that includes sensors for detection of Carbon Dioxide, Carbon Monoxide, Formaldehyde, Nitrogen Dioxide, Volatile Organic Compounds (VOC) sensor, Temperature and Humidity sensor as shown in the figure 2 [18].

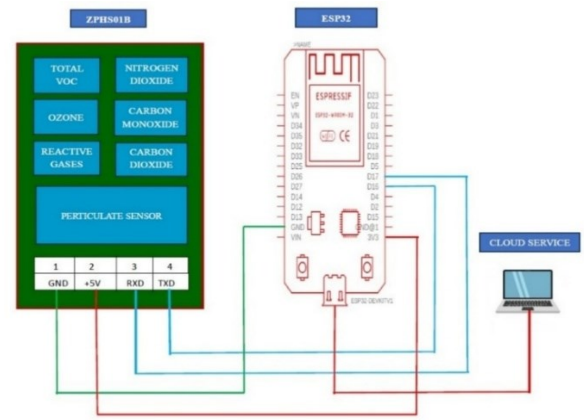

Figure 1- Block Diagram of Tolune Detection and Analysis

It also has a UART port to initiate the communication between MSA and the ESP-32 Controller. The commands are given by the ESP-32 controller to the MSA to enquire about the sensor data and MSA generates the response providing the sensor data to ESP-32.

ESP32 processor is employed to read the sample values of the VOC and transmit it over the cloud using the inbuilt Wi-Fi capability of ESP-32. The data from the sensor are fetched by initiating the dialogue between ESP-32 and MSA in the form of Command - Response (question answer).

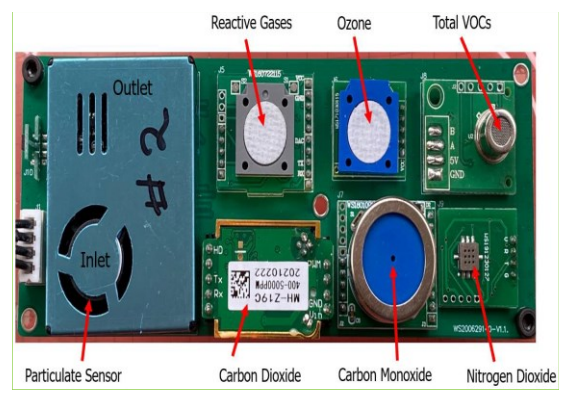

Figure 2- Sensitivity Characteristics of ZP07 VOC Sensor

The byte commands are issued from the ESP-32 to MSA and MSA responds to these commands by sending the appropriate data to the ESP-32. The data received by the ESP-32 is then transmitted to the cloud, which is configured with the help of Amazon Web Services (AWS) and ESP-32 IoT Controller to form the data base.

The framework utilizes a VOC sensor (ZP07 Module) Air Quality Detection that Module detects acetone, toluene, benzene, methane, and alcohol vapours using a semiconductor gas sensor that alters resistance in reaction to these volatile organic compounds. The module's heating element helps to maintain the appropriate sensor temperature for efficient gas interaction, while the signal conditioning circuit and microprocessor process the sensor's resistance changes to provide usable output data. Figure-3 below the sensitive characteristics of the VOC sensor to various hazardous compounds [19], [20]. This sensor module is used for monitoring and detection of Toluene in an ambient air. This characteristic clearly indicates that the sensor is highly sensitive to the Toluene and other VOCs.

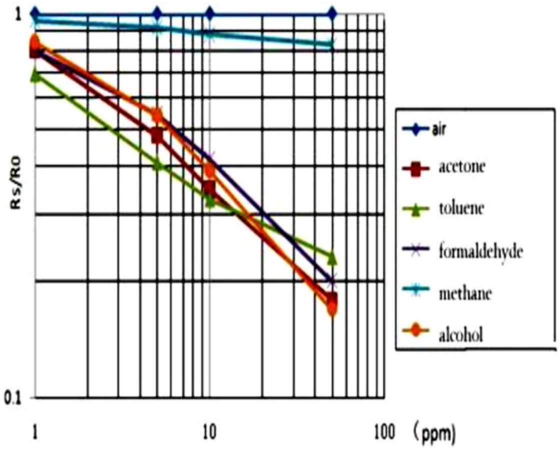

Figure3. Sensitivity Characteristics of VOC Sensor

Figure-4, below shows the experimental set up for monitoring and detection of Toluene using ZP-07 sensor. it consists VOC sensor that has four terminals: two terminals for power supply and two signal outputs i. e. Signal A & Signal B.

