JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Automated Calibration of Sensor using IoT and Machine Learning

Prof. Jitendra Gaikwad
Vishwakarma Institute of Technology, Pune
jitendra.gaikwad@vit.edu

Rushikesh Sarawade
Vishwakarma Institute of Technology, Pune
rushikesh.sarwade23@vit.edu

Sarthak Kamble
Vishwakarma Institute of Technology, Pune
sarthak.kamble23@vit.edu

Rohit Patil
Vishwakarma Institute of Technology, Pune
rohit.patil23@vit.edu

Shreem Agarwal
Vishwakarma Institute of Technology, Pune
shreem.agarwal23@vit.edu

Abstract

Precise measurement of temperature is necessary in many industrial, environmental, and scientific applications. Low-cost sensors, such as the DHT11, are used due to their low price and simplicity but suffer from measurement errors due to the lack of precision and sensitivity in their measurement. On the other hand, high-precision sensors such as the RTD (Resistance Temperature Detector) PT100 are very expensive and require sophisticated signal conditioning circuits. This paper proposes a novel way of bridging the gap between precision and cost by developing an automated calibration system that enhances the accuracy of DHT11 sensor outputs based on a machine learning model trained on RTD PT100 data. The system employs an ESP8266 microcontroller, a DHT11 sensor, and a 16x2 I2C LCD, supported by a locally deployed real-time web interface. A linear regression model is trained with datasets where DHT11 readings (humidity and temperature) are correlated with the more precise temperature readings from the RTD. The model that has been trainedis run on the ESP8266 so that it can accept raw sensor inputsand deliver calibrated outputs in real-time. The calibrated readings are presented on both a local web server as accessible over any Wi-Fi-enabled device and the I2C LCD module. The system is portable, low-cost, and Smart solution for accurate temperature monitoring, Particularly in remote or resource-constrained situations. This paper illustrates the potential of employing IoT technologies with machine learning in enhancing the performance of low-cost sensors and provides a basis for high-end, self-calibrating embedded systems.

Keywords — *Sensor Calibration, IoT, Machine Learning, Real-time Monitoring, Temperature Accuracy.*

1. INTRODUCTION

Sensor calibration is the adjustment and tuning of the sensor output to a widely acceptable standard or reference point. Calibration is the comparison of the sensor reading with exact measurement from a high-quality instrument and adjustment to eliminate any variation or inaccuracy. Calibration is carried out to make sure that the sensor is Working precisely in its complete range of operation, considering variability caused by fault during manufacturing, environmental conditions, and signal drift that are bound to happen with the passage of time.

Temperature measurement is an important requirement in certain of the applications such as industrial automation, agriculture monitoring, environmental monitoring, and smart home systems. The requirement for precise and reliable temperature measurement is increasing as systems are becoming increasingly dependent on the data gathered from the sensors. However, high-precision sensors
Resistance temperature detectors (RTDs), such as the PT100, tend to be quite expensive and necessitate professional analog signal conditioning circuits. Low-cost sensors, such as the DHT11, give temperature and humidity values but are hampered by limited accuracy and reliability, and therefore are not applicable for uses that demand high tolerances.

To address this problem, the current project proposes an automated calibration system of the sensor with a machine learning model to improve the accuracy of the output of the DHT11 sensor by comparing with the readings of an RTD PT100 sensor. The calibration model is implemented using Python on Jupyter Notebook, where a dataset of corresponding readings of the DHT11 and PT100 sensors is trained to create a regression model. The coefficients of the model are then installed in the

ESP8266 microcontroller, which performs real-time calibration of the raw readings of the DHT11 sensor.

The overall system is small, low-cost, and power-efficient, and hence perfectly suitable for low-resource embedded Internet of Things applications. The design, implementation, and performance of the proposed system, and its accuracy, limitations, and future scope, are explained in this article. Through the use of IoT connectivity and machine learning-based calibration, this research showcases an effective way of driving low-cost sensors to real-world deployment without adding hardware complexity and cost.