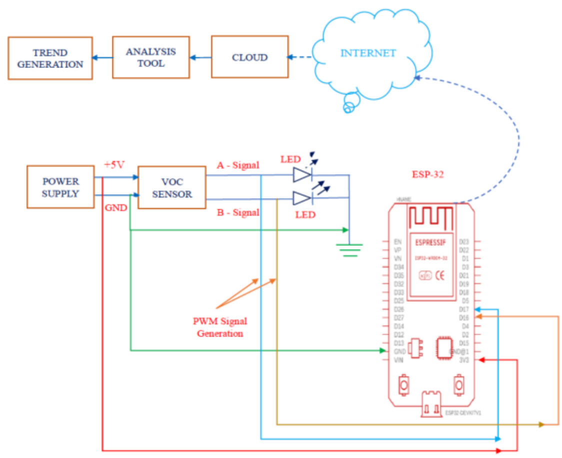

Figure-4. Detailed Block Diagram of Toluene Detection

The sensor's resistance changes based on gas concentration. It detects pollutant gases by interacting with air molecules, causing a change in electrical resistance. This change is processed and converted into a PWM output signal. As concentration increases (ppm), Rs decreases, indicating higher detection sensitivity. A higher concentration (lower Rs) leads to a higher duty cycle on Output A. This increases LED ON time, making it glow brighter for more pollution. A microcontroller measures the PWM duty cycle. It maps the duty cycle to pollution classes (Table 1) as prescribed in the ZP07 datasheet. Thus, the monitoring and detection of toluene is carried out.

Pollution Class	High (ms)	Low (ms)	Duty Cycle (%)
1	10	90	10%
2	20	80	20%
3	30	70	30%
4	40	60	40%
5	50	50	50%
6	60	40	60%
7	70	30	70%
8	80	20	80%
9	90	10	90%
10	100	0	100%

Table 1: Output Signal – Pollution Class Mapping

Experimental Results & Discussion

The device developed for the detection and monitoring of toluene was installed and the experimentation was carried out at Ameya Paper Mills situated at MIDC, Hingna, Distt, Nagpur (India). The samples of toluene were read from the device and collected on the cloud using AWS for over a span of seven days. The frequency of the data collection was 12 data samples per minute. Figure-5 below shows the experimentation site, where the experimentation was carried out.

Figure-5. Device installation at Experimentation Site

The samples from the device were read using the UART protocol and these samples were collected on the cloud to form the data set using IoT Controller i. e. ESP-32. The analysis of the data set is carried out to develop the relation between Toluene and VOC. Also, the effect of temperature and humidity on the toluene is studied. The following section deals with the analysis of the toluene.

VOC & Toluene vs Temperature

The 3D plot of VOC & Toluene vs Temperature is shown in the Fig-6 that provides a visualization of how VOC (Volatile Organic Compounds) levels and Toluene concentrations vary with Temperature. As VOC readings increase (from 0 to ~3), Toluene concentrations also tend to increase. Suggests a strong correlation between general VOC detection and Toluene presence. At higher temperatures (30°C to 40°C), there's a denser clustering of high Toluene concentrations, indicating Either increased emission rates of Toluene at higher temperatures. Or, sensor sensitivity improvement with rising temperature. Such a 3D dataset would be ideal for training regression or classification models. A model could predict Toluene concentration based on real-time VOC and temperature data, improving detection accuracy without needing a Toluene-specific sensor at every point.

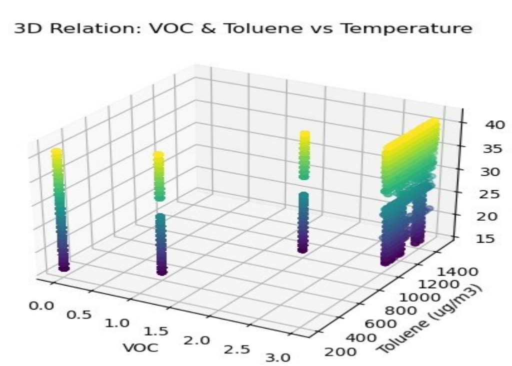

Fig-6 -3D plot of VOC & Toluene vs Temperature

VOC & Toluene vs Humidity

Figure -7 below shows the 3D relation between VOC, Toluene and Humidity. With reference to this graph, Toluene levels increase with VOC values, reaffirming a reliable correlation. Toluene values cluster more tightly at higher humidity levels (around 80–110%) indicating that sensor performance or gas behaviour is affected by moisture content. High humidity may enhance Toluene detection or cause interference. Continuous variation along humidity and Toluene axes implies these are continuous input/output features useful for training robust regression models.

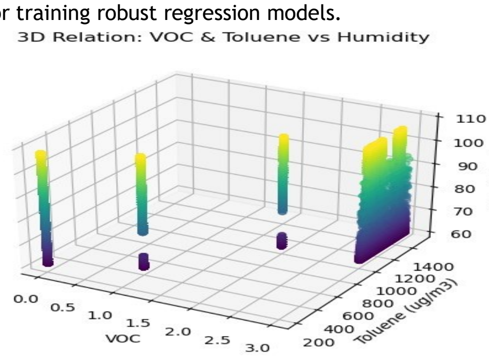

Fig-7 -3D plot of VOC & Toluene vs Humidity

To 2-D relation between Toluene and temperature is shown in the figure-8 as shown below. The provided image showcases relationship between Temperature (°C) and the concentration of a Toluene (bottom right corner). A red trendline is included in each plot, likely representing a regression line, which helps visualize the direction of correlation. The parameter in consideration i. e. toluene readings (200 to 1500 µg/m^3) show a slight decline with temperature. The negative trend suggests that sensor readings for Toluene decrease as temperature rises, potentially due to reduced gas adsorption or sensor sensitivity at high temp. The ML model can be trained with temperature as a key feature to predict actual Toluene levels even when sensor output is affected by heat.

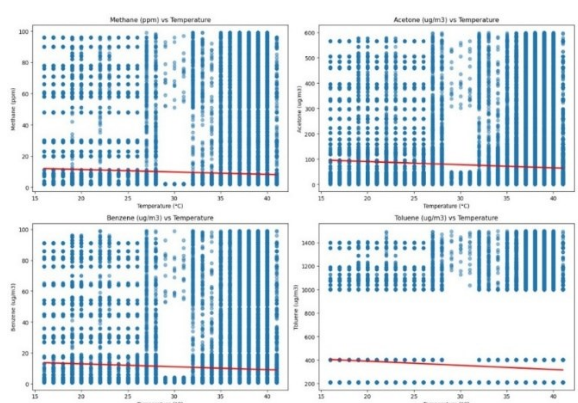

Fig-8 -2D plot of Toluene vs Temperature

The 2-D graph shown in the figure -9 is critical in evaluating how environmental humidity impacts gas sensor readings, which is essential for machine learning models in Real-Time Detection and Analysis of Toluene System. This image shows a relationship between Humidity (%) and gas concentrations for Toluene (Bottom right corner), with a blue trendline indicating linear correlation. It is observed that toluene readings increasing with humidity.

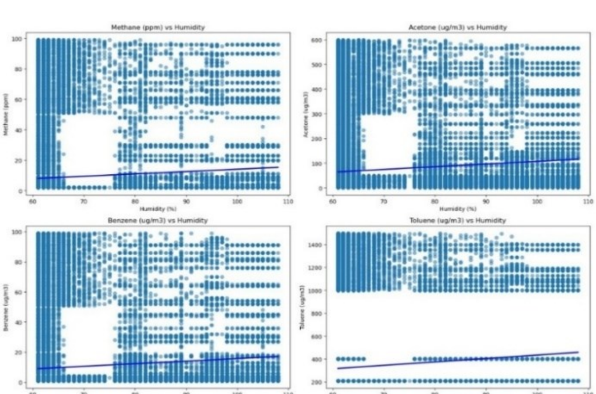

Fig-9 -2D plot of Toluene vs Humidity

This suggests Toluene detection is influenced by environmental moisture, either enhancing sensor response or causing baseline drift. Including humidity as a feature in the ML model is critical for accurate and consistent Toluene predictions, especially in fluctuating weather or indoor environments.