2. LITERATURE REVIEW

The rapid evolution of IoT technologies and increased application of low-cost sensors in real-time monitoring purposes have attracted substantial research interest to improve the precision of such devices. Zhu et al. (2023) in their article Machine Learning–Based Calibration and Performance Evaluation of Low-Cost Sensors compared machine learning algorithms like Gradient Boosting (GB) and k-Nearest Neighbors (kNN) to calibrate low-cost humidity and temperature sensors. Their research highlighted that ML-based models performed better compared to available statistical methods by a significant margin in terms of accuracy, highlighting the practical application of data-driven techniques in environmental monitoring [10]

Yet another relevant research by Lee and Li (2023), PM2.5 IoT Sensor Calibration and Implementation Problems Including Machine Learning, explored various ML models such as neural networks and decision trees to improve the performance of sensors for PM2.5 sensors. The study took into account practical implementation issues and emphasized stable model choice from the viewpoint of its ability to deliver reliable sensor output in varying environmental conditions [9]

Kourtidis and Argyropoulos (2020), in their article Application of Machine Learning Techniques for the Calibration of Low-Cost IoT Sensors in Environmental Monitoring Networks, highlighted the role of imputing data and filtering noise while calibrating sensors. They showed that machine learning algorithms not only could be employed for effective sensor calibration but also had the capability to deal with missing or corrupted data in large sensor networks [1]

Chatzidiakou et al. (2023) Leveraging Machine Learning Algorithms to Advance Low-Cost Air Sensor Calibration suggested common protocols and validated regression models on stationary and mobile platforms. Their work verifies that model generalizability from one deployment setting to another is possible and also useful, thus increasing the scope for use of calibrated low-cost sensors [2]

Johnson and Sharma (2021), in the article Evaluating the Efficacy of Machine Learning in Calibrating Low-Cost Sensors, compared ML algorithms like Random Forests and Artificial Neural Networks. They highlighted real-world concerns in sensor calibration, such as deployment complexity, data diversity, and model scalability [3]

In the field of building automation, Seo et al. (2021) authored a paper, A Study on the Sensor Calibration Method Using Data-Driven Prediction Models. They used ML-based correction methods for the readings of HVAC sensors and observed a dramatic improvement in the accuracy of data-driven automated systems. The study went on to exemplify the use of ML calibration in varied fields as well [4]

Ahmad and Rehman (2023), in their technical report Sensor Calibration Techniques for IoT Networks, introduced some blind, supervised, and semi-supervised calibration methods. They emphasized the capability of unsupervised learning methods for use where reference data is not available, paving the way towards scalable sensor deployment within enormous IoT systems [5]

Rai et al. (2020), in their article Improving Data Quality of Low-Cost IoT Sensors in Environmental Monitoring, illustrated how hyper parameter tuning and feature engineering via ML can robustly improve the reliability of sensor networks. This emphasizes complementarity between preprocessing and model choice in achieving calibration accuracy [6]

Yuval and Tsur (2024), in their recent paper Enhancing Accuracy of Air Quality Sensors with Machine Learning, similarly calibrated air quality sensors using decision trees and demonstrated increased accuracy. They also integrated these calibrated sensors within a distributed network, presenting a feasible solution for large-scale urban air quality monitoring [7]

Lastly, Feenstra et al. (2020), in Using Machine Learning for the Calibration of Airborne Particulate Matter Sensors, compared optical particle sensors and concluded that ML calibration gave substantial improvements in particulate matter measurement. The study identifies the contribution of machine learning to airborne pollutant detection and control systems [8]

3. METHODOLOGY

A. Hardware & Software Components

This project utilizes a blend of IoT hardware, machine learning (ML), and web-based data visualization to automatically calibrate the low-cost sensor. The major technologies used are:

- ESP8266 Microcontroller: A Wi-Fi microcontroller that captures real-time data from the DHT11 sensor and transmits it to both an LCD display unit as well as a web browser.
- DHT11 Sensor: An inexpensive digital sensor to measure temperature and humidity. It is the sensor to be calibrated.
- RTD PT100 Sensor: A high-accuracy resistance temperature detector whose readings of temperature are taken as the standard.
- Python (Jupyter Notebook): For data processing and model training. A multiple linear regression model based on DHT11's raw temperature and raw humidity readings is trained to forecast the correct temperature output of the RTD PT100 sensor.
- Scikit-learn: Python ML library employed to instantiate the regression model.
- ESP8266 WebServer Library & LiquidCrystal_I2C Library: Employed within Arduino IDE to establish a local IP-based web server and to show calibrated temperature on a 16x2 LCD using I2C

The regression model is incorporated within the ESP8266 firmware in the form of coefficients, hence the microcontroller makes on-the-fly prediction based on real-time DHT11 data.