This correlation heatmap shows the Pearson correlation coefficients between the different gas concentrations (Methane, Acetone, Benzene, Toluene) and environmental variables (Temperature and Humidity). It provides a clear view of interdependencies among features that are crucial for feature selection, model design, and sensor calibration in a real-time detection system. Strong Positive Correlations Among VOCs: Toluene vs Benzene: 0.94, Toluene vs Acetone: 0.94, Toluene vs Methane: 0.92. Toluene concentration increases slightly with higher humidity. Sensor readings are marginally affected by RH, possibly due to surface adsorption effects. Humidity should still be included in real-time models to prevent baseline shift or false alarms under humid conditions. Toluene Vs Temperature (-0.07) and Toluene Vs Humidity (0.11) These correlations are weak, indicating that environmental variables do not strongly influence toluene concentration directly.

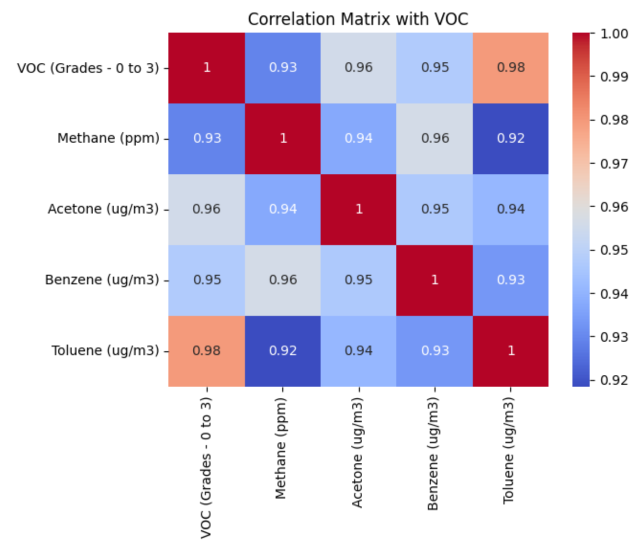

Fig-10 -Correlation Heatmap

CONCLUSION

The real time, IoT based system developed for the monitoring and detection of toluene was installed in the paper mill industry in MIDC Hingna, Distt, Nagpur (India) and an experimentation was carried out. The samples of the toluene were collected over the cloud from the sensor module with the help of an IoT controller. The analysis was carried out with the help of LSTM algorithm which suggest a strong positive correlation between VOC grade and Toluene concentration (0.98). The analysis also shows that, there is either little or no effect of temperature and humidity as far as detection of toluene is concerned.

The strong correlation between VOC grade and Toluene concentration shows that the developed device is highly capable of detecting the toluene concentration in an air. The dataset created over the cloud can be useful to the external agencies for the pollution audit in industrial settings. The device is developed using the low cost and portable sensors, collects the real time data using the AWS cloud. Therefore, this device can be easily installed in industrial settings that can ensure increased workplace safety and regulatory compliance by providing an affordable and scalable environmental monitoring solution.

Author Statements:

Conflict of interest: There are no conflicts of interest, according to the authors.

Acknowledgement: Mr. Dilip Andhare, Manager, Ameya Paper Mills, Hingna MIDC, Nagpur (Maharashtra), India for permitting for the experimentation in their industry. Also, Sincere gratitude is extended by the researcher to the research supervisor, colleagues, and affiliated institutions for their helpful advice and priceless contributions. Their direction and encouragement have been crucial in improving the data collecting process, letting thorough evaluation, and shaping the complete review paper.

•**Author contributions:** SAB: Development and experimentation of the system, NKC: Ideation and conceptualization, Writing, ARC: Review and Editing, SAM: methodology, ATB: Analysis and Conclusion.

•**Funding information:** No financial assistance was given to the research.

•**Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

REFERENCES

- [1] 7 million Premature Deaths Annually Linked to
- Air Pollution. Accessed: Apr. 27, 2019. [Online].
- Available:https://www.who.int/phe/eNews_63.pdf
- [2] World Health Organization, "Ambient Air
- Pollution: A global assessment of exposure and burden of disease," Geneva, 2016.
- [3] Zahra Soltanpour, Sajjad Barzegar, "Health risk
- assessment of exposure to benzene, toluene, ethyl benzene, and xylene in shoe industry-related workplaces", Toxicology and Industrial Health, Volume 40, Issue 1-2, https://doi.org/10.1177/07482337231212693, November 7, 2023.
- [4] Occupational Safety and Health Administration's
- OSHA Safety and Health Program Management Guidelines November 2015.
- [5] Mohamed Arabi, Majed Alghamdi, Khalid Kabel,
- Ahmed Labena, Walaa S. Gado, Bhoomi Mavani, Alison J. Scott, Alexander Penlidis, Mustafa Yavuz, Eihab Abdel-Rahman, "Detection of Volatile Organic Compounds by Using MEMS Sensors", Sensors 2022, 22, 4102, doi.org/10.3390/s22114102.
- [6] JunHo Jo, ByungWan Jo, JungHoon Kim, SungJun
- Kim, WoonYong Han, "Development of an IoT- Based Indoor Air Quality Monitoring Platform", Hindawi Journal of Sensors Volume 2020, https://doi.org/10.1155/2020/8749764
- [7] Armaan Suhel Shaikh, V. B. Gaikwad, "Volatile
- Organic Compound (VOCs) Monitoring and Alerting for Indoor Settings with MOS Sensor and Wireless Sensor Nodes", International Journal of Software & Hardware Research in Engineering ISSN-2347-4890, Volume 4 Issue 10 October, 2016.
- [8] Dan Zhang, Simon S. Woo, "Real Time Localized
- Air Quality Monitoring and Prediction Through Mobile and Fixed IoT Sensing Network", IEEE Access, Multidisciplinary, Rapid Review, *DOI 10.1109/ACCESS.2020.2993547.*
- [9] Rajib Saha, M M R Manu, Aminul Hoque, S N M
- Azizul Hoque, "Monitoring Air Quality of Dhaka
- using IoT: Effects of COVID-19", IEEE Xplore 2021, 2nd International Conference on robotics, Electrical and Signal Processing Techniques (ICREST).
- [10] Dylan Wall, Paul McCullagh, Ian Cleland, and
- Raymond Bond, "Development of an Internet of
- Things (IoT) solution to monitor and analyse indoor air quality", Internet of Things14,100392, w.elsevier.com/locate/iot, doi.org/10.1016/j.iot.2021.100392.
- [11] S. Du, T. Li, Y. Yang, and S.-J. Horng, "Deep air
- quality forecasting using hybrid deep learning framework", IEEE Trans. Knowl. Data Eng., vol. 33, no. 6, pp. 2412–2424, Jun. 2021.
- [12] W. Huang, T. Li, J. Liu, P. Xie, S. Du, and F. Teng,
- "An overview of air quality analysis by big data techniques: Monitoring, forecasting, and traceability, Inf. Fusion, vol. 75, pp. 28–40, Nov. 2021.
- [13] Rahul Bagai, "Comparative Analysis of AWS
- Model Deployment Services", International Journal of Computer Trends and Technology Volume 72 Issue 5, 102-110, May 2024 ISSN: 2231–2803 / doi.org/10.14445/22312803/IJCTT-V72I5P113
- [14] M. Soncin, "Extending the Capabilities of
- Amazon SageMaker: ACase Study," *IEEE Software*, vol. 40, pp. 59-67, 2023.
- [15] Gastiadi, M. Nanak Zakaria, Ahmad Wilda
- Yulianto, "Design and Build a System to Minimize the Impact of Toluene Exposure on IoT-Based Workshop Workers", Journal of Telecommunication Network (Jurnal Jaringan Telekomunikasi) Vol. 12, No.4 (2022) E-ISSN: 2654-6531 P- ISSN: 2407-0807 218.
- [16] Galang Gastiadi, M. Nanak Zakaria, Ahmad
- Wilda Yulianto, "Design and Build a System to Minimize the Impact of Toluene Exposure on IoT-ased Workshop Workers", Journal of Telecommunication Network, 2407-0807 (print), 2654-6531 (online).
- [17] Jungmo Ahn, Hyungi Kim, Eunha Kim, JeongGil
- Ko, "VOCkit: A low-cost IoT sensing platform for volatile organic compound classification", IEEE International Conference on Distributed Computing in Sensor Systems 2019), doi.org/10.1016/j.adhoc.2020.102360
- [18] Multi-in-One Sensor Module （Model：
- ZPHS01B) Manual Version：1.3 Valid From: 2020.07.13 Zhengzhou Winsen Electronics Technology Co., Ltd
- [19] MULTI-SENSOR ARRAY (VERSION 4) Design
- Document Team: Michael Petersen Jen Stoddard Wesley Mahurin Project: MSAv4 Date Prepared: November 20, 2013
- [20] Air-Quality Detection Module Model：ZP07
- MP503）User's Manual Version:1.3 Valid from 2018.08.14 Zhengzhou Winsen Electronics Technology Co., Ltd.