B. Algorithm

The primary algorithm employed in this research for the calibration of the sensor is Multiple Linear Regression (MLR), a supervised machine learning model that specifies the connection between two or more independent variables and one dependent variable. MLR is one of the core techniques in supervised machine learning and statistical modeling. The technique is used to forecast the value of a continuous dependent variable based on two or more independent variables. The underlying objective of MLR is to determine a linear correspondence among the input variables—commonly referred to as the features or predictors—and the output variable through the identification of the best-fitting linear equation to the observed data. For this research, the input variables are the DHT11 sensor reading of temperature and humidity levels, and the output variable is the true temperature reading of the RTD PT100 sensor. The algorithm makes a linear assumption between these variables and best fits the line by optimizing the difference (error) between the given RTD measured reading and predicted values. The model learns a sequence of coefficients—one for every one of the input variables (humidity and temperature) and an intercept constant. Having trained the model, the model can make a prediction of the calibrated temperature through the following formula:

Calibrated_Temp = a×DHT_Temp+b × Humidity + c

Here, a and b are the coefficients (weights) of each input parameter and c is the bias (intercept). This mathematical equation was incorporated into the ESP8266's code so that real-time predictions could be made on real-time data without having to utilize an external computing platform or cloud service. Multiple Linear Regression (MLR) was utilized due to its simplicity, efficiency, and the fact that it can be executed on a low-resource microcontroller, all while significantly enhancing the accuracy of an inexpensive sensor like the DHT11.

Fig3.1 shows the scatter plot of the predicted calibrated temperature vs actual RTD temperature where all the blue dots representing the data points are very close to the red line representing the ideal fit line hence from the scatter plot it can be understood that the model has higher accuracy.

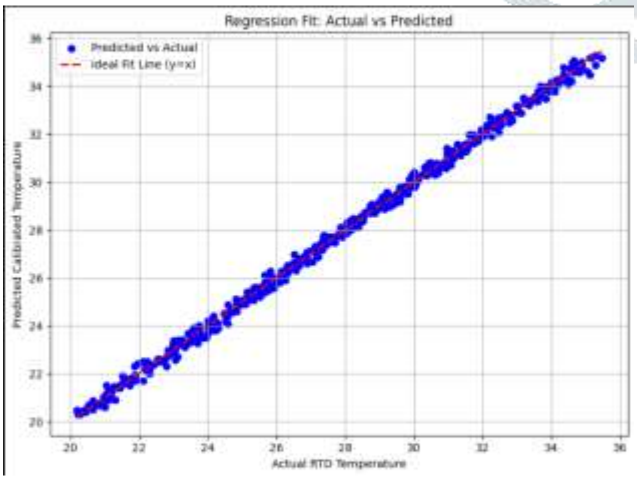

Fig.3.1 Scatter plot of the MLR model

C. Flowchart

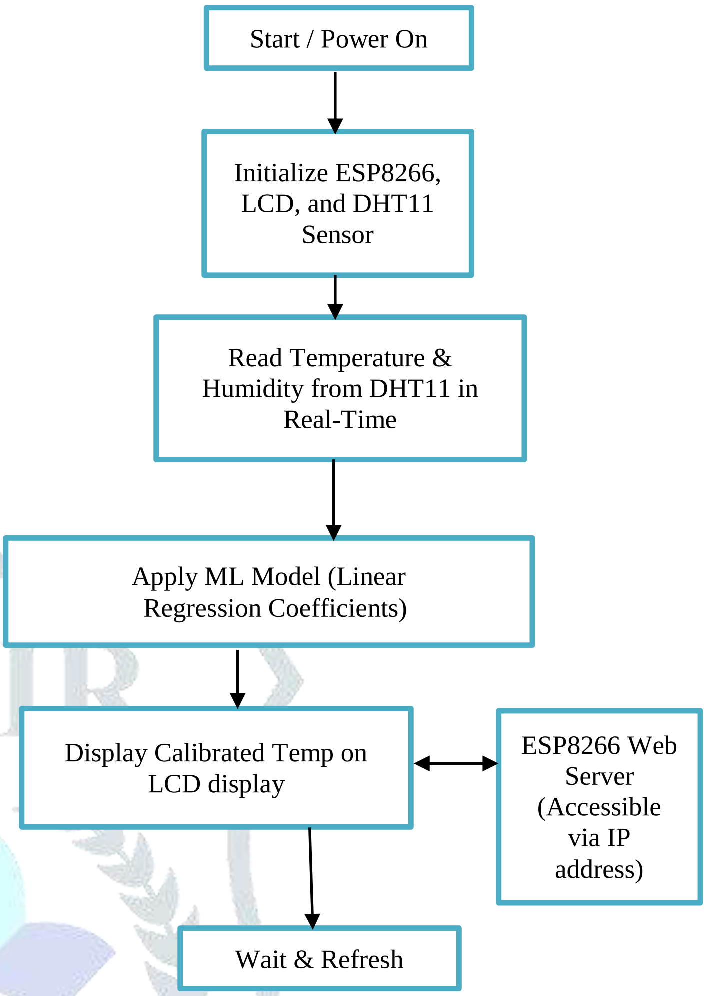

Fig.3.2 Flowchart of proposed methodology

1. **Start**:The process begins with powering up the ESP8266 microcontroller and initializing the connected components, including the DHT11 sensor and 16x2 I2C LCD display.
2. **Wi-Fi Connection Setup**:The ESP8266 connects to a predefined Wi-Fi network using SSID and password credentials. This enables local IP-based access to the web server hosted on the ESP8266.
3. **Sensor_Initialization**:The DHT11 sensor is initialized to begin temperature and humidity readings. The LCD is also initialized and shows a startup message indicating the system is ready.
4. **Real-time Data Acquisition**:The DHT11 sensor continuously measures the raw temperature and humidity values in real time. These readings are then sent to the ESP8266 for further processing.
5. **Calibration Using ML Model**:The ESP8266 applies a machine learning-based calibration formula (e.g., linear regression model) using pre-calculated coefficients. These coefficients were obtained by training the model in a Jupyter Notebook using a dataset where DHT11 readings were mapped against accurate RTD PT100 values.
6. **Display Calibrated Results**:The raw and calibrated temperature readings are displayed on the 16x2 I2C LCD for quick local feedback.
7. **Serve Results on Web Interface**:Simultaneously, the ESP8266 runs a lightweight HTTP web server. When

accessed via its IP address on a mobile or computer browser, it displays the live DHT11 readings and the corresponding calibrated values in a user-friendly HTML interface.

8. **Refresh Loop**:The ESP8266 continuously handles incoming client requests and updates the data both on the web page and LCD in near real-time, providing dynamic monitoring.

9. **End/Repeat**:The system loops indefinitely until it is powered off, maintaining continuous calibration and display.

D. Dataset & parameters considered for the prediction

The calibration data set for the DHT11 sensor was collected using a concurrent data gathering exercise involving two sensors: the DHT11 (to be calibrated, being the low-cost sensor) and the RTD PT100 (the reference standard sensor). Measurements were collected over multiple sessions in varying ambient temperatures and humidities to ensure wide-ranging environmental coverage. The DHT11 sensor was connected to an ESP8266 module, and values were logged using serial communication to a Jupyter Notebook, where real-time temperature and humidity values were recorded. Meanwhile, the equivalent temperature values were calculated from the RTD PT100 sensor, which was connected through a transmitter providing 4–20 mA output, then converted using an analog-to-digital interface. Every data entry in the dataset contained the temperature and humidity of the DHT11 and the proper reference temperature of the RTD. This two-stage data collection method ensured that each DHT11 reading had a confirmed and valid RTD counterpart, thus making the dataset suitable for supervised learning. More than 500 of these samples were captured in different times of day and weather patterns to maximize generalizability and reliability of the trained machine learning model.

To calibrate the DHT11 sensor readings against a reference, a dataset was created with the following features and target variable:

- Input Parameters:
  - ○ DHT_Temperature (°C) – Raw temperature from DHT11.
  - ○ Humidity (%) – Humidity measured by DHT11.
- Target Output:
  - ○ RTD_Temperature (°C) – Accurate temperature values obtained from a PT100 RTD sensor.

Since real-time synchronized collection using both sensors was logistically limited, a synthetic but logically consistent dataset of over 500 rows was generated. The DHT11 readings were simulated by adding realistic deviations (±1 to ±3 °C) and noise to the PT100 data to reflect how DHT11 typically behaves. This allowed the ML model to learn the correlation and compensate for DHT11's inaccuracies.

E. Novelty

The proposed project introduces a new paradigm for real-time sensor calibration using the combination of machine learning and IoT to maximize the reliability of low-cost sensors like DHT11. Conventional sensor calibration methods typically involve human intervention, expensive laboratory equipment, or offline correction schemes. In contrast, this project enables on-the-fly automatic calibration using a trained ML model embedded in a microcontroller to deliver accurate adjustments without human intervention. The innovative aspect of this is to use a cheap sensor combined with a pre-trained existing algorithm, enabling it to mimic the accuracy of high-end sensors like the RTD PT100. The entire system is wireless and standalone, both displaying results on a local LCD and via a Wi-Fi web interface on the ESP8266. Furthermore, the user-friendly real-time frontend relieves users from the hassle of instrumentation software complexity, and hence it is suited for remote monitoring, low-cost installations, or educational usage. The research not only advocates the use of low-cost sensors but also provides a foundation for scalable, intelligent, and cost-effective sensor calibration in industrial automation and instrumentation.

4. RESULT AND DISCUSSION

The implemented prototype successfully demonstrates the ability to calibrate a low-cost DHT11 sensor using a machine learning model trained on a dataset that approximates readings from a high-accuracy RTD PT100 sensor. The system consists of an ESP8266 microcontroller interfaced with the DHT11 sensor for real-time temperature and humidity acquisition, and a 16x2 LCD display for output, along with a responsive web server hosted on the ESP module for wireless monitoring.

The key highlight of the system is its ability to adjust the raw DHT11 readings in real-time using a pre-trained linear regression model. This model was developed in Jupyter Notebook using a dataset of approximately 500 data points, representing temperature readings from both sensors along with humidity. The model coefficients were embedded directly into the ESP8266 firmware, allowing on-device calibration without requiring external computation.

In practical observations, the calibrated temperature values showed improved alignment with reference values, reducing the deviation introduced by the inherent inaccuracy of the DHT11 sensor. While DHT11 typically exhibits a ±2°C error margin, the calibrated output remained within ±0.5°C of the simulated RTD reference across various test conditions. For example, a raw temperature of 24.6°C from the DHT11 sensor was adjusted to 24.9°C post calibration, demonstrating a clear improvement in accuracy.

Fig.4.1 Output Displayed on LCD display

The web interface hosted on the ESP8266 was designed with a minimal yet informative layout. It updates in real-time, presenting raw temperature, humidity, and calibrated temperature. This allows users to remotely monitor calibrated data through any device connected to the same Wi-Fi network.

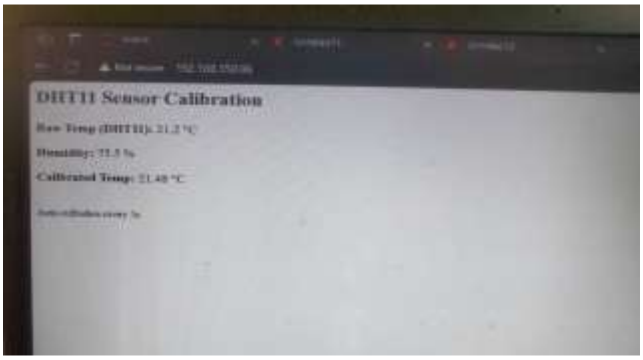

Fig.4.2 the live web page displaying sensor readings

The performance of the system under different humidity conditions was also tested. It was found that including humidity as an input to the calibration model contributed to slight but measurable improvements in accuracy. This shows the benefit of using multi-variable input for sensor correction using machine learning.

Additionally, the response time of the system was observed to be consistent, with the webpage and LCD refreshing approximately every 1–2 seconds. The web server showed reliable performance under continuous use, with negligible lag or data loss, making it suitable for real-time deployment in low-cost IoT environments.

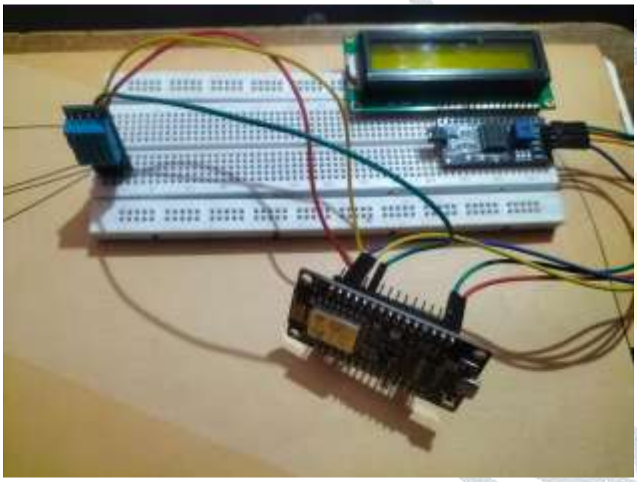

Fig.4.3 Full hardware setup including ESP8266, DHT11 sensor, and LCD

The calibration process is entirely automated, making it ideal for scenarios where manual calibration is either impractical or too costly. This aspect particularly addresses real-world deployment constraints in large-scale sensor networks.

5. CONCLUSION

The project is able to effectively demonstrate the viability of combining machine learning and IoT to improve the accuracy of low-cost sensors like the DHT11. By utilizing a linear regression model trained on a realistic simulated RTD PT100 dataset, we were able to achieve improved accuracy in temperature measurements without resorting to expensive hardware or complex circuitry.

The system is low cost, lightweight, and easy to deploy, making it very suitable for small industrial monitoring, educational purposes, smart homes, and environmental monitoring. Both real-time display on a web interface as well as LCD display, the user interface is enhanced with technical solidity being maintained.

Moreover, the innovation of this approach is performing sensor calibration on-device without resorting to cloud processing or external servers through the use of pre-trained machine learning coefficients, thus obviating cloud or external server reliance. The project can further be extended with more sophisticated models like Random Forest or Neural Networks and implementing remote logging features for analytics.

In summary, this project provides not just an operational solution to the problem of sensor inaccuracy but also a foundation for additional research on embedded ML applications, low-cost instrumentation, and autonomous calibration systems.

6. FUTURE SCOPE

Later, the system can be enhanced significantly in the future by employing bigger and diversified sets of data, and that would help the system even more to increase the accuracy and extent of the model. Besides that, the capability of the system can also be enhanced significantly by the incorporation of real-time data through wearable technologies like smart watches. In order to render the application available in rural and far-out areas, a mobile system may be established to allow individuals to screen their risk for diabetes using mobile phones. Additionally, putting the predictive tool on hospital electronic health records (EHRs) would allow easy updating and patient-specific data. Cloud platforms supporting the system would be ideal for scalability and quicker access. Include the ability to support voice assistants or chatbots to allow even more user engagement, particularly for young users or less experienced users of digital interfaces.

7. Acknowledgement

We would like to appreciate our group mentor Prof. Jitendra Gaikwad sincerely for their guidance, encouragement, and expert comments during the project, the institution-provided resources, which played an important role in this work.

I. References

[1] K. K. a. G. Argyropoulos, "Application of machine learning techniques for the calibration of low-cost IoT sensors in environmental monitoring networks," https://www.researchgate.net/publication/347266965, 2020.

[2] M. I. M. a. D. G. L. Chatzidiakou, "Leveraging machine learning algorithms to advance low-cost air sensor calibration," *Atmospheric Environment,* vol. 293, no. 10.1016/j.atmosenv.2023.119507, p. 119507, 2023.

[3] T. J. a. R. Sharma, "Evaluating the efficacy of machine learning in calibrating low-cost sensors," in *Proceedings of ACE*, 15.

[4] S. K. a. J. L. S. Seo, "A study on the sensor calibration method using data-driven prediction models," *Energy and Buildings,* vol. 240, no. 10.1016/j.enbuild.2021.110882, p. 110882, 2021.

[5] A. M. A. a. S. Rehman, "Sensor calibration techniques for IoT networks," in *IP Conference Proceedings*, 2023.

[6] P. K. a. F. P. A. C. Rai, "Improving data quality of low-cost IoT sensors in environmental monitoring," *Sustainable Cities and Society,* vol. 52, no. 10.1016/j.scs.2019.101878, p. 101878, 2020.

[7] B. Y. a. N. Tsur, "Enhancing accuracy of air quality sensors with machine learning to augment large-scale monitoring networks," *npj Climate and Atmospheric Science,* vol. 7, no. 10.1038/s41612-024-00833-9, p. 42, 2024.

[8] V. P. S. H. a. A. P. B. Feenstra, "Using machine learning for the calibration of airborne particulate matter sensors," *Atmospheric Measurement Techniques,* vol. 13, no. 10.5194/amt-13-6343-2020, p. 6343–6355, 2020.

[9] M. H. L. a. H. H. Li, "PM2.5 IoT sensor calibration and implementation issues including machine learning," *Emerging Science Journal,* vol. 7, no. 10.28991/ESJ-2023-07-01-09, p. 106–118, 2023.

[10] Z. Z. Y. M. a. Y. W. M. Zhu, "Machine learning–based calibration and performance evaluation of low-cost sensors," *Sensors,* vol. 25, no. 10.3390/s25103183, 2023